The New Encyclopaedia Britannica

in 30 Volumes

MACROPEDIA
Volume 10

Knowledge in Depth

FOUNDED 1768
15 TH EDITION

Encyclopædia Britannica, Inc.
William Benton, Publisher, 1943–1973
Helen Hemingway Benton, Publisher, 1973–1974

Chicago
Auckland/Geneva/London/Manila/Paris/Rome
Seoul/Sydney/Tokyo/Toronto

First Edition	1768-1771
Second Edition	1777-1784
Third Edition	1788-1797
Supplement	1801
Fourth Edition	1801-1809
Fifth Edition	1815
Sixth Edition	1820-1823
Supplement	1815-1824
Seventh Edition	1830-1842
Eighth Edition	1852-1860
Ninth Edition	1875-1889
Tenth Edition	1902-1903

Eleventh Edition
© 1911
By Encyclopædia Britannica, Inc.

Twelfth Edition
© 1922
By Encyclopædia Britannica, Inc.

Thirteenth Edition
© 1926
By Encyclopædia Britannica, Inc.

Fourteenth Edition
© 1929, 1930. 1932. 1933. 1936. 1937. 1938. 1939, 1940, 1941, 1942, 1943,
1944, 1945, 1946, 1947, 1948, 1949, 1950, 1951, 1952, 1953, 1954,
1955, 1956, 1957, 1958, 1959, 1960, 1961, 1962, 1963, 1964,
1965. 1966. 1967, 1968. 1969, 1970, 1971, 1972, 1973
By Encyclopædia Britannica, Inc.

Fifteenth Edition
© 1974. 1975, 1976, 1977. 1978, 1979, 1980, 1981, 1982, 1983
By Encyclopædia Britannica, Inc.

© 1983
By Encyclopædia Britannica. Inc

Copyright under International Copyright Union
All rights reserved under Pan American and
Universal Copyright Conventions
by Encyclopædia Britannica, Inc.

No part of this work may be reproduced or utilized
in any form or by any means, electronic or mechanical
including photocopying. recording, or by any
information storage and retrieval system, without
permission in writing from the publisher.

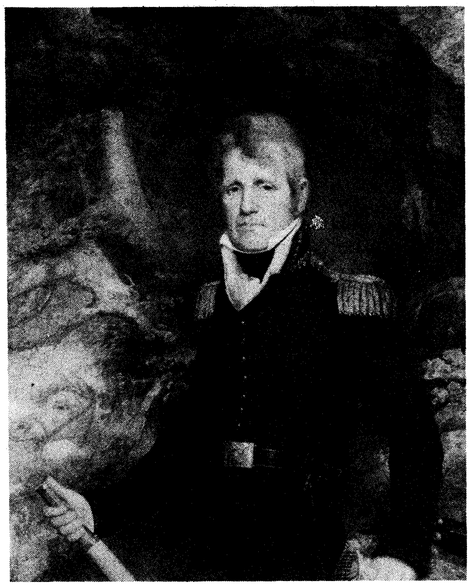
Printed in U.S.A.

Library of Congress Catalog Card Number: 81-70041
International Standard Book Number: 0-85229-400-X

Jackson, Andrew

A military hero and the seventh president of the United States, Jackson was the first man to win the presidency primarily through his appeal to the mass of the voters and by means of a skillfully planned national campaign rather than through the support of the political leaders of the seaboard states. As president, he initiated little significant legislation, but he was a forceful president who exercised vigorously the power and authority of his office; and, with the collaboration of a coterie of astute advisers, he transformed a political faction into a truly national political organization, the Democratic Party, which may be described as the first modern political party in the history of the United States.

By courtesy of the Metropolitan Museum of Art, New York, Harris Brisbane Dick Fund, 1964



Andrew Jackson, oil painting by John Wesley Jarvis, c. 1819. In the Metropolitan Museum of Art, New York.

He was born on March 15, 1767, at the Waxhaw settlement of the western frontier of the Carolinas. The area was in dispute between North Carolina and South Carolina, and both states have claimed him as a native son. Jackson maintained that he was born in South Carolina, and the weight of evidence supports his assertion. The Waxhaw area offered little opportunity for formal education, and what little schooling Jackson received was interrupted by the British invasion of the western Carolinas in 1780–81. In 1781 he was captured and imprisoned for a time by the British. Shortly after he was captured, he refused to shine the boots of a British officer and was struck across the face with a sabre. His mother and two brothers, Hugh and Robert, died during the closing years of the war, direct or indirect casualties of the invasion of the Carolinas. This sequence of tragic experiences fixed in Jackson's mind a lifelong hostility to Great Britain.

Early career

After the end of the U.S. war of independence, Jackson studied law in an office in Salisbury, North Carolina, and was admitted to the bar of that state in 1787. In 1788 he went to the Cumberland region as prosecuting attorney of the western district of North Carolina—the region west of the Appalachians, which was soon to become the state of Tennessee.

When Jackson arrived in Nashville, the judicial seat of the district, the community was still a frontier settlement exposed to the danger of Indian raids, with a fluid social structure, no well-established political institutions, and a volatile economy in which virtually every investment had the character of a speculation. As prosecuting attorney, Jackson was principally occupied with suits for the collection of debts. He was so successful in these litigations that he soon had a thriving private practice and had gained the friendship of landowners and creditors. For almost 30 years Jackson was allied with this group in Tennessee politics.

Marriage

Jackson boarded in the home of Col. John Donelson, where he met and quickly became attracted to the Colonel's daughter, Mrs. Rachel Robards. Her marriage with Robards was a stormy one, which seemed destined to end in separation or divorce even before she met Jackson. In 1790 Robards sought a divorce by legislative enactment in Virginia, but the legislature merely empowered Robards to sue for divorce. Jackson and Mrs. Robards believed the legislature had granted the divorce and in 1791 they were married. Robards, however, did not sue for divorce until 1793 and then obtained it on the grounds of desertion and adultery. When Jackson and his wife learned the true state of affairs, they were remarried quietly in Nashville. Until the death of Mrs. Jackson, nearly 40 years later, political opponents did not hesitate to make unflattering insinuations about events leading to his marriage. Although it appears to have been a happy union, gossip concerning it kept Jackson perpetually on the defensive.

Jackson was a successful lawyer, but he was too restless and his interests too varied to be confined to a career as an attorney. The expanding economy of frontier Tennessee encouraged adventures in trade and in land speculation, and Jackson became involved in both. He invested heavily in land and horses and also raised cotton. The course of his investments was an unsteady one. At times he appeared to be financially secure; on other occasions, he was hard pressed by debt.

Tennessee politics. His interest in public affairs and in politics had always been keen. He had gone to Nashville as a political appointee, and in 1796 he became a member of the convention that drafted a constitution for the new state of Tennessee. In the same year he was elected as the first representative from Tennessee to the national House of Representatives. He refused to seek re-election and served only until March 4, 1797. These were the closing months of President George Washington's administration, and the chief event of Jackson's brief service in the House was the debate over the response to Washington's farewell address. In a speech in the House, Jackson criticized Washington because of the latter's support of Jay's Treaty, which Jackson believed did not contain adequate guarantees that Great Britain would not violate the rights of American ships on the high seas during the wars of the French Revolution. He was one of 12 representatives who voted against sending a cordial reply, thanking Washington for the address.

Jackson returned to Tennessee, vowing never to enter public life again, but before the end of the year he was elected to the United States Senate. His willingness to accept the office reflects his emergence as an acknowledged leader of one of the two political factions contending for control of the state.

Jackson resigned from the Senate in 1798 after an uneventful year. Personal financial pressures were the immediate cause of his resignation, but he appeared ill at ease in national politics and was unhappy in being separated from his family. Soon after his return to Nashville he was appointed a judge of the Superior Court (in effect, the supreme court) of the state, serving until **1804,** when financial difficulties again forced him out of public life.

In **1802** Jackson was elected major general of the Tennessee militia, a position he still held when the War of **1812** opened the door to a command in the field and a hero's role. In the intervening years, though he was nominally out of politics, Jackson maintained his old associations with many of the active political leaders of the state. It was during this period that he renewed his acquaintance with Aaron Burr, whom he had met and admired when they served together in Congress. For a time, Jackson appeared to be willing to aid Burr in the latter's nebulous scheme for the conquest of Texas. Later, when he began to suspect that Burr entertained treasonable intentions of establishing an independent republic, Jackson withdrew his support. He was not fully convinced of Burr's guilt, however, and he appeared as a defense witness in Burr's trial for treason in **1807.**

Military feats. In March **1812,** when it appeared that war with Great Britain was imminent, Jackson issued a call for **50,000** volunteers from the state to be ready for an invasion of Canada. After the declaration of war, in June **1812,** Jackson offered his services and those of his militia to the United States. The government was slow to accept this offer, and, when Jackson finally was given a command in the field, it was to fight against the Creek Indians, who were allied with the British and who were threatening the Southern frontier. In a campaign of about five months, in **1813–14,** Jackson crushed the Creeks, the final victory coming in the Battle of Tohopeka (or Horseshoe Bend) in Alabama. The victory was so decisive that the Creeks never again menaced the frontier, and Jackson was able to impose upon them a treaty whereby they surrendered to the federal government an estimated **23,000,000** acres **(9,000,000** hectares) of land, comprising about one-fifth of the state of Georgia and more than three-fifths of the present state of Alabama. The campaign against the Creeks thus opened a vast new area to settlement and established Jackson as the hero of the West.

War of **1812**

In August **1814,** Jackson moved his army south to Mobile. Though he was without specific instructions, his real objective was the Spanish post at Pensacola. The motive was to prepare the way for U.S. occupation of Florida, then a Spanish possession. Jackson's justification for this bold move was that Spain and Great Britain were allies in the wars in Europe. At Mobile, Jackson learned that an army of British regulars had landed at Pensacola. In the first week in November, he led his army into Florida and, on November **7,** he occupied that city just as the British evacuated it to go by sea to Louisiana. Jackson then marched his army overland to New Orleans, where he arrived early in December. A series of small skirmishes between detachments of the two armies culminated in a major battle on January **8, 1815,** in which Jackson's forces inflicted a decisive defeat upon the British army and forced it to withdraw from Louisiana. The news of this victory reached Washington at a time when morale in the capital was at a low point. A few days later, news of the signing of a peace treaty between the United States and Great Britain on December **24, 1814,** at Ghent, Belgium, reached Washington. The twin tidings of victory and peace brought joy and relief to the American people and made Jackson the hero not only of the West but of a substantial part of the nation.

Florida campaign

After the close of the war, Jackson retained his commission in the army and was named commander of the southern district. The chief duty of the army in that district was to defend southern Georgia against possible raids by Indians from the Spanish colony of Florida; and Jackson entrusted the command of the troops in the field to subordinates while he retired to his home at the Hermitage, near Nashville. From that congenial environment, he was ordered back to active service at the end of December **1817,** when unrest along the border appeared to be reaching critical proportions. The instructions given Jackson were vague, probably purposely so, and it was well-known in Washington that Jackson hoped to gain Florida for the United States. As could have been expected, he ordered an invasion of Florida immediately after taking active command of the troops. He captured two Spanish posts and, before returning to the United States, appointed one of his subordinates military governor of Florida. These bold actions brought an immediate and sharp protest from Spain and precipitated a Cabinet crisis in Washington. The staunch defense of Jackson by Secretary of State John Quincy Adams, both in the Cabinet and in a note to Spain, saved Jackson from censure and hastened the acquisition of Florida by the United States.

Presidential prospects. Jackson's military triumphs led to suggestions that he become a candidate for president, but he disavowed any interest in seeking the presidency, and political leaders in Washington assumed that the flurry of support for him would prove transitory. The campaign to make him president, however, was kept alive by his continued popularity and was carefully nurtured by a small group of his friends in Nashville, who combined devotion to the general with a high degree of political astuteness. In **1822** these friends manoeuvred the Tennessee legislature into a formal nomination of their hero as a candidate for president. In the following year this same group persuaded the legislature to elect him to the United States Senate—a gesture designed to demonstrate the extent of his popularity in his home state. By December **1823,** when he reached Washington to assume his new duties, Jackson had become an open and formidable contender for the presidency.

In the election of **1824** four candidates received electoral votes. Jackson received the highest number, the other three being John Quincy Adams, William H. Crawford, and Henry Clay. Since no one had a majority of the electoral votes, the House of Representatives was required to elect a president from the three with the highest number of votes. Crawford was critically ill and the actual choice was between Jackson and Adams. Clay, as speaker of the House, was in a strategic and perhaps decisive position to determine the outcome. Adams was elected on the first ballot. When he appointed Clay secretary of state—then regarded as the second most powerful office in the government—it seemed to admirers of Jackson to confirm rumours of a "corrupt bargain" between Adams and Clay. Jackson's friends persuaded him that the popular will had been thwarted by intrigues, and he thereupon determined to vindicate himself and his supporters by becoming a candidate again in **1828.**

Election of **1828**

The campaign to elect Jackson president was resumed almost as soon as Adams took office. Skillful politicians flocked to the support of Jackson, whom they recognized as the man of the future, and Adams' efforts to provide leadership were frustrated by factional opposition. In **1828** Jackson defeated Adams by an electoral vote of **178** to **83** after a campaign in which personalities and slander played a larger part than in any previous national election in the history of the United States. But Jackson's hour of triumph was soon overshadowed by personal tragedy. During the campaign his opponents had resurrected the charges that his marriage had been irregular and this may have hastened a decline in his wife's health. Worn-out and defeated in her long effort to avoid the glare of public life, Mrs. Jackson died at the Hermitage on December **22,1828.**

"Jacksonian Democracy." The election of **1828** is commonly regarded as a turning point in the political history of the United States. Jackson was the first president from the area west of the Appalachians, but it was equally significant that the initiative in launching his candidacy and much of the leadership in the organization of his campaign also came from the West. The victory of Jackson indicated a westward movement of the centre of political power. He was also the first man to be elected president through a direct appeal to the mass of the voters rather than through the support of a recognized political organization. His victory was regarded by contemporaries and by historians as the triumph of political democracy, and the political movement of which he was the central figure is commonly described as the "Jacksonian Democracy."

Jackson was the beneficiary of a rising tide of democratic sentiment which was evident in nearly every part of the United States. The trend toward greater political democracy, aided by the admission of six new states to the union, five of which had manhood suffrage, along with the extension of the suffrage laws by many of the older states, weakened the power of the older political organizations and opened the way for the rise of new political leaders skilled in appealing to the mass of voters.

Not the least remarkable triumph of the Jacksonian organization was its success in picturing its candidate as the embodiment of democracy, despite the fact that Jackson had been aligned with the conservative faction in Tennessee politics for 30 years and that in the financial crisis that swept the West after 1819 he had vigorously opposed legislation for the relief of debtors. In terms of personal qualities, however, he was well cast in the role of the champion of democracy. He had been born on the frontier and had risen to fame as the defender of the frontier against its traditional enemies—the Indians, the British, and the Spanish. He was a man of limited education and no pretense to inherited social position. He was the epitome of the self-made man, the ideal admired by hundreds of thousands of his fellow citizens and one they hoped to emulate.

The new presidential personality

As the victory of Jackson reflected the emergence of new forces in U.S. politics, so Jackson himself brought to the presidency a new set of personal qualifications that were to become the standard by which presidential candidates would be judged for the remainder of the 19th century. He was the first president since Washington who had not served a long apprenticeship in public life and had no personal experience in the formulation or conduct of foreign policy. His brief periods of service in Congress provided no clue to his stand on the public issues of the day, except perhaps on the tariff. On even that issue, his position was ambiguous, for he would only say that he favoured a "judicious" tariff and that manufactures "properly fostered and protected" were important to the defense of the nation.

Jackson approached the problems of the presidency as he had approached all other problems in life. He met each issue as it arose, and in the White House he exhibited the same vigour and determination in carrying out decisions that had characterized his conduct as commander of an army in the field. He made it clear from the outset that he would be the master of his own administration, and, at times, he was so strong-willed and decisive that his enemies referred to him as "King Andrew I" and to his administration as the "reign" of Andrew Jackson. Strong leadership did not guarantee consistency in planning a comprehensive policy. In retrospect, certain fundamental principles that affected his decisions may be discerned, such as the maintenance of the strength of the Union, a distrust of inflation and of paper money, and a conviction that concentrations of power—particularly of "moneyed power"—were dangerous. His practice of acting upon each question as he met it often made it difficult for even his friends to predict the decisions he would make; and on at least one major issue—that of states' rights—his actions on successive problems in which this principle was involved were clearly inconsistent.

The first term. When Jackson was inaugurated on March 4, 1829, it was the first time in more than a quarter of a century that the advent of a new president reflected the repudiation of his predecessor. Hundreds who had worked for the election of Jackson hoped this would mean that incumbent officeholders would be replaced by friends of the new president, and within a few weeks the process of removing opponents of Jackson to make way for supporters had begun. Some years later, in the U.S. Senate, William L. Marcy of New York defended the principle of "rotation of office" with the aphorism, "To the victors belong the spoils." Rotation in office, however, did not begin with Jackson, nor did he utilize this practice so extensively as was charged. In eight years as president, Jackson removed fewer than one-fifth of all federal officeholders.

Jackson was in poor health when he became president, and few believed that he would have the strength or inclination to seek a second term. The question of the succession was, therefore, certain to attract early attention. One obvious candidate was Vice President John C. Calhoun from Jackson's native state of South Carolina. Another was Martin Van Buren, Jackson's first secretary of state. The harmony of the new administration was marred from the outset by the rivalry between Calhoun and Van Buren. When Jackson learned, in 1830, that during the Cabinet debates in 1818 Calhoun had urged that he be censured for his invasion of Florida, he concluded that he could no longer trust Calhoun. From that time, Van Buren was generally recognized as the probable successor of Jackson as president.

Threats to federal authority

The feud between Jackson and Calhoun assumed crucial importance in 1830 when Calhoun openly espoused the cause of South Carolina in its opposition to a high protective tariff. Feeling in South Carolina was so intense that there were covert threats that the state would attempt to prevent collection of the tariff within its borders. To make his own position clear, Jackson used the occasion of a Jefferson Day dinner to offer the toast, "Our federal union, it must be preserved." The meaning of the toast was clear to all who heard it, including Calhoun. When Calhoun, obviously shaken, followed with the toast, "The union, next to our liberty most dear," the conflict between the two men was evident to all. The issue of the tariff drifted unresolved, however, until 1832, when congressional leaders sought a compromise in the form of a moderate reduction of the tariff. South Carolina was not satisfied and in reply adopted a resolution declaring the tariffs of 1828 and 1832 null and void and prohibiting the enforcement of either within its boundaries after February 1, 1833. Jackson accepted the challenge, denounced the theory of nullification, and asked Congress for authority to send troops into South Carolina to enforce the law. The President believed the tariff to be too high, however, and urged Congress to reduce the tariff rates it had enacted a few months earlier. On March 1, 1833, Congress sent to the President two companion bills. One reduced tariff duties on many items. The other, commonly called the Force Bill, empowered the president to use the armed forces to enforce federal laws within a state. Public opinion outside South Carolina appeared to favour the compromise. Later in the month, South Carolina repealed its nullification ordinance, but at the same time it declared the Force Act null and void. The vigour with which Jackson responded to nullification in South Carolina may be explained, in part, by his deep dislike of Calhoun, but personalities alone do not explain the position of either man. Jackson was devoted to the Union, Calhoun to South Carolina and the doctrine of states' rights.

Whatever the motives, Jackson had preserved the integrity of the Union against the most serious threat it had yet faced. In contrast, he was remarkably complacent when Georgia defied the federal government. In 1829 Georgia extended its jurisdiction to about 9,000,000 acres of land that lay within its boundaries but was still occupied by Indians. The Indians' title to the land had been guaranteed by a treaty with the United States. The Indians appealed to the federal courts. In two separate cases, the Supreme Court ruled against Georgia, but Georgia ignored those decisions and continued to enforce its jurisdiction within the territory claimed by the Indians. In contrast to his strong reaction against South Carolina's defiance of federal authority, Jackson made no effort to restrain Georgia; and those close to him felt certain that he sympathized with the position taken by that state. He was spared, however, the embarrassment of making any public statement about the controversy between the court and Georgia, because neither party asked him to intervene.

Re-election in 1832. In the meantime, Jackson acquiesced to the pressure of friends and sought a second

Bank issue

term as president. As the election of 1832 approached, Jackson's opponents hoped to embarrass him by presenting him with a new dilemma. The charter of the Bank of the United States was due to expire in 1836. The President had not clearly defined his position, but he was increasingly uneasy about the bank as then organized. More significant in an election year was the fact that large blocs of voters who favoured Jackson were openly hostile to the bank. In the summer of 1832, Jackson's opponents rushed through Congress a bill to recharter the bank, thus forcing Jackson either to sign the measure and alienate many of his supporters or to veto it and appear as a foe of sound banking. Jackson's Cabinet was divided between friends and critics of the bank, but the obviously political motives of the recharter bill reconciled all of them to the necessity of a veto. The question before Jackson actually was whether the veto message should leave the door open to future compromise.

Few presidential vetoes have caused as much controversy in their own time or later as the one Jackson sent to Congress on July 10, 1832. A long document of more than 7,000 words, much of it argued that the bill was premature or attacked certain details in the legal status of the bank. At the opening of the message and again near the end, Jackson conceded the value of a national bank, "so organized as not to infringe on our delegated powers or the reserved rights of the states." Judged by comments made prior to the introduction of this bill in Congress, there were many friends of the bank who could have agreed with a large part of the veto message. Portions of the message, however, took more controversial ground. Jackson began with the charge that the stockholders would gain a windfall of $7,000,000 through increased value of their stock if the bill became law. In a closing paragraph, he revealed the basic cause for his opposition to the existing bank, asserting that "the rich and powerful too often bend the acts of government to their selfish purposes."

As anticipated, the Senate upheld the veto. Jackson's opponents rejoiced, believing that he had given them an issue on which they could defeat him. They had seriously underestimated the strength of the coalition which opposed the bank. Westerners, many of whom had been suspicious of all banks since the Panic of 1819, were joined in their opposition to the bank by bankers, especially in New York, who were jealous of the special advantages enjoyed by the bank, by businessmen and speculators who desired an expansion of credit to encourage a more rapid development of the economy, and by labouring men, especially in New York, who looked upon Jackson as their champion. That motley coalition rallied behind Jackson; and in November they re-elected him by a margin which was more decisive than the one by which he had been elected in 1828.

Subsequent financial policy

The veto of the bill to recharter the bank was the prelude to a conflict over financial policy that continued through Jackson's second term. After Jackson's re-election, Nicholas Biddle, president of the bank, initiated a policy of calling in loans and contracting the volume of notes issued by the bank, ostensibly because it was necessary to prepare for the time when the charter would expire. Jackson and his associates were convinced that Biddle was seeking to create a panic in order to force Congress to recharter the bank. The administration met this threat by refusing to deposit funds of the federal government in the bank and, instead, deposited them in banks all over the country, an action that Jackson believed would minimize any effort by the bank to control the volume of bank notes in circulation but that also stimulated an inflationary trend. Friends of the bank countered with congressional resolutions affirming belief in the solvency of the bank and censuring Jackson; but these had no other effect than to persuade Jackson that Biddle was using political influence to preserve the bank.

While the "bank war" raged, Jackson and his advisers had developed a theory of business cycles, in which paper money was the villain that caused alternate periods of inflation and depression. If the circulation of paper could be restricted and the proportion of gold and silver to paper increased, the cycle could be brought under some measure of control. Wage earners and those with fixed incomes would then receive some protection against periodic disaster. These were the very groups, Jackson believed, that usually were not adequately represented in the government and whose champion he should be.

Efforts to persuade Congress to enact legislation limiting the circulation of bank notes failed, but there was one critical point at which Jackson was free to apply his theories. Nearly all purchasers of public lands paid with bank notes, many of which had to be discounted because of doubts as to the continuing solvency of the banks that issued them. Partly to protect federal revenues against loss and partly to advance his concept of a sound currency, Jackson issued the Specie Circular in July 1836, requiring payment in gold or silver for all public lands. This measure created a demand for specie that many of the banks could not meet; the effect of bank failures in the West spread to the East, and by the spring of 1837 the entire country was gripped by a financial panic. The panic did not come, however, until after Jackson had had the pleasure of seeing Van Buren inaugurated as president on March 4, 1837.

Later years

Upon leaving the presidency, Jackson retired to his home, the Hermitage. He was virtually an invalid during the remaining eight years of his life, but he continued to have a lively interest in public affairs. Friends of earlier years, party leaders, and aspiring young men made the Hermitage a mecca where they hoped to renew friendships or to receive the counsel and blessings of the former President. He died at the Hermitage on June 8, 1845.

Jackson's influence. Jackson had left office more popular than when he entered. The widespread approval of his actions exercised a profound effect on the character of U.S. politics for half a century. The success of Jackson appeared to be a vindication of the new democracy. Powerful voices still questioned the wisdom and even the morality of democracy in 1829; there were few who would question it openly in 1837. Jackson had likewise established a pattern that future candidates for the presidency attempted to imitate. Birth in humble circumstances, experience on the frontier, evidence of being close to the mass of the people, a devotion to democracy, and, if possible, some military exploits were all valuable assets for any candidate who hoped to succeed.

The intensity of the political struggles from 1825 to 1837 led to the revival of the two-party system, which had been in abeyance since 1817. Jackson never thought of himself as a master politician, but he and his associates proved themselves the most skillful political leaders of that generation. When Jackson was elected president in 1828, he was the candidate of a faction rather than of a party. When he retired from the presidency he left a vigorous and well-organized Democratic Party as one of his legacies to the future.

BIBLIOGRAPHY. The standard biographies of Jackson are J.S. BASSETT, ***The Life of Andrew Jackson,*** 2 vol. ***(1911;*** rev. ed., 1916; 1 vol., 1931, reprinted 1967); and a two-volume work by MARQUIS JAMES: ***Andrew Jackson: The Border Captain*** (1933) and ***Andrew Jackson: Portrait of a President*** (1937). The best study of the election of 1828, emphasizing the political skill of the men who organized Jackson's campaign, is ROBERT V. REMINI, ***The Election of Andrew Jackson*** (1963). An excellent study of politics in the Jacksonian era is R.P. MCCORMICK, ***The Second American Party System: Party Formation in the Jacksonian Era*** (1966).

(H.W.Br.)

Jackson, "Stonewall"

One of the leading generals of the Confederacy during the U.S. Civil War, Stonewall Jackson ranks among the most skillful tacticians in military history.

Thomas Jonathan Jackson was born at Clarksburg, Virginia (now West Virginia), on January 21, 1824. The early death of his father, who left little support for the family, and his mother's subsequent death shortly after her remarriage, caused the boy to grow up in the homes of relatives. He had little opportunity for formal education in his early years, but he received an appointment, in 1842, to the U.S. Military Academy at West Point. After a slow start, he graduated 17th in his class and was commissioned as a second lieutenant assigned to artillery. He joined his regiment in Mexico, where the United States was then at war. In the Mexican War he first met Gen. Robert E. Lee, who later became the commanding general of the Confederate armies, and it was here that Jackson first exhibited the qualities for which he later became famous: resourcefulness, the ability to keep his head, bravery in the face of enemy fire. At the end of the fighting in Mexico, having been promoted to first lieutenant and to the brevet rank of major, he was assigned to the occupation forces in Mexico City.

By courtesy of the Library of Congress, Washington, D.C.

"Stonewall" Jackson, 1883.

"Deacon Jackson"

Finding service in the peacetime army tedious, he resigned his commission and became professor of artillery tactics and natural philosophy at the Virginia Military Institute in **1851.** Though he worked hard at his new duties, he never became a popular or highly successful teacher. A stern and shy man, he earned a reputation for eccentricity that followed him to the end of his career. During the ten years at VMI, his first wife, Eleanor, died. She was the daughter of the Rev. George Junkin, president of Washington College. Three years later Jackson married Mary Anna Morrison, daughter of a North Carolina minister. Meanwhile, he began to investigate religion, a search that began with Roman Catholicism, evolved into a personal code of ethics, and culminated in an affiliation with the Presbyterian Church. His strong sense of duty and moral righteousness, coupled with great devotion to the education of cadets, earned for him the derisive title "Deacon Jackson" and comparison with Oliver Cromwell.

Upon the outbreak of the Civil War he offered his services to his state of Virginia and was ordered to bring his VMI cadets from Lexington to Richmond. Soon after, he received a commission as colonel in the state forces of Virginia and was charged with organizing volunteers into an effective Confederate army brigade, a feat that rapidly gained him fame and promotion. His untimely death only two years later cut Jackson down at the height of an increasingly successful career, leaving unanswered the question of his capacity for independent command, which his rapid rise suggests he might have achieved.

Jackson's first assignment in the Confederate cause was the small command at Harpers Ferry, Virginia (now West Virginia) where the Shenandoah River flows into the Potomac. His mission was to fortify the area and hold it if possible. When Gen. Joseph E. Johnston took over the Confederate forces in the valley, with Jackson commanding one of the brigades, Jackson withdrew to a more defensible position at Winchester.

Battle of Manassas

In July **1861** the invasion of Virginia by Federal army troops began, and Jackson's brigade moved with others of Johnston's army to unite with Gen. P.G. Beauregard on the field of Bull Run in time to meet the advance of Gen. Irvin McDowell's Federal army. It was here that he stationed his brigade in a strong line, withholding the enemy against overwhelming odds—an incident that earned him the famous sobriquet "Stonewall." The spring of **1862** found Jackson again in the Shenandoah Valley, where his diversionary tactics prevented reinforcements being sent to Federal army general George B. McClellan, who was waging the peninsular campaign against Richmond, the Confederate capital. Jackson's strategy possibly accounted for Lee's victory later in the Seven Days' Battles. Lee, then chief military adviser to Confederate president Jefferson Davis, suggested to Jackson that he use his troops to attack Federal troops in the valley and thus threaten Washington. By rapid movement, Jackson closed separately with several Federal units and defeated them. In April he struck in the mountains of western Virginia; then on May **24–25** he turned on Gen. Nathaniel P. Banks and drove him out of Winchester and back to the Potomac River.

He then quickly turned his attention to the southern end of the valley, defeating the Federals at Cross Keys, Virginia, on June **8,** and at Port Republic on the next day. Lee then brought Jackson's troops by road and railroad to Richmond to envelop the right wing of McClellan's army. But Jackson arrived a day late and his reputation lost some of its lustre, possibly because of his lack of experience in large-scale action; nevertheless, McClellan was beaten back and was ordered to evacuate the peninsula.

Lee at once joined Jackson against the Federal forces regrouping under Gen. John Pope. He sent Jackson, by a wide encircling movement to attack the rear of Pope's forces and bring on the Second Battle of Bull Run, in which Pope was soundly beaten. Lee next crossed the Potomac for the "liberation" of Maryland. To protect Richmond, Lee detached Jackson to capture Harpers Ferry, which he did in time (September **13–15)** to rejoin Lee at Antietam. After his return to Virginia, Lee divided his army into two corps, Gen. James Longstreet commanding the first and Jackson, now a lieutenant general, the second. At Fredericksburg, Virginia, in December, Jackson was in command of the Confederate right when Federal general Ambrose E. Burnside's rash attack was easily repulsed and he was crushingly defeated.

Death

In April, Gen. Joseph Hooker, Burnside's successor, attempted to turn the Confederate position on the Rappahannock River, south of Washington. There the seemingly invincible team of Lee and Jackson made its boldest move. Leaving a small detachment to meet Federal troops on the Rappahannock, Lee moved his main body, including Jackson's corps, to meet Hooker's threatened envelopment in the woods of Chancellorsville. He then divided his army again, keeping only **10,000** men to demonstrate against Hooker's front, and he sent Jackson to move secretly around Hooker's right with his entire corps. The manoeuvre was completely successful. On the evening of May **2,** Jackson rolled up the flank of, the unsuspecting Federal forces. Then, in the moment of victory, tragedy struck. Jackson, who had ridden forward to organize the pursuit, was accidentally shot down by his own men when he returned at dusk and was seriously, but not mortally, wounded. Although his left arm was amputated successfully, pneumonia set in and he died on May **10, 1863,** at Guiney's Station. Lee could not replace him; for while Jackson had lost his left arm, Lee had, indeed, lost his right arm.

That Jackson was the ablest of Gen. Robert E. Lee's generals is rarely questioned. The qualities of the two men complemented each other, and Jackson cooperated most effectively. In him were combined a deep religious fervour and a fiercely aggressive fighting spirit. He was a stern disciplinarian, but his subordinates and his men trusted him and fought well under his leadership. A master of rapid movement and surprise tactics, he kept his intentions sometimes so veiled in secrecy that often his own officers did not fully know his plans until they were ordered to strike.

BIBLIOGRAPHY. DOUGLAS S. FREEMAN, *Lee's Lieutenants: A Study* in Command, 4 vol. (1942–46), especially vol. 2, which deals with Jackson's potential for independent command—the most exhaustive use of all available sources; F.E.

VANDIVER, *The Mighty Stonewall* (1957), the best treatment of his early life and the most objective and penetrating analysis of Jackson the man, with less detail of military activities; G.F.R. HENDERSON, *Stonewall Jackson and the American Civil War*, 2 vol. (1898; reprinted in 1 vol., 1937), the classic life of Jackson and story of the war in Virginia until Freeman's work appeared; H.K. DOUGLAS, *I Rode with Stonewall* (1940), an uncritical personal view of Jackson's youngest staff officer, who kept notes and wrote this account 30 years later.

(C.W.T.)

Ja'far ibn Muhammad

Ja'far ibn Muhammad, often called Ja'far aṣ-Ṣādiq (the Trustworthy), was the sixth *imdm*, or spiritual head, of the Shi'ah (those Muslims who regard ʿAlī, Muhammad's son-in-law, and his descendants as divinely appointed successors of the Prophet Muhammad), and he was the last to be recognized as *imām* by all the Shi'ite sects. He was a leading authority on Hadith (sayings attributed to Muhammad), much quoted by the orthodox (Sunnī) as well as Shi'i Muslims. He is also credited with having had an outstanding knowledge of the physical and the occult sciences.

Life

Born in Medina (Arabia) in AH 80 (AD 699–700) or AH 83 (AD 702–703), he was the son of Muhammad al-Bāqir, the fifth *imdm*, and great-grandson of the fourth caliph, ʿAlī, who is considered to have been the first *imām* and founder of Shi'ism (Arabic Shi^{c}at ʿAlī, the party of ʿAlī). On his mother's side, Ja'far was descended from the first caliph, Abii Bakr, whom Shi'ites usually consider a usurper. This may explain why he would never tolerate criticism of the first two caliphs.

There is some doubt whether the Shi'ite conception of an infallible religious leader, or *imdm*, was really formulated before the 10th century, except possibly in some sort of "underground movement." But the Shi'ites certainly felt that the political leadership of Islām exercised by the caliph should belong to the direct descendants of ʿAlī. Moreover, this political leadership was not clearly separated from religious leadership; and to the end of the Umayyad regime, the caliphs sometimes preached in the mosque, using the sermon to reinforce their authority. Consequently, after his father's death, sometime between 731 and 743, Ja'far became a possible claimant to the caliphate and a potential danger to the Umayyads.

The Umayyad regime was already threatened by other hostile elements, including the Persians, who resented Arab domination and the spread of Shi'ism throughout Iran through a mixture of religious, racial, and political motives compounded the opposition. The successful revolt of 749–750 that overthrew the Umayyads, however, was under the leadership of the ʿAbbāsid family, descended from one of the Prophet's uncles; and they, not the family of ʿAlī, founded the new ruling dynasty.

The new caliphs were, understandably, worried about Ja'far. Al-Manṣūr (ruled 754–775) wanted him in his new capital, Baghdad, where he could keep an eye on him. Jaʿfar preferred to stay in Medina and reportedly justified this by quoting a saying he ascribed to the Prophet that, though the man who leaves home to make a career may achieve success, he who remains at home will live longer! After the defeat and death of the ʿAlid rebel Muhammad ibn ʿAbd Allāh in 762, however, he thought it prudent to obey the Caliph's summons to Baghdad. After a short stay, however, he convinced al-Manṣūr that he was no threat and was allowed to return to Medina, where he died in 765.

Character, ideas, and significance

A just assessment of Ja'far is made difficult by later Shi'ite accounts, which depict every *imdm* as a sort of superman. He undoubtedly was both politically astute and intellectually gifted, keeping out of politics and not openly claiming the imamate. A pious family man, Ja'far had ten children—seven by two wives and three by concubines. He entertained generously, gathering around him learned pupils including Abii Hanifah and Mālik ibn Anas, founders of two of the four recognized Islāmic legal schools, the Hanafite and Mālikite; and Wāṣil ibn ʿAtaʾ, founder of the Mu'tazilite school—which advocated that grave sinners could not be judged to be either believers or nonbelievers and used methods borrowed from Hellenistic philosophy. Equally famous was Jābir ibn Hayyān, the alchemist known in Europe as Geber, who credited Ja'far with many of his scientific ideas and indeed suggested that some of his works are little more than records of Jaʿfar's teaching or summaries of hundreds of monographs written by him. One scholar has made a convincing attempt to show that the bulk of the works ascribed to Jābir were actually written some centuries later. As to the manuscripts of half a dozen religious works bearing Jaʿfar's name, scholars generally regard them as spurious. It seems likely that he was a teacher who left writing to others.

Various Muslim writers have ascribed three fundamental religious ideas to him. First, he adopted a middle road about predestination, asserting that God decreed some things absolutely but left others to human agency—a compromise that was widely adopted. Second, in the science of Hadith, he proclaimed the principle that what was contrary to the Qurʾān (Islāmic scripture) should be rejected, whatever other evidence might support it. Third, he described Muhammad's prophetic mission as a ray of light, created before Adam and passed on from Muhammad to his descendants.

Shi'ite divisions date from Jaʿfar's death. His eldest son, Ismāʿīl, predeceased him, but the "Seveners," represented today chiefly by the Ismāʿīlīs (followers of Ismāʿīl)—argued that Ismāʿīl merely disappeared and would reappear one day. Three other sons also claimed the imamate; of these, Mūsā al-Kāẓim gained widest recognition. Shi'ite sects not recognizing Ismāʿīl are mostly known as "Twelvers"; they trace the succession from Ja'far to the 12th *imām*, who disappeared and is expected to return at the Last Judgment.

BIBLIOGRAPHY. D.M. DONALDSON, *The Shi'ite Religion* (1933), contains the only biographical account of any length on the subject. FAZLUR RAHMAN, *Islam* (1966), is an excellent, general study of Islām that puts Shi'ism and Ja'far ibn Muhammad in proper perspective—written in clear, simple language.

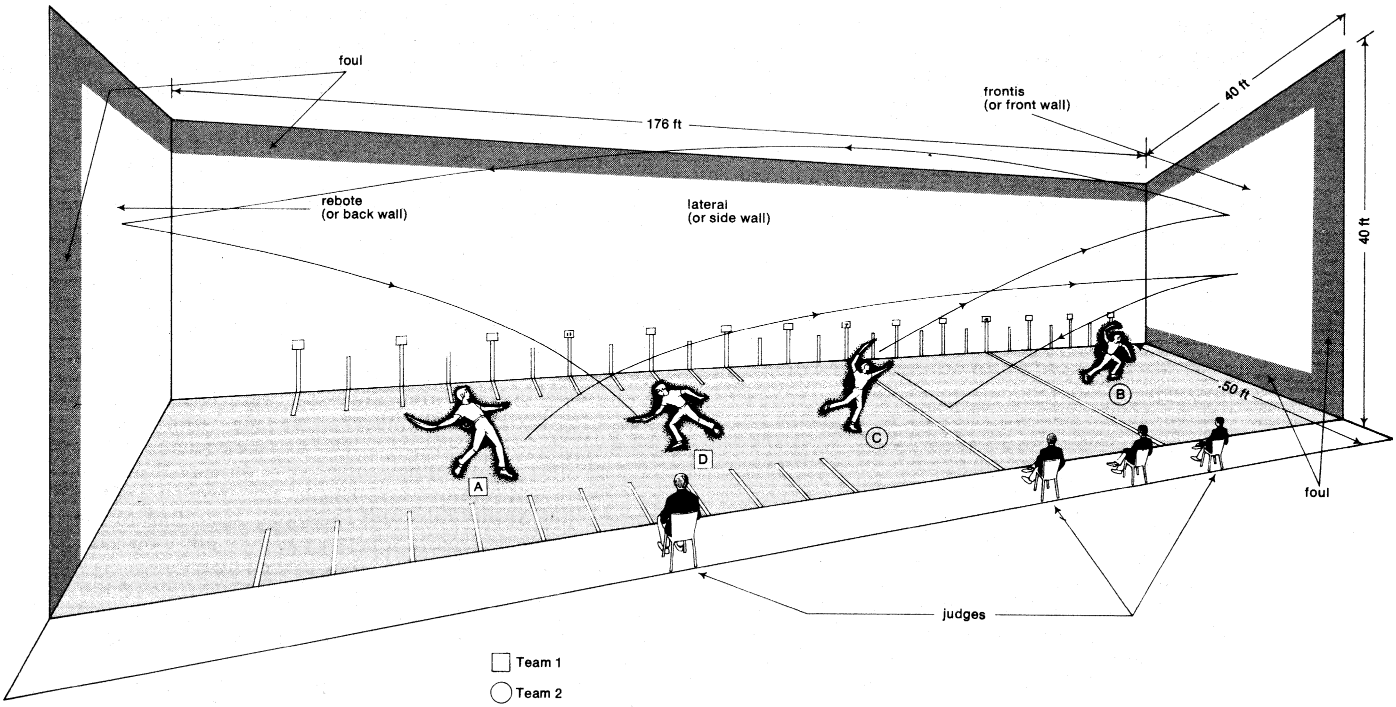
(J.A.H.)

Jai Alai

Jai alai is a ball game of Basque origin played in a three-walled court with a hard rubber ball that is caught and thrown with a cesta, a long, curved wicker scoop strapped to one arm. Called *pelota vasca* in Spain, the Western Hemisphere name jai alai (Basque "merry festival") was given the game when it was imported to Cuba in 1900. It is generally accepted that jai alai developed from a form of handball, and it is presumed that the availability of a lively ball made possible by the introduction of rubber to Europe from South America permitted players to speed up the game. The next step, it is thought, was the introduction of the *guante*, a simple leather glove worn on the right hand, which in turn led to the use of a flat wooden bat, or *pala*. A tapestry by Goya (1746–1828) in the Prado Museum, Madrid, called "Juego de Pelota," depicts such a bat in use on a one-walled court. Later the *guante* developed into a catching and throwing device leading finally to the evolution of the cesta, or wicker basket, at first a short implement but now about 2½ feet (76 centimetres) long, gracefully curved and efficient, with which the player can catch the ball and hurl it with tremendous power and speed. Each cesta is custom made of Pyrenees Mountain reeds woven over a light, ribbed frame of Spanish chestnut; a leather glove sewn to the outside holds the player's hand securely. The *pelota* (Spanish "ball") is a little smaller than a baseball and harder and heavier than a golf ball. It is made of hand-wound virgin rubber with a few final turns of linen or nylon thread and covered with two layers of hardened goat skin, the outer layer of which can be replaced. In play, speeds of 150 miles per hour are not uncommon. Professional players have worn protective helmets in the United States since 1967.

History and development of the game

The modern three-walled playing court, or cancha, averages about 175 feet long by 50 feet wide and is 40 feet or more high. The walls and the floor are made of special high-impact material to withstand the pounding of the ball.



Jai alai court.
Adapted from F. Menke, *Encyclopedia of Sports*

The spectators sit in tiers along the open side with the front of the court to their right, the side wall directly in front of them, and the back wall to their left. The entire plant is the fronton; some Basque frontons date from as early as 1785. The game is played professionally in ten frontons in Spain: five in the Basque country, of which the one in Guernica is the finest; two in Barcelona; and one each in Palma de Mallorca, Zaragoza, and Madrid. It is also played in southern France and in Milan, Italy. In the Western Hemisphere it is popular in Mexico, where there are two frontons in Mexico City and one each in Acapulco and Tijuana. It was played in the United States in the 1920s and 1930s, first in Chicago and then in New Orleans, but in both places it was withdrawn after the gambling laws were changed to ban betting on the game; it was demonstrated without wagering in New York in 1938. The Miami fronton, with 5,100 seats, the world's largest, was the only one in the United States when Florida adopted a law in 1935 permitting pari-mutuel wagering on the sport. Since that time professional courts have been established at Tampa, West Palm Beach, Daytona Beach, Orlando, and Dania. The game formerly was popular in Cuba and there are two frontons in the Philippines. Except for the United States, where there are only two frontons for amateurs, it is a popular amateur sport wherever it is played professionally.

First-class players traditionally have come from the Basque country. Outstanding contemporaries include Juan Cruz Bustinduy (playing name Juaristi) and Francisco Asis in the front court and Ramon Soroa and Tomás Cortajarena in the back court. Most players reach their top form in their late 20s or early 30s but one of the greatest of all time was Erdoza Menor who played until he was in his 50s, dropping dead on the court of a heart attack. Management of the fronton in Miami supports training schools in Spain for the development of young players, the one at Guernica having graduated many of the world's greatest. There is also a school at Miami where promising young amateurs may receive four or five years of training for a professional career.

Principles of play

The principles of the game are basically those of handball and are very simple. The ball is served against the front wall and must land in a designated serving zone; the opposing player in the case of singles, or one of the opposing partners in doubles games, must catch and return the ball before it touches the floor more than once. The ball must be caught and thrown in one continuous motion. The object is to bounce the ball off the front wall with such speed and English (spin) that the opposition cannot return it and loses the point. Play continues until the ball is missed or goes out of bounds. The ball is out of bounds if it strikes the area clearly marked in red around the front wall, strikes the overhead screen above the court or any other area marked in red or outside the foul line. If a player stops his throw because another player is in front of him, interference may be called, and the point will be played over.

The basic game played outside the United States, *partidos,* is a match singles or doubles game to 10 to 40 points. Betting is on the eventual outcome of the game, at any time during the game. As the points fluctuate, so does the spread of the odds. The American game is adapted to the pari-mutuel system. Six to eight one- or two-man teams compete for five to seven points. Two players or teams play for one point, the losing side retiring from the court and the winning side continuing to play until it loses a point and is retired or wins enough points to win the game; playoffs determine second (place) and third (show) positions. Betting is the same as at horse races, each player being identified by a number that is called his "post position." Under the American Qualifying Point System eight post positions (players or teams) play for five points. Play follows the usual elimination system until three positions have made three points each; the three then play off for win, place, and show places. The popular ***quiniela*** wager, in which the bettor picks two players or teams to finish first and second, in either order, was originated for jai alai wagering in the 1930s by Richard I. Berenson (1893–1967), former president of the Miami fronton. The *quiniela* has since flourished at horse races. dog races, and other events throughout the world. The ***quiniela*** was followed by the *perfecta,* in which the bettor must pick first and second in that order. Horse racing's daily double, the selection of the winners in two different events, has become popular in jai alai; and a Big Q., picking the ***quiniela*** in two events, has been introduced. (L.S.B.)

BIBLIOGRAPHY. F.G. MENKE, ***The Encyclopedia of Sports, 4th*** rev. ed., pp. 662–664 (1969), and (comp.), ***All-Sports Record Book*** (1931), a brief history; E. BLAZY, ***La Pelote Basque*** (1929); P. LEGARRAIDE, "La pelote Basque," in ***Encyclopedic des Sports,*** pp. 362–373 (1961); E.T. SACHS, "Pelota," in ***Baily's Magazine of Sports and Pastimes*** (May 1906).

(R.W.He.)

Jainism

Jainism is both a philosophy and a religion of India. Along with Hinduism and Buddhism, Jainism is one of the three major religions that developed within the Indian civilization. The name Jainism derives from the term Jina ("conqueror," or "victor^{m}")—applied to the 24 great religious figures on whose examples the religion is centred—and its followers are known as Jainas or Jains. There are two principal sects: Svetiimbaras and Digambaras (see below History and background).

NATURE AND EXTENT OF JAINISM

Both Jainism and Buddhism arose in about the 6th century BC in protest against the orthodox Vedic (early Hindu) ritualistic cult of the period. Jainism was founded by Vardhamiina Mahiivira, a contemporary of Siddhārtha Gautama, the founder of Buddhism; according to Jaina traditions, however, he was but the 24th of the Jinas, or Tirthankaras (makers or builders of a crossing), and was preceded by Pārśvanātha, who died 250 years before Mahiivira's Nirvāṇa (final extinction). Pārśvanātha is historical; hardly any solid historical information regarding early doctrines and religious organization survives. Jainism continues to be a living faith in India, with more than 2,000,000 adherents, who extend over nearly all the Indian states but are concentrated mainly at present in western India and Uttar Pradesh, Mysore, Madhya Pradesh, and Mahārāshtra. Jainism has practically no following outside the country of its birth.

The core of Jaina ethics is the doctrine of ahimsd, or noninjury to all living creatures. Jainism does not espouse belief in a creator god but does divide the universe into two independent eternal categories of *jīva* and ajiva ("life" and "nonlife"). The ideal is that perfection of man's nature, to be achieved predominantly through the monastic and ascetic life. Its distinctive philosophic contribution is the doctrine of naya ("point of view") and the logical position of anekdntavdda (the "many sidedness of reality"), which holds that it is impossible to make an absolute assertion about anything. In the fields of art and architecture, languages and literature (in Priikrit, Sanskrit, Apabhramia, Tamil, Kannada, and Old Gujarati), astronomy, logic, mathematics, grammar, and dramaturgy, Jaina accomplishments are significant.

HISTORY AND BACKGROUND

Concepts of Jina and Tirthankara

Early background and traditional accounts (before 6th century AD). The terms Jina ("conqueror"), Buddha ("enlightened one"), and arhat ("deserving worship") were commonly used by both early Buddhists and Jainas to refer to those who had achieved spiritual liberation. The founders of Jainism were also called Tirthankaras (makers of a tirtha, "ford," or a place to cross over).

The 24 *Tīrthaṅkaras.* Jainas believe that 23 Tirthankaras preceded Mahāvīra in this cosmic age. Similarly in the next cosmic age, which will be an ascending order of happiness, another 24 Tirthankaras will live. Thus in the Jaina conception of time as endless, Jaina doctrine has no beginning or end. Modern scholarship, however, regards only the last two Tirthankaras, Pārśvanātha and Mahāvīra, as historical figures.

Belief in the 24 Tīrthaṅkaras evolved gradually. The *Kalpa-sūtra*, a Jain canonical text, describes the lives of Ṛṣabhanātha (first Jina), Ariṣṭaneminātha (22nd), Pārśvanātha, and Mahāvīra in detail, whereas the lives of the remaining 20 Tirthankaras are given in a standard form probably not of historical value. This portion (*Jinacarita*, "Lives of Jinas") of the above text seems to have been added and edited about the 4th century AD.

The role of Jainas in Indian society. Since, traditionally, most Jainas have been engaged in trade and commerce, wealthy and influential Jaina merchants have played important roles throughout Indian history in many fields. Jaina emphasis on ahimsd and prohibition of the eating of meat and certain other kinds of food may also have gradually influenced other Indian religions, including both Hinduism and Buddhism.

Early sectarian *and* other institutional developments. Jamāli, Mahiivira's son-in-law (according to later literature), led the first schism (nihnava) in Jainism during the Jina's lifetime. Seven more schisms followed, and out of the last, led by Śivabhūti (c. AD 80), gradually evolved the two principal sects, the Śvetāmbara ("white-robed") and the Boṭika or Digambara ("sky-clad," or naked). The split occurred 609 years after Mahiivira (in AD 83 according to the Śvetāmbaras, and in AD 80 according to the Digambaras; others give the dates AD 82 and 79), mainly over the question of whether monks should wear clothes or remain naked. A further difference was the Digambaras' belief that women cannot attain salvation.

'Sectarian schisms

Digambara traditions enumerate 28 different *ācārya*s (spiritual leaders) up to 683 years after Mahāvīra (*i.e.*, AD 156). Sometime afterward Ācārya Arhadabali convened a council of monks, and various groups (samghas) of monks were then organized.

During Mahiivira's lifetime his followers lived mainly in the ancient kingdoms of Videha, Magadha, and Anga in east India and westward as far as Kāśī (modem Vārānasi) and Kosala. Jaina influence soon spread also to Daiapura (Mandasor) and Ujjain. Bhadrabāhu I (170 years after Mahāvīra) went to Nepal, according to the Svetiimbaras, and to Sravana-Belgola in South India during 12 years of famine, according to the Digambaras.

In about the late 2nd century BC, King Khiiravela of Kaliṅga (modem Orissa state) professed Jainism, excavated Jaina caves, and set up Jaina images and memorials to monks.

Images of the Tirthankaras found at Mathura and datable to the Kuṣāṇa period either depict the Jina in a standing attitude and unclothed or, if seated in the crossed-legged posture, are sculptured in such a way that neither garments nor genitals are visible. Though the Śvetiimbara–Digambara differences had already originated in the Kuṣāṇa age, it would appear that at this time both sects worshipped nude images of Tīrthaṅkaras. The earliest known image of a Jina with a lower garment, the standing Ṛṣabhanātha discovered at Akota in Gujarāt state, has been dated to the latter part of the 5th century AD, the age of the last council at Valabhi. This suggests that the Valabhi council marked the final separation of the two sects.

King Samprati, the grandson of the Mauryan emperor Aioka, is said to have given great patronage to Jainism by facilitating Jaina travel in South India beyond the Deccan. Tamil literary classics such as *Maṇimekhalai* and *Silappadikaram* attest to the high degree of Jaina influence in South India itself.

Early teachers

In about the 1st century BC the renowned Jaina teacher Kālakācārya (the "black teacher") invited the Scythian tribe known as the Śakas to western India and Ujjain (possibly from Śakasthāna in Sindh) to overthrow King Gandabhilla, abductor of his sister, who was a Jaina nun. The same Kālakācārya is said to have travelled as far abroad as Suvarṇadvīpa in Southeast Asia (probably Annam in Vietnam) and is credited with having written the *Mūlaprathamānuyoga* texts, now lost, which were re-edited by Arya Raksita in about the 2nd century AD.

Arya Vajra was another renowned teacher who appears to have supported *caityavāsa* (dwelling by monks in temples), a practice that later led to corruption among the Śvetāmbaras. An inscription of about the 1st century AD at Son Bhaṇḍāra cave (Rājgīr, Bihār state) shows that Ācārya Vaira (Vajra) excavated two caves that were suitable for dwellings of monks and in which Jaina images were installed for worship.

By the Gupta age (AD 320–600), Jainas had become stronger in central and western India than in their homeland, Magadha; but they were also patronized by the Gupta emperors of Magadha.

Further historical developments. From the 5th century to the 12th, the Ganga, Kadamba, Caulukya, and Rāṣṭrakūṭa dynasties of South India accorded royal patronage to Jainism. Many Jaina poets of great repute flourished under patronage of Rāṣṭrakūṭās of Mānyakheṭa (Malkhed). Virasena wrote his monumental commentaries on *Ṣaṭkhaṇḍāgama* under Jagattuṅga and his successor; Jinasena and Gunabhadra composed the great epic *Mahāpurāna* at the time of King Amoghavarṣa, a follower of

Jainism; and Mahāvīrācārya wrote a work on mathematics. King Amoghavarṣa was himself the author of Ratnamdlikii, a work that became popular with all sects. Puṣpadanta composed his famous *Mahāpurāṇa* in Apabhraṃśa under the patronage of ministers of Kṛṣṇa III.

From about 1100 Jainism gained ascendancy in the court of the Caulukyas of Gujarāt, especially during the reigns of Siddharāja and Kumārapāla, both of whom held in high respect their contemporary, Hemacandra Aciirya, a versatile and prolific Jaina author and monk.

From about the 7th century the Śvetāmbara order gained strength in Gujarāt and Rājasthān due to royal patronage and the activities of such great monks as Haribhadra Siiri, Udyotana Siiri, Vādin Deva Siiri, Abhayadeva Siiri, Hemacandra Aciirya, and others. The Śvetāmbara order of monks divided into several gacchas (subgroups) during the 11th and later centuries. Hiravijaya Sūri and his pupils of the Tapāgaccha received special respect from the Mughal emperors Akbar and Jahāngīr.

The Sthānakavāsīs

In about the 16th century Loṅkāsāha started a subsect of Śvetāmbaras known as Sthānakavāsīs in western India and claimed that image worship is not sanctioned by the scriptural *Āgama*s (canons); furthermore, he denied the authority of certain canonical works of later origin that do refer to image worship. In the 18th century Aciirya Bhikṣu organized Teriipantha, a subsect of Loṅkāsāha's *ga*ccha. Tulasi Ganin, the present leader of Teriipantha, has organized a group known as the Anuvrata Samgha. In the 16th century a Digambara named Tāraṇuasvāmin organized the Tiiranapantha sect, which repudiated image worship.

Of 84 gacchas of Śvetāmbara monks and lay followers that developed after the 7th–8th century AD, few have survived. The most important today are the Kharatara (mainly in Rājasthān state), Tapā, and Añcala gacchas. The Digambaras are also divided into Vīsāpanthīs and Terāpanthīs (founded by Banārasīdāsa in 1626). Other important Digambara *saṃgha*s, already noted, are Nandi, Kāṣṭhā, Drāviḍa, and Sena. In modern times the great Svetiimbara leader was Ācārya Vijayānanda Siiri, called Ātmārāmajī, whose great-grandpupil, Muni Punyavijaya, reorganized Jaina bhandaras (manuscript libraries) and planned critical editions of all the Jaina canonical texts on modern principles. He published three texts and was preparing others but died in 1971. Malvania, his co-editor, was expected to edit the remaining texts.

MYTHOLOGY

Lesser gods are classified into four main groups: *bhavanavāsī*s (gods of the house), vyaniaras (intermediaries), *jyotiṣka*s (luminaries), and vaimdnikas (astral gods). These are each subdivided into several groups with indras (chiefs) at the head, *lokapāla*s (guardians of the cardinal points of the universe), armies of gods, and queens of indras. Vyantaras, for example, are divided into *yakṣa*s (vegetation spirits), *bhūta*s (ghosts), *piśāca*s (fiends), *rāksaṣa*s (demonical beings), *kinnara*s (half-horse, half-human), gandharvas (celestial musicians), and others. Such deities also played an important role in ancient Indian folk worship; the Jaina, Buddhist, and Hindu religions assimilated them into their pantheons and rituals.

Besides these, certain other gods and goddesses are mentioned in various Jaina texts, including four gatekeepers of the rampart of the Jambudvipa and four goddesses—Jayā, Vijayā, Jayantā, and Aparājitā. A list of 64 *dikkumārī*s (maidens of the directions), who act as nurses when the Tirthankara is born, includes several goddesses that suggest Hindu influence or borrowing from some common ancient Indian heritage.

These deities are, however, assigned a position subordinate to the Tirthankaras and other liberated souls (siddhas) who are called devddhidevas (Lords of Gods). Next in order to the Tīrthaṅkaras (arhats) and siddhas are the Jaina ascetic souls called *ācārya*s (leaders of groups of monks), *upādhyāya*s (readers who teach sacred texts), and sddhus (monks in general). These five constitute the *pañca-parameṣṭhin*s (five chief divinities). The 24 Tirthankaras along with certain other souls including 12 cakravartins (world conquerors), 9 *vāsudeva*s (counterparts of Hindu Kṛṣṇa [Krishna] Vāsudeva), 9 *bala*devas (counterparts of Hindu Balarāma), constitute a list of 54 *śalākāpuruṣa*s or *mahāpuruṣa*s (great souls) to which are later added 9 prati-vdsudevas (enemies of *vāsu*devas), making a grand total of 63. Jaina texts often deal with the lives of these *śalākāpuruṣa*s. Other figures such as nine ndradas, 11 rudras, and 24 *kāmadeva*s (gods of love) show Hindu influences. Bāhubali, the son of the first Tirthankara, Ṛṣabhanātha, is said to be the first of the *kāmadeva*s (see also HINDU MYTHOLOGY).

DOCTRINES OF JAINISM

The notion of the wheel of time

Time and the universe. Time is eternal and formless. It is conceived as a wheel with 12 spokes called *ārās* or ages: six make an ascending cycle, the other six a descending one. The ascending cycle (*utsarpiṇī*) shows man's progress in knowledge, age, stature, and happiness in each age, while the descending cycle (*avasarpiṇī*) reveals a gradual deterioration in the state of everything. The two cycles joined together make one rotation of the wheel of time, or one halpa (20×10^{14} *sāgaropama* years).

The world is infinite. It was never created. Its constituent element—the six substances (dravyas) or five magnitudes (astikdyas) together with time—are soul, matter, time, space, and the principles of motion and rest. These are eternal and indestructible but their conditions change constantly.

Space (*ākāśa*), all pervasive and formless, provides accommodation to all objects of the universe and is divided into the space of the universe (*lokākāśa*) and that of the non-universe (*a-lokākāśa*), the latter having no substance in it. The universe is conceived as a figure with legs apart, arms akimbo, slender, circular, and flat at the waist. The whole is enveloped in three atmospheres called *vāta-valaya*s (wind-shields) and described as dense and thin.

Through the centre of the universe runs the region of mobile souls (trasa-jiva *nāḍī*) in which all living beings, including men, animals, gods, and devils, live. At the lowermost point of this region is the seventh hell of most acute pains and tortures. Above it is the sixth hell and so on until the first and mildest hell is reached. In the lower world (adholoka) the uppermost region, called *ratna*prabhd, is also the seat of gods of two classes, *bhavanovāsī*s and vyantaras or *vāṇamantara*s.

Above the central region (*madhyaloka*) is the upper world (*ūrdhvaloka*) of two parts, containing 16 heavens and 14 celestial regions. On top of the universe (*lokākāśa*) is the *siddha-śīlā*, a crescent-shaped flexible slab or place of large dimensions on which liberated souls live after leaving earthly bodies.

Jiva and ajiva. Jainism is dualistic. It sees reality as constituted of two separate eternal entities: soul or living substance (*jīva*) and nonsoul, or nonliving substance (ajiva). The *ajīva* or matter is made up of five basic factors (*astikāya*s): motion (*dharma*), rest (adharma), space (*ākāśa*), gross matter (pudgala), and time (*kāla*). All these are eternal (without beginning); all but the souls (jiva) are without life; all but gross matter (*pud*gala) are noncorporeal. Motion and rest do not exist by themselves, are prerequisites of the other substances, with respect to matter, and effect the conditions of motion and rest. These five substances are accompanied by time, which is eternal and one.

The essential characteristic of the souls is consciousness or mental function (*cetanā*). Nonliving substance (ajiva) causes souls to assume bodies and become involved in corporeal functions.

Kinds of life

Jiva or the life principle, in its pure state, possesses the qualities of unending perception, limitless knowledge, infinite bliss, and infinite power. The souls, infinite in number, are divisible in their embodied state, into two main classes, immobile (*sthāvara*) and mobile (trasa), according to the number of sense organs they possess. The first group consists of immeasurably small particles of earth, water, fire, and air, as also the vegetable kingdom, and has only one sense, that of touch; the second group includes bodies that have two, three, four, or five sense organs. The Jainas believe that the four elements (earth, water, air, and fire) also are animated by souls. Besides, the universe is full of an infinite number of minute beings, *nigodas,* which are slowly evolving.

The soul is formless and cannot be perceived by senses. A soul is not all-pervasive but can, by contraction or expansion, occupy various proportions of space. Like the light of a lamp in a small or a large room, it can fill both the smaller and larger bodies it occupies. But the soul is not identical with the body.

Matter (pudgala) has the characteristics of touch, taste, smell, and colour. Its essential characteristic is lack of consciousness. The smallest unit of matter is the atom (*paramāṇu*), which is eternal and indivisible.

The elements of nature--earth, water, fire, and air—are gross manifestations of matter, the subtlest form of which is the atom, paramdnu. Heat, light, and shade are forms of fine matter.

Dharma, or (the medium of) motion, though filling the entire universe, is imperceptible; adharma is similarly the medium of rest.

Ākāśa, or space, another ajiva substance, is nonmaterial and infinite and provides space for the existence of all other entities. Time (*kāla*), also an *ajīva* substance, consists of innumerable eternal and indivisible particles of "noncorporeal substance" that never mix with one another but fill the whole universe.

Karman. All phenomena are said to be linked together in a universal chain of cause and effect. Every event has a definite cause behind it. By nature each soul is pure, possessing infinite knowledge, bliss, and power. But these faculties are restricted from the beginningless time by foreign matter coming in contact with the soul. Fine foreign matter producing the chain of cause and effect, of birth and death, is karman, a substance, a fine atomic particle (*sūkṣma-pudgala-paramāṇu*), and not a process as in Hinduism. To be free from the shackles of karman, one must stop the influx (samvara) of new karmans and eliminate the acquired ones.

Acquired karmans can be annihilated through a process called *nirjarā,* which consists of fasting, not eating certain kinds of food, control over taste, resorting to lonely places, mortifications of the body, atonement and expiation for sins, modesty, service, study, meditation, and renunciation of the ego.

*Guṇa-*sthdnas: stages of the soul's growth

A soul passes through various stages of spiritual development before becoming free from all karman bondages. These stages of development (gunasthdnas) involve progressive manifestations of the innate faculties of knowledge and power and ate accompanied by decreasing sinfulness and increasing purity.

Due to the connection of *jīva* with karmic matter, the souls obtain different kinds of bodies. These embodied souls bear different colours or tints (*leśyā*), varying according to the merits or demerits of the particular being. This doctrine of *leśyā*s, peculiar to Jainism, seems to have been borrowed from the Ājīvika doctrine of six classes (*jātis*) of bodies, expounded by Gośāla Maṅkhaliputta, in the age of Mahāvīra. The six *leśyā*s in Jainism are, in the ascending order of man's spiritual progress: black, blue, gray, fiery red, lotus-pink (or yellow), and white.

Theories of knowledge as applied to salvation. In Jaina thought, four stages of perception-observation, will to recognize, determination, and impression—lead to a subjective cognition (*mati-jñāna*), the first of five kinds of knowledge (jtidna). The second kind of knowledge is *śruta-jñāna*, derived from both scriptures and general information. Both *mati-* and *śruta-jñāna* are mediate cognition (*parokṣa-jñāna*), based on external conditions. Immediate knowledge (*pratyakṣa*) is of three kinds: avadhi (supersensory perception), *manaḥ-paryāya* (thought reading), and kevala (the final cognition revealing past, present, and future), a stage of omniscience accompanied by freedom from the karmans obstructing knowledge and by direct experience of the soul's pure form unblemished by matter.

Knowledge is said to be the innate attribute of the soul, whose chief characteristic is *cetanā* (consciousness), which consists of both knowledge (*jñāna*) and intuition (*darśana*). But worldly souls obscured by the veil produced by destructive karmans do not enjoy infinite knowledge, power, and bliss.

Yoga (physical and meditative discipline), according to Jainism, is the cause of *mokṣa*, or liberation. Yoga consists of knowledge of reality as it is (*jñāna*), faith in teachings of the Tīrthaṅkaras (*śraddhā*), and character or cessation from doing evil—right conduct leading to release from karman bondage (cdritra). Yoga is thus equated with the three jewels (ratna-traya) of right belief, right knowledge, and right conduct (see INDIAN PHILOSOPHY; LOGIC, HISTORY OF).

Concept of *ahiṃsā*

Jaina ethics. Jainism advocates an attitude of equality and designates the whole pattern of religious conduct and philosophical thought that helps to develop such an attitude as bambhacerdim (brahmacarydni, conduct leading to the Brahma or the pure soul) religious life. This attitude finds expression in the principle of nonviolence or noninjury (ahimsd). Even in the domain of philosophical thought, it is this principle of equality and noninjury to others (both psychologically and physically) that has given rise to the attitude of nonabsolutism in thought (*anekāntavāda* or syddvdda), which respects opposite viewpoints regarding the same object.

Two separate courses of conduct are laid down for the ascetics and the householders (laymen). In both cases, the code of morals is based on the doctrine of *ahiṃsā*. Since thought is considered the father of action, violence in thought (*bhāva-hiṃsā*) merely precedes physical injury (*dravya-hiṃsā*). Violence in thought is then the greater and more subtle form of violence, arising from passionate ideas of attachment and aversion, due to men's negligence (pramdda) in their behaviour. Jainism enjoins avoidance of all forms of injury (himsd), physical or otherwise, whether committed by mind, body, or speech.

The three jewels. Right faith (*samyag-darśana*), right knowledge (*samyag-jñāna*), and right conduct (*samyak-*cdritra) are the three jewels (ratna-traya). Right faith includes freedom from doubts and desires, steadfastness, brotherhood toward fellow believers, and the propagation of Jaina principles among others (prabhdvand).

Right faith leads to calmness or tranquillity; detachment; kindness; renunciation of pride of birth, beauty of form, wealth, scholarship, prowess, and fame. Right faith leads to perfection only when followed by right conduct.

There can be no virtuous conduct without right knowledge, which consists of clear distinction between the self and the nonself. Knowledge of scriptures is distinguished from inner knowledge. Knowledge without faith and conduct is futile. Without purification of mind, all austerities are mere body torture. Right conduct is thus spontaneous, not a forced mechanical quality. Attainment of right conduct is a gradual process, and a householder can observe only partial self-control; when he becomes a monk, he is further able to observe more comprehensive rules of conduct.

Consequences of doctrines on social practices of Jainas. Emphasis on *ahiṃsā* is responsible for Jains' avoiding such professions as agriculture and the sale and manufacture of arms, instruments, and intoxicants. Jaina involvement in trade and commerce has resulted in a fairly high percentage of literacy among them. Belief in the equality of all souls allows Jains, unlike Hindus, to eat with anyone. Admission to the ascetic orders was open to everyone irrespective of caste, class, or nationality.

Jain castes

Perhaps during the medieval period, several castes arose among the Jains as among the Hindus. It must however, be noted that the caste system is not followed by Jaina monks but by the laity only. Some caste names (mainly occupational) are common with Hindus, some are named after places, while others are exclusively Jain. Though features of the Hindu caste system such as hierarchy also appear in Jaina castes, social differentiation is not nearly so marked. Some castes are common to both Śvetāmbaras and Digambaras, while others are exclusive to one or the other (see also CASTE SYSTEMS).

RITUAL PRACTICES AND RELIGIOUS AND SOCIAL INSTITUTIONS

The monastic orders and their practices. Śvetāmbaras acknowledge two classes of monks: (1) Jina-kalpins, who wander naked and use the hollows of their palms as alms bowls, and (2) sthavira-kalpins, who retain minimal possessions such as a robe, an alms bowl, *rayoharaṇa* (sweeping duster), and mukha-vastrikii (a piece of cloth held over the mouth to strain out insect life). A monk must obey "the great vows" (mahdvratas) to avoid injuring any form of life, lying, stealing, sexual intercourse, possessions, and taking meals at night (which increases the possibility of injuring minute insects and small creatures); and his conduct is regulated in all details by specific ordinances. Monks are expected to suffer with equanimity such hardships as those imposed by travel, theft, famine, and political disturbances. Exceptions are allowed in emergencies, since a monk who survives a calamity can purify himself by confession and by practicing even more rigorous austerities.

A junior monk must pay full respect to elders. An *uvajjhāya* (upddhyiiya) is the chief reader and instnictor of a group of monks. An dyariya (dcrirya), or *sūri*, is the head of a group of monks. A *gaṇin,* a monk of a few years' standing, is next in order of rank.

Among the Digambaras, a full-fledged monk remains naked, though there are lower grade monks who wear a loincloth and keep with them one piece of cloth not more than one and a half yards long. Digambara monks use a rayoharana (duster) of peacock feathers, live in monasteries or apart from human habitations, beg and eat food only once a day, and do not use an alms bowl.

Eight essentials noted for the conduct of monks include three guptis—protection of mind, speech, and action—and five kinds of vigilance over behaviour (samiti). The six *āvaśyaka*s (obligations) include equanimity (*sāmāyika*); praise of 24 Jinas (*caturviṃśatistavana*); obeisance to Jinas, teachers, and scriptures (vandana); atonement (pratikramana); resolution to avoid sinful activities (*pratyākhyāna*); and meditation in a straight but not rigid posture, without moving (kdyotsarga).

Kinds of vows

Religious disciplines of laity. The life of a lay votary or a householder is only a preparatory stage to the rigorous life of an ascetic. The lay votary (*śrāvaka*) is enjoined to observe eight *mūla-guṇa*s (primary qualities, variously listed and usually including the avoidance of meat, wine, honey, fruits, roots, and night eating), and 12 vows: five *aṇuvrata*s ("little vows"), three gunavratas, and four *śīkṣāvrata*s. The *aṇuvrata*s are primary vows of abstinence from gross violence, gross falsehood, and gross stealing; contentment with one's own wife; and limitation of possessions.

The *guṇavrata*s and *śīkṣāvrata*s (collectively called *sila*-vratas) are supplementary vows to strengthen and protect the five anuvratas. They involve avoidance of unnecessary travel, harmful activities, and the pursuit of pleasure; fasting and diet control; offering of gifts and service to monks, poor people, and fellow believers; and voluntary death if observance of vows becomes impossible.

The *sāmāyika* (equanimity), a very important obligatory observance for both monks and laity, is for the attainment of equanimity of mind through atonement and resolve not to slacken in spiritual practice, especially cessation of blameworthy activities and concentration on the blameless ones. It may be performed in one's own house, in a temple, in a *poṣadha-śālā* (fasting hall), or before a monk.

Eleven *pratimā*s, or stages of spiritual progress of a householder, are given. The word pratimd (literally, "statue") is also used in specific sense to designate the meditation posture of standing at ease and avoidance of movement, night eating, and sexual pleasure. The *pratimās* represent partly a theoretical graduation of spiritual level and partly the possibility of choice. Medieval writers conceive pratimii as a regular progressing series, a ladder leading to higher and higher stages. The last two stages imply renunciation of the world and manhood.

Sacred times and places. Festivals and fairs. The principal Jaina festivals are connected with the five auspicious events (paiica-kalyiinakas) in the life of each Tīrthaṅkara. Especially popular are those commemorating the lives of Mahāvīra, Ṛṣabhanātha, and Pārśvanātha; these mark the occasions of (1) descent in the mother's womb (*garbhādhāna,* cyavana), (2) birth (janma), (3) renunciation (*dīkṣā*), (4) attainment of omniscience (*kevala-jñāna*), and (5) death and final emancipation of a Jina.

The festival of Pajjuṣaṇa

The most popular festival of the Jainas is Paryuṣaṇa, or Pajjusana, in the month of Bhiidrapada (August–September). Pajjuṣaṇa literally means (1) pacification by forgiving and service with wholehearted effort and devotion and (2) staying at one place during the monsoon. On the last day of the festival, the Jainas distribute alms to the poor and take out a Jina image in a car that moves through the streets in a procession. During the festival, annual pratikramana, or confession, is performed to remove all old ill-feelings about conscious or unconscious misdeed.

Twice a year, for nine days, during the months of Caitra (March–April) and Āśvina (September–October), a fasting ceremony known as *olī* is observed. These are also the *aṣṭāhnika* (eight days) festivals corresponding to mythical celestial worship of Jina images.

On the full-moon day in the month of Kiirttika (October–November), on the same day as the Hindus celebrate Dīwālī (the festival of lights), Jainas also commemorate the Nirvāṇa of Mahiivira by lighting lamps. Five days later is Jñāna-Pañcamī, which the Jainas celebrate with temple worship and especially with worship of the scriptures in manuscript form.

The full-moon day of Caitra (March–April) is celebrated with great pomp by a large number of pilgrims on Mt. Satrunjaya, sacred to the first Tīrthaṅkara. Mahiivira Jayanti, marking the birth date of Mahiivira, is celebrated all over India by the Jainas.

In common with Hindus, Jainas celebrate Holi (the festival of spring), the Makara-Saṃkrānti (January 14, sun's entry into Capricorn), and the Navariitri ("nine nights"). In South India Jainas also often celebrate such Hindu festivals as Poxigal, Kiirttika, Yugādi, Gauri, and Navarātrī (see also ***HINDUISM***).

Pilgrimages and shrines. The erection of shrines and donation of copies of religious manuscripts are regarded as pious acts. Almost every town or village inhabited by Jainas has at least one Jaina shrine; some of them have become places of pilgrimage. Lists of such Jaina shrines have been composed, and the most noteworthy shrines are offered adoration in daily worship (caitya-vandana).

Places of pilgrimage were created at sites marking the principal events (kalyanakas) in the lives of Tīrthaṅkaras. Sameta Sikhara or Parasnath Hill in Bihiir, Satrunjaya and Girnar hills in Kāthiāwār, are among such famous ancient places of pilgrimage. Riijgir in Bihiir is an old temple site associated with Mahāvīra's travels, and Sravana-Belgola in Mysore is especially famous for the 57-foot-high stone statue of Bāhubali (called Gommateśvara) installed in the 10th century.

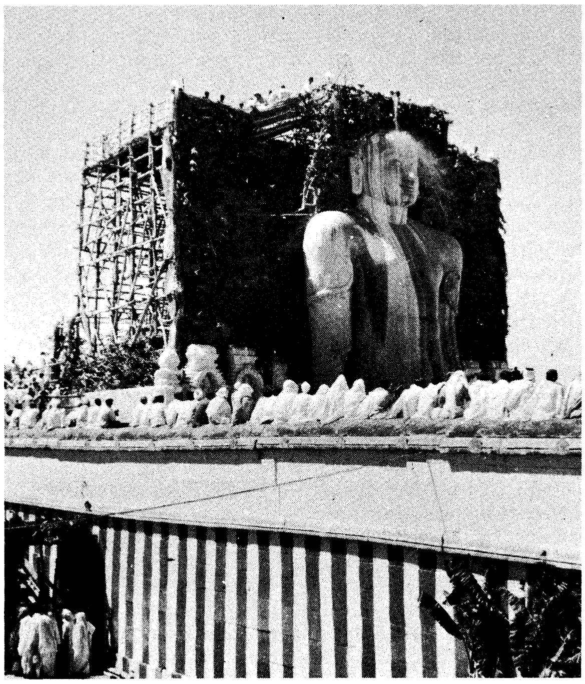
The shrines at Dilwāra, Mt. Abii, are examples of very fine chiselling of marble (the earliest built in the 11th century). An imposing shrine built on a complex plan was erected at Ranakpur in the 15th century. Kesariaji in Rājasthān and Antarikśa Pārśvanātha in Akola district, Miihiirashtra, are famous places of pilgrimage.

Cave temples

Cave temples in Udayagiri and Khandagiri, Orissa, were excavated by Khāravla in about the late 2nd century BC. At Riijgir, Bihiir state, are two adjacent cave temples, one known as Son Bhandiira.(c. 1st century BC or AD). Some Jaina caves exist at Bāwā Pyārā Math (Junagadh), Dhāṅk, and Talājā in Saurashtra, at Ellora in Mahārāshtra (c. 9th–10th century AD), at Aihole in Karṇataka (c. 7th century, along with a beautiful shrine built by Ravikirti), and at Mangi Tungi and Ter in Mahārāshtra. Caves at Sittānnavāsal in Tamil Nadu dating from the 7th century AD are famous; they contain some beautiful wall paintings probably completed during the 10th century AD.

The Jainas, especially the wealthy, also maintain small domestic shrines. some of which contain excellent examples of wood carving and bronze casting.

Temple worship and observance. Temple worship is mentioned in early texts that describe gods worshipping Jina images and relics in heavenly eternal shrines (*śāśvata*-caitya). Worship (*pūjā*), closely associated with *āvaśyaka* (obligatory rites), is not merely restricted to ado-



Ceremony of anointing the colossal image of the Jaina saint Bāhubali (called Locally Gommateśvara) at Sravana-Belgola, India.

James Burke—LIFE Magazine © Time Inc.

ration of the Jina image but is offered to all liberated souls, monks, and the holy writ (*śruta*). Though Tīrthaṅkaras remain unaffected by offerings and worship, the person offering *pūjā*, remembering the virtues of the Jina, imbibes them, and attains to a degree of tranquillity. Daily worship includes recitation of the names of the Jinas; and idol worship through bathing the idol and making offerings of flowers, fruits, perfumes, and lamps. Only Svetiimbaras decorate images with clothing and ornaments. The worshipper also chants hymns of praise and prayers and mutters the sacred formula (*japa*). In front of the Jina, eight auspicious symbols (*aṣṭamaṅgalas*) are designed with rice and other things. Jaina rituals including idol worship, passage rites (see below), and consecrations of images and shrines show considerable Hindu influence. Rituals connected with different *vratas* ("vows") and penances are detailed in later Jaina texts.

The Digambara author Kundakunda and others laid more emphasis on mental culture (*bhāva-pūjā*) than on idol worship (*dravya-pūjā*). Meditation and observance of the various vows of either a householder or an ascetic lead to man's spiritual uplift through successive stages of perfection called the 14 *guṇasthānas*.

Other practices and institutions. Domestic rites and rites of passage. Ancient Jaina literature is silent about domestic rites and rites of passage (marking the main points in human life). The passage ritual is modelled mainly on the 16 Hindu *saṃskāras* (rites of passage). The rites include conception, birth, first meal, naming, tonsure, wearing of sacred thread, beginning of study, marriage, and death. Post-cremation ceremonies, especially the draddhd of *pitṛ-tarpaṇa*, offerings of water to deceased ancestors, are not uniformly practiced by Jainas all over India.

Status of women. Even though monks are warned against contacts with women, who are said to be the source of many impurities, the position of women was never very low in Jaina society. Mahiivira allowed females to enter the ascetic order; rules of conduct and spiritual practice were practically identical for monks and nuns. Women were, however, given a position subordinate to monks in the church hierarchy. Originally womanhood was no bar to salvation; the Svetiimbaras believe that the 19th Tīrthaṅkara, Mallinātha, was a princess. Women as well as men perform Jaina rituals.

Welfare institutions. Jainas are renowned for such various types of munificence as taking a large group on a pilgrimage, famine relief, relief to Jaina widows and the poor, and maintaining shelters (*pañjarapola*) for old animals to save them from being butchered (an act of *ahiṃsā*). Vijaya Vallabha Siiri, the centenary of whose birth occurred in 1970, organized charities to promote the education of poor Jainas through founding the Mahiivira Jaina Vidyalaya with several branches. The institution has also undertaken to publish critical editions of canonical works. Some leading houses of Digambara and Śvetāmbara merchants have encouraged research and publication in various branches of knowledge, of which the Bhiiratiya Jñānapīṭhā at Vārānasi and the L.D. Institute of Indology at Ahmadābād are especially noteworthy. The latter institution, engaged in publishing *Āgamakośa*, a collection of the canonical writings of Jainism, and other works, collects, preserves, and centralizes in the institute various Jaina manuscript collections and is also building up a corpus of microfilms of rare manuscripts from other collections.

JAINA LITERATURE

Canonical literature. Jaina canonical scriptures do not belong to a single period, nor can a single text be deemed free from later revision or additions. The sacred literature preserved orally from the time of Mahiivira was first systematized in a council at Pātaliputra (Patna) about the end of the 4th century BC, and again in two later councils, at Mathura under Ārya Skandila and at Valabhi under Ārya Nāgārjuna, in the early 3rd century AD. The fourth and last Jaina council, at Valabhi c. AD 454 or 467, is said to be the source of the existing Svetiimbara Jaina canon, though some textual commentators insist that the present reading is in accordance with the Mathura council of Ārya Skandila.

The Svetiimbaras follow an extensive canon (*Āgama*) as the source of their system, said to be based upon compilations from discourses of Mahiivira by his direct disciples. The Digambaras believe that the genuine canon is lost but that the substance of Mahāvīra's teachings is contained in the writings of ancient religious figures such as Vattakera, Umāsvāti, Śivārya, Kuṇḍakuṇḍa, Samantabhadra, and Virasena.

Śvetāmbara canon

The Svetiimbara canon consists of 45 *Āgamas*: 11 Ariga ("parts") texts—a 12th, the *Dṛṣṭivāda*, is not extant—12 *Upāṅga* (subsidiary) texts, four *Mūla-sūtras*, six *Cheda-sūtras*, two *Cūlikā-sūtras*, and ten *Prakīrṇakas* (mixed, assorted texts). The Ariga texts contain several dialogues mainly between Mahāvīra and his disciple Gautama presumably recorded by another disciple Ārya Sudharmā, who transmitted the teachings to his own disciples.

According to modern scholars, the *Ācārāṅga-* and *Sūtrakṛtāṅga-sūtras* of the Aiiga texts and the *Uttarādhyayana* of the *Mūla-sūtras* are among the oldest parts of the canon. One of the Cheda-siitras, the *Daśāśrutaskandha*, concludes with *Pajjosavaṇakappo*, or *Kalpa-sū*tra, which recounts the lives of Jinas with an appendix offering rules for monastic life and a list of eminent monks.

Bhadrabāhu I (170 years after Mahiivira) is credited with authorship of some metrical commentaries called Niryuktis. Other commentaries on the *Āgamas*, known as *Bhaṣyas*, were composed mainly during the Gupta age (AD 320–600) and also during the 6th and 7th centuries, in Prākrit language. During the medieval period, explanatory commentaries in Sanskrit were also composed. Haribhadra, Śīlāṅka, Abhayadeva, and Malayagiri are the best known authors of such commentaries.

Digambaras give canonical status to two works in Prākrit: the Karmaprdbhyta—"Chapters on *Karman*"—and the *Kaṣāyaprābhṛta*—"Chapters on *Kaṣāyas*." *Karmaprābhṛta* (also called *Ṣaṭkhaṇḍūgama*) was composed by Puṣpandanta and Bhiitabalin on the basis of the now lost *Dṛṣṭivāda*. It deals with the doctrine of *karman* and is said to have been composed in the 7th century after Mahiivira. The *Kaṣāyaprābhṛta* by Gunadhara, also based on *Dṛṣṭi*vdda, composed in the same age, deals with *kaṣāyas* (passions), such as attachment and aversion, which defile and bind the soul. A later commentary by Virasena (9th century AD) on the first five books of the *Karmaprābhṛta* and another commentary by Virasena and his pupil Jinasena on the *Kaṣāyaprābhṛta* receive similar respect from the Digambaras.

Digambaras also value the Prākrit works of Kuṇḍakunda (c. 1st or 2nd century AD), including the *Pravacanasāra* (on ethics), the *Samayasāra* (on fine entities), the *Niyamasāra* (on Jaina monastic discipline), and the six *Prabhṛtas* ("chapters") on various religious topics.

Of similar importance are the *Mūlācāra* of Vaṭṭakera, the *Kārttikeyānuprekṣā* of Kumāra, or Kārtikeya (between 6th and 13 centuries AD), and the *Tattvarthādhigama-sūtra* of Umāsvāmin, or Umāsvāti, whose work is claimed by both the Śvetāmbaras and Digambaras. The *Tattvārthādhigama-sūtra*, composed early in the Christian era, is the first work in Sanskrit on Jaina philosophy dealing with logic, epistemology, ontology, ethics, cosmography, cosmogony, etc., and had commentaries on it including one by Umāsvāti himself.

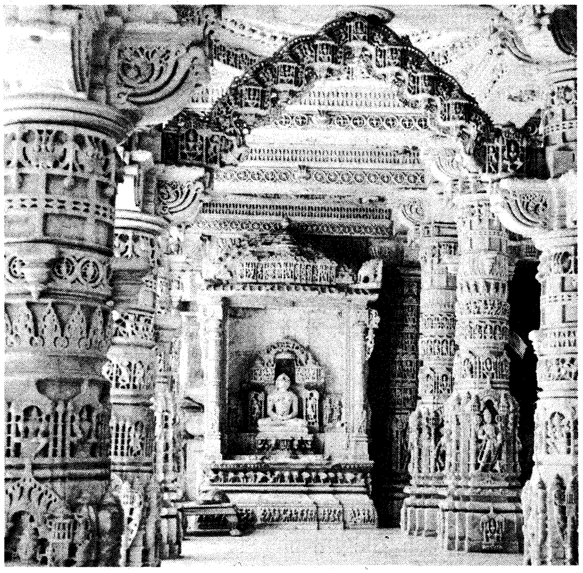
Philosophical and other literature. Among noncanonical Jaina writers on philosophy, besides Umāsvāti and Kundakunda, may be noted Mallavādin I (4th century AD); Siddhasena Divākara (c. 5th century AD); Haribhadra Sūri (*c.* 8th century); Samantabhadra (before *c.* 5th century); Akalanka (c. 8th century); Siddharṣi Ganin (10th century); Śāntisūri (11th century); Vidyānandin (*c.* 8th–9th century); Anantakirti (10th century); and Māṇikyānandin (11th century); Prabhācandra (11th century); Vādi Deva Sūri (12th century); and Ācārya Hemacandra (12th century). Among later authors, especially noteworthy is Upādhyāya Yaśovijaya (c. 17th century), a versatile scholar and author of several works on philosophy.

Noncanonical Śvetāmbara and Digambara compositions are very extensive, covering a very wide range of subjects and composed in Prākrit, Sanskrit, and Apabhraṃśa as well as Old Western Rājasthānī or Gujarati, and Old Kannada. They include works on dramaturgy, epics and *purāṇas*, plays, poems, works on yoga, dictionaries, music, medicine, grammar, mathematics, astrology, fables and fiction, ballads and hymns, and rituals.

RELIGIOUS SYMBOLISM AND ICONOGRAPHY

Image worship was introduced at an early stage, perhaps even during the first century following the death of Mahāvīra. He himself apparently neither prohibited nor prescribed worship of images of himself or of other Tīrthaṅkaras.

By courtesy of the Government of India Tourist Office, London



Interior of a Jaina temple, showing the Tirthankara image enshrined. Dilwāra Temple, Mt. Abū, India, 13th century AD.

Descriptions of stiipas (reliquaries for the bones and ashes of saints), commemorative pillars, and tree shrines (*caityavṛkṣa*s) first appear in Jaina texts, which also refer to the worship in the heavens by gods of images of the four legendary Śāśvata-Jinas (eternal victors) and of costly relic boxes. Mention is also made of *śilāpaṭa*s, which apparently were stone plaques or reliefs placed on lion thrones underneath trees, such as those associated with *yakṣa* worship, and also depicted on Buddhist reliefs from Bharhut (2nd century BC). The *śilāpaṭa*s appear to be the prototypes of the later Jaina *āyāgapaṭa*s (tablets of homage) from Mathura, which show representations of *stūpa*s, caitya-pillars surmounted by elephants, and dharmacakras (the wheel of the law) and the *aṣṭamaṅgala*s (eight auspicious symbols). Later *āyāgapaṭa*s show a Jina attended by two nude disciples, the figure of the monk Kanha Samana with his devotees, or the figure of a noblewoman with attendants.

*Āyāgapaṭa*s and *aṣṭamaṅgala*s

The earliest known Tirthankara image that can be positively identified as such is the highly polished Mauryan period torso from Lohanipur, near Patna (Bihār). The Mauryan ruler Samprati (grandson of the emperor Aśoka) was well-known for his patronage of Jainism. Numerous Tīrthaṅkara images in the sitting (*padmāśana*) and the standing (kdyotsarga) postures dating from the early Christian centuries have been uncovered in the excavations of a Jaina *stūpa* at Kaṅkālī Ṭīlā, Mathura. The earliest images of Tirthankaras are all nude. The various Jinas are distinguished by inscriptions giving their names carved on the pedestals, but later iconographic devices such as cognizances (*lāñchaña*s) or recognizing symbols for each Jina did not evolve until about the 5th century AD.

Worship of the 16 principal Jaina Tantric goddesses, the Mahāvidyās, was introduced probably in the Gupta age. From the 6th to the 11th century a common pair of attendants was employed in sculpture for all the Tirthankaras, but from about the 9th century the 24 different *śāsanadevatā*s (*yakṣa-yakṣiṇī*s) were evolved to attend each Tirthankara. The names of many of the attendants (such as Īśvara, Brahmā, and Kali—also the names of Hindu deities) suggest Hindu influence, while others (Vajraśṛṅkhala, Vajrāṅkuśa, and Bhrfikuti—the names of figures from Buddhist mythology) show Buddhist influence.

Illustrated manuscripts

The religious merit that accrues from hearing and reading Jaina texts encouraged the careful and loving preservation of illustrated manuscripts. The miniature paintings on palm-leaf and paper manuscripts preserved in the Jaina *bhāṇḍāra*s (monastery libraries) provide a continuous history of the art of painting in western India from the 11th century up to the present. The lives of the Jinas and legends of Jaina saints provide a framework for the artists to depict gods and goddesses, throne rooms and village interiors, gardens, and temples. Religious symbols such as the *aṣṭamaṅgala*s (eight auspicious symbols) and the 14 dreams of the mothers of the Tirthankaras frequently appear in paintings.

In addition to the miniatures, and to painted wooden book covers that often show mythological scenes, paintings on cloth are also known. Wall paintings are found on cave shrines at Sittānavāsal (Tamil Nadu state) and at Ellora.

Jaina temples generally contain a number of metal images of various types and metal plaques showing auspicious symbols. Metal images of the Jinas are also kept by pious Jains for home devotion. Among the earliest known bronzes are one of Pārśvanātha in the Prince of Wales Museum of Western India, Bombay, which can be dated c. 1st century BC, and a group of bronzes from Chausa in Bihār in the Patna Museum (1st-3rd centuries AD).

JAINISM AND OTHER RELIGIONS

Jainism and Hinduism. Hindus regard Jainism along with the beliefs of Buddhists and *cārvākas* (materialists) as ndstika-darfana (nonorthodox points of view) since they do not believe in either Vedic (Hindu scriptural) authority or the existence of a creator god.

Both the Jainas and the Hindus believe in the doctrine of karman, as the basic principle of *saṃsāra* (the chain of birth and death) and rebirth. Late Hindu dietary restrictions, especially among the Vaiṣṇavas (followers of the Hindu god Viṣṇu), may have been the result of Jaina influence on Indian society. The application of *ahiṃsā* to the political sphere in India's struggle for independence by the great saint and political leader Mahatma Gandhi was a further extension of Mahāvīra's doctrine of *ahimsd* to all spheres of life. Gandhi had great regard for Shrimad RZjachandra, a versatile genius of high spiritual attainment. Born in 1867, in a small village in Saurashtra, Shrimad RZjachandra died in 1900 at the age of 33. A great saint, thinker, and writer, RZjachandra was perhaps the greatest and the best known Jaina layman of modern times, who believed more in practice than in theory. Spiritually a very advanced soul, he was above narrow sectarian differences and dogmas and believed in the equality of all religions which led to realization of the Pure Self (dtman). Though he preached *ahimsd,* one of his published letters shows that he attached supreme value to the principle of truth. Mahatma Gandhi regarded him as one of his gurus. Both preached truth and nonviolence, both often practiced self-mortification (fasts, etc.), but while RZjachandra was eager to leave this miserable body and the world, Gandhi wanted to realize God through service of humanity, especially of the poor and the downtrodden.

Jainism and Buddhism. Both Jainism and Buddhism revolted against Brahmanical division of class by birth and the sacrificial killing of animals. Both assimilated common Indian conceptions of deities, such as Indra, Brahmā, and the *yakṣa*s and symbol worship of such objects as the *stūpa, dharma-cakra,* caitya-trees, *ratna-traya.*

Among the elements in common with Buddhism should be noted the 32 attributes of a *mahapuruṣa* ("great man"), the iconic similarity in the type of a seated Buddha with a seated Jina, the worship of the *stūpa,* and the *dharma-cakra.* Both Mahiivira and Buddha obtained the highest knowledge while meditating under a tree.

Jainism and Islām. In reference to Muslim influence on Jainism, one scholar has suggested that the conception of *āśātanā*s—activities that are unfitting or indecent in a temple—reveals a notion of the sanctity of the temple that is more evocative of Muslim *barakah* (holiness) than of any traditional Jaina attitude. The most obvious influence of Islām is seen, however, in the Śvetāmbara Loṅkāsāha's repudiation of image worship as something without canonical support. A parallel sect, the Tāraṇapantha, also arose among the Digambaras.

The Jaina influence at the Mughal court of Akbar is a bright chapter in Jaina history. Hiravijaya Sūri, then leader of the Śvetāmbara Tapā *gaccha,* was invited and honoured by Akbar. His disciples and some other monks also received respect from the Mughal emperors Jahāngīr, Shāh Jahān, and even Aurangzeb. Akbar issued a *firmān* (decree) prohibiting animal slaughter near the Jaina sites at Gimar, Satrunjaya, Ābū, Rājgīr, and Parasnath hill during Pajjuṣaṇa festival. Jahiingir also issued *firmdns* for protection of Satrunjaya and Aurangzeb gave a *firmān* in favour of Shantidas Sheth of Ahmadābād, donating proprietor-rights over Mt. Satrunjaya. Mughal influence in different schools of Indian painting is well-known to art historians, and Jaina miniature painting was not free from it.

Jainism and other religions. Zoroastrian influence on Jainism is not certain. The Jainas in South India came in contact with Christianity in the early centuries AD, though there was no significant modification of either religion as a result of such contact.

BIBLIOGRAPHY. H. JACOBI, "Jainisrn," *Encyclopedia of Religion and Ethics,* vol. 7 (1955), and *Jaina Sūtras,* "Sacred Books of the East Series," vol. 22 and 45 (1884, 1895), introductions to the volumes esp. noteworthy; G. BUHLER, *On the Indian Sect of the Jainas* (Eng. trans. 1903); H. VON GLASENAPP, *Der Jainismus* (1925, reprinted 1964), the most comprehensive text on Jainism, and *The Doctrine of Karman in Jain Philosophy* (Eng. trans. 1942); W. SCHUBRING, *The Doctrine of the Jainas* (rev. Eng. trans. 1962), standard, scholarly work, and *The Religion of the Jainas* (Eng. trans. 1966); AMULYACHANDRA SEN, *Schools and Sects in Jaina Literature* (1931); JAGAMANDARLAL JAINA, *Outlines of Jainism* (1940); C. CHOPRA, *A Short History of the Terapanthi Sect of the Swetamber Jains and Its Tenets* (1937); B.C. LAW, *Mahavira: His Life and Teachings* (1937), a lucidly written introduction to the subject; NATHAL TATIA, *Studies in Jaina Philosophy* (1951), see esp. the discussion on the problem of Avidyā in various systems; SATKARI MOOKERJEE, *The Jaina Philosophy of Non-Absolutism* (1944), a standard work by an authority on Indian philosophy; MOHANLAL MEHTA, *Outlines of Jaina Philosophy* (1954), *Jaina Culture* (1969); S.B. DEO, *History of Jaina Monachism from Inscriptions and Literature* (1956); R.H.B. WILLIAMS, *Jaina Yoga* (1963), a masterly analysis of the Jain ethics concerning the laity, with critical notes on authors of different sourcebooks; DAYANAND BHARGAVA, *Jaina Ethics* (1968); K.C. SOGANI, *Ethical Doctrines in Jainism* (1967); V.A. SANGAVE, *Jaina Community: A Social Survey* (1959); C.R. JAINA, *Jaina Law* (1926); M. WINTERNITZ, *History of Indian Literature,* vol. 2 (Eng. trans. 1933), see the chapter on Buddhist and Jaina literature; H.R. KAPADIA, *History of the Canonical Literature of the Jainas* (1941), a very good descriptive account of the Jaina canon; B.C. BHATTACHARYA, *Jaina Iconography* (1939), a brief outline of the subject; U.P. SHAH, *Studies in Jaina Art* (1955), a review of Jaina art in North India, with a discussion of various symbols in Jaina worship and a good bibliography, and *Akota Bronzes* (1959), a description of rare Jaina bronzes from a site in Gujarāt that throws light on some aspects of Jaina iconography; KLAUS FISHER, *Caves and Temples of the Jainas* (1956); MOTI CHANDRA, *Jain Miniature Paintings from Western India* (1949), a standard textbook for the study of Jaina miniature paintings in western India, with exhaustive references; W. NORMAN BROWN, *The Story of Kālaka* (1933), a well-known work on Kālakācārya as well as on miniature Jaina paintings; P.B. DESAI, *Jainism in South India and some Jaina Epigraphs* (1957), a useful compilation on Jainism in South India.

(U.P.S.)

Jalāl ad-Din ar-Rūmī

Jaliil ad-Din ar-Riimi, called Mawlānā (Mevlāna; "Our Master"), was the greatest Ṣūfī—Islāmic mystic—poet in the Persian language; he is famous for his lyrics and for his didactic epic *Masnavi* ("The Spiritual Couplets"), which has been called "the Qur'ān in Persian." The work widely influenced Muslim mystical thought and literature—the number of commentaries and translations in the different Muslim languages is almost beyond number.

Jalāl ad-Din was born—probably on September 30, 1207, in Balkh (modern Afghanistan). His father, Bahā' ad-Din Walad, was a noted mystical theologian, author, and teacher. Mainly because of the threat of the approaching Mongols, Bahā' ad-Din and his family left their home town in about 1218. According to a legend, in Nīshāpūr, Iran, the family met Farīd od-Din 'Aṭṭār, a Persian author of mystical epics, who blessed young Jaliil ad-Din. After a pilgrimage to Mecca and journeys through the Near East, Bahā' ad-Din and his family reached Anatolia (Riim, hence the surname Riimi), a country that enjoyed peace and prosperity under the rule of the Turkish Seljuq dynasty. After a short stay at Laranda (Karaman), where Jaliil ad-Din's mother died and his first son was born, they were called to the capital, Konya, in 1228. Here, Bahā' ad-Din Walad taught at one of the numerous *madrasahs* (religious schools); after his death in 1231 he was succeeded in this capacity by his son. Early life

A year later, Burhān ad-Din Muhaqqiq, one of Bahā' ad-Din's former disciples, arrived in Konya and introduced Jaliil ad-Din deeper into some mystical theories that had developed in Iran. Burhān ad-Din, who contributed considerably to Jalāl ad-Din's spiritual formation, left Konya about 1240. Jalāl ad-Din is said to have undertaken one or two journeys to Syria (unless his contacts with Syrian Ṣūfī circles were already established before his family reached Anatolia); there he may have met Ibn al-'Arabi, the leading Islāmic theosophist whose interpreter and stepson Ṣadr ad-Din al-Qunawi was Riimi's colleague and friend in Konya.

The decisive moment in Rūmī's life occurred on November 30, 1244, when in the streets of Konya he met the wandering dervish—holy man—Shams ad-Din ("Sun of Religion") of Tabriz, whom he might have first encountered in Syria. Shams ad-Din can not be connected with any of the traditional mystical fraternities; his overwhelming personality, however, revealed to Jalāl ad-Din the mysteries of divine majesty and beauty. For months the two mystics lived closely together, and Rūmī neglected his disciples and family so that his scandalized entourage forced Shams to leave the town in February 1246. Jalāl ad-Din was heartbroken; his eldest son, Sultan Walad, eventually brought Shams back from Syria. The family, however, could not tolerate the close relation of Jalāl ad-Din with his beloved, and one night in 1247 Shams disappeared forever. It has recently been established that he was indeed murdered, not without the knowledge of Riimi's sons, who hurriedly buried him close to a well that is still extant in Konya.

Principal works

This experience of love, longing, and loss turned Riimi into a poet. **His** mystical poems—about 30,000 verses and a large number of *robāʿīyāt* ("quatrains")—reflect the different stages of his love, until, as his son writes, "he found Shams in himself, radiant like the moon." The complete identification of lover and beloved is expressed by his inserting the name of Shams instead of his own pen name at the end of most of his *ghazal*s (love poems). The *Dīvān-e Shams* (The Collected *Poetry* of Shams) is a true translation of his experiences into poetry; its language, however, never gets lost in lofty spiritual heights or nebulous abysses of speculation. The fresh language excels by its strong rhythms and sometimes assumes forms close to popular verses. There would seem to be cause for the belief, expressed by chroniclers, that most of this poetry was composed in a state of ecstasy, induced by the music of the flute or the drum, the hammering of the goldsmiths, or the sound of the water mill in Meram, where Riimi used to go with his disciples to enjoy nature. He found in nature the reflection of the radiant beauty of the "Sun of Religion" and felt flowers and birds partaking in his love. He often accompanied his verses by dancing.

A few years after Shams ad-Din's death, Riimi experienced a similar rapture in his acquaintance with an illiterate goldsmith, Ṣalāḥ ad-Din Zarkiib. It is said that on one day, hearing the sound of a hammer in front of Ṣalāḥ ad-Din's shop in the bazaar of Konya, Riimi began a whirling dance. The shop owner became one of Riimi's closest and most loyal disciples, and his daughter became the wife of Riimi's eldest son. This love again inspired Jalāl ad-Din to write some poetry. After Ṣālāḥ ad-Din's death, Ḥusām ad-Din Chelebi became his spiritual love and deputy. Riimi's main work, the *Masnavī-ye* Ma'navi ("Spiritual Couplets"), was composed under his influence. Ḥusām ad-Din had asked him to follow the model of the poets ʿAṭṭār and Sanāʾi who had laid down mystical teachings in long poems, interspersed with anecdotes, fables, stories, proverbs, and allegories. Their works were widely read by the mystics and by Riimi's disciples. Jalāl ad-Din followed Ḥusām ad-Din's advice and composed nearly 26,000 couplets of the Masnavi during the following years. It is said that he would recite his verses even in the bath or on the roads accompanied by Ḥusām ad-Din, who committed them to paper. The Masnavi, which shows all the different aspects of Ṣūfism in the 13th century, often carries the reader away with loose associations of thought, so that one understands what subjects the master had in mind at a particular stage of his life. The work reflects the experience of divine love; both Ṣalāḥ ad-Din and Ḥusām ad-Din were, for Riimi, renewed manifestations of Shams ad-Din, the all-embracing light. Thus he called Ḥusām ad-Din Diyn' al-Haqq (Light of the Truth); *ḍiyāʾ* is the Arabic term for sunlight.

After completing the Masnavi, Riimi continued writing for a short while. He always remained a respected member of Konya society and his company was sought by the leading officials as well as by Christian monks. He died on December 17, 1273. Ḥusām ad-Din became his successor and was in turn succeeded by Sulṭān Walad, who organized the loose fraternity of Riimi's disciples into Mawlawiyah, known in the West as the Whirling Dervishes because of the mystical dance that constitutes their principal ritual. Sulṭān Walad's poetical accounts of his father's life are the most important source of knowledge of Riimi's spiritual development.

Besides his famous poetry, Riimi has left a small collection of occasional talks as they were noted down by his friends; in the collection, known as *Fīhi mā* fihi ("There is in it what is in it"), the main ideas of his poetry recur. There also exist some letters directed to different persons. It is impossible to systematize his ideas, which at times contradict each other; and changes in the use of symbols often puzzle the reader. His poetry is one of the most human expressions of mystical experiences in which each reader can find his own favourite ideas and feelings—from enthusiastic flights into the heavens to matter-of-fact descriptions of daily life. Riimi's influence on Turkish cultural life can scarcely be overrated; his mausoleum, the Green Dome, today a museum, is still a place of pilgrimage for thousands.

BIBLIOGRAPHY. AFZAL IQBAL, *The Life and Thought of Mohammad Jalal-ud-din Rumi*, 2nd ed. rev. (1964), is the only English biography of Rūmī; unfortunately it is not too critical. A full evaluation in a Western language is still lacking. Biographical notes are found in the numerous translations of Rūmī's works by R.A. NICHOLSON and A.J. ARBERRY, chiefly in the latter's *Discourses of Rūmī*, a translation of the *Fīhi mā fihi* (1961).

(An.Sc.)

Jamaica

Jamaica, a parliamentary state, is the third largest island in the Caribbean Sea. It has a total land area of 4,244 square miles (10,991 square kilometres) and is about 146 miles long, 51 miles wide at its greatest width, and 22 miles wide at its narrowest point. It is situated some 100 miles due west of Haiti, 90 miles south of Cuba, and 310 miles northeast of Cape Gracias a Dios, Nicaragua, the nearest point on the American continent. The island's population was over 1,800,000 in the early 1970s. The national capital is Kingston.

Christopher Columbus, who discovered the island in 1494, called it Santiago, but the original Amerindian name of Jamaica, or Xaymaca, has persisted. Columbus considered it to be "The fairest isle that eyes have beheld," and many travellers still regard it as one of the most beautiful islands in the Caribbean.

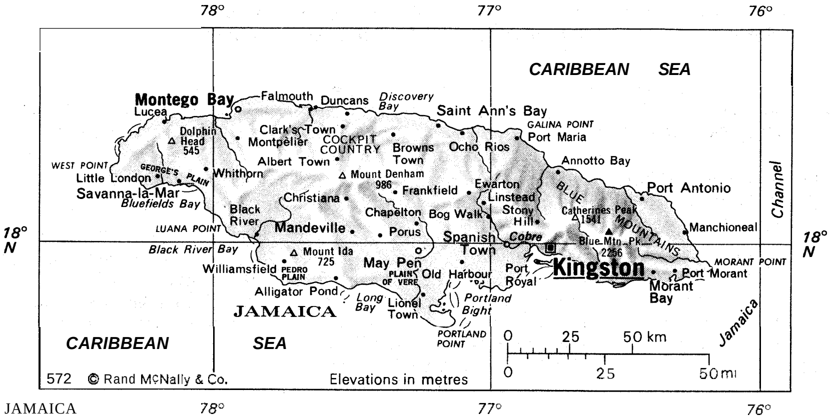
Although agriculture remains the major employer of labour, industry—notably bauxite mining—and tourism are the main contributors to the national income. Jamaica has made great strides since its independence in 1962 and is one of the most rapidly developing countries in the world. Adherence to constitutional government and the rule of law are traditional, and the national motto, "Out of Many, One People," describes a multiracial society whose integration is profound and enviable. (For an associated feature, see CARIBBEAN SEA.)

THE LANDSCAPE

The natural environment. Relief features. In general, the topography consists of coastal plains encircling an island that is bisected from east to west along its length by mountains and plateaus. The mountains form the chief physical feature. Almost half of Jamaica's surface is over 1,000 feet above sea level. The chief range rises in the east to Blue Mountain Peak at 7,402 feet (2,256 metres) and then loses altitude as it sweeps westward. The mountains contribute to the great diversity of scenery for which the island is famous, ranging from the stunted, elfin forests of the highest peaks to the dry, sandy, cactus-growing areas of the south. There are damp, tree-fern rain forests and wide, flat, alluvial plains. Located chiefly on the south side of the island, the principal plains are: Liguanea Plain in Kingston and St. Andrew; Rio Cobre and St. Dorothy plains in St. Catherine; the Vera Ma Hollis Savanna (plains of Vere) in Clarendon; George's Plain in Westmoreland; and Pedro Plains in St. Elizabeth. The rolling limestone hills and plateaus in the central and western areas include the unusual, trackless karst (a limestone region broken by ridges, depressions, and caverns) region of the Cockpit Country, covering 500 square miles of the interior.

Jamaica's mountains

Drainage and soils. There are some 120 rivers and streams, with numerous tributaries issuing from ravines in the mountains. Few are navigable for any great distance, because of their rapid descent from the mountains; in times of flood they are violent torrents. Some of the larger rivers have alluvial plains in their lower valleys and some have deltas. The Black River is navigable by small boats for about 17 miles from its mouth.



JAMAICA

More than half of the island's surface is covered with white limestone that overlies yellow limestone, beneath which are older metamorphic rocks (compact rocks formed by heat and pressure) and igneous rocks (formed by the cooling of molten material). The upland areas are mostly covered with bare rock or soils of little depth and are very susceptible to erosion. The alluvium of the coastal plains is composed chiefly of deep loam (a friable mixture of clay, silt, and sand) and clay. The valley floors are covered with residual clays.

The climate. The tropical climate is influenced by the sea and is characterized by little change in seasonal temperature, although the mountains cause regional variations. Because the island lies between the subtropical high-pressure and the equatorial low-pressure belts of the Atlantic Ocean, the northeast trade winds are dominant and blow throughout the year. Along the coasts, breezes blow onshore by day, and offshore at night. During the winter months, from December to March, cold winds known locally as "northers" reach the island through the wide, open trough of the North American plains.

Variations in temperature range from 90° F (32" C) on the coasts to 40° F (4° C) on the peaks. Kingston, at sea level, has an average daily maximum temperature of 88" F (31° C) and an average daily minimum of 71° F (22" C). At Stony Hill, at 1,400 feet, the maximum and minimum means are 86° F (30" C) and 68° F (20" C).

Rainfall

Rains are seasonal, falling chiefly in May and October, although thunderstorms in the summer months, from June to September, can bring heavy showers. The average annual rainfall is 77 inches, but regional variations are considerable. The mountains force the trade winds to deposit more than 200 inches a year on the northern parishes of Portland and St. Thomas, while little precipitation occurs on the hot, dry savanna lands of the southern and southwestern plains. Jamaica is susceptible to hurricanes during the summer but has not been struck since 1951. Earthquakes have caused serious damage only twice—in 1692 and 1907.

Vegetation and animal life. The richness and diversity of Jamaica's trees and plants constitute one of its chief glories, though it has changed considerably through the centuries. The island was completely forested in the 15th century, except for small agricultural clearings. The great timber trees were cut down for building purposes by the European settlers, and the plains, savannas, and mountain slopes were cleared for cultivation. Many new plants were introduced: the food plants—including sugarcane, bananas, and citrus and other tropical fruits—were almost all introduced to the island.

Jamaica has few indigenous quadrupeds. The coney, a member of the rodent family, once very numerous and prized as food before the European immigration, is now extremely scarce. The mongoose, a small carnivore that feeds on rats and snakes, is widespread but was introduced in 1872 from India. The 25 species of bats are the most numerous of the mammals. The native crocodile is now in danger of extinction. There are no poisonous snakes. The main freshwater fish is the mountain mullet; there are four species of crayfish.

Rich birdlife

More than 200 species of birds have been recorded, including 25 endemic species, such as the streamertail hummingbird, which is the national bird. The bird population also includes species found both in the West Indies and in America and birds that migrate to Jamaica, either in the winter or in the summer.

The landscape under human settlement. ***Rural settlement.*** With the emancipation of the island's black slaves in 1838, a large proportion of the freed population—often with the aid of nonconformist missionaries—left the large plantations. They moved to the hills where land was plentiful and created their own villages and communities. This migration laid the foundation of the present pattern of rural settlement, and many of the villages retain their original character.

Urban settlement. Most of the cities and chief towns are located on the coastal plains, where the main commercial crops are grown. Kingston, the national capital, is located on the southeastern coast. It stands on the Liguanea Plain with the sea to the south and the St. Andrew Mountains, which form part of the ranges of the parish of St. Andrew, to the north. It is the commercial, administrative, and cultural centre of the island and the focus of its transportation services. Spanish Town, the old capital, is 11 miles west of Kingston. The other important towns—Montego Bay, Ocho Rios, and Port Antonio—are centred on the north coast. Their fine white-sand beaches and unparalleled mountain scenery make them popular tourist resorts.

THE PEOPLE

Origin of the population

The aboriginal Arawak Indians were exterminated by the Spanish colonists by the time the English invaded the island in 1655. The Spaniards themselves disappeared as a population element shortly afterwards. With the large-scale introduction of African slaves to work the sugar estates, the English settlers were soon greatly outnumbered. Today the population is predominantly African and Afro-European in origin, with minority elements originating from the United Kingdom, India, China, Syria, Portugal, and Germany.

Languages. English is the official language, but a local dialect, Creole, is also widely spoken. It is basically English in vocabulary and grammar but contains features derived from a variety of African languages, as well as from Spanish and French. Creole is not bad English but a dialect of much value and charm.

Religions. Freedom of worship is entrenched in Jamaica's constitution. The majority of the population belongs to the Church of England, which was the established church of the country until 1870. There is also a large number of Baptists and Roman Catholics. Nearly every Christian denomination and sect is represented, and the Jewish community is one of the oldest in the Americas. There is a Hindu community, a Muslim mosque, and a branch of the Ethiopian Orthodox church. Some of the popular and revivalist sects base their beliefs on Christianity, but their forms of worship differ widely from those accepted by most orthodox churches. The central feature of the Pocomania sect, for example, is spirit possession; the Cumina sect has rituals characterized by heavy drumming, dancing, and spirit possession.

Population trends. The population is unevenly distributed over the island. The coastal lowlands are the most densely populated regions, especially in the Kingston metropolitan area, which has a population of over 500,000.

Jamaica, Area and Population

	area		population	
	sq mi	sq km	1960 census	1970 census*
Counties†				
Cornwall				
Parishes				
Hanover	174	450	54,000	59,000
Saint Elizabeth	468	1,212	117,000	127,000
Saint James	230	595	83,000	104,000
Trelawny	338	875	56,000	61,000
Westmoreland	312	807	110,000	113,000
Middlesex				
Parishes				
Clarendon	462	1,196	164,000	177,000
Manchester	321	830	112.000	124,000
Saint Ann	468	1,213	114,000	121,000
Saint Catherine	460	1,192	154,000	186,000
Saint Mary	236	611	94,000	100,000
Surrey				
Parishes				
Kingston	8	22	123,000	117,000
Portland	314	814	65.000	69.000
Saint Andrew	166	431	296,000	433.000
Saint Thomas	287	743	69,000	71,000
Total Jamaica	4,244	10,991	1,610,000‡	1,861,000‡

*Preliminary. †The counties are historical divisions and have no administrative function. ‡Figures do not add to total given because of rounding.
Source: Official government figures.

The rate of natural increase of the population continued to decline in the 1960s. In 1971 the birth rate fell to 34.8 per thousand from its previous level of 42.4 per thousand in 1960. This rapid decline was due in part to an intensified family-planning program. The death rate also continued to fall, from 8.8 per 1,000 in 1960 to 7.4 per 1,000 in 1971, largely because of increased access to medical facilities and the spread of public-health education. The infant mortality rate fell by more than one-third during the 1960s.

Emigration

During the 20th century there has been considerable emigration to Panama, Cuba, Central and South America, and the United States. Emigration during the 1960s was particularly high and reached a level of 30,000 a year. In the first half of this period there was much emigration to the United Kingdom, but numbers declined sharply after the passage of restrictive immigration legislation in 1967. The United States is now the chief outlet for Jamaican emigrants, followed by the United Kingdom and Canada.

HISTORY

The Spanish period. Jamaica was discovered by Christopher Columbus in 1494, during his second voyage to the New World. More than 15 years were to pass, however, before colonization was attempted. The Spanish government was disappointed in the country's lack of gold, and Jamaica became a neglected part of the property of the Columbus family. Its chief value to Spain was as a supply base; its settlers were mainly engaged in cattle-ranching.

In 1655 the island fell to a British invasion force and became the first colony in the Americas to be captured by a formal British expedition.

The British period. Two important features of the early British period were the activities of the Maroons and of the buccaneers. The Maroons—whose name probably derives from the Spanish *cimarrón,* meaning "wild" or "untamed"—were freed or escaped slaves who had taken to the thick woods and mountains at the time of the British invasion. Organized, armed, and encouraged by the Spanish, they harried the invaders with guerrilla-warfare tactics and were to prove a thorn in the side of the British for almost 150 years. The buccaneers—who preyed on Spanish ships—operated mainly from their base at rich and corrupt Port Royal. By their relentless attacks on Spanish Caribbean cities, they kept the Spaniards occupied at a time when Britain was unable to spare a fleet for the protection of its West Indian colonies.

The importation of African slave labour, begun by the Spaniards, continued under the new regime. It grew steadily in volume as sugar production increased in extent and value. By the 18th century, Jamaica, like other sugar colonies, had become one of the most valuable of colonial possessions.

Crisis of 1865

The abolition of the slave trade in 1807 and full emancipation 31 years later caused the collapse of the plantation system. This, together with other factors, including unemployment, heavy taxation, and droughts, produced a crisis in 1865 that changed the old social and economic patterns for all time. In October 1865 an uprising in the easternmost parish of St. Thomas, known as the Morant Bay Rebellion, was put down with terrible severity. The scandalous handling of the crisis by the British governor, Edward Eyre, led to his recall. Before leaving, however, he induced the frightened House of Assembly to vote for its own extinction. In its place a crown-colony form of government, in which the governor wielded the only real executive or legislative power, was established by an act of the British Parliament in 1866.

By 1938 dissatisfaction with the crown-colony system, sharpened by the hardships and suffering brought on by a worldwide economic depression, erupted in serious and widespread rioting. These events resulted in the formation of the first lasting labour unions as well as of political parties linked to them. A growing demand for self-determination also became apparent.

Transition to independence. The constitution of 1[illegible] provided for a two-party House of Representatives, the appointment of ministers, and universal adult suffrage. Further constitutional advances took place in 1953 and 1957, and full internal self-government was obtained in 1959. In 1958 Jamaica had become a founding member of the Federation of the West Indies, from which it seceded in 1961. On August 6, 1962, Jamaica became an independent nation with full dominion status within the Commonwealth.

THE NATIONAL ECONOMY

Jamaica's economy is essentially an open one, with heavy dependence on primary exports and on imports of manufactures and capital goods. There is active foreign economic participation, especially in the export sector (bauxite and alumina) and in tourism.

Natural resources. Among the minerals found on the island, bauxite, gypsum, silica sand, ceramic clays, marble, and limestone are of commercial interest. The bauxite is found in an area of about 1,000 square miles in central Jamaica; the gypsum and marble are in eastern Jamaica; clays are in the west; and limestone is found throughout the island. An estimated 19 percent of the island's total area consists of forests, about half of which are government owned.

Sources of national income. ***Agriculture, forestry, and fishing.*** Agriculture continues to be the main basis of the island's economy, engaging about 36 percent of the total labour force. The two major crops are sugar—with its by-products of rum and molasses—and bananas. Other important crops are citrus fruit, coffee, pimentos, cocoa, and ginger.

Local forestry production is insufficient to meet the country's needs; about 90 percent of the wood, cork, and paper consumed is imported. It is anticipated that afforestation programs now being pursued will provide the basis for eventual self-sufficiency.

The annual catch of fish amounts to about 35,000,000 pounds (16,000,000 kilograms); the island shelf is the traditional fishing area. Mechanized boats sail about 60 miles southwest of Jamaica to Pedro Bank.

Mining and quarrying. Since 1952 the mineral industry has played an increasingly significant role in the country's economic development and is considered to be the most rapidly expanding sector of the economy. Jamaica is the world's largest producer of bauxite; production in 1970 equalled 11,800,000 tons. Bauxite was mined in the early 1970s by one Canadian and three United States companies, which have also undertaken the manufacture of alumina (a product made from bauxite that is used for making aluminum). In 1969 mining, quarrying, and refining accounted for about 13 percent of the national income. The production of silica sand is absorbed by local glass-container manufacture, while most of the gypsum is mined for export. Cement is largely used in local construction.

Bauxite production

Industry. Manufacturing is increasingly important, both in providing employment and in satisfying the increasing demand for manufactured goods. Food processing is the most important manufacture, followed by metal products. Other significant categories are sugar, rum, and molasses processing, textile manufacturing, printing, chemical production, and cement and clay products. Industrial growth has been stimulated by the activities of the Jamaica Industrial Devellopment Corporation, a statutory body that administers the incentive laws that provide concessions such as the duty-free importation of machinery and equipment and income tax exemption to foreign and local investors.

Increasing reliance was placed on tourism in the 1960s. In 1969 this industry was second only to bauxite as a dollar earner. The traditional attraction of Jamaica to the tourist is the warm climate and good beaches.

Electricity is supplied from both private and public sources. Public generation of electricity is mainly by steam turbines. Privately owned generating plants supply the power needs of the sugar, cement, bauxite, and alumina industries.

Finance. Financial services are dominated by commercial banks that are mainly subsidiaries of Canadian, British, and American banks. Savings and credit services are also offered by life-insurance companies, building societies, and credit unions. The central bank, called the Bank of Jamaica, founded in 1960, is empowered to control money and credit and to promote economic development. The Development Finance Corporation provides loans for industry, housing, and tourism. There is a Small Business Loan Board, and local securities are also sold. The monetary unit is the Jamaican dollar (Jam$1 = $1.20 U.S.; Jam$2 = £1 sterling on May 1, 1971).

Foreign trade. The principal exports are bauxite, alumina, agricultural products, and clothing. Widely varied imports include food, beverages, tobacco, manufactured goods, fuel oil, fertilizer, and machinery. The United States, the United Kingdorn, and Canada are the leading trade partners. The European Economic Community is an important source of imports. Jamaica is a member of the sterling area and of the Commonwealth preferential-trading system.

Management of the economy. ***The Public and private sectors.*** The economy is based on private enterprise. The government uses monetary and fiscal policies to achieve maximum benefits for the economy as a whole. The mining and tourist industries, as well as much of the manufacturing sector, are financed by foreign capital. The government follows economic: policies that serve to foster the highest possible rate of growth consistent with the country's social development.

The largest sources of government revenue are income tax, customs duties, and excise duties. Development loans are raised by the government in London, New York, and locally.

Trade unions and employer associations. In 1971 there were more than 40 trade-unions, of which the most important were the Bustamente Industrial Trade Union (affiliated with the Jamaican Labour Party) and the National Worker's Union (affiliated with the People's National Party); there are about ten employers' associations.

Contemporary economic policies. The government is attempting to achieve the highest possible rate of growth consistent with the attainment of the goals of its social development policy. Emphasis is being laid on attempts to achieve a wider distribution of the benefits of development, on fiscal and monetary stability, and on the maintenance of conditions conducive to a high level of investment by the private sector.

Transportation. Generally, the transport systems either follow the coastline or cut across the central mountains from north to south.

Roads and highways. The road system at the start of the 1970s consisted of more than 2,680 miles of main roads, about 1,900 of which were paved. There are also about 7,000 miles of secondary roads. The main roads encircle the island, loop into the plains areas, and cross the mountains at three major north-to-south crossings. Public passenger services are available outside the capital which itself has a regular bus service. There are also taxi and limousine services.

Railways. There are 205 miles of railway routes, parts of which have been in existence since 1845. The main line runs northwest from Kingston to Montego Bay via Spanish Town, May Pen, and Montpelier. From Spanish Town, a branch line runs to Annotto Bay and Port Antonio by way of Bog Walk, where a short line branches to Ewarton.

Air and water transport. There are scheduled international air services at the two major airports of Palisadoes, near Kingston, and Montego Bay. These airports also handle scheduled domestic flights and an air-taxi service. Port Antonio and Ocho Rios have licensed aerodromes, and there are 40 other airstrips throughout the island.

Jamaica has 16 seaports in use. Kingston, Montego Bay, and Port Antonio are the principal ports. Regular passenger and cargo services are maintained with the United Kingdom, Canada, and the United States. There are also shipping lines to continental Europe, South America, the Caribbean, Australia, and New Zealand.

ADMINISTRATION AND SOCIAL CONDITIONS

Government structure. Under the Jamaica (Constitution) Order in Council of 1962, by which the island achieved independence, the monarch of the United Kingdom is titular head of state. A Jamaican governor general is chosen by the monarch on the advice of the prime minister. The prime minister is appointed by the leading political party from its parliamentary members. The legislature is a bicameral parliament consisting of a House of Representatives and a Senate. The House has 45 to 60 members, who are elected by universal adult suffrage. The speaker and deputy speaker are elected by the House from its members. The Senate has 21 members, who are appointed by the governor general—13 in accordance with the advice of the prime minister and eight on the advice of the leader of the opposition party. The president and deputy president are elected by the Senate from among such of its members as are not ministers or parliamentary secretaries. The principal instrument of policy is the Cabinet, which consists of the prime minister and at least 11 other ministers, of whom at least two or three must be members of the Senate. The Privy Council is limited to advising the governor general on the exercise of the royal prerogative of mercy and on the discipline of government officers.

Legislature

The island is divided into 14 parishes, two of which are amalgamated as the Kingston and St. Andrew Corporation. Local affairs in the other parishes are administered by individual parish councils. Their size varies between 13 and 33 members, who are elected by universal adult suffrage. Members of the House of Representatives hold ex-officio seats on parish councils, and the mayors of those parish capitals that enjoy mayoral status are the chairmen of their councils.

The two political parties are the Jamaica Labour Party and the People's National Party. General elections are held every five years.

Justice. The judiciary comprises a supreme court, a court of appeal, and resident magistrate's (parish), petty sessional, coroner's, and juvenile courts. There is also a traffic court. The legal and judicial system is based on English common law. The attorney general is the government's principal legal adviser. There is a director of public prosecutions.

The police. The constabulary force consists of about 80 officers and 3,000 other men and women. Its responsibilities include immigration and the registration of aliens. There is a criminal investigation department and a telecommunications branch, as well as water police and mounted detachments. Among the various police auxiliaries are the special constabulary force and the district constables.

The armed forces. The Jamaica Defense Force is organized into regular and part-time, or national reserve, elements. The regular forces include army, air force, coast guard, and logistics units. The air wing operates both helicopters and light aircraft; the Coast Guard is equipped with fast patrol boats. The National Reserve includes army units, an air squadron (using mainly privately owned aircraft), and small marine units.

Education. A substantial part of the annual budget goes to the Ministry of Education. Considerable sums are devoted to the School of Agriculture, the School of Arts and Crafts, and the University of the West Indies, the main campus of which is at Mona, a northeastern section of Kingston. Education is provided by government-owned, government-aided, and private schools, some of which are run by religious bodies. The Ministry operates more than 750 primary schools, 16 junior secondary schools, and 40 secondary schools, as well as a number of vocational centres, technical institutes, technical high schools, and teacher-training colleges.

Health and welfare services. Medical care is provided by 27 public hospitals, including the University College Hospital, and about 150 health centres and clinics. There are four private hospitals. Highly successful programs of insect control and malaria eradication have been undertaken, and periodic inoculations against tuberculosis, polio, diphtheria, whooping cough, and tetanus are given.

Welfare programs

The Ministry of Youth and Community Development undertakes community services through the Social Welfare Commission and various voluntary groups. There are a number of commissions and committees concerned with social services, including the Social Welfare Commission and the Council of Voluntary Social Services, which is a coordinating body affiliated with 35 voluntary organizations. The government operates a compulsory National Insurance Scheme that covers all gainfully employed persons between 18 years old and retirement age.

Housing. Much attention has been paid to housing, and there are many large development schemes in both urban and rural areas, especially in the Kingston and St. Andrew suburbs. Although the government undertakes many types of housing schemes, its chief concern is with low-income projects.

Social and economic divisions. A study of family income distribution in the early 1960s indicated that 20 percent of the families enjoyed about 56 percent of the total household income.

The outstanding feature of the social structure is the degree of harmonious coexistence among people of different ethnic and cultural backgrounds. This is largely explained by the wide degree of racial intermixture, as well as by the common exposure to a Western-oriented value system. Rigidities in the class system are being gradually erased by measures which seek to increase social mobility through the expansion of opportunities in education and in ownership of property.

CULTURAL LIFE AND INSTITUTIONS

The arts. There is a vigorous and productive art movement in Jamaica, stemming from the 1930s and 1940s. The works of Jamaican novelists may be read in several languages. Jamaican artists have exhibited successfully abroad, and local art shows are a regular part of life. The Institute of Jamaica, an early patron and promoter of the arts, sponsors exhibitions and awards. It also runs the Jamaica School of Art.

The Jamaica Library Service and the University of the West Indies contribute to the promotion of the arts. There are many successful commercial galleries and one run by the Contemporary Jamaican Artists Association. The steadily increasing number of Jamaican art patrons indicates the involvement of the society with its own cultural identity.

There are many active theatre and musical groups. The National Dance Theatre Company, formed in 1962, has earned international recognition; its annual dance season is one of the most eagerly awaited events in the cultural calendar. Much of the country's artistic expression finds an outlet in Festival, sponsored annually by the government as part of the independence celebrations. While the festival has many features of the traditional Caribbean type of carnival, it is much wider in scope. In addition to street dancing and parades, there are also exhibitions of arts and crafts, and literary, theatrical and musical competitions.

Folk traditions

The concern with Jamaica's cultural tradition is evident in an artistic and cultural awakening accompanied by a keen search for roots in folk forms, which are based chiefly on the colourful, rhythmic intensity of an African heritage, with overtones of unique multiracial influences. Folk music, stories, and dances are being systematically sought out and recorded. The important aesthetic elements in some of the revivalist cults, notably Pocomania, are recognized, and modern dance and drama employ many of the folk expressions that might otherwise have disappeared.

Press and broadcasting. Jamaica has one daily morning paper and one afternoon tabloid. There are a number of periodicals, magazines, and journals and five weekend newspapers. There are two radio stations, one of which is a publicly owned corporation that also operates a television service.

PROSPECTS FOR THE FUTURE

The future development of the Jamaican economy depends to a large extent upon agricultural development. Agriculture, in the early 1970s, was expected to continue to employ the largest percentage of the labour force and to remain for some time one of the mainstays of the economy. More and more of the unused land was expected to be put into production, although further agricultural investment was needed for this to take place. As technology is increasingly applied to Jamaican industry, it was anticipated that future development would tend to require larger amounts of capital rather than of labour. The government, meanwhile, has emphasized the expansion of educational opportunities and in the course of the 1970s was expected to pay particular attention to primary education and to adult literacy training. Tourism and light manufacturing were both estimated to have a potential for further expansion.

BIBLIOGRAPHY. R.M. BENT and ENID L. BENT-GOLDING, ***A Complete Geography of Jamaica*** (1966), the standard work on the subject; CLINTON V. BLACK, ***The Story of Jamaica from Prehistory to the Present,*** rev. ed. (1965), the island's story from early times to independence; FREDERIC G. CASSIDY, ***Jamaica Talk*** (1961), a sound and entertaining study of the country's dialect; and with R.B. LE PAGE (eds.), ***Dictionary of Jamaican English*** (1967), the standard work on the subject; FRANK CUNDALL, ***The Governors of Jamaica in the Seventeenth Century*** (1936) and ***The Governors of Jamaica in the First Half of the Eighteenth Century*** (1937); and with JOSEPH PIETERSZ, ***Jamaica Under the Spaniards*** (1919), based on abstracts from the Archives of Seville; W.J. GARDNER, ***A History of Jamaica*** (1873), includes useful information on trade, agriculture, customs, and religion; PHILIP HENRY GOSSE, ***A Naturalist's Sojourn in Jamaica*** (1851); WALTER JEKYLL (ed.), ***Jamaican Song and Story*** (1907, reprinted 1966), an old but valuable work on the island's folklore; EDWARD LONG, ***The History of Jamaica,*** 3 vol. (1774), a badly arranged but invaluable 18th-century history; FRANCISCO MORALES PADRON, ***Jamaica Española*** (1952), a scholarly study; GEORGE W. ROBERTS, ***The Population of Jamaica*** (1957), an analysis of its structure, growth, and changes in internal distribution; CAREY ROBINSON, ***The Fighting Maroons of Jamaica*** (1969), a stirring account of the Maroons. Important serials include the ***Handbook of Jamaica*** (annual); the ***Jamaica Gazette*** (weekly); the publications of the Central Planning Unit of the Ministry of Finance, especially the ***Economic Survey;*** the publications of the Department of Statistics, especially the ***Population Census;*** and annual reports of various departments and ministries.

(C.V.B.)

Jamāl ad-Din al-Afghānī

Jamāl ad-Din al-Afghani (1838/39–97), Muslim politician, political agitator, and journalist whose belief in the potency of a revived Islamic civilization in the face of European domination significantly influenced the development of Muslim thought in the 19th and early 20th centuries. Afghānī considered Islām the potential source of Muslim solidarity and political strength and to this end devoted most of his life.

Early career in Afghanistan, Istanbul, and Egypt

Very little is known about Afghānī's family or upbringing. Despite the appellation Afghani, which he adopted and by which he is known, some scholars believe that he was not an Afghan but a Persian Shi'ite (*i.e.*, a member of the Shi'ah, one of the two major divisions of Islām), born in Asadābād near Hamadan in Persia. An appreciable part of Afghani's activities took place in areas where Sunnī Islām—the other major division of Islām—was predominant, and it was probably to hide his Persian and Shi'ite origin, which would have aroused suspicion among Sunnis, that he adopted the name Afghani. As a young man he seems to have visited, perhaps in order to extend and perfect his theological and philosophical education, Karbalā' and an-Najaf, the Shi'ite centres in southern Mesopotamia, as well as India and perhaps Istanbul. The intellectual currents with which he came in contact remain obscure, but whatever they were they made him into a religious skeptic at an early date. Only from about November 1866, when Afghani appeared in Qandahār, Afghanistan, can evidence he pieced together to form a consecutive and coherent picture of his life and activities. From the death in 1863 of the famous Dōst Mohammad Khān, who had ruled for more than 20 years, Afghanistan had been the scene of civil wars occasioned by the quarrels of his sons over the succession. In 1866 one of these sons, Shir ʿAlī Khān, was established in the capital, Kābul, but two of his brothers, Mohammad Afdal Khān and Mohammad A'zam Khan, were threatening his tenure. In January 1867 Shir ʿAlī was defeated and ousted from Kābul, where Afdal and, upon his death shortly afterward, A'zam reigned successively in 1867–68. At the end of 1866 A'zam captured Qandahār and Afghani immediately became Aʿẓam's confidential counsellor, following him to Kābul. He remained in this position until A'zam was in turn ousted by Shir ʿAlī, who succeeded in regaining his throne in September 1868.

That a foreigner should have attained so quickly such a position was remarked upon in the contemporary accounts; some scholars speculate that Afghani (who then called himself Istanbuli) was, or represented himself to be, a Russian emissary able to obtain for A'zam Russian money and political support against the British, with whom Aʿẓam was on bad terms. When Shir 'Ali succeeded in regaining the throne, he was naturally suspicious of Afghani and expelled him from his territory in November 1868.

Afghani next appeared in Istanbul in 1870, where he gave a lecture in which he likened the prophetic office to a human craft or skill. This view gave offense to the religious authorities, who denounced it as heretical. Afghani had to leave Istanbul and in 1871 went to Cairo, where for the next few years he attracted a following of young writers and divines, among them Muhammad 'Abduh, who was to become the leader of the modernist movement in Islām, and Sa'd Pasha Zaghlūl, founder of the Egyptian nationalist party, the Wafd. Again, a reputation of heresy and unbelief clung to Afghani. The ruler of Egypt then was the khedive Ismāʿīl, who was both ambitious and spendthrift. By the mid-1870s his financial mismanagement led to pressure by his European creditors and great discontent among all his subjects. Ismāʿīl tried to divert their wrath from himself to the creditors, but his manoeuvres were clumsy, and, in response to French and British pressure, his suzerain, the Ottoman sultan, deposed him in June 1879. During this period of political effervescence, Afghani attempted to gain and manipulate power by organizing his followers in a Masonic lodge, of which he became the leader, and by delivering fiery speeches against Ismāʿīl. He seems to have hoped to attract thereby the favour and confidence of Tawfiq, Ismāʿīl's son and successor; but the latter, reputedly fearing that Afghani was propagating republicanism in Egypt, ordered his deportation in August 1879.

Activities in India and elsewhere

Afghani then went to Hyderābād and later, via Calcutta, to Paris, where he arrived in January 1883; his stay there contributed greatly to his legend and posthumous influence as an Islamic reformer and a fighter against European domination. In Paris, Afghānī, together with his former student 'Abduh, published an anti-British newspaper, *al-ʿUrwat al-wuthqā* ("The Indissoluble Link"), which claimed (falsely) to be in touch with and have influence over the Sudanese Mahdi, a messianic bearer of justice and equality expected by some Muslims in the last days. He also engaged Ernest Renan, the French historian and philosopher, in a famous debate concerning the position of Islām versus science. He tried unsuccessfully to persuade the British government to use him as intermediary in negotiation with the Ottoman sultan, Abdülhamid II, and then went to Russia, where his presence is recorded in 1887, 1888, and 1889 and where the authorities seem to have employed him in anti-British agitation directed to India. Afghani next appeared in Iran, where he again attempted to play a political role as the Shāh's counsellor and was yet again suspected of heresy. The shāh, Nāṣer od-Din Shāh, became very suspicious of him, and Afghani began a campaign of overt and violent opposition. Again, in 1892, his fate was deportation. For this, Afghani revenged himself by instigating the Shāh's murder in 1896. It was his only successful political act.

From Iran, Afghani went to London, where he stayed briefly, editing a newspaper attacking the Shāh and urging resistance to him and particularly to the tobacco concession that had been granted to a British subject. He then went to Istanbul, in response to an invitation made by an agent of the Sultan's. The Sultan may have hoped to use him in pan-Islāmic propaganda, but he soon aroused suspicion and was kept inactive, at arm's length and under observation. He died in Istanbul. His burial place was kept secret, but in 1944 what was claimed to be his body, owing to the mistaken impression that he was an Afghan, was transferred to Kābul, where a mausoleum was erected for it.

BIBLIOGRAPHY. A. ALBERT KUDSI-ZADEH, *Sayyid Jamāl al-Din al-Afghani: An Annotated Bibliography* (1970). NIKKI R. KEDDIE, *An Islamic Response to Imperialism* (1968), includes the main writings of Afghani in translation together with a long introduction dealing with his life and ideas. ELIE KEDOURIE, *Afghani and 'Abduh: an Essay on Religious Unbelief and Political Activism in Modern Islam* (1966), sets forth and examines much new evidence on which a radical re-evaluation of Afghānī's role and significance is based. HOMA PAKDAMAN, *Djemal-ed-Din Assad Abadi, dit Afghani* (1969), contains a large number of documents by and about Afghānī and is a useful supplement to the works of Keddie and Kedourie.

(E.Ke.)

James I the Conqueror, of Aragon

James I of Aragon, one of the great Spanish monarchs of the Middle Ages, took a leading part in the reconquest (Reconquista) of Spanish lands from the Muslims, adding new territories within the peninsula to his own dominions and opening up the sea routes for overseas conquests. He thus initiated the Catalan–Aragonese expansion in the Mediterranean that was to reach its zenith in the last decades of the 14th century.

Birth and early life

The son of Peter II of Aragon and Mary of Montpellier, James was born at Montpellier, France, on February 2, 1208. When Peter, allied with the Albigensian heretics, died fighting against the crusaders sent against them at the Battle of Muret, James was only five years old and was at Carcassonne, France, in the hands of the crusaders' leader, Simon de Montfort. James was released in April 1214 and recognized as sovereign in Aragon and Catalonia; placed under the protection of the Knights Templars at Monzón, he was cared for and educated by them. The regency was exercised by his great-uncle, Count Sancho of Roussillon (in Aragon, now in France), until 1218, when Sancho resigned in the face of opposi-

James I, detail of an illumination from the *Crónica de Jaime I.* In the University of Barcelona.
Archivo Mas, Barcelona

tion from some Aragonese and Catalonian nobles. The ensuing rebellions, during which the King often found himself in great danger, formed a hard school for the forging of his character. Fearless already as a youth, he fought an Aragonese noble in hand-to-hand combat, took part in the siege of the port of Castejón in 1222, and three years later tried to seize another port.

In 1227 James took over the effective government of his kingdoms and at once began the first of his great campaigns of reconquest—that of the Balearic Islands. Majorca was captured in December 1229, and the occupation was rounded off in 1235 by the conquest of Ibiza by the Bishop of Saragossa. Thenceforward, the islands were a bulwark to defend the Catalan coasts and a base from which trade and political expansion could be launched eastward.

Wars of reconquest against the Muslims

In 1233 James began a second war of reconquest—against the Saracen rulers of the Kingdom of Valencia. The campaign lasted three long years and suffered various interruptions before the capital itself was captured in 1238. The occupation of the kingdom was completed later with the capture of other towns, and in 1244 James signed, with Alfonso X of Castile, a treaty by which the boundaries of their respective kingdoms were delimited in the newly conquered areas.

James I married twice. In 1221 he married Leonor, daughter of Alfonso VIII of Castile, but he later divorced her and in 1235 married the daughter of Andrew II of Hungary, Yolande, by whom he had many children. In 1248 and 1262 he divided his realms among his sons but only succeeded in causing virulent civil strife. In the second division, his elder son, Peter, received Aragon, Valencia, and Catalonia, and his younger son, James, received the Balearic Islands, Roussillon, and other Pyrenean counties that he was to hold in fief from Peter. This division of realms among his heirs was not James's only political blunder. By the Treaty of Corbeil (1258) he renounced his claims to territories in the south of France, thus abandoning the traditional policy that the Catalan dynasty had hitherto pursued across the Pyrenees. He was, however, able to develop relations and promote trade with the states of North Africa; and, with a clear view of the future, he married his principal heir, Peter, to Constance of Sicily, thus making it easy for the latter kingdom to be added in later years to the crown of Aragon. Always a chivalrous soldier, James helped his son-in-law Alfonso X of Castile to suppress the rebellion of the Moors in the Kingdom of Murcia (1266); he also set out on a crusade to the Holy Land (1269), though this was a failure. He died at Valencia on July 27, 1276.

Assessment

A soldier of extraordinary courage and great gifts of leadership, James was a stout man, strong and handsome; he has been criticized for his many love affairs that caused him to be described as an "home de fembres" ("lady's man"). On balance, his reign was very beneficial. The important code of maritime law called the *Llibre del consolat del mar* was compiled; the Kingdom of Valencia received its own legal system; various cities, including Barcelona, acquired their own civic administrations; and the Cortes—the representative assembly—came into being. The King protected men of letters, inspired the chronicle that bears his name (though he did not himself write it), and brought his different peoples to a degree of political and cultural maturity that can reasonably be described as admirable.

BIBLIOGRAPHY. The chief source for James I's reign is his chronicle, called the *Libre dels feyfs esdeuenguts en la uida del. molt alt senyor Rey, en Jacme lo Conqueridor*. The latest edition is by J.M. DE CASACUBERTA, *Crdnica*, 9 vol. (1926–62); and there is an English translation by JOHN FORSTER, *The Chronicle of James I., King of Aragon, Surnamed the Conqueror*, 2 vol. (1883), for which PASCUAL DE GAYANGOS provided a historical introduction and footnotes. FERRAN SOLDEVILA has written three fundamental works on James I: *Els primers temps de Jaume I* (1968), an exhaustive biographical study of the King's early years; *Vida de Jaume I el Conqueridor* (1958), covering the King's whole life; and *Jaume I, Pere Gran: Els grans reis del segle XIII.*, 2nd ed. (1965), a general study of the reign.

(E.Sa.)

James I of Great Britain

England's first Stuart king, reigning from 1603 to 1625, James I also ruled Scotland as James VI from 1567. A strong advocate of royal absolutism, he promoted peace in Europe and colonial expansion in America and began disputes with Parliament that finally erupted into civil war in the reign of his son, Charles I.

By courtesy of the National Maritime Museum, Greenwich, England

James I, painting attributed to John de Critz, *c.* 1620. In the National Maritime Museum, Greenwich, England.

Proclaimed king

James was born in Edinburgh Castle on June 19, 1566, the only son of Mary, Queen of Scots, and her second husband, Henry Stewart, Lord Darnley. Eight months later his father died when his house was destroyed by an explosion. After her third marriage, to James Hepburn, earl of Bothwell, Mary was defeated by the rebel lords and abdicated the throne. James, one year old, became king of Scotland on July 24, 1567; Mary left the kingdom on May 16, 1568, and never saw her son again. During his minority James was surrounded by a small band of the great Scottish lords, from whom emerged the four successive regents, the earls of Moray, Lennox, Mar, and Morton. There did not exist in Scotland the great gulf between rulers and ruled that separated the Tudors and their subjects in England. For nine generations the Stuarts had in fact been merely the ruling family among many equals, and James all his life retained a feeling for those of the great Scottish lords who gained his confidence.

The young King was kept fairly isolated but was given a good education until the age of 14. The course had been rather narrow, mainly three languages, Greek, French, and Latin, always set within a biblical (chiefly Old Testament) framework. His mother's books provided him with an introduction to the French poets. More useful than all his tutors' teaching was the library, of classical, religious, and, to a lesser extent, historical writings that his tutors, George Buchanan and Peter Young, assembled for him.

In his youth King James wrote poetry, with energy but slender talent. His works included a long translation from the French of Guillaume de Salluste, Seigneur du Bartas, a Gascon Huguenot, and in the whole series of the King's verse there is only one reference to Scottish scenery, to the cloud-topped Cheviot Hills. The young sovereign impressed his Scottish subjects considerably with his great knowledge.

When he grew to manhood he was slender and of average height. He was constantly on horseback and loved the chase, hunting dogs, and horses, activities that endeared him to the people of Scotland. Later, however, his English subjects noticed his weak and spindly legs and his narrow jaws, which made it difficult for him to eat.

Even as a young man, he liked to influence those around him, an outgrowth of his conception of his great position. From the time he grew up his most salient quality was his pleasure in giving, accompanied, unfortunately, by an absence of any money sense. Through his life he also had a deep desire for peace.

Dispute with Presbyterians

Though he opposed the attempts of Presbyterian ministers to undermine royal authority in Scotland, James was always a convinced doctrinal Presbyterian. In his later years he would be described in general terms as an Anglican; but he never desired to introduce the English creed into his native country. It is clear that he considered his differences with the Presbyterian ministers in the years of his Scottish reign as political rather than doctrinal. Their form of rule was ultimately theocratic and therefore incompatible with his idea of monarchy.

He retained through life a vivid belief in the power and enmity of the devil and was deeply harassed by this belief. Uncommunicative and brooding, he always kept to himself. He was very sure of his decisions. He was only 15 when he arranged for the execution of the Regent Morton on the charge of murdering his father, Darnley. In 1600 he became annoyed at the opposition of John Ruthven, the earl of Gowrie, to his tax proposals and was, in fact, disturbed by the very proximity of the Gowrie family, and, in August of the year, at the Gowrie home, the earl and his brother, Alexander, were killed in mysterious circumstances. James claimed that his life had been threatened and the killings occurred in the ensuing struggle, but the Gowrie conspiracy has never been satisfactorily explained.

In August 1589, James was married to Anne, the daughter of Frederick II of Denmark who, in 1594, gave birth to their first son, Prince Henry. Throughout his life, however, James exhibited little interest in the opposite sex. It seems that he never had a mistress and was interested in women only as the wives and mothers of his male friends. The first friend who influenced him greatly was his cousin Esmé Stuart, seigneur of Aubigny, whom he created duke of Lennox. James was no traveller and relied on Lennox for information about the French court.

From early manhood James's policy was determined by his need to remain on good terms with the childless Elizabeth of England, in order to ensure his claim as heir to the English throne. When his mother, who had been a prisoner in England for 20 years, was executed at Fotheringhay in 1587, he lodged only a formal protest. He realized that he was in no position to attack the English polity and that it would not be in his best interest to do so.

Daemonologie

In 1597, in Scotland, he published a treatise *Daemonologie*, which gives a clear idea of his fear of the spirit world, and two years later he brought out a first small, private edition of the ***Basilikon Doron,*** described as "His Majesties instructions to his dearest sonne Henry the Prince." Somewhat narrow in scope, the examplars are drawn almost entirely from Scotland, and the theological passages are of low intensity.

In this part of the reign the administration of the realm was carried on with the King's full consent by a group known as the Octavians. There were certain difficulties with the Kirk, the (Presbyterian) Church of Scotland, but the King showed himself both cool and crafty. Because the Kirk had no power to reward, its influence in the King's circle was slight. It was also clear that his conflict with the Church of Scotland was without danger; it did not call upon James's small store of physical courage.

In the last years of the reign of Queen Elizabeth, a private correspondence began between King James and Sir Robert Cecil, Queen Elizabeth's secretary of state, who was to serve him after his accession. A sense of the glory of kingship bound James to the King of France and to the King of Spain, England's traditional enemies. This highly defined sense of identity with others of his estate led James to believe that he could bring England peace, as indeed he was to do in 1604 by speedily ending the war with Spain.

King of England

When Elizabeth died in 1603, King James set out for England, leaving the ministers to whom he was long accustomed in their posts in Scotland. During the later part of his reign he returned to Scotland only once, in 1617. Throughout his reign in England he received Scotsmen at his court, but they were not liked in his southern kingdom. Shortly after he moved to England, he ceased marital relations with his wife. Though he appreciated the Queen's concern for masques and pageantry, he disliked her recurring interest in Roman Catholicism and did not wish her to influence his children's minds. Still he always remained courteous to her as the King of Denmark's sister. She died in 1619.

The happiest period of James's rule was the nine years between his accession to the English throne and the death of Cecil, whom he created earl of Salisbury. He never had another minister who was both experienced and powerful. In the later portion of his reign his two successive favourites, Robert Carr, who became Viscount Rochester and earl of Somerset, and George Villiers, who was created duke of Buckingham, exerted an extraordinary influence over the King.

But James was not appreciated in England as he had been in Scotland. He hardly understood the rights or the temper of the English Parliament, and thus came into conflict with it. He had little contact with the English middle classes, and he suffered from the narrowness of his horizons. Elizabeth had travelled throughout her country and played her part before the nation; the crowds along the roadside all could hear her calculated loud asides. But James was domestic; perhaps his outdoor servants, his huntsmen, and the keepers of his hunting lodges were the only members of the rural working class with whom he came in contact. In his later years his judgment faltered. The King had a sympathy, which his countrymen found inexplicable, for the Spanish ambassador, Diego Sarmiento de Acuña, the count of Gondomar. When Sir Walter Raleigh, who had been released from the Tower of London and a death sentence and had gone to Guiana in search for gold, came into conflict with the Spaniards, then at peace with England, Gondomar persuaded James to have Raleigh beheaded. The King's last decisive action, also a result of Gondomar's influence, was the arrangement of Prince Charles's journey to Madrid to seek a Spanish princess in marriage.

Death

The Duke of Buckingham began an enmity with Prince Charles, who became the heir when his brother Prince Henry died in 1612, but in the course of time the two formed an alliance from which the King was quite excluded. In the last 18 months of the reign the King, in effect, exercised no power; Charles and Buckingham decided most issues. James died at his favourite country residence, Theobalds, in Hertfordshire, on March 27, 1625.

BIBLIOGRAPHY. The ***Poems of James VI of Scotland,*** 2 vol., were edited by J. CRAIGIE for the Scottish Texts Society in 1955–58; and the 1616 edition of the ***Political Works of James VI*** was edited by C.H. MCILWAIN in 1918. The ***Correspondence of James VI of Scotland with Sir R. Cecil and Others in England*** was edited by JOHN BRUCE for the Camden Society in 1861.

A good recent biography is D.H. WILLSON, ***King James VI and I*** (1956). See also DAVID MATHEW, ***The Jacobean Age*** (1938) and ***James I*** (1967); CHARLES WILLIAMS, ***James I*** (1934); and H. ROSS WILLIAMSON, ***King James I*** (1935).

(D.Ma.)

James II of Great Britain

The last Stuart monarch of the direct male line, James II was king of Great Britain from 1685 until his forced "abdication" in the so-called Glorious or Bloodless Revolution of 1688. That revolution', engendered by James's Roman Catholicism, permanently established the supremacy of Parliament as the ruling power of England.

James II was born at St. James's Palace, London, on October 14, 1633, the second surviving son of Charles I and Henrietta Maria. He was created duke of York in January 1634. During the Civil War he lived at Oxford—from October 1642 until the city surrendered in June 1646. He was then removed by order of Parliament to St. James's Palace, from which he escaped to the Netherlands in April 1648. He rejoined his mother in France in early 1649. Joining the French army in April 1652, he served in four campaigns under the great French general Turenne, who commended his courage and ability. When his brother, Charles II, concluded an alliance with Spain against France in 1656 he reluctantly changed sides, and he commanded the right wing of the Spanish army at the Battle of the Dunes in June 1658.

By courtesy of the National Portrait Gallery. London

James II, painting by Sir Godfrey Kneller, c. 1685. In the National Portrait Gallery, London.

Heir presumptive

At the restoration of his brother Charles to the English throne in 1660, James was lord high admiral and did much to maintain the efficiency and improve the organization of the navy. He also showed considerable interest in colonial ventures; it was on his initiative that New Amsterdam was seized from the Dutch in 1664 and renamed New York in his honour. He commanded the fleet in the opening campaigns of the Second and Third Dutch wars. This was to be his last taste of active military command until 1688. In politics he was a strong supporter of the Earl of Clarendon, whose daughter Anne he married in September 1660. Both before and after marriage he had the reputation of being as great a libertine as his brother. But in 1668 or 1669 he was admitted to the Roman Catholic Church, though, on his brother's insistence, he continued to take the Anglican sacraments until 1672, and he attended Anglican services until 1676. Charles II also insisted that James's daughters, Mary and Anne, be raised in the Protestant faith.

James's conversion had little effect on his political views, which were already formed by his reverence for his dead father and his close association with the High Church Party. James, in fact, was always more favourable to the Anglican Church than was his Protestant brother. He welcomed the prospect of England's re-entering the European war on the side of the Dutch; and he consented to the marriage of his elder daughter, Mary, to William of Orange in 1677. For most of his life James was the spokesman of the conservative Anglican courtiers, who believed that his views on monarchy and Parliament coincided with theirs, who found his formal and humourless nature more congenial than Charles's slippery geniality, and who respected his frank acknowledgment of his religious beliefs.

In view of the Queen's childlessness, however, the conversion of the heir presumptive to the throne roused great alarm in the general public. James resigned all of his offices in 1673 rather than take an anti-Catholic oath imposed by the so-called Test Act and thus made his position known publicly. Later that year, his first wife having died, he gave further offense by marrying a Roman Catholic princess, Mary of Modena. By 1678 James's Catholicism had created a climate of hysteria in which the fabricated tale of a "Popish Plot" to assassinate Charles and put his brother on the throne was generally believed. From 1679 to 1681 three successive Parliaments strove to exclude James from the succession by statute. During this crisis James spent long periods in exile at Brussels and Edinburgh. But owing largely to his own tenacious defense of his rights, the exclusionists were defeated. In 1682 he returned to England and resumed the leadership of the Anglican Tories, whose power in local government was re-established and increased by the "remodelling" of the borough corporations and the government of the counties in their favour. By 1684 James's influence on state policy was paramount, and when he finally came to the throne on February 16, 1685, with very little overt opposition or even criticism, it seemed likely that the strong support of the Anglicans would make him one of the most powerful of the 17th-century British kings.

Accession

The new royalist Parliament that assembled in May 1685 voted James a large income and there seemed no reason why he should not in time secure adequate toleration for his coreligionists. But unsuccessful rebellions led by the Duke of Monmouth in England and the Duke of Argyll in Scotland, in the summer of 1685, marked a turning point in his attitude. James's distrust of his subjects, conceived in the turbulent 1670s, was at once sharpened. The rebellions were put down with great ferocity, the army was considerably increased, and the new regiments were granted to Roman Catholic officers who had had military experience abroad and whose loyalty was undoubted. This last act of policy provoked a quarrel between King and Parliament, which was prorogued in November 1685, never to meet again. In 1686 the division between the King and his former allies, the Anglican Tories, deepened. After many of them had been replaced, the judges of King's Bench in the collusive action *Godden* v. ***Hales*** found in favour of the king's power to excuse individuals from the Test oath; Roman Catholics were admitted to the privy council and later to the high offices of state. A commission for ecclesiastical causes was set up to administer James's powers as supreme governor of the Anglican Church, and its first act was to suspend Henry Compton, bishop of London, one of the most outspoken critics of royal policy.

Religious policy

In 1687 James intensified his Roman Catholic policy and dismissed his Anglican brothers-in-law Clarendon and Rochester. Magdalen College, Oxford, was given over for the use of Roman Catholics, and a papal nuncio was officially accredited to St. James's. In April James issued the so-called Declaration of Indulgence suspending the laws against Roman Catholics and Protestant dissenters alike; in July he dissolved Parliament, and in September he launched an intensive campaign to win over the Protestant dissenters and with their aid secure a new Parliament more amenable to his wishes.

What those wishes were is still not clear: some of his utterances suggest a genuine belief in religious toleration as a matter of principle; others point to the establishment of Roman Catholicism as the dominant if not the exclusive religion of the state. This confusion may well reflect the state of James's own mind, which undoubtedly deteriorated in the years 1687–88, and some of his assertions, accusations, and threats at this time verge on the insane.

The unexpected news that the Queen was pregnant (November 1687), establishing the prospect of a Roman Catholic succession, had great effect on most Protestants; while a wholesale "remodelling" of borough corporations, lord lieutenancies, deputy lieutenancies, and magistracies that winter inflamed the majority of the nobility and gentry, whose political and social power suffered by it. Ever since the spring of 1687 many English leaders had been in touch with William of Orange, the husband of the heiress presumptive and the champion of Protestant Europe against Louis XIV. The spark was touched off by James himself, when he re-issued his Declaration of Indulgence on May 7, 1688, and on May 14 ordered it to be read in the churches. The Archbishop of Canterbury and six of his bishops petitioned James to withdraw the order. Their petition was subsequently published, and James made the mistake of prosecuting its authors for seditious libel. Meanwhile, on June 20, in slightly mysterious circumstances, the Queen gave birth to a son.

The "Glorious Revolution"

On July 10 the seven bishops were acquitted—a tremendous defeat for the government—and that same day seven leading Englishmen sent a letter inviting William of Orange to lead an army to England and call a free parliament to arbitrate on the legitimacy of the Prince of Wales. By September William's intentions were obvious, but James declined Louis XIV's offer of assistance for fear of the reaction in England; in any case he was confident in the ability of his forces to repel invasion. William sailed in late October, under cover of the general war that had by then broken out in Europe, evaded the English fleet, and landed at Tor Bay on November 15, 1688. In the subsequent "campaign," James's Protestant officers deserted to the enemy in such large numbers that he dared not commit the army to a pitched battle. This, together with the defection of his daughter Anne, finally shattered his nerve. He attempted to flee to France but was intercepted at Kent on December 21; 12 days later he was allowed to escape. On February 22, 1689, the Convention Parliament declared that James had abdicated, and next day offered the crown to William and Mary. The Scots Parliament followed suit in May.

In March 1689 James landed in Ireland, and a Parliament summoned to Dublin acknowledged him as King. But his Irish–French army was defeated by William at the Boyne (July 11, 1690), and he returned to France. William's generals reconquered Ireland the following year. In Ireland James had shown none of his former military ability and he now aged rapidly, falling increasingly under the influence of his pietistic wife. He became daily more absorbed in his devotions, and his more aggressive supporters soon came to regard him as something of a liability. The Treaty of Rijswijk between England and France (1697) removed his last hopes of restoration, and he died at Saint-Germain on September 16, 1701.

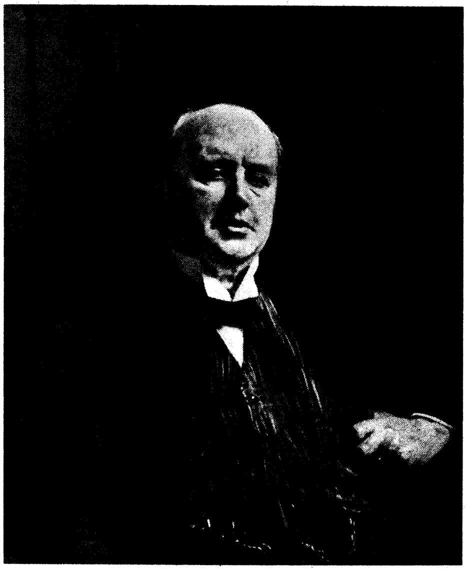
BIBLIOGRAPHY. F.C. TURNER, *James II* (1948), is the definitive biography and a very good one. MAURICE ASHLEY, *The Glorious Revolution of 1688* (1966), is brief but the best modern account. LORD MACAULAY, *History of England,* ch. *4–10* (1849–65), is biassed and in some respects out-of-date, but it is still the most detailed factual account.

(J.P.K.)

James, Henry

Henry James's career was one of the longest and most productive—and most influential—in American letters. A master of prose fiction from the first, he practiced it as a fertile innovator, enlarged the form, and placed upon it the stamp of a highly individual method and style. He wrote for 51 years—20 novels, 112 tales, 12 plays, several volumes of travel and criticism, and a great deal of literary journalism. Beyond his craft, he made himself a formidable figure in the transatlantic culture. He recognized and helped to fashion the myth of the American abroad—the encounter of the New World with the Old—and incorporated this myth in the "international novel" of which he was the artisan and master. His fundamental theme was the innocence of the New World, the corruption—and wisdom—of the Old, and the clash between the two. In his vision of an exuberant and democratic America confronting an alien aristocratic culture he had a troubled and prophetic sense of America's coming-of-age and the moral problems the United States would encounter as a world power in the 20th century.

By courtesy of the National Portrait Gallery, London



Henry James, oil painting by John Singer Sargent, 1913. In the National Portrait Gallery. London.

Early life and works. James was born April 15, 1843, in New York City, just off Washington Square, which would give its name to one of his celebrated short novels. Although often characterized as a New Englander, he was actually a child of the Hudson Valley and, unlike the rural writers of Concord, was firmly rooted in an early urban America. His grandfather, an Irish immigrant who went to Albany just after the American Revolution, amassed one of the largest fortunes in the state; and while this was divided among numerous children, there was sufficient to give the future novelist a comfortable childhood. He was named for his father, Henry James (1811–82), a brilliant conversationalist and religious visionary, much influenced by Swedenborg and concerned in his writings with man's relation to God. The younger Henry was a second son; and he and his older brother, William (1842–1910), founder of psychological study and research in America and philosopher of pragmatism, were taken abroad as infants. In the ensuing two decades they crossed and recrossed the Atlantic, were schooled by tutors and governesses, and spent their preadolescent years in the then still parochial Manhattan. Returned to Geneva, Paris, and London during their teens, the James children acquired languages and an awareness of Europe vouchsafed to few Americans in their time. This phase of the novelist's life is vividly recorded in the autobiographies he wrote late in life, ***A Small Boy and Others*** (1913) and ***Notes of a Son and Brother*** (1914), in which he showed the roots of his precocious cosmopolitanism as well as early, vivid impressions of "old New York."

Education and travel

Henry James was a shy, book-addicted boy who assumed the role of quiet observer beside his active elder brother. From the first he was content to be "simply literary." On the eve of the Civil War the James family

Early short stories

settled at Newport, R.I., and there, and later in Boston, Henry came to know New England intimately. A back injury suffered as a volunteer fireman at Newport caused him to lead a sedentary life for some years and diminished his sense of guilt at his recoil from the fratricidal Civil War. (Years later, in a short story, he would portray a young pacifist, prepared to die for his beliefs, thereby proving himself a good soldier.) For a few months he entertained the illusion that he should have a professional career; in 1862, when he was 19, he enrolled at the Harvard Law School, but devoted his study time to Sainte-Beuve, Balzac, and Hawthorne. His first story appeared anonymously two years later in the New York *Continental Monthly.* His second tale, dealing with civilian life during the Civil War, appeared in *The Atlantic Monthly* in 1865 when he was 22. By this time he was reviewing books regularly in the *North American Review;* and when William Dean Howells became editor of the *Atlantic,* James found in him a friend and mentor who published him regularly. Between them, James and Howells inaugurated the era of American "realism." By his mid-20s James was regarded as one of the most skillful writers of short stories in America. Critics, however, deplored his tendency to write of the life of the mind, rather than of action. The stories of these early years show the leisurely existence of the well-to-do at Newport and Saratoga. They contain within them echoes of George Sand, Balzac, and Mérimée. His English model was George Eliot, his American, Hawthorne. His apprenticeship was thorough. He wrote stories, reviews, and articles for almost a decade before he attempted a full-length novel. There had to be also the traditional "grand tour," and James went abroad for his first adult encounter with Europe in 1869. In England he met Ruskin, Rossetti, William Morris, Leslie Stephen, and Darwin; he journeyed thence to Italy, which he had never visited as a boy. The Italian experience was profound, and he returned there repeatedly during his lifetime. His first important novel was set in Rome; and the Italian cities figure in his major works. While he was in England his beloved cousin Minny Temple, to whom he had formed a deep but reticent attachment, died. This shock, and the high emotion of his "passionate pilgrimage," provided much of the experience that would figure in works such as *The Portrait of a Lady* of his middle period and *The Wings of the Dove* of his final years. His year's wandering in England, France, and Italy set the stage for a lifetime of travel, always along the familiar and well-beaten paths of these countries. James never married. By nature he was friendly and even gregarious, but while he was an active observer and participant in society, and in his own literary world, he tended, until late middle age, to be "distant" in his relations with people and was careful to avoid "involvement."

Career—first phase. Recognizing the appeal of Europe, given his cosmopolitan upbringing, James made a deliberate effort to discover whether he could live and work in the United States. He spent 1870 to 1872 in Boston, during which he published his first significant American-European tale, "A Passionate Pilgrim," recording in it a deep yearning for historic England; and an early novel of Boston life, *Watch and Ward.* Two more years in Europe, mainly in Rome, and a winter of unremitting hackwork in New York convinced him that he could write better and live more cheaply abroad. Thus began his long expatriation—heralded by publication in 1875 of *Roderick Hudson,* the story of an American sculptor's struggle by the banks of the Tiber between his art and his passions. It was in effect James's "portrait of the artist as a young man." Simultaneously with this novel, James brought out *Transatlantic Sketches,* his first collection of travel writings, and a collection of tales. With these three substantial books, he inaugurated a career of authorship (hitherto confined to magazine writing) that saw some 100 volumes through the press during the next 40 years.

Paris and London international novels

During 1875–76 James lived in Paris, writing literary and topical letters for the *New York Tribune* and working on his novel *The American.* He sought out the Russian novelist Ivan Turgenev, whose work appealed to him, and through Turgenev was brought into Flaubert's *cénacle,* where James got to know Edmond de Goncourt, Zola, Daudet, and the then still unpublished Maupassant. James admired the intensity of the French writers, deplored their "unclean" subjects, and felt closest to Turgenev. From the Russian cosmopolite he received confirmation of his own view that a novelist need not worry about "story," that in focussing on character, he would arrive at the life experience of his protagonist. Much as he liked France, James felt he would be an eternal outsider there, and late in 1876 he crossed to London. There, in small rooms in Bolton Street off Piccadilly, he wrote the major fiction of his middle years. In 1878 he achieved international renown with his story of an American flirt in Rome, *Daisy Miller.* In England he was promptly taken up by the leading Victorians and became a regular at Lord Houghton's breakfasts, where he consorted with Tennyson, Gladstone, Browning, and others. A great social lion, James dined out 140 times during 1878 and 1879 and visited in many of the great Victorian houses and country seats. Elected to London clubs, publishing his stories simultaneously in English and American periodicals, mingling with Meredith, Robert Louis Stevenson, Edmund Gosse, and other writers, a familiar in the studios of Whistler and Sargent, he established himself as a significant figure in Anglo-American literary and artistic relations, along with his friend the poet and essayist James Russell Lowell, who in the 1880s was Washington's minister to the court of St. James's.

James's reputation was founded on his versatile studies of the "American girl." In a series of witty tales he pictured the "self-made" girl, the flirt, the bold, brash American innocent who insists upon American standards in European society. James ended this first phase of his career by producing his masterpiece, *The Portrait of a Lady,* a study of a young woman from Albany who brings to Europe her narrow provincialism and pretensions, but also her sense of her own sovereignty, her "free spirit," her refusal to be treated, in the Victorian world, merely as a marriageable object. As a picture of Americans moving in the expatriate society of England and of Italy this novel has no equal in the history of modern fiction. It is a remarkable study of a band of egotists while at the same time offering a shrewd appraisal of the American character. It embodies the national myth: an ideal of freedom and equality hedged with historical blindness and pride; a self-interest which often takes generous form; a sense of hurt when this generosity is interpreted as a wielding of power. James's understanding of power in personal relations was profound.

Novels of the 1880s

Short fiction of the second phase

Second phase—reformers and revolutionaries. Feeling that he had shot his bolt with his international novels and tales, James turned to new subjects during his second phase. In the 1880s he wrote two novels dealing with social reformers and revolutionaries, *The Bostonians* and *The Princess Casamassirna.* In the novel of Boston life James analyzed the struggle between conservative masculinity embodied in a Southerner living in the North and an embittered man-hating suffragette. *The Bostonians* remains the fullest and most rounded American social novel of its time in its study of cranks, faddists, and "do-gooders." In *The Princess Casamassima* James exploited the anarchist violence of the decade and depicted the struggle of a man who toys with revolution and is destroyed by it. These novels were followed by *The Tragic Muse,* in which James projected a study of the London and Paris art studios and the stage, the conflict between art and "the world." The novel raised the curtain on his own "dramatic years," 1890–95, during which he tried to win success on the stage. His dramatization of *The American* in 1891 was a modest success; but an original play, *Guy Domville,* produced in 1895, was a failure, and James was booed at the end of the first performance. Crushed and feeling that he had lost his public, he spent several years seeking to adapt his dramatic experience to his fiction. The result was a complete change in his storytelling methods. In *The Spoils of Poynton* (1897), *What Maisie Knew* (1897), *The Turn of the Screw* and *In the Cage* (1898), and *The Awkward Age* (1899) James began to use the methods of alternating "picture" and dramatic scene, close adherence to a given angle of vision, a withholding of information from the reader, making available to him only that which the characters see. The subjects of this period are the developing consciousness and moral education of children, principally young girls—in reality James's old international theme of innocence in a corrupting world, transferred to the English setting.

Final phase. The experiments of this "transition" phase led James to the writing of three grandiose novels at the beginning of the new century, which represent his final—his "major"—phase, as it has been called. In these novels James pointed the way for the 20th-century novel. He had begun as a Balzacian realist who describes minutely his crowded stage. He ended by leaving his stage comparatively bare, and showing a small group of characters in a tense situation, with a retrospective working out, through multiple angles of vision, of their drama. In addition to these technical devices he resorted to an increasingly allusive style, which became dense and charged with symbolic imagery. His late "manner" derived in part from his dictating directly to a typist; and in part from his unremitting search for ways of projecting subjective experience in a flexible prose, the sort of writing Proust would do after him in France. At the same time he extended his original subject from the narrow study of personal and interpersonal dilemmas to a view of society and civilization within which his characters struggle for self-identity and survival.

The three novels of the "major" phase

The first of the three novels was ***The Ambassadors,*** written in 1901 but published in 1903. This is a high comedy of American-European manners, of a middle-aged American who goes to Paris to bring back to a Massachusetts industrial town a wealthy young man who, in the view of his affluent family, has lingered too long abroad. The "ambassador" in the end is captivated by civilized Parisian life. The novel is a study in the growth of perception and awareness in the elderly hero, and it balances the relaxed moral standards of the Continent against the parochial rigidities of New England. The work is important not only for the story it tells, but for the way in which it is told, largely through the perceptions of the hero and the unfolding of his experiences.

The second of this series of novels was ***The Wings of the Dove,*** published in 1902, before ***The Ambassadors,*** although written after it. This novel, dealing with a melodramatic subject of great pathos, that of an heiress doomed by illness to die, avoids its cliche subject by focussing upon the characters surrounding the unfortunate young woman. They intrigue to inherit her millions. Told in this way, and set in London and Venice, it becomes a powerful study of well-intentioned humans who, with dignity and reason, are at the same time also birds of prey. In its shifting points of view and avoidance of scenes that would end in melodrama, ***The Wings of the Dove*** demonstrated the mastery with which James could take a tawdry subject and invest it with grandeur. His final novel was ***The Golden Bowl*** (1904), a study of adultery, with four principal characters; the first part of the story is seen through the eyes of the aristocratic husband and the second through the developing awareness of the wife.

The three novels transcend mere "plot" and "character" or the melodrama of their themes; in the comedy of ***The Ambassadors,*** in the soap-opera plot of the ***Wings,*** and in the high drama of the ***Bowl,*** James wrote a continuous picture of a flawed civilization that could preserve itself only by recognizing human privacy and freedom and the significance of moral and spiritual "controls." James's late novels are "philosophical" studies not only of Western society's modes of survival; they embody a social ethic that suggests that man in the West must cultivate the molds and traditions—however defective—that have shaped him out of chaos, and cherish the artifacts and forms by which he lives.

While many of James's tales were potboilers written for the current magazines, he achieved high mastery in the ghostly form, notably ***The Turn of the Screw,*** and in such remarkable narratives as "The Aspern Papers" and "The Beast in the Jungle"—his prophetic picture of dissociated 20th-century man lost in an urban agglomeration. As a critic James tended to explore the character and personality of writers as revealed in their creations; his essays are a brilliant series of studies, moral portraits, of the most famous novelists of his century, from Balzac to the Edwardian realists. His travel writings, ***English Hours, Italian Hours,*** and ***A Little Tour in France,*** portray the backgrounds James used for his fictions. Both in his light comedies and his tragedies, James's sense of the human scene was sure and vivid; and in spite of the mannerisms of his later style he was one of the great prose writers and stylists of his century.

James as a critic

Visit to America

In his later years, James lived in retirement in an 18th-century house at Rye in Sussex; he was called "Master" by Conrad, Ford Madox Ford, Max Beerbohm, and the younger Edwardians; and while he had a limited public after the 1880s he became the symbol of the literary establishment, the fount and oracle of the "modern novel." On completion of ***The Golden Bowl*** James in 1904–05 revisited the United States. He had lived abroad for 20 years, and in the interval America had become a great industrial and political power. James returned to the scenes of his childhood and youth in New York and in New England, travelled south to Florida and west to California; he lectured as he went on "The Lesson of Balzac," and was much in demand by literary and cultural organizations. His observation of the land and its people led him to write, on his return to England, a poetic volume of rediscovery and discovery, ***The American Scene*** (1907), prophetic in its vision of urban doom, spoliation and pollution of resources, and filled with misgivings over the anomalies of a "melting pot" civilization. The materialism of American life, the new age of advertising and public relations, the annihilation of America's modest past—the constant emphasis on "here-and-now"—deeply troubled James. On his return to England he set to work to shore up his own writings, and his own career, against this ephemeral world. He devoted three years to rewriting and revising his principal novels and tales for the highly selective "New York Edition" published in 24 volumes. He included 12 of his 20 novels and 66 of his 112 tales. The tales are grouped thematically rather than chronologically, as in Balzac's *Comédie humaine.* For this edition James wrote 18 significant prefaces, which contain both reminiscence and exposition of his theories of fiction. As theorist, he explains his use of a "central intelligence" in certain of his novels to provide maximum illumination of the human and psychological history he is recording; and he explains also his preference for a realism in which the story is narrated through successive scenes with minimal use of the old-fashioned omniscient author. In both his theory and practice James was a votary of aesthetic beauty infused with moral value; he believed in man's intellectual and reflective power, his sentience and "awareness," his ability to explore and grow in the understanding of his inner world. In this way he anticipates some of the insights of modern psychology. In his late work he is a dedicated solipsist, determined to reach as deeply as possible into man's inner life, and arrive at the integration of the self.

Travel writings and essays

During the Edwardian years James also revised his travel writings. The most important are included in ***English Hours*** and ***Italian Hours.*** His literary essays remain grouped in the volumes in which they originally appeared, ***French Poets and Novelists*** (1878), ***Partial Portraits*** (1888), ***Essays in London*** (1893), and ***Notes on Novelists*** (1914). Some of his art essays are included in ***Picture and Text*** (1893). Many of his fugitive pieces remain uncollected in the journals in which they originally appeared. He was also a remarkable letter writer, believing that correspondence, in the hands of a stylist, belongs to the highest forms of literary expression. Thousands of his holograph letters survive in the major libraries of the West, and notably in the Houghton Library at Harvard where the James family papers were deposited. While many letters are social and trivial, a certain proportion are highly documentary in their picture of life abroad and as a record of literary relations.

Honours came to James belatedly. He received degrees from Oxford and Harvard; he was elected to the National Institute of Arts and Letters and in 1905 to the American

Academy of Arts and Letters; and on becoming a British subject in 1915—throwing his moral weight into Britain's struggle in World War I—he received the Order of Merit from King George V. James died Feb. 28, 1916, and his ashes were taken to America and buried in the James family plot in Cambridge Cemetery, where the headstone describes him as "interpreter of his generation on both sides of the sea."

Reputation. James's public remained limited during his lifetime. Since the revival of interest in his work following centenary observances of 1943 he reached an ever-widening audience; his works were translated in many countries and he was recognized in the late 1960s as one of the subtlest craftsmen who ever practiced the art of the novel. His rendering of the inner, the "phantasmagoric," life made him a forerunner of the "stream of consciousness" movement in the 20th century. In his originality, distinction of style, and inventiveness he is the prototype of the disinterested American artist and a "novelist's novelist." Since mid-20th century he has acquired the status of a classic and also of a "culture-hero." Earlier hostility in the United States, because of his expatriation and abandonment of citizenship, has largely disappeared; and there exists a singular unanimity in English and American criticism in seeing him as a large creative figure who began as a "representationalist" and ended in the camp of the moderns as impressionist and symbolist. Writers as different as James Joyce and Graham Greene, Conrad and Virginia Woolf, were influenced by him; and the modern school of fictional criticism regards him as its supreme theorist, deriving from him much of its fundamental terminology.

MAJOR WORKS

NOVELS: *Roderick Hudson,* 1875; *The American,* 1877; *Watch and Ward,* 1878; *The Europeans,* 1878; *Daisy Miller,* 1878; *Confidence,* 1879; *Washington Square,* 1880; *The Portrait of a Lady,* 1881; *The Bostonians,* 1886; *The Princess Casamassima,* 1886; *The Reverberator,* 1888; *The Tragic Muse,* 1890; *The Other House,* 1896; *The Spoils of Poynton,* 1897; *What Maisie Knew,* 1897; *In the Cage,* 1898; *The Awkward Age,* 1899; *The Sacred Fount,* 1901; *The Wings of the Dove,* 1902; *The Ambassadors,* 1903; *The Golden Bowl,* 1904; *The Ivory Tower,* 1917 (unfinished); *The Sense of the Past,* 1917 (unfinished).

SHORT STORIES: *A Passionate Pilgrim and Other Tales,* 1875; *An International Episode,* 1879; *The Madonna of the Future and Other Tales,* 1879; *The Diary of a Man of Fifty,* 1880; *The Siege of London,* 1883; *Tales of Three Cities,* 1884; *The Author of Beltraffio,* 1885; *Stories Revived,* 1885; *The Aspern Papers,* 1888; *A London Life,* 1889; *The Lesson of the Master,* 1892; *The Real Thing and Other Tales,* 1893; *The Private Life,* 1893; *Terminations,* 1895; *Embarrassments,* 1896; *The Two Magics: The Turn of the Screw, and Covering End,* 1898; *The Soft Side,* 1900; *The Better Sort,* 1903; *Julia Bride,* 1909; *The Finer Grain,* 1910; *The Outcry,* 1911 (based on an unproduced play).

PLAYS: *Daisy Miller,* 1883; *The American,* 1891; *Theatricals,* 1894 (two comedies, *Tenants* and *Disengaged); Guy Domville,* 1895; *Theatricals: Second Series,* 1894 *(The Album* and *The Reprobate); The High Bid,* 1908.

TRAVEL SKETCHES: *Transatlantic Sketches,* 1875; *Portraits of Places,* 1883; *A Little Tour in France,* 1884; *English Hours,* 1905; *The American Scene,* 1907; *Italian Hours,* 1909.

LITERARY AND ART STUDIES: *French Poets and Novelists,* 1878; *Hawthorne,* 1879; *The Art of Fiction,* 1884; *Partial Portraits,* 1888; *Picture and Text,* 1893; *William Wetmore Story and His Friends,* 1903; *The Question of Our Speech,* 1905 (and "The Lesson of Balzac," two lectures); *Views and Reviews,* 1908; *Notes on Novelists,* 1914; *Notebooks,* 1947.

AUTOBIOGRAPHICAL: *A Small Boy and Others,* 1913; *Notes of a Son and Brother,* 1914; *The Middle Years,* 1917.

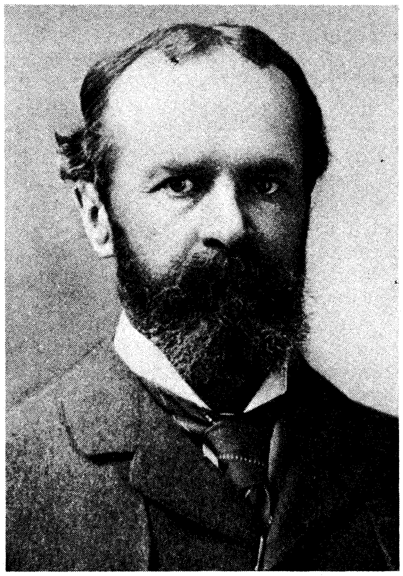
BIBLIOGRAPHY. LEON EDEL, *The Life of Henry James,* 5 vol. (1953–72), is the only full-length biography of the novelist. It is based on his unpublished correspondence, diaries, and other documents and treats all the stages of his development as artist and his career in the United States and abroad. A concise biography, largely critical, by F.W. DUPEE, *Henry James,* was published in 1951. A valuable collection of biographical anecdotes is SIMON NOWELL-SMITH, *Legend of the Master* (1947). S. GORLEY PUTT, *Henry James: A Reader's Guide* (1966), is a useful survey of James's novels and tales. A complete bibliography of James's writings, listing his unsigned work in periodicals as well as all editions of his novels and tales, *A Bibliography of Henry James,* by LEON EDEL and DAN H. LAURENCE, was issued in the Soho Bibliographies (1957; rev. ed., 1961). A valuable early survey of James's work is PELHAM EDGAR, *Henry James: Man and Author* (1927). The first important treatment of the late novels is in PERCY LUBBOCK, *The Craft of Fiction* (1921); a later work, based on the then unpublished *Notebooks,* was F.O. MATTHIESSEN, *Henry James: The Major Phase* (1944); see also FREDERICK C. CREWS, *The Tragedy of Manners: Moral Drama in the Later Novels of Henry James* (1957); and STEPHEN SPENDER, *The Destructive Element: A Study of Modern Writers and Beliefs* (1935). Important essays on James may be found in EDMUND WILSON, *The Triple Thinkers* (1938); R.P. BLACKMUR, *The Lion and the Honeycomb* (1955); F.R. LEAVIS, *The Great Tradition* (1948), which places James with Conrad and George Eliot as the major novelists of the English tradition; and in LIONEL TRILLING, *The Liberal Imagination* (1950), containing a study of *The Princess Casamassima.* A comprehensive evaluation of James as critic was published by RENE WELLEK, "Henry James's Literary Theory and Criticism," in *American Literature,* 30:293–321 (1958). DOROTHEA KROOK, *The Ordeal of Consciousness in Henry James* (1962), is a measured study of James's mental processes in his fiction. A great many specialized studies of James are listed in the annual bibliographies of the *Publications of the Modern Language Association.*

(L.Ed.)

James, William

William James, eminent American psychologist and philosopher, was leader of the movement known as Pragmatism, which stresses the role of action in the formulation of belief. Although deeply concerned with experiential religious faith, James represents a reaction against all monistic absolutisms, such as that of his friend, Josiah Royce, an American Idealist.

By courtesy of the Harvard University News Service



William **James.**

Home environment

Early life and education. James was born in New York City on January 11, 1842. He was the eldest son of Henry James, an idiosyncratic and voluble man of pungent wit and vigorous writing style, whose philosophical interests attracted him to the theology of Emanuel Swedenborg. One of William's brothers was the distinguished Henry James, novelist and critic and one of the most influential theorists of fiction in the English-speaking world. The elder Henry James had begun as a theological student but had developed at Princeton Theological Seminary an "antipathy to all ecclesiasticisms which he expressed with abounding scorn and irony throughout all his later years." Both his physical and spiritual life were marked by restlessness and wanderings, largely in Europe, that affected the training of his children at school and their education at home. Building upon the works of Swedenborg, which had been proffered as a revelation from God for a new age of truth and reason in religion, the elder James had constructed a system of his own that seems to have served him as a vision of spiritual life. This philosophy provided the permanent intellectual atmosphere of William's home life, to some degree compensating for the undisciplined irregularity of his schooling, which ranged from New York to Boulogne, France, and to Geneva and back. The habits acquired in dealing with his father's views at dinner and at tea carried over into the extraordinarily sympathetic yet critical manner that William displayed in dealing with anybody's views on any occasion.

Medical training

When James was 18 years of age he tried his hand at studying art, under the tutelage of William M. Hunt, an American painter of religious subjects. But he soon tired of it and the following year entered the Lawrence Scientific School of Harvard University. From courses in chemistry, anatomy, and similar subjects there, he went to the study of medicine in the Harvard Medical School; but he interrupted this study in order to accompany the eminent naturalist Louis Agassiz, in the capacity of assistant, on one of his most important explorations, an expedition to the Amazon. There James's health failed, and his duties irked him. He returned to the medical school for a term and then during 1867–68 went to Germany for courses with the physicist and physiologist Hermann von Helmholtz, who formulated the law of the conservation of energy; with Rudolf Virchow, a pathologist; and with Claude Bernard, the foremost experimentalist of 19th-century medicine; and with others. At the same time he read widely in the psychology and philosophy then current, especially the writings of Charles Renouvier, a Kantian Idealist and relativist.

The acquaintance with Renouvier was a focal point in James's personal and intellectual history. He seems from adolescence to have been a delicate boy, always ailing, and at this period of his stay in Germany he suffered a breakdown, with thoughts of suicide. When he returned home in November 1868, after 18 months in Germany, he was still ill. Though he took the degree of M.D. at the Harvard Medical School in June 1869, he was unable to begin practice. Between that date and 1872 he lived in a state of semi-invalidism in his father's house, doing nothing but reading and writing an occasional review. Early in this period he experienced a sort of phobic panic, which persisted until the end of April 1870. It was relieved, according to his own statement, by the reading of Renouvier on free will and the decision that "my first act of free will shall be to believe in free will." The decision carried with it the abandonment of all determinisms—both the scientific kind that his training had established for him and that seems to have had some relation to his neurosis and the theological, metaphysical kind that he later opposed in the notion of "the block universe." His revolutionary discoveries in psychology and philosophy, his views concerning the methods of science, the qualities of men, and the nature of reality all seem to have received a definite propulsion from this resolution of his poignant personal problem.

Interest in psychology. In 1872 James was appointed instructor in physiology in Harvard College, in which capacity he served until 1876. But he could not be diverted from his ruling passion, and the step from teaching physiology to teaching psychology—not the traditional "mental science" but physiological psychology—was as inevitable as it was revolutionary. It meant a challenge to the vested interests of the mind, mainly theological, that were entrenched in the colleges and universities of the United States; and it meant a definite break with what Santayana called "the genteel tradition." Psychology ceased to be mental philosophy and became a laboratory science. Philosophy ceased to be an exercise in the grammar of assent and became an adventure in methodological invention and metaphysical discovery.

With his marriage in 1878, to Alice H. Gibbens of Cambridge, Massachusetts, a new life began for James. The old neurasthenia practically disappeared. He went at his tasks with a zest and an energy of which his earlier record had given no hint. It was as if some deeper level of his being had been tapped: his life as an originative thinker began in earnest. He contracted to produce a textbook of psychology by 1880. But the work grew under his hand, and when it finally appeared in 1890, as *The Principles of Psychology*, it was not a textbook but a monumental work in two great volumes, from which the textbook was condensed two years later.

Major work in psychology

The *Principles*, which was recognized at once as both definitive and innovating in its field, established the functional point of view in psychology. It assimilated mental science to the biological disciplines and treated thinking and knowledge as instruments in the struggle to live. At one and the same time it made the fullest use of principles of psychophysics (the study of the effect of physical processes upon the mental processes of an organism) and defended, without embracing, free will.

Interest in religion. The *Principles* completed, James seems to have lost interest in the subject. Creator of the first U.S. demonstrational psychological laboratory, he disliked laboratory work and did not feel himself fitted for it. He liked best the adventure of free observation and reflection. Compared with the problems of philosophy and religion, psychology seemed to him "a nasty little subject" that he was glad to have done with. His studies, which were now of the nature and existence of God, the immortality of the soul, free will and determinism, the values of life, were empirical, not dialectical; James went directly to religious experience for the nature of God, to psychical research for survival after death, to fields of belief and action for free will and determinism. He was searching out these things, not arguing foregone conclusions. Having begun to teach ethics and religion in the late 1880s, his collaboration with the psychical researchers dated even earlier. Survival after death he ultimately concluded to be unproved; but the existence of divinity he held to be established by the record of the religious experience, viewing it as a plurality of saving powers, "a more of the same quality" as oneself, with which, in a crisis, one's personality can make saving contact. Freedom he found to be a certain looseness in the conjunction of things, so that what the future will be is not made inevitable by past history and present form; freedom, or chance, corresponds to Darwin's "spontaneous variations." These views were set forth in the period between 1893 and 1903 in various essays and lectures, afterward collected into works, of which the most notable is *The Will to Believe and Other Essays in Popular Philosophy* (1897). During this decade, which may be correctly described as James's religious period, all of his studies were concerned with one aspect or another of the religious question.

His natural interest in religion was reinforced by the practical stimulus of an invitation to give the Gifford Lectures on natural religion at the University of Edinburgh. He was not able to deliver them until 1901–02, and their preparation focussed his labours for a number of years. His disability, involving his heart, was caused by prolonged effort and exposure during a vacation in the Adirondacks in 1898. A trip to Europe, which was to have taken up a sabbatical year away from university duties, turned into two years of invalidism. The Gifford Lectures were prepared during this distressful period. Published as *The Varieties of Religious Experience* (1902), they had an even greater acclaim as a book than as articles. Cautious and tentative though it was, the rich concreteness of the material and the final summary of the evidence—that the varieties of religious experience point to the existence of specific and various reservoirs of consciousness-like energies with which we can make specific contact in times of trouble—touched something fundamental in the minds of religionists and at least provided them with apologetic material not in conflict with science and scientific method. The book was the culmination of James's interest in the psychology of religion.

Career in philosophy. James now explicitly turned his attention to the ultimate philosophic problems that had been at least marginally present along with his other interests. Already in 1898, in a lecture at the University of California on philosophical conceptions and practical results, he had formulated the theory of method known as Pragmatism. Originating in the strict analysis of the logic of the sciences that had been made in the middle

Early work on Pragmatism

1870s by Charles Sanders Peirce, the theory underwent in James's hands a transforming generalization. He showed how the meaning of any idea whatsoever—scientific, religious, philosophical, political, social, personal—can be found ultimately in nothing save in the succession of experiential consequences that it leads through and to; that truth and error, if they are within the reach of the mind at all, are identical with these consequences. Having made use of the pragmatic rule in his study of religious experience, he now turned it upon the ideas of change and chance, of freedom, variety, pluralism, and novelty, which, from the time he had read Renouvier, it had been his preoccupation to establish. He used the pragmatic rule in his polemic against monism and the "block universe," which held that all of reality is of one piece (cemented, as it were, together); and he used this rule against internal relations (*i.e.*, the notion that you cannot have one thing without having everything), against all finalities, staticisms, and completenesses. His classes rang with the polemic against absolutes, and a new vitality flowed into the veins of American philosophers. Indeed, the historic controversy over Pragmatism saved the profession from iteration and dullness.

Meanwhile (1906), James had been asked to lecture at Stanford University, in California, and experienced there the earthquake that nearly destroyed San Francisco. The same year he delivered the Lowell Lectures in Boston, afterward published as *Pragmatism: A New Name for Old Ways of Thinking* (1907). Various studies appeared —"Does Consciousness Exist?" "The Thing and Its Relations," "The Experience of Activity"—chiefly in *The Journal of Philosophy;* these were essays in the extension of the empirical and pragmatic method, which were collected after James's death and published as *Essays in Radical Empiricism* (1912). The fundamental point of these writings is that the relations between things, holding them together or separating them, are at least as real as the things themselves; that their function is real; and that no hidden substrata are necessary to account for the clashes and coherences of the world. The Empiricism was radical because until this time even Empiricists believed in a metaphysical ground like the hidden turtle of Hindu mythology on whose back the cosmic elephant rode.

James was now the centre of a new life for philosophy in the English-speaking world. The continentals did not "get" Pragmatism; if its German opponents altogether misunderstood it, its Italian adherents—among them, of all people, the critic and devastating iconoclast Giovanni Papini—travestied it. In England it was championed by F.C.S. Schiller, in the United States by Dewey and his school, in China by Hu Shih. In 1907 James gave his last course at Harvard. In the spring he repeated the lectures on Pragmatism at Columbia University. It was as if a new prophet had come; the lecture halls were as crowded on the last day as on the first, with people standing outside the door. He was feted and photographed; he himself describes the visit as the "high tide of my existence." Shortly afterward came an invitation to give the Hibbert Lectures at Oxford University. These lectures, published in 1909 as *A Pluralistic Universe,* state, in a more systematic and less technical way than the *Essays,* the same essential positions. They present, in addition, certain religious overbeliefs of James's, which further thinking—if the implications of the posthumous *Some Problems in Philosophy* may be trusted—was to mitigate. These overbeliefs involve a panpsychistic interpretation of experience (one that ascribes a psychic aspect to all of nature) that goes beyond radical Empiricism and the pragmatic rule into conventional metaphysics.

Home again, James found himself working, against growing physical trouble, upon the material that was partially published after his death as *Some Problems in Philosophy* (1911). He also collected his occasional pieces in the controversy over Pragmatism and published them as *The Meaning of Truth* (1909). Finally, his physical discomfort exceeded even his remarkable voluntary endurance. After a fruitless trip to Europe in search of a cure, he returned, going straight to the country home in Chocorua, New Hampshire, where he died on August 26, 1910.

Philosophy of world and man

Significance and influence. It is still too early to estimate James's significance and influence. In psychology, his work is of course dated, but it is dated as is Galileo's in physics or Darwin's in biology because it is the originative matrix of the great variety of new developments that are the current vogue. In philosophy, his positive work is still prophetic. The world he argued for was soon reflected in the new physics, as diversely interpreted, with its resonances from Charles Peirce, particularly by Albert Einstein, Bertrand Russell, and the Danish quantum physicist Niels Bohr—a world of events connected with one another by kinds of next to next relations, a world various, manifold, changeful, originating in chance, perpetuated by habits (that the scientist calls laws), and transformed by breaks, spontaneities, and freedoms. In human nature, James believed, these visible traits of the world are equally manifest. The real specific event is the individual, whose intervention in history gives it in each case a new and unexpected turn. But in history, as in nature, the continuous flux of change and chance transforms every being, invalidates every law, and alters every ideal.

James lived his philosophy. It entered into the texture and rhythms of his rich and vivid literary style. It determined his attitude toward scientifically unaccepted therapies, such as Christian Science or mind cure, and repugnant ideals, such as militarism. It made him an anti-imperialist, a defender of the small, the variant, the unprecedented, the weak, wherever and whenever they appeared. His philosophy is too viable and subtle, too hedged, experiential, and tentative to have become the dogma of a school. It has functioned rather to implant the germs of new thought in others than to serve as a standard old system for others to repeat.

For the James–Lange theory of emotions, see EMOTION: *Emotion as bodily sensation: James–Lange theory;* for his philosophy, see PRAGMATISM: *James.*

MAJOR WORKS

PSYCHOLOGY: *The Principles of Psychology,* 2 vol. (1890); *Talks to Teachers on Psychology: and to Students on Some of Life's Ideals* (1899).

PHILOSOPHY AND RELIGION: *Is Life Worth Living?* (1896); *The Will to Believe and Other Essays in Popular Philosophy* (1897); *Human Immortality* (1898); *The Varieties of Religious Experience* (1902); *The Sentiment of Rationality* (1905); *Pragmatism: A New Name for Old Ways of Thinking* (1907); *A Pluralistic Universe* (1909); *The Meaning of Truth* (1909); *Some Problems in Philosophy* (1911); *Essays in Radical Empiricism* (1912).

OTHER WORKS: *The Literary Remains of Henry James* (1885), containing some of his father's views on life.

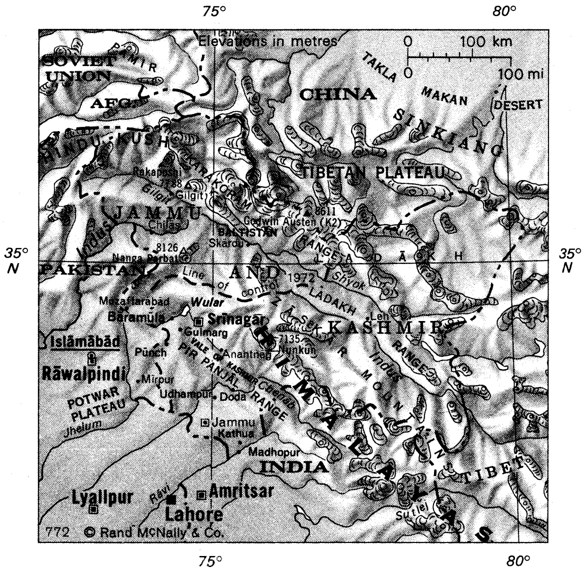
BIBLIOGRAPHY. The best and most comprehensive biography, richly documented, is still RALPH B. PERRY, *The Thought and Character of William James,* 2 vol. (1935), which should be supplemented by the following: *The Letters of William James,* 2 vol., ed. by his son, HENRY JAMES (1920), valuable as a historical and biographical source; *The Philosophy of William James, Drawn from His Own Works,* with an introduction by HORACE M. KALLEN (1925); GAY WILSON ALLEN, *William James* (1967), a sympathetic popularization; and HENRY JAMES, *Notes of a Son and Brother* (1914). See also HENRI BERGSON'S provocative chapter on James in *The Creative Mind* (1946); T. FLOURNOY, *La Philosophie de William James* (1911; Eng. trans., 1917); and C.H. GRATTAN, *The Three Jameses* (1932, reprinted 1962).

(H.M.K.)

Jammu and Kashmir

The territory of Jammu and Kashmir, situated in the northern part of the Indian subcontinent in the vicinity of the western Himalayan mountain range, has been in dispute between India and Pakistan since the partition of India in 1947. Formerly one of the largest princely states of India, it is bounded on the northwest by Afghanistan, on the north by the Sinkiang-Uighur Autonomous Region of China, on the east by Tibet (the Tibetan Autonomous Region), on the south by the Indian states of Himachal Pradesh and the Punjab, and to the west by Pakistan's Northwest Frontier Province and Punjab Province.

The territory has an area of 86,023 square miles (222,798 square kilometres); it is divided by a "line of control,"



JAMMU AND KASHMIR

agreed on in 1972, that leaves an area of 32,358 square miles (83,806 square kilometres) in the north and west to Pakistan and the remainder, amounting to 53,665 square miles (138,992 square kilometres), under Indian control. Repeated United Nations attempts to settle the dispute by holding a plebiscite have been unsuccessful: a plebiscite has not been held, and the dispute has been further complicated by Chinese incursions into the territory from the north.

The population of Jammu and Kashmir is believed to be about 5,900,000; the inhabitants of the Indian sector were estimated to number 4,600,000 in 1971, and those of the Pakistani sector were estimated at 1,300,000. The administrative capital of the Indian sector is Srinagar in summer and Jammu in winter; the administrative capital of the Pakistani sector is Muzaffarābād. The Vale of Kashmir, located in the Indian sector, is renowned for its beauty and fertility. (For associated physical features, see HIMALAYAN MOUNTAIN RANGES; INDUS RIVER; and KARAKORAM RANGE; for further historical background, see INDIAN SUBCONTINENT. HISTORY OF THE.)

History. Buddhism and Hinduism penetrated Jammu and Kashmir at an early date; from the 9th to the 12th century AD, the region appears to have achieved considerable prominence as a centre of Hindu culture. It was brought under Muslim rule in the 14th century, and it remained so for some five centuries until Sikh and then Dogra (hill Rājput) rulers assumed control in the 19th century. (The Sikhs are a religious community of northern India.)

Raja Gulab Singh

The state of Jammu and Kashmir, in its contemporary form, dates from 1846, when, by the treaties of Lahore and Amritsar at the conclusion of the First Sikh War, Raja Gulab Singh, the Dogra ruler of Jammu, was created maharaja (ruling prince) of an extensive but somewhat ill-defined Himalayan kingdom "to the eastward of the River Indus and westward of the River Rāvi." For the British the creation of this princely, protected state helped to safeguard their northern flank in their advance to the Indus and beyond during the latter part of the 19th century; the state thus formed part of a complex political buffer zone interposed by the British between their Indian empire and the empires of Russia and China to the north. For Gulab Singh, confirmation of title to these mountain territories marked the culmination of almost a quarter of a century of campaigning and diplomatic negotiation among the petty hill kingdoms along the northern borderlands of the Sikh empire of the Punjab.

Boundaries. Some attempts were made in the 19th century to define the boundaries of the state, but precise definition was in many cases defeated by the nature of the country and by the existence of huge tracts lacking permanent human settlement. In the far north, for example, the maharaja's authority certainly extended to the Karakoram Range, but beyond lay a debatable zone on the borders of the Turkistan and Sinkiang regions of Central Asia, and the boundary was never demarcated. There were similar doubts about the alignment of the frontier where this northern zone skirted the district known as Aksai Chin, to the east, and joined the better known and more precisely delineated boundary with Tibet, which had served for centuries as the eastern border of the Ladākh region of what is now Jammu and Kashmir. The pattern of boundaries in the northwest became clearer in the last decade of the 19th century, when Britain delimited boundaries in the Pamir region in negotiations with Afghanistan and Russia. At this time Gilgit, always understood to be part of Kashmir, for strategic reasons was constituted as a special agency in 1889 under a British agent.

Counterclaims of India and Pakistan. As long as the existence of the state was guaranteed by the United Kingdom, the weaknesses in its structure and along its peripheries were not of great consequence; following the withdrawal of the United Kingdom from South Asia in 1947, however, they became apparent. By the terms agreed upon for the partition of the Indian subcontinent between India and Pakistan, the rulers of princely states were given the right to opt for either Pakistan or India or—with certain reservations—to remain independent. The maharaja of Kashmir, Hari Singh, initially believed that by delaying a decision he could maintain the independence of Kashmir, but, caught up in a train of events that included a revolution among his Muslim subjects along the western borders of the state and the intervention of Pathan tribesmen. he signed an Instrument of Accession to the Indian Union on October 26, 1947. This was the signal for intervention both by Pakistan, which considered that the state was a natural extension of Pakistan, and by India, which intended to confirm the act of accession. Localized war continued during 1948 and was terminated through the agency of the United Nations in a cease-fire, which took effect in January 1949. In July of the same year, India and Pakistan defined a cease-fire line that divided the administration of the territory. This left Jammu, together with part of the districts of Pūnch and Mirpur, the Vale of Kashmir, and the LadHkh region, under Indian control and western Pūnch (which, together with Muzaffarābād and part of Mirpur, formed Azad Kashmir), Gilgit, and Baltistān under Pakistan's authority. Regarded merely as a temporary expedient, this partition along the cease-fire line still existed in the early 1970s, though warfare between the two contestants was briefly resumed in 1965 and again in 1971, despite the many proposals made to end the dispute.

The 1947 to 1949 hostilities

Chinese interests. The conflicts within the state were heightened by tensions at the periphery when a manifestation of Chinese power created an additional problem. After the Chinese Communist conquest of Tibet and the resurgence of Chinese power in Sinkiang, on the border with the Soviet Union, Chinese forces penetrated into the northeastern parts of Ladākh; by 1956–57 they had completed a military road through Aksai Chin to provide better communication between Sinkiang and western Tibet. The belated discovery of the existence of this road by Indian patrols led to border clashes. Indian refusal to negotiate with China on the alignment of the Ladākhi boundary in this area was in contrast with the willingness of Pakistan to reach a settlement with China on the course of the Karakoram boundary that lies within Pakistani-administered Kashmir.

The seven geographic zones

Physical geography. Physiographically, the territory comprises a number of zones that are closely associated with the structural components of the western Himalayas. From southwest to northeast these zones consist of the plains, the foothills, the Pir Panjāl Range, the Vale of Kashmir, the central Himalayan zone, the Upper Indus Valley, and the Karakoram Range.

The plains. The narrow zone of plains country in Jammu and Pūnch is characterized by interlocking sandy alluvial fans (fan-shaped accumulations of debris) that have been deposited by streams discharging from the foothills and by a much-dissected pediment (eroded bedrock surface) covered by loams and loess (wind-deposited silt) of the Pleistocene Epoch (about 10,000 to 2,500,000 years old). Rainfall is low, amounting to about 15 or 20 inches (380 to 500 millimetres) a year, and it occurs mainly in the form of heavy but infrequent showers in the summer (June to September), when the monsoon winds blow. The countryside has been almost entirely denuded of trees, and thorn scrub and coarse grass are the dominant forms of vegetation.

The foothills. The foothills of the Himalayas, rising from 2,000 to 7,000 feet (600 to 2,100 metres) comprise an outer zone and an inner zone. The outer zone consists of sandstones, clays, silts, and conglomerates, influenced by Himalayan folding movements and eroded to form long ridges and valleys (duns). The inner zone consists of more massive sedimentary rock, including red sandstones of the Miocene Epoch (about 7,000,000 to 26,000,000 years old), that has been folded, fractured, and eroded to form steep spurs and plateau remnants. River valleys are deeply incised and terraced, and faulting has produced a number of alluvium-filled basins, such as those of Udhampur and Pūnch. Rainfall increases with elevation, and the lower scrubland gives way to pine forests at higher altitudes.

The Pīr Panjāl Range. The Pir Panjiil Range constitutes the first mountain rampart associated with the Himalayas. It has an average crest line of 12,500 feet, with individual peaks rising to 15,000 feet. Consisting of an ancient rock core of granites, gneisses, quartz rocks, and slates, it has been subject to considerable uplift and fracturing and was heavily glaciated during the Pleistocene Epoch. The range receives considerable winter snowfall as well as summer rain and has extensive areas of pasture above the tree line.

The Vale of Kashmir. The Vale of Kashmir is a deep, asymmetrical basin lying between the Pir Panjiil and the central Himalayan ranges at an average altitude of 5,300 feet (1,600 metres). During the Pleistocene Epoch it was occupied at times by a lake, known as Lake Karewa, and is now filled by lacustrine sediments and alluvium deposited by the Upper Jhelum River at elevations of up to some 6,000 feet above sea level. Soil and water conditions vary considerably, and the climate is characterized by an annual rainfall of about 30 inches, some of which is derived from the summer monsoon winds and some from winds associated with winter depressions. Temperatures are modified by altitude; average minimum temperatures of about 11" F (—12" C) occur in January and average maximum temperatures of 99" F (37" C) in July. Up to 7,000 feet, woodlands of deodar (Indian cedar), blue pine, walnut, willow, elm, and poplars occur; from 7,000 to 10,500 feet coniferous forests with fir, pine, and spruce are found; from 10,500 to 12,000 feet birch is dominant; and above 12,000 feet there are meadows with rhododendrons, dwarf willows, and honeysuckle.

Forests and woodlands in the Vale of Kashmir

The central Himalayan zone. Geologically complex and topographically immense, the central Himalayan zone has ranges reaching more than 20,000 feet and deeply entrenched, remote valleys. The region was heavily glaciated in the Pleistocene Epoch, and remnant glaciers and snowfields are still present. The zone receives some rain from the Indian monsoon winds in the summer months, and the lower slopes are forested, but the mountains, nevertheless, constitute a climatic divide, representing a transition from the monsoon climate of the Indian subcontinent to the dry, continental climate of Central Asia.

The Upper Indus Valley. The valley of the Upper Indus is a well-defined feature that follows the geological strike (structural trend) westward from the Tibetan border to the point where it rounds the great mountainous mass of Nringa Parbat to run southward in deep gorges cut across the strike. In its upper reaches it is flanked by gravel terraces; each tributary builds an alluvial fan out into the main valley. The town of Leh in the Indian sector stands on such a fan, 11,500 feet above sea level, with a climate characterized by almost total lack of rainfall, intense insolation (exposure to the Sun's rays), and great diurnal and annual ranges of temperature. Life depends on meltwater from the surrounding mountains, and vegetation is alpine (above the limit of tree growth), growing on thin skeletal soils.

The Karakoram Range. The great granite-gneiss massifs (mountain masses) of the Karakoram contain some of the world's highest peaks, including K2, or Mt. Godwin Austen, with an altitude of 28,250 feet (8,611 metres); at least 30 other peaks exceed 24,000 feet. The range, which is still being glaciated, rises starkly from dry, desolate plateaus characterized by extremes of temperature and shattered rock debris. The Karakoram region is aptly named the "roof of the world."

The great mountain peaks

Animal life. Wild animals include ibex, urial (a species of wild sheep with a reddish coat), the Kashmir stag, black and brown bears, and many game birds, including vast numbers of migratory duck.

The people. *Rural settlements.* Physiographic diversity is matched by the considerable differences that occur in the forms of human occupation of the country.

In the plains and foothills, agricultural settlements have resulted from colonization movements from the Punjab extending over a long period; both the people and their culture are similar to those of adjacent parts of the Punjab and other lowlands to the west. Where alluvial soils and the availability of water for irrigation make agriculture possible in the *dūn*s and lower valleys, population densities of 300 to 400 per square mile (120 to 150 per square kilometre) are sustained by crops of wheat and barley, which are gathered in the spring (*rabīʿ*) harvest, and of maize and millet, gathered in the late summer (*kharif*) harvest, as well as by the keeping of livestock. The upper sections of the valleys support a sparser population who depend on a maize, cattle, and forestry economy in which seasonal migration to higher pastures in the spring is necessary for the production of milk and clarified butter, or ghee, for southern lowland markets. In winter the hillmen return to lower areas to work in state forests and timber mills. Agricultural hamlets and nucleated villages predominate, and towns such as Jammu and Udhampur function essentially as market centres and administrative headquarters for the rural populations and estates in the vicinity.

Jammu and Pūnch. Jammu, winter capital of the maharajas and second city of the state, with a population in 1971 of more than 155,000, was traditionally the seat of Dogra power and serves a largely Hindu population. Of the total population of Jammu—1,500,000 in 1961—approximately 1,000,000 were classed as Hindus; of these the greater part live in the southeastern territories. Culturally, ethnically, and linguistically, they are closely related to the Hindus in the Punjab state of India, many speaking the Dogri language. The majority of the 63,000 Sikhs recorded at the 1961 census also live in Jammu. To the northwest, however, the proportion of Muslims increases, with Muslims constituting a dominant majority in western Pūnch. Here cultural and economic ties are closer with Riiwalpindi (the interim capital of Pakistan since 1959) and the Punjab province of Pakistan, and the inhabitants have consistently opposed Hindu–Dogra rule.

Kashmiris of the Vale and highlands. The Vale of Kashmir with its surrounding highlands (Kashmir proper) has always retained something of its own individuality. Of the 2,000,000 Kashmiris in the Vale, the vast majority—about 90 percent—are Muslims: culturally and ethnically, links are closest with the northwestern highlands. The Kashmiri language is influenced by Sanskrit and belongs to the Dardic group of Aryan languages, spoken by the various hill tribes of Gilgit in the northwest; physical traits of the population suggest early connections in the same direction.

The Kashmiri language

The people of Ladākh and Baltistān. Across the main Himalayan divide, the people are as different as the landscape. Ladākh (sometimes called "little Tibet"), to the northeast, is thinly populated; its 10,000 inhabitants are of Tibetan racial stock and language (Ladakhi) and hold the Buddhist faith, while Baltistrin to the north, with its administrative centre at Skiirdu, is inhabited by people of similar stock who have been converted to Islām.

Administration and social conditions. Before its division, the territory of the state as a whole was composed of the provinces of Jammu, Kashmir, and the border states—Ladiikh, Baltistān, and the Gilgit Agency. Jammu, with Pūnch, is divided into the districts of Jammu, Kathua, Udhampur, Riijauri, Mirpur, and Doda. Kashmir is divided into the districts of Anantniig, Bāramūlla, Astor, Muzaffariibiid, and Srinagar. Muzaffariibiid and part of Mirpur and Pūnch comprise the state of Azad (meaning "free") Kashmir, which is under the protection of Pakistan, while Baltistiin, Astor, and the Gilgit Agency are administered directly by Pakistan. Ladiikh, Jammu, Srīnagar, Kathua, Udhampur, Rājauri, Anantnāg, Bāramūlla, Doda, and part of Pūnch form the Indian state and are represented in the Lok Sabha (House of the People, or lower house) of the Indian Parliament as the state of Jammu and Kashmir.

The bicameral legislature

The Indian state. Under the constitution of Jammu and Kashmir, as amended in 1965, the governor of the state is appointed by the president of India. The legislature consists of two houses: the Legislative Assembly of 100 members, representing constituencies; and the Legislative Council of 36 members. The state sends four elected representatives to the Lok Sabha and two members, nominated by the governor, to the Rajya Sabha (Council of States, or upper house).

The High Court consists of a chief justice and two or more other judges appointed by the president of India.

Education is free, and there were more than 4,000 educational institutions in the early 1970s. The university at Srinagar was founded in 1948. There is an efficient medical service with dispensaries scattered throughout the state. In spite of its height, the Vale of Kashmir is not healthful, and in the past there have been considerable outbreaks of cholera and other infectious diseases, although the incidence of these has been much reduced.

Azad Kashmir. The government, with headquarters at Muzaffarābād, was first set up in 1947. It is under a president, assisted by a State Council of 12. The state has its own High Court. Pakistan maintains control over the state through the Ministry of Kashmir Affairs.

Education at its lower stages is free. In addition to primary, middle schools, and high schools, the state has several colleges, at one of which students follow degree courses. There are several hospitals and many more dispensaries and first-aid posts.

The Gilgit Agency and Baltistān. Although both are administered by the government of Pakistan through political agents, these two areas are governed as separate units because of the difficult mountain terrain and the lack of communications between them. Gilgit town is the headquarters of the Gilgit Agency, and Skārdu is the administrative centre in Baltistiin. In the two areas there are a number of primary schools, middle schools, and high schools, as well as several hospitals, dispensaries, and first-aid posts.

Agricultural crops

The economy. The majority of the Kashmiri population is occupied in agriculture of diverse kinds, each adapted to local conditions. Rice, the staple crop, is planted in May and harvested in late September. Maize, millet, pulses (plants such as peas, beans, and lentils), cotton, and tobacco are—with rice—the main summer crops; wheat and barley are the chief spring crops; many temperate fruits and vegetables are grown in areas adjacent to urban markets or in well-watered areas with rich organic soils. Lake margins are particularly favoured, and intensive cultivation of vegetables is practiced in reclaimed marshland or on artificial floating gardens. The pressure of population on land is everywhere apparent, and all available resources are utilized. The lakes and rivers provide fish, water chestnuts, power (for example, near Bāramūlla, 25 miles east of Srinagar), and transport and are a tourist attraction. The mountains supply many kinds of timber, pasture for sheep and dairy cattle, and such snow-sports centres as that at Gulmarg, 14 miles south of Bāramūlla.

The region has mineral resources, including several small coalfields and bauxite deposits in the Udhampur district, but these are virtually unexploited.

Between 1947 and the early 1970s the installed electric-power capacity, originally 4,000 kilowatts, was increased tenfold. All the principal towns, including Leh, were electrified, and hydroelectric and thermal generating plants have been constructed to provide power for industrial development based on local raw materials. Land reforms have been undertaken, grain production was increased, and the quantity of major exports—timber, fruits, and handicraft products—was vastly expanded since 1947. Improvements have also been made in facilities for tourist traffic, although here the potential has barely been tapped.

Srinagar possesses many specialized agricultural markets, retail shopping centres, and associated industries. The industries are essentially developments from rural crafts—with handloom weaving of local silk, cotton, and wool, carpet weaving, wood carving, leatherwork, and other village industries organized on a factory basis in an urban context. Such industries, together with silverwork and copperwork and jewelry, were stimulated by the presence of the royal court and the tourist trade but also owe something to the important position achieved by Srinagar in west Himalayan trade. The city acts as an entrepôt for the products of the Punjab on the one hand and of High Asia (the region east of the Karakoram, Pamirs, and Ladākh ranges) on the other; routes run northwestward into Gilgit via the Raj Diangan Pass and northeastward via the Zoji La (pass) to Leh, Ladakh and beyond.

The economy of Ladākh and Baltistān

Cultivation in Ladiikh and Baltistiin is restricted to such main valleys as those of the Indus and Shyok rivers, where it consists of small irrigated plots of barley, buckwheat, roots, and sown grass situated amid desert surroundings. Pastoralism is a vital feature of the economy since animals are necessary for human survival. Herds of yak and flocks of goatlike sheep provide not only animal products but also means of transport between the widely dispersed settlements and monasteries. Grazing rights are subjects of important litigation; small townships are basically caravansaries (inns for the overnight accommodation of caravans) located at points where herders and traders can obtain provisions and grazing prior to undertaking the next stage of their journey through some of the loftiest passes and over some of the most barren terrains on Earth. Political troubles have unfortunately reduced the volume of this transit trade, as well as the benefit that Ladakhi settlements such as Leh once derived from it.

Transportation. Under three five-year plans, the Indian central government has made a huge investment in developing almost 3,700 miles of highways and communications. As a result of the India–Pakistan dispute over Kashmir, the blocking of the Jhelum Valley route, which runs for 200 miles from Srinagar to Riiwalpindi, made it necessary to transform the longer and more difficult Banihāl Pass cart road into an all-weather motor road linking Jammu with the Vale. This necessitated the construction of one of the longest tunnels in Asia—the Jawahar Tunnel. In the early 1970s there were schemes for extending rail communications from the railhead at Miidhopur in the Indian Punjab into the Vale.

The Kashmir problem. In the early 1970s, after a quarter of a century, the "Kashmir problem" remained intractable. Although there was a clear Muslim majority in the state before the 1949 partition and its economic, cultural, and geographic contiguity with the Muslim-majority area of the Punjab could be convincingly demonstrated, the accidents of history have resulted in a division of territory that has no rational basis. Pakistan has been left with territory that, although basically Muslim in character, was thinly populated, difficult of access, and economically underdeveloped. The largest Muslim group, situated in the Vale of Kashmir and estimated to number more than half the population of the entire state, lies in Indian-administered territory, with its former outlets via the Jhelum Valley route being blocked.

Pakistan has been frustrated in its attempts to use the waters of the Indus, Jhelum, and Chenāb rivers granted to it under the Indus Waters Treaty (1960), since their upper catchments lay in territories outside its control. Indeed, apart from the improvement of communications linking Gilgit and Baltistān with military bases in the Northwest Frontier Province and the extension of administration into these highland areas, not a great deal of development has been possible.

India, on the other hand, has acquired the lion's share of both territory and population and with them substantial linguistic, racial, and religious problems. Only 18 percent of the population in the Indian sector may be classed as literate. There were still movements in the early 1970s seeking independence for the state from both India and Pakistan. Thus, a strong military presence became necessary in the Indian sector, in part to support the administrative structure of the state, in part to confront the forces of Pakistan along the cease-fire line, and in part to confront Chinese forces on the northeastern frontiers.

BIBLIOGRAPHY. A good introduction to the physical landscapes of the western Himalayas is provided by HELMUT DE TERRA and T.T. PATERSON, *Studies on the Ice Age in India and Associated Human Cultures* (1939); and by KENNETH MASON, "The Glaciers of the Karakoram and Neighbourhood," *Rec. Geol. Surv. India*, 63:214–278 (1930). General geographical accounts are contained in KAZI S. AHMAD, *A Geography of Pakistan* (1964); D.N. WILBER *et al.*, *Pakistan: Its People, Its Society, Its Culture* (1964); RAM NANDAN, "Jammu and Kashmir," *Focus*, 13:1–6 (1962); and HARALD UHLIG, "Hill Tribes and Rice Farmers in the Himalayas and South-East Asia," *Trans. Inst. Br. Geogr.*, 47:1–23 (1969)—the latter develops an interesting account of the complementary relationship between hill folk and valley folk in the mountainous environment of Kashmir and of interlocking systems of transhumance. P.N.K. BAMZAI, *A History of Kashmir* (1962); and K.M. PANIKKAR, *The Founding of the Kashmir State: A Biography of Maharajah Gulab Singh, 1792–1858* (1953), provide sound accounts of the formation of the state of Jammu and Kashmir under British political protection. The development of political conflicts in the postwar period is dealt with by a number of authors. ALASTAIR LAMB, *Crisis in Kashmir, 1947–1966* (1966), analyzes the recent history leading to partition and contains a comprehensive bibliography; the recurrent discussions between India and Pakistan are chronicled in the *United Nations Official Records of the General Assembly* and . . . *of the Security Council;* the historical background to the Chinese occupation of parts of Ladākh is presented in MARGARET W. FISHER, LEO E. ROSE, and ROBERT A. HUTTENBACK, *Himalayan Battleground: Sino-Indian Rivalry in Ladakh* (1963). The influence of the diverse physical environments on the political problems are dealt with in R.C. MAYFIELD, "A Geographic Study of the Kashmir Issue," *Geogrl. Rev.*, 45:181–196 (1955); ALI TAYYEB, *Pakistan: A Political Geography* (1966), and by NANDAN (*op. cit.*); while the economic issues and consequences are indicated in P.N. DHAR, "The Kashmir Problem: Political and Economic Background," *India Quarterly*, 7:143–162 (1951); and in INDIA, MINISTRY OF INFORMATION AND BROADCASTING, *India: A Reference Manual* (annual).

(W.Ki.)

Jansen, Cornelius Otto

Cornelius Jansen, leader of the Roman Catholic reform movement known as Jansenism, by the force and scholarship of his writings brought about a reappraisal of the church's piety and role in national and international politics. His emphasis on the doctrine of grace and a scholarly interpretation of the Bible, as well as his antipathy to the Jesuits (who had become influential politically), earned for him the unmerited reputation as being sympathetic to Protestantism.

Education and early desires for reform

Born at Acquoi, near Leerdam, Holland, in 1585, the son of Jean Otto, he entered the University of Louvain, in Flanders, in 1602 to study theology. According to the custom adopted by the humanists of the Renaissance, Jansen took a Latin name, Cornelius Jansenius. His teacher, Jacques Janson, taught the doctrine of the theologian Michel de Bay, called Baïus, who died at Louvain in 1589. According to the latter, man is affected from his birth by the sin of Adam, his ancestor. His instincts lead him necessarily to evil. He can be saved only by the grace of Christ, accorded to a small number of the elect who have been chosen in advance and destined to enter the Kingdom of Heaven. This doctrine, inspired by certain writings of the great North African theologian St. Augustine (354–430), attracted Jansen, as well as another student who had come to study at Louvain, a Frenchman named Jean Duvergier de Hauranne (1581–1643), who was to become a leader of the Jansenist movement. The two young men became friends in Paris, where Jansen went in 1604. They decided to revive the theology that they believed the doctors of the Sorbonne (theological faculty) had reduced to subtle and vain discussions of Scholastics (medieval philosopher-theologians). Jansen and Duvergier thought that it was necessary to render to God the homage owed by men. According to them, the pride of the Renaissance savants had alienated Christians from the Jesus who loved the simple and the humble.

Jansen, engraving by Jean Morin (1600–1650)

By courtesy of the trustees of the British Museum; photograph, J.R. Freeman & Co. Ltd.

In 1611 Jansen followed his friend Duvergier to the home of the latter's parents, located in the outskirts of Bayonne. The bishop of the city entrusted to Jansen the direction of the episcopal college there from 1612 to 1614. For three years afterward the young Dutchman, with Duvergier, dedicated himself to the study of the writings of the early Church Fathers. In 1617 Jansen returned to Louvain, where he directed the college of Ste.-Pulchérie, created for Dutch students.

A violent dispute had arisen at Louvain between the disciples of Baïus and the Jesuits, who considered as dangerous the doctrines of this theologian condemned by Pope Pius V (reigned 1566–72) in 1567. Jansen then undertook a thorough study of the works of Augustine by which Baïus had been inspired. He read them, he declared, ten times consecutively. But he devoted himself most particularly to the texts drafted by Augustine to combat the doctrine of the British monk Pelagius (fl. 405–418) who had upheld that, in spite of the fault committed by Adam, man continues to be entirely free to do good and to obtain salvation by means of his own merits. Jansen then began his great work, the *Augustinus*. For him, the divine grace that alone can save man is not due at all to his good actions. It is, he claimed, a gratuitous gift by means of which Christ leads the elect to eternal life, but the multitude, "the mass of perdition," is doomed to damnation. Thus, men are predestined to obtain grace or to suffer condemnation. In reality, Augustine had not envisaged the fate of human beings with such great rigour. He had even proclaimed the power of man's free will at the time when he was engaged in the struggle against the Manichaeans, religious dualists originating in Persia who believed in the omnipotence of Satan. But Augustine had then been constrained to enclose the liberty of man within strict limits, in order to refute the affirmations of Pelagius, who had radically reduced the value of the divine grace obtained by Jesus on the cross.

The basis of Jansenism: the *Augustinus*

Jansen was so fascinated by Augustine's treatise against the Pelagians that he apparently lost sight of Augustine's works against the Manichaeans.

Jansen also wrote his commentaries on the evangelists and on the Old Testament—notably on the Pentateuch, the first five books of the Old Testament—as well as a "Discourse on the Reformation of the Inner Man." He was likewise the author of pamphlets directed against the Protestants.

Having acquired the degree of doctor in theology at Louvain, Jansen became the rector of that university in 1635, and in 1636 he became bishop of Ypres. The Jesuit scholar René Rapin asserted in his book on the history of Jansenism, *Histoire du Jansénisme* (Paris, 1861), that Jansen had obtained his mitre (bishop's headpiece) as a result of the personal intervention of the king of Spain, Philip IV (1605–65). This sovereign had recognized him for having published a pamphlet entitled *Mars Gallicus* in which he violently criticized the policy of the French cardinal and prime minister de Richelieu (1585–1642), who had contracted an alliance with the Dutch Protestants against Spain. A short time after his elevation to the episcopate, on May 6, 1638, Jansen died of the plague. After his death, in 1640 the friends of the deceased published at Louvain the work he had dedicated to St. Augustine, under the title *Augustinus Cornelii Jansenii, Episcopi, seu Doctrirza Sarzcti Augustini de Humanae Naturae, Sanitate, Aegritudine, Medicina adversus Pelagianos et Massilienses* ("The Augustine of Cornelius Jansen, Bishop, or on the Doctrines of St. Augustine Concerning Human Nature, Health, Grief, and Cure against the Pelagians and Massilians"—the latter being Semi-Pelagians).

This book had cost its author 22 years of effort. In the epilogue Jansen declared: "I leave my work to the judgment of the Roman Church . . . I retract all that she will decide that I ought to (must) retract."

Condemnation of Jansen

In the terms of a bull (papal document) of March 6, 1642, Pope Urban VIII (reigned 1623–44) forbade the reading of the *Augustirzus*, which had been published without the authorization of the Holy See and was based on the doctrine of Baïus, already condemned. Five propositions that Nicolas Cornet, syndic (lay administrator—representing the papacy—of property held by the Franciscan monastic order that, according to its rules, was not allowed to hold property) of the Faculty of Theology of Paris, claimed to have discovered in the *Augustinus* were condemned by a bull of Pope Innocent X (reigned 1644–55), published on June 9, 1653, and by a constitution of his successor, Alexander VII (reigned 1655–67). The bishops of France were required to make all of the priests, monks, and nuns sign a formulary conforming to the pontifical decisions. But Duvergier de Hauranne, who had become an abbt (one who has the right to wear the dress of a parish priest) of Saint-Cyran, had taught the doctrine of Jansen to the nuns of the abbey of Port-Royal. This convent (a base for Jansenism) became a focus of resistance against the Jesuits, who, having obtained the pontifical decisions in their favour, intended to impose them. From that time, a conflict began between the Jesuits and Antoine Arnauld, a disciple of "Monsieur de Saint-Cyran" (Duvergier de Hauranne) who called himself an Augustinian. The Jesuits, however, called him a Jansenist. According to them, the doctrine of Arnauld was that of Jansen and not of St. Augustine. Blaise Pascal wrote *Les Provinciales* ("Provincial Letters") in 1656 and 1657 to defend Antoine Amauld. The latter was condemned by the Faculty of Theology of Paris, which taught at the Sorbonne.

In 1709 Louis XIV ordered the dispersal of the nuns of Port-Royal des Champs into diverse convents, and he had the abbey destroyed in 1710. In 1723 followers of Jansen's views established an autonomous Jansenist Church at Utrecht, Holland, which still existed in the late 20th century. (See also OLD CATHOLIC CHURCHES).

BIBLIOGRAPHY. SAINTE-BEUVE, *Port-Royal* (in French), 3rd ed., vol. 1, ch. 11 (1867), studies Jansen's theological concepts and those of the abbé of Saint-Cyran, and includes the correspondence between the two men, which later inspired the editing of the *Augustinus*. MARC ESCHOLIER, *Port-Royal* (1965; Eng. trans., 1968), explains the relationship between Jansen and Saint-Cyran, and discusses the theological theses supported in the *Augustinus*, by Jansen.

(M.M.E.)

Japan

The island country of Japan (Nihon) lies off the east coast of Asia. It consists of four main islands—from northeast to southwest these are Hokkaidd, Honshii, Shikoku, and Kyūshū—and of numerous smaller islands, having a total land area of 143,818 square miles (372,488 square kilometres), with a total population of over 105,000,000 people. Japan is bounded to the west by the Sea of Japan, which separates it from the eastern shores of the Soviet Union and North and South Korea; to the north by La Perouse Strait (separating it from the Soviet island of Sakhalin) and the Sea of Okhotsk; to the northeast by the Kuril Islands (part of the Soviet Union); to the east and south by the Pacific Ocean; and to the southwest by the East China Sea, which separates it from the People's Republic of China. The national capital, Tokyo, which has a population of 9,000,000 inhabitants, is the most populous city in the world.

Aspects of Japanese life

Complexity is the keynote of life in Japan—a nation possessing an intricate and ancient cultural tradition, yet one that, especially since World War II, has emerged as a modern industrial giant. The juxtaposition of old and new is apparent in all phases of Japanese life. A characteristic sensitivity to natural beauty and a concern with form and balance are evident in such cities as Kyōto and Nara, as well as in Japan's ubiquitous gardens. Tokyo, however, symbolizes the impact of rapid westernization upon many aspects of Japanese life. The agricultural regions are characterized by well-ordered rice fields and fruit orchards, whereas the industrial belt along the Pacific coast of Honshū produces smoke and other wastes that pollute Japan's air and rivers. The time-honoured code of social behaviour is being challenged by new attitudes and values, creating an element of tension in city life.

Japan's spectacular industrial growth—the greatest of any nation since the 1940s—has brought the country to the forefront of the world economy. It is the world's principal shipbuilder and is one of the major producers and exporters of manufactured goods, including steel, synthetic rubber, and electrical products. An important feature of the burgeoning economy is the prevalence of large, quasi-monopolistic industrial companies. Although Japan is capable of becoming the world's most prosperous nation, in the early 1970s it was plagued by inflation and a growing shortage of skilled labour. It was hoped that continued cooperation between government and business interests would introduce solutions to these problems and that Japan would maintain its rate of growth to the benefit of all its citizens (see city articles KYOTO, NAGOYA, OSAKA-KOBE METROPOLITAN AREA, and TOKYO-YOKOHAMA METROPOLITAN AREA). For related physical features see CHINA SEA; JAPAN, SEA OF; and OKHOTSK, SEA OF. For history, see JAPAN, HISTORY OF. For cultural aspects, see ASIAN PEOPLES AND CULTURES; EAST ASIAN CULTURES; DANCE AND THEATRE, EAST ASIAN; LITERATURE, EAST ASIAN; MUSIC, EAST ASIAN; VISUAL ARTS, EAST ASIAN; JAPANESE LANGUAGE; JAPANESE MYTHOLOGY; JAPANESE PHILOSOPHY; and JAPANESE RELIGION.

This article is divided into the following sections:

I. The landscape

THE NATURAL ENVIRONMENT

Relief features. The mountainous character of the country is the outcome of recent orogenic (mountain-building) forces, as evidenced by the frequent occurrence of violent earthquakes, volcanic activity, and signs of change in levels along the coast. There are no sizable structural plains and peneplains (large land areas shaped by erosion), which usually occur in more stable regions of the Earth. The mountains—which cover about 85 percent of Japan's land surface—are for the most part in a youthful stage of dissection in which steep slopes are incised by dense river-valley networks. Rivers are mostly torrential, and their valleys are accompanied by series of river terraces that are the result of movements in the Earth's crust. Recent volcanoes are juxtaposed with old and highly dissected ones. The shores are characterized by elevated and depressed features such as headlands and bays, which display an incipient stage of development.

The mountains are divided into many small land blocks that are separated by lowlands or deep saddles; there is no long or continuous mountain range. These land blocks are the result of intense faulting (movement of adjacent rock masses along a fracture) and warping (bending of the Earth's crust), the former process being regarded as dominant. One consequence is that mountain blocks are often bounded by fault scarps and flexure slopes that descend in step formation to the adjacent lowlands.

Coalescing alluvial fans—cone-shaped deposits of alluvium that run together—are formed where rivers emerge from the mountains. When the rivers are large enough to extend their courses to the sea, low deltaic plains develop in front of the fans; this occurs most frequently where the rivers empty into shallow and sheltered bays, as in the deltas of eastern Kantō, Nōbi, and Osaka. In most places, however, fan surfaces plunge directly into the sea and are separated by low, sandy beach ridges.

Dissected plains are common in the lowlands. Intense disturbances have caused many of the former alluvial fans, deltas, and sea bottoms to be substantially uplifted to form flat-topped uplands such as those found in the Kwanto Plain (Kantb-heiya). Not infrequently the uplands have been overlain with volcanic ash, as in the Kwanto and Tokachi plains.

In addition, Japan is truly a land of volcanoes. Violent eruptions are frequently experienced, and at least 40 volcanoes have been active within historical time. New volcanoes born during the 20th century include Shōwa-shinzan (Shdwa Volcano) on Hokkaidb and Myōjin-shō off the Beyoneisu-retsugan (or Bayonnaise Rocks) in the Pacific. The abundant hot springs are mostly of volcanic origin. Many of the gigantic volcanoes are conical in shape (*e.g.*, Fujiyama [Fuji-san]), while others form steep lava domes (*e.g.*, Dai-sen and Unzen). Conspicuous shield volcanoes (broad, gently sloping volcanic cones such as Mauna Loa on Hawaii Island) are rare, and extensive lava plateaus are lacking. One of the characteristics of the volcanic areas is the prevalence of calderas (large, circular, basin-shaped volcanic depressions), especially in the northeast and southwest, many of which are occupied by lakes, such as Kutcharo, Towada, and Ashino.

Japan's mountains have been influenced by the orogenic formation of six mountain arcs off the Pacific coast of Asia. They are, from northeast to southwest, the Chishima Range of the Kuril Islands; the Karafuto Mountain system of Sakhalin Island; the Northeast, Southwest, and Shichito–Mariana ranges of Japan; and the Ryukyu Island formations. The four major landform areas of Japan—Hokkaidō Island, Northeastern. Central, and Southwestern regions—have developed as a result of the formation of these arcs.

Four major landform regions

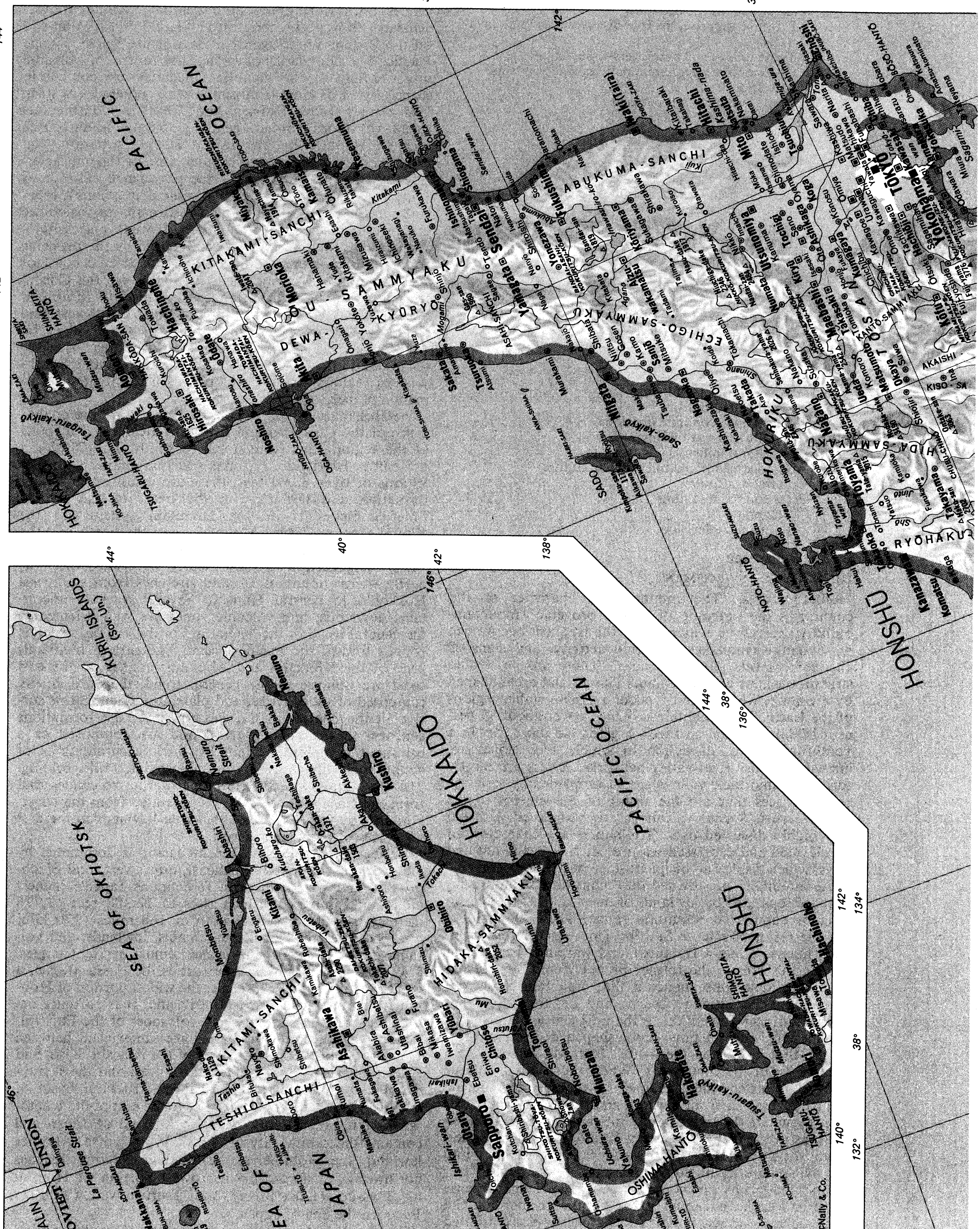
The Hokkaidb Island region was formed by the coalescence of the Chishima and Karafuto arcs. The backbone of the region runs from north to south. The Chishima arc enters Hokkaidb as three volcanic chains with elevations of over 6,000 feet; these are arranged in ladder formation and terminate in the heart of the region. Chief components of the mountain system are the Kitami-sanchi (Kitami Mountains) in the north and the Hidaka-sammyaku (Hidaka Mountains) in the south.

The Northeastern Region nearly coincides with the northeastern mountain arc and stretches from southwest Hokkaidō to central Honshii. Several rows of mountains, lowlands, and volcanic zones are well oriented to the general trend of the insular arc of this region, which is convex toward the Pacific Ocean. The eastern highlands zone of the Kitakami and Abukuma ranges on the east coast are oblique to the general trend; they are chiefly composed of older rocks, and plateau-like landforms survive in the centre. In the western zone, the formations conform to the general trend and are composed of a basement complex overlain by thick accumulations of young rocks that have been subjected to mild folding. The Ōu-sammyaku, capped with towering volcanoes that form the Nasu volcanic zone, is separated from the coastal ranges by the Kitakami–Abukuma lowlands to the east and a row of basins in the west.

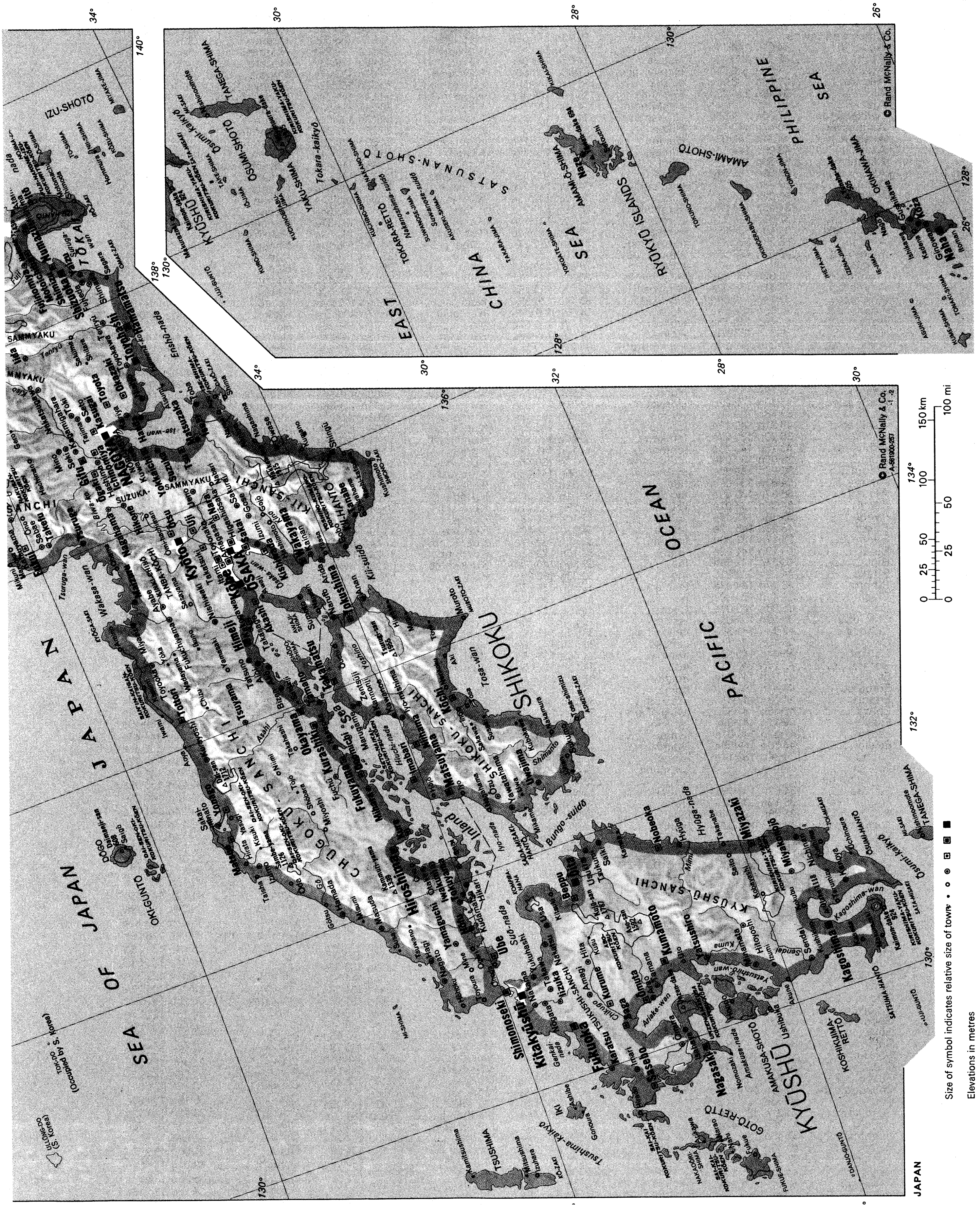
The Central Region of Honshii Island is dominated by the overlapping of the Northeast, Southwest, and Shichito–Mariana mountain arcs. It contains Japan's highest mountain, Fujiyama, which rises to 12,388 feet (3,776 metres), and its broadest width of 168 miles (270 kilometres). The trend of the mountains, lowlands, and volcanic zones intersects the island almost at right angles. The most notable physical feature is the Fossa Magna, a great rift lowland that traverses the widest portion of Honshii from the Japan Sea to the Pacific. It is partially occupied by hills, mountains, and volcanoes of the Fuji volcanic zone. Intermontane basins are sandwiched between the lofty central mountain knots of the Akaishi, Kiso, and Hida (Japanese Alps) ranges to the west and the Kantō-sammyaku to the east. The shallow structural basin of the Kwanto Plain, which stretches to the east of the Kantō-sammyaku, is the most extensive lowland of Japan and contains the capital city of Tokyo in its centre. Another lowland, the Ndbi-heiya, surrounding Nagoya, reaches northwestward across the narrowest portion of Honshū to Biwa-ko (Lake Biwaj and Tsuruga-wan (Tsuruga Bay) on the Sea of Japan.

The Southwestern Region of southern Honshii, Shikoku, and northern Kyiishii generally coincides with the southwest mountain arc, and the general trend of highlands and lowlands is roughly convex toward the Japan Sea. The region is divided into the Inner Zone, which was formed by complex faulting, and the Outer Zone, which

Japan



© Rand McNally & Co.


SEA OF JAPAN
PACIFIC OCEAN
EAST CHINA SEA
PHILIPPINE SEA
KYŪSHŪ
SHIKOKU
CHUGOKU SANCHI
SHIKOKU-SANCHI
KYUSHU-SANCHI
RYUKYU ISLANDS
SATSUNAN-SHOTO
AMAMI-SHOTO
OSUMI-SHOTO
TOKARA-RETTO
IZU-SHOTO
OKI-GUNTO
GOTO-RETTO
TSUSHIMA
AMAKUSA-SHOTO
KOSHIKIJIMA-RETTO
Inland Sea
ULLUNG-DO (S. Korea)
TOK-DO (Occupied by S. Korea)
DOGO
OSAKA
Kyoto
Kobe
NAGOYA
Hiroshima
Okayama
Kurashiki
Fukuyama
Himeji
Takamatsu
Matsuyama
Kochi
Tokushima
Wakayama
Nara
Kitakyushu
Fukuoka
Shimonoseki
Ube
Yamaguchi
Kumamoto
Kagoshima
Nagasaki
Sasebo
Saga
Kurume
Iizuka
Omuta
Yatsushiro
Oita
Beppu
Nobeoka
Miyazaki
Miyakonojo
Tottori
Matsue
Yonago
Tsuyama
Imabari
Niihama
Uwajima
Toyota
Okazaki
Toyohashi
Hamamatsu
Shizuoka
Gifu
Tsu
Naha
Tsushima-kaikyo
Genkai-nada
Suo-nada
Iyo-nada
Bungo-suido
Hyuga-nada
Tosa-wan
Kii-suido
Ise-wan
Enshu-nada
Wakasa-wan
Tsuruga-wan
Ariake-wan
Yatsushiro-kai
Kagoshima-wan
Osumi-kaikyo
Tokara-kaikyo
YAKU-SHIMA
TANEGA-SHIMA
AMAMI-O-SHIMA
OKINAWA-JIMA
34°
32°
30°
28°
26°
128°
130°
132°
134°
136°
138°
140°
0 25 50 100 150 km
0 25 50 100 mi
© Rand McNally & Co.
A-561900-257 -1 -2
Size of symbol indicates relative size of town
Elevations in metres
JAPAN

MAP INDEX

Cities and towns

Abashiri	44·01n	144·17e
Abuta	42·33n	140·46e
Ago	34·20n	136·51e
Aikawa	38·02n	138·15e
Aioi	34·48n	134·28e
Aizu-wakamatsu	37·30n	139·56e
Ajigasawa	40·47n	140·12e
Akabira	43·34n	142·03e
Akan	43·06n	144·10e
Akashi	34·38n	134·59e
Aki	33·30n	133·54e
Akita	39·43n	140·07e
Akkeshi	43·02n	144·51e
Akō	34·45n	134·24e
Akune	32·01n	130·11e
Amagasaki	34·43n	135·25e
Amagi	33·25n	130·39e
Amarume	38·50n	139·55e
Amatsu-kominato	35·07n	140·10e
Amino	35·41n	135·02e
Anamizu	37·14n	136·54e
Anan	33·55n	134·39e
Aomori	40·49n	140·45e
Aoya	35·31n	133·59e
Arai	37·01n	138·15e
Arao	33·00n	130·28e
Arida	34·05n	135·07e
Arikawa	32·59n	129·07e
Asahi	35·12n	137·02e
Asahigawa	43·46n	142·22e
Ashibe	33·48n	129·46e
Ashibetsu	43·31n	142·11e
Ashikaga	36·20n	139·27e
Ashiyoro	43·15n	143·30e
Atami	35·05n	139·04e
Atsugi	35·27n	139·22e
Atsumi	34·37n	137·07e
Atsumi	38·37n	139·35e
Awaji	34·35n	135·01e
Ayabe	35·18n	135·15e
Bekkai	43·23n	145·17e
Beppu	33·17n	131·30e
Bibai	43·19n	141·52e
Biei	43·35n	142·28e
Bifuka	44·29n	142·21e
Bihoro	43·49n	144·07e
Bizen	34·44n	134·09e
Chiba	35·36n	140·07e
Chichibu	35·59n	139·05e
Chitose	42·49n	141·39e
Chizu	35·16n	134·14e
Chōshi	35·44n	140·50e
Date	42·27n	140·51e
Ebetsu	43·07n	141·34e
Embetsu	44·44n	141·47e
Engaru	44·03n	143·31e
Eniwa	42·45n	141·33e
Esashi	39·12n	141·09e
Esashi	41·52n	140·07e
Esashi	44·56n	142·35e
Fuchū	34·34n	133·14e
Fuji, see Motoichiba		
Fujieda	34·52n	138·16e
Fujinomiya	35·12n	138·38e
Fuji-yoshida	35·38n	138·42e
Fukagawa	43·43n	142·03e
Fukaya	36·12n	139·17e
Fukuchiyama	35·18n	135·07e
Fukue	32·41n	128·50e
Fukui	36·04n	136·13e
Fukuoka	33·35n	130·24e
Fukuoka	40·16n	141·18e
Fukushima	37·45n	140·28e
Fukushima	41·29n	140·15e
Fukuyama	34·29n	133·22e
Funabashi	35·42n	139·59e
Furano	43·21n	142·24e
Furukawa	36·14n	137·11e
Furukawa	38·34n	140·58e
Gero	35·48n	137·14e
Gifu	35·25n	136·45e
Ginowan	26·17n	127·46e
Gobō	33·53n	135·10e
Gojō	34·21n	135·42e
Gojōme	39·56n	140·07e
Gōnoura	33·45n	129·41e
Gose	34·27n	135·44e
Gosen	37·44n	139·11e
Goshogawara	40·48n	140·27e
Gōtsu	35·00n	132·14e
Gushikawa	26·21n	127·52e
Haboro	44·22n	141·42e
Hachiman	35·45n	136·57e
Hachinohe	40·30n	141·29e
Hachiōji	35·39n	139·20e
Hagi	34·24n	131·25e
Hakodate	41·45n	140·43e
Hakui	36·53n	136·47e
Hamada	34·53n	132·05e
Hamamatsu	34·42n	137·44e
Hamanaka	43·05n	145·10e
Hamatombetsu	45·07n	142·23e
Hanamaki	39·23n	141·07e
Hanawa	40·11n	140·47e
Handa	34·53n	136·56e
Haranomachi	37·38n	140·58e
Hasaki	35·44n	140·50e
Hashima	35·19n	136·42e
Hashimoto	34·19n	135·37e
Higashibsaka	34·39n	135·35e
Hikari	33·58n	131·56e
Hikone	35·15n	136·15e
Himeji	34·49n	134·42e
Himi	36.51 n	136·59e
Hirado	33·22n	129·33e
Hiraizumi	38·59n	141·07e
Hirata	35·26n	132·49e
Hiratsuka	35·19n	13921e
Hiroo	42·17n	143·19e
Hirosaki	40·35n	140·28e
Hiroshima	34·24n	132·27e
Hita	33·19n	130·56e
Hitachi	36·36n	140·39e
Hitachi-ōta	36·32n	140·31e
Hitoyoshi	32·13n	130·45e
Hiwasa	33·44n	134·32e
Hófu	34·03n	131·34e
Hōhoku	34·17n	130·57e
Hōjō	33·58n	132·46e
Hommura	34·22n	139·15e
Honbetsu	43·07n	143·37e
Hondo	32·27n	130·12e
Honjō	36·14n	139.11e
Honjō	39·23n	140·03e
Horoizumi	42·01n	143·09e
Hyūga	32·25n	131·38e
Ichihara	35.31n	140·05e
Ichikawa	35·44n	139·55e
Ichinohe	40·13n	141·17e
Ichinomiya	35·18n	136·48e
Ichinoseki	38·55n	141·08e
Iida	35·31n	137·50e
Iiyama	36.51 n	138·22e
Iizuka	33·38n	130.41e
Ikuno	35·10n	134·48e
Imabari	34·03n	133·00e
Imaichi	36·43n	139·41e
Imari	33·16n	129·53e
Ina	35·59n	140·03e
Ino	33·33n	133·26e
Isahava	32·50n	130·03e
Ise	34·29n	136·42e
Isesaki	36·19n	139·12e
Ishikawa	26·25n	127·50e
Ishinomaki	38·25n	141·18e
Ishioka	36·11n	140·16e
Itami	34·46n	135·25e
Itō	34·58n	139·05e
Itoigawa	37·02n	137·51e
Itoman	26·08n	127·40e
Iwaizumi	39·50n	141·48e
Iwaki (Taira)	37·03n	140·55e
Iwakuni	34·09n	132·11e
Iwami	35·35n	134·20e
Iwamizawa	43·12n	141·46e
Iwanai	42·58n	140·30e
Iwanuma	38·06n	140·52e
Iyo	33·46n	132·42e
Iyo-mishima	33·58n	133·33e
Izuhara	34·12n	129·17e
Izumi	32·05n	130·22e
Izumi	34·29n	135·26e
Izumo	35·22n	132·46e
Jbhen	32·57n	132·35e
Kadena	26·22n	127·45e
Kaga	36·18n	136·18e
Kagamigahara	35·24n	136·54e
Kagoshima	31·36n	130·33e
Kainan	34·09n	135·12e
Kakizaki	37·15n	138·25e
Kakogawa	34·46n	134·51e
Kakuda	37·58n	140·47e
Kamae	32·48n	131·56e
Kamaishi	39·16n	141·53e
Kamakura	35·19n	139·33e
Kameda	41·48n	140·36e
Kamiiso	41·49n	140·39e
Kamikawa	43·49n	142·45e
Kaminokuni	41·48n	140·06e
Karnioka	36·16n	137·18e
Karnitsushima	34·50n	129·28e
Kamo	37·39n	139·03e
Kanagawa	37·35n	139·55e
Kanazawa	36·34n	136·39e
Kanonji	34·07n	133·39e
Kanoya	31·23n	130·51e
Kanuma	36·33n	139·44e
Karatsu	33·26n	129·58e
Kariya	34·59n	136·59e
Karumai	40·19n	141·28e
Kasama	36·23n	140·16e
Kasaoka	34·30n	133·30e
Kaseda	31·25n	130·19e
Kashima	35·58n	140·38e
Kashiwa	35·52n	139·59e
Kashiwazaki	37·22n	138·33e
Kasugai	35·14n	136·58e
Kasumi	35·38n	134·38e
Katsuta	36·24n	140·30e
Katsuura	35·08n	140·18e
Katsuyama	36·03n	136·30e
Kawagoe	35·55n	139·29e
Kawaguchi	35·48n	139·43e
Kawanoe	34·01n	133·34e
Kawasaki	35·32n	139·43e
Kesennuma	38·54n	141·35e
Kiryū	36·24n	139·20e
Kisakata	39·13n	139·54e
Kisarazu	35·23n	139·55e
Kishiwada	34·28n	135·22e
Kisuki	35·17n	132·54e
Kitaibaraki	36·48n	140·45e
Kitakami	39·18n	141·07e
Kitakata	37·39n	139·52e
Kitakyūshū	33·53n	130·50e
Kitami	43·48n	143·54e
Kitsuki	33·25n	131·37e
Kobayashi	31·59n	130·59e
Kōbe	34·41n	135·10e
Kōchi	33·33n	133·33e
Kōfu	35·39n	138·35e
Koga	36·11n	139.43e
Koide	37·13n	138·57e
Kokubu	34·56n	132·07e
Komatsu	36·24n	136·27e
Komoro	36·19n	138·26e
Kōnosu	36·03n	139·31e
Kōriyama	37·24n	140·23e
Kosaka	40·19n	140·44e
Koza	26·20n	127·50e
Kubokawa	33·12n	133·08e
Kudamatsu	34·00n	131·52e
Kuji	40·11n	141·46e
Kumagaya	36·08n	139·23e
Kumaishi	42·08n	139·59e
Kumamoto	32·48n	130·43e
Kumano	33·54n	136·05e
Kurashiki	34·35n	133·46e
Kurayoshi	35·26n	133·49e
Kure	34·14n	132·34e
Kurobe	36·51n	137·26e
Kuroishi	40·38n	140·36e
Kuroiso	36·58n	140·03e
Kurume	33·19n	130·31e
Kushikino	31·44n	130·16e
Kushima	31·29n	131·14e
Kushimoto	33·28n	135·47e
Kushiro	42·58n	144·23e
Kusu	33·16n	131·09e
Kutchan	42·54n	140·45e
Kuwana	35·04n	136·42e
Kyōto	35·00n	135·45e
Maebashi	36·23n	139·04e
Maizuru	35·27n	135·20e
Maki	37·45n	138·53e
Makurazaki	31·16n	130·19e
Marugame	34·17n	133·47e
Maruoka	36·09n	136·16e
Mashike	43·51n	141·31e
Masuda	34·40n	131·51e
Matsudo	35·47n	139·54e
Matsue	35·28n	133·04e
Matsumae	41·26n	140·07e
Matsumoto	36·14n	137·58e
Matsushima	38·22n	141·04e
Matsuyama	33·50n	132·45e
Matsuzaka	34·34n	136·32e
Mihara	34·24n	133·05e
Mikame	33·25n	132·27e
Mikasa	43·20n	141·40e
Mikuni	36·13n	136·09e
Mimmaya	41·12n	140·26e
Minamata	32·13n	130·24e
Mine	33·17n	130·26e
Mino	35·32n	136·55e
Misawa	40·41n	141·24e
Mishima	35·07n	138·55e
Misumi	34·46n	131·58e
Mito	36·22n	140·28e
Mitsuke	37·32n	138·56e
Mitsushima	34·10n	129·20e
Miura	35·08n	139·37e
Miyako	39·38n	141·57e
Miyakonojō	31·44n	131·04e
Miyazaki	31·54n	131·26e
Miyazu	35·32n	135·11e
Miyoshi	34·48n	132·51e
Mizusawa	39·08n	141·08e
Mobara	35·25n	140·18e
Mōka	36·26n	140·01e
Mombetsu	44·21n	143·22e
Monzen	37·17n	136·46e
Mori	42·06n	140·35e
Morioka	39·42n	141·09e
Motoichiba	35·09n	138·39e
Motomachi	34·45n	139·21e
Mugi	33·40n	134·25e
Murakami	38·14n	139·29e
Muroran	42·18n	140·59e
Muroto	33·18n	134·09e
Mutsu	41·17n	141·10e
Nabari	34·37n	136·05e
Nachi-Katsuura	33·30n	135·55e
Nagahama	33·36n	132·29e
Nagahama	35·23n	136·16e
Nagai	38·06n	140·02e
Nagano	36·39n	138·11e
Nagaoka	37·2	7138n51e
Naaasaki	32·48n	129·55e
Nagashima	34·12n	136·20e
Nagato	34·21n	131·10e
Nago	26·35n	127·59e
Nagoya	35·10n	136·55e
Naha	26·13n	127·40e
Nakajd	38·03n	139·24e
Nakama	33·50n	130·43e
Makaminato	36.21 n	140·36e
Nakamura	32·59n	132·56e
Nakano	36·45n	138·22e
Nakashibetsu	43·33n	144·59e
Nakatsu	33·34n	131·13e
Nakatsugawa	35·29n	137·30e
Namie	37·29n	141·00e
Nanao	37·03n	136·58e
Nandan	34·15n	134·43e
Nankoku	33·39n	133·44e
Nansei	34·22n	136·41e
Nanto	34·20n	136·31e
Nanyb	38·03n	140·10e
Naoetsu	37.11 n	138·15e
Nara	34.41 n	135·50e
Narao	32·50n	129·04e
Narita	35·47n	140·19e
Naru	32·49n	128·56e
Naruko	38·44n	140·43e
Naruto	34·11n	134·37e
Natori	38·08n	140·55e
Nayoro	4421n	142·28e
Naze	28·23n	129·30e
Nemuro	43·20n	145·35e
Nichinan	31·36n	131·23e
Niigata	37·55n	139·03e
Niihama	33·58n	133·16e
Niimi	34·59n	133·28e
Niitsu	37·48n	139·07e
Nikkd	36·45n	139·37e
Nishinoomote	30·44n	131·00e
Nishio	34·52n	137·03e
Nishiwaki	34·59n	134·58e
Nobeoka	32·35n	131·40e
Noboribetsu	42·27n	141·1l e
Nbgata	33·44n	130·44e
Noheji	40·52n	141·08e
Nomozaki	32·35n	129·45e
Noshiro	40·12n	140·02e
Noto	37·18n	137·09e
Numata	36·38n	139·03e
Numata	43·48n	141·57e
Numazu	35·06n	138·52e
Nyūzen	36·56n	137·30e
Ōarai	36·18n	140·34e
Obarna	35·30n	135·45e
Obihiro	42·55n	143·12e
Obira	44·00n	141·35e
Ōda	35·11n	132·30e
Odaka	37·34n	141·00e
Ōdate	40·16n	140·34e
Odawara	35·15n	139·10e
Ōfunato	39·04n	141·43e
Oga	39·53n	139.51e
Ōgaki	3521n	136·37e
Ogatsu	38·41n	141·28e
Ogōri	33·22n	130·32e
Ōhara	35·15n	140·23e
Ōhata	41·24n	141·10e
Ōita	33·14n	131·36e
Ojiya	37·18n	138·48e
Okaya	36·03n	138·03e
Okayama	34·39n	133·55e
Okazaki	34·57n	137·10e
Okushiri	42·10n	139·31e
Ōma	41·32n	140·55e
Ōmagari	39·27n	140·29e
Ōme	35·47n	139·15e
Ōmi-hachiman	35·08n	136·06e
Ōmiya	35·54n	139·38e
Ōmu	44·34n	142·58e
Ōmura	32·54n	129·57e
Ōmuta	33·02n	130·27e
Onagawa	38·26n	141·27e
Ōno	34·18n	132·17e
Ōno	35·59n	136·29e
Onoda	33·59n	131·11e
Onomichi	34·25n	133·12e
Ōsaka	34·40n	135·30e
Oshamambe	42·30n	140·22e
Oshika	38·16n	141·32e
Ōta	36·18n	139·22e
Ōtake	34·12n	132·13e
Otaru	43·13n	141·00e
Otawara	36·52n	140·02e
Ōtsu	35·00n	135·52e
Ōtsuchi	39·21n	141·54e
Ōtsuki	35·36n	138·57e
Ōuchi	34·16n	134·18e
Owase	34·04n	136.12e
Oyabe	36·40n	136·52e
Oyama	35·36n	139·22e
Ōzu	33·30n	132·33e

MAP *INDEX* *(continued)*

Rausu	44·01n	145·12e
Rikuzen-takata	39·01n	141·38e
Rubeshibe	43·47n	143·38e
Rumoi	43·56n	141·39e
Ryōtsu	38·05n	138·26e
Sabae	35·57n	136·11e
Saga	33·15n	130·18e
Sagae	38·22n	140·17e
Sagamihara	35·32n	139·23e
Sagara	34·41n	138·12e
Saidaiji	34·39n	134·02e
Saigō	36·12n	133·20e
Saiki	32·57n	131·54e
Saito	32·06n	131·24e
Sakai	34·35n	135·28e
Sakai-minato	35·33n	133·15e
Sakata	38·55n	139·50e
Sakawa	33·30n	133·17e
Sakuma	35·05n	137·48e
Sakurai	34·30n	135·51e
Samani	42·07n	142·56e
Sanjō	37·37n	138·57e
Sano	36·19n	139·35e
Sapporo	43·03n	141·21e
Sarufutsu	45·16n	142·12e
Sasayama	35·04n	135·13e
Sasebo	33·10n	129·43e
Sawara	35·53n	140·30e
Sawata	38·00n	138·16e
Seki	35·29n	136·55e
Sendai	31·49n	130·18e
Sendai	38·15n	140·53e
Setana	42·26n	139·51e
Seto	35·14n	137·06e
Setouchi	28·10n	129·15e
Shari	43·55n	144·50e
Shibata	37·57n	139·20e
Shibecha	43·17n	144·36e
Shibetsu	43·40n	145·08e
Shibetsu	44·10n	142·23e
Shima	34·59n	135·20e
Shimada	34·49n	138·11e
Shimizu	43·01n	142·53e
Shimizu	35·01n	138·29e
Shimminato	36·47n	137·04e
Shimoda	34·40n	138·57e
Shimodate	36·18n	139·59e
Shimokawa	44·18n	142·39e
Shimonoseki	33·57n	130·57e
Shingū	33·44n	135·59e
Shinjō	38·46n	140·18e
Shiogama	38·19n	141·01e
Shiojiri	36·06n	137·58e
Shirahama	33·40n	135·20e
Shirakawa	37·07n	140·13e
Shiranuka	42·57n	144·05e
Shiraoi	42·33n	141·21e
Shiroishi	38·00n	140·37e
Shitara	35·05n	137·35e
Shizugawa	38·40n	141·27e
Shizunai	42·20n	142·22e
Shizuoka	34·58n	138·23e
Shōbara	34·51n	133·01e
Sōma	37·48n	140·57e
Sukagawa	37·17n	140·23e
Sukumo	32·56n	132·44e
Sumoto	34·21n	134·54e
Sunagawa	43·29n	141·55e
Susa	34·37n	131·36e
Susaki	33·22n	133·17e
Suttsu	42·48n	140·14e
Suwa	36·02n	138·08e
Suzaka	36·39n	138·19e
Suzu	37·25n	137·17e
Suzuka	34·51n	136·35e
Tachikawa	35·42n	139·25e
Tadami	37·21n	139·19e
Tagajō	38·20n	141·00e
Taira, see Iwaki		
Taisha	35·24n	132·40e
Tajimi	35·19n	137·08e
Takada	37·06n	138·15e
Takahagi	36·43n	140·43e
Takahashi	34·47n	133·37e
Takamatsu	34·20n	134·03e
Takanabe	32·08n	131·30e
Takanosu	40·13n	140·22e
Takaoka	36·45n	137·01e
Takasago	34·45n	134·48e
Takasaki	36·20n	139·01e
Takatsuki	34·51n	135·37e
Takawa	33·38n	130·49e
Takayama	36·08n	137·15e
Takefu	35·54n	136·10e
Takikawa	43·33n	141·54e
Tamana	32·55n	130·33e
Tamano	34·30n	133·56e
Tanabe	33·44n	135·22e
Taneichi	40·26n	141·43e
Tarumizu	31·29n	130·42e
Tateyama	34·59n	139·52e
Tatsuno	34·52n	134·33e
Tendō	38·21n	140·22e
Tenryū	34·52n	137·49e
Teshikaga	43·29n	144·28e
Teshio	44·53n	141·44e
Toba	34·29n	136·51e
Tōbetsu	43·13n	141·31e
Tochigi	36·23n	139·44e
Tōgane	35·33n	140·22e
Togi	37·08n	136·44e
Tohoku	40·40n	141·15e
Tōjō	34·53n	133·16e
Tōkamachi	37·08n	138·46e
Toki	35·21n	137·11e
Tokushima	34·04n	134·34e
Tokuyama	34·03n	131·49e
Tōkyō	35·42n	139·46e
Tomakomai	42·38n	141·36e
Tomioka	36·15n	138·54e
Tonami	36·38n	136·54e
Tōno	39·19n	141·32e
Tosa	33·29n	133·25e
Tosa-shimizu	32·46n	132·57e
Tottori	35·30n	134·14e
Towada	40·37n	141·13e
Toyama	36·41n	137·13e
Tōyō	33·22n	134·18e
Toyohashi	34·46n	137·23e
Toyokawa	34·49n	137·24e
Toyonaka	34·47n	135·28e
Toyooka	35·32n	134·50e
Toyota	35·05n	137·09e
Toyoura	34·08n	130·58e
Tsu	34·43n	136·31e
Tsubame	37·39n	138·56e
Tsuchiura	36·05n	140·12e
Tsukumi	33·04n	131·52e
Tsuruga	35·39n	136·04e
Tsuruoka	38·44n	139·50e
Tsuwano	34·28n	131·46e
Tsuyama	35·03n	134·00e
Ube	33·56n	131·15e
Uchinoura	31·16n	131·05e
Ueda	36·24n	138·16e
Ueno	34·45n	136·08e
Uji	34·53n	135·48e
Uozu	36·48n	137·24e
Urahoro	42·48n	143·39e
Urakawa	42·09n	142·47e
Urawa	35·51n	139·39e
Usa	33·31n	131·22e
Ushibuka	32·11n	130·01e
Usuki	33·08n	131·49e
Utashinai	43·31n	142·03e
Uto	32·41n	130·40e
Utsunomiya	36·33n	139·52e
Uwa	33·20n	132·30e
Uwajima	33·13n	132·34e
Wadayama	35·19n	134·52e
Wajima	37·24n	136·54e
Wakayama	34·13n	135·11e
Wakayanagi	38·46n	141·08e
Wakkanai	45·25n	141·40e
Yaizu	34·52n	138·20e
Yakumo	42·15n	140·16e
Yamada	39·28n	141·57e
Yamagata	38·15n	140·15e
Yamaguchi	34·10n	131·29e
Yamasaki	35·00n	134·33e
Yanai	33·58n	132·07e
Yasugi	35·26n	133·15e
Yatsuo	36·34n	137·08e
Yatsushiro	32·30n	130·36e
Yawata	35·43n	139·56e
Yawatahama	33·27n	132·24e
Yoichi	43·12n	140·41e
Yōka	35·24n	134·46e
Yōkaichiba	35·42n	140·33e
Yokkaichi	34·58n	136·37e
Yokohama	35·27n	139·39e
Yokosuka	35·18n	139·40e
Yokote	39·18n	140·34e
Yonago	35·26n	133·20e
Yonezawa	37·55n	140·07e
Yuasa	34·02n	135·11e
Yūbari	43·04n	141·59e
Yūbetsu	43·13n	144·05e
Yuda	39·20n	140·50e
Yukuhashi	33·44n	130·59e
Yuza	39·01n	139·54e
Yuzawa	39·10n	140·30e
Zentsūji	34·14n	133·47e

Physical features and points of interest

Abukuma, river	38·02n	140·56e
Abukuma-sanchi, mountains	37·30n	140·45e
Agano, river	37·57n	139·08e
Aguni-jima, *island*	26·35n	127·14e
Akaishi-*sammvzku, mountain range*	35·18n	138·07e
Akan-kokuritsu-*kben, national park*	43·30n	144·15e
Akuseki-shima, island	29·27n	129·37e
Arnakusa-nada, sea	32·35n	130·05e
Amakusa-shotb, isiands	32·20n	*130.15e*
Amami-Ō-shima, *island*	28·15n	129·20e
Amami-shotb, islands	28·16n	*129.21e*
Ara, river	35·39n	*139.51e*
Ariake-wan, bay	31·24n	131·08e
Asahi, river	34·36n	133·58e
Asahi-dake, mountain	43·40n	142·51e
Asahi-sanchi, mountains	38·25n	139·50e
Asama-yama, mountain	36·24n	138·31e
Ashizuri-zaki, cape	32·44n	133·01e
Aso-kokuritsu-*kben, national park*	33·00n	131·07e
Aso-san, mountain	32·53n	131·06e
Awaji-shima, *island*	34·21n	134·51e
Awa-shima, island	38·27n	139·14e
Bandai-asahi-*kokuritsu-kben, national park*	38·16n	139·57e
Bandai-san, mountain	37·36n	140·04e
Biwa-ko, lake	35·15n	136·05e
Bōsō-hantō, *peninsuia*	35·18n	140·10e
Bungo-suidō, *strait*	33·00n	132·13e
Chichibu-tama-*kokuritsu-kben, national park*	35·52n	139·00e
Chikugo, river	33·09n	130·21e
Chbbu-chihb, physical region	36·00n	137·30e
Chūbu-sangaku-*kokuritsu-kben, national park*	36·30n	137·41e
Chūgoku-sanchi, *mountains*	34·58n	132·57e
Daimanji-san, *mountain*	36·15n	133·19e
Daiō-zaki, *cape*	34·17n	136·54e
Dai-sen, mountain	35·22n	133·33e
Dai-sen-oki-kokuritsu-kōen, *national park*	35·20n	133·35e
Daisetsu-zan, *see Asahi-dake*		
Daisetsu-zan-kokuritsu-kōen, *national park*	43·30n	142·57e
Danjo-guntō, *islands*	32·02n	128·23e
Dewa-kyūryō, *hills*	39·05n	140·10e
Dōgo, *islands*	36·15n	133·16e
East China *Sea* (Tō-Kai)	29·00n	128·30e
Echigo-*sammyaku, mountain range*	37·50n	139·50e
Enshū-nada, *sea*	34·27n	137·38e
Erimo-misaki, cape	41·55n	143·15e
Esan-saki, point	41·49n	141·11e
Ezo-fuji, *see* Shiribeshi-yama		
Fuji, river	35·07n	138·38e
Fuji-hakone-izu-*kokuritsu-kben, nationai park*	35·21n	138·44e
Fuji-yama, *volcano*	35·22n	138·44e
Fukue-shima, island	32·40n	128·45e
Gō, *river*	35·02n	132·13e
Gotō-rettō, *islands*	32·50n	129·00e
Hajiki-saki, *cape*	38·19n	138·31e
Hakken-san, mountain	34·10n	135·54e
Hakkōda-san, *mountains*	40·40n	140·53e
Hako-dake, mountain	44·40n	142·25e
Hakone-yama, mountain	35·14n	139·02e
Haku-san, mountain	36·09n	136·46e
Haku-san-*kokuritsu-kben, national oark*	36·12n	136·47e
Hamana-Ko, *lake*	34·45n	137·34e
Hayachine-san, mountain	39·34n	141·29e
Hidaka-*sammyaku, mountain range*	42·35n	142·45e
Hida-sammyaku, *mountain range*	36·25n	137·40e
Hiuchiga-dake, mountain	36·57n	139·17e
Hiuchi-nada, sea	34·05n	133·20e
Hodaka-dake, mountain	36·17n	137·39e
Hokkaidb, island	44·00n	143·00e
Hokuriku, historic region	37·00n	138·00e
Honshb, isiand.	36·00n	138·00e
Horoshiri-dake, *mountain*	42·43n	*142.41e*
Hybga-nada, sea	32·00n	131·35e
Ie-jima, *island*	26·43n	127·47e
Iheya-jima, *isiand*	27·00n	127·56e
iki, isiand	33·47n	129·43e
Inawashiro-ko, *lake*	37·29n	140·06e
Inland Sea (Seto Uchi; Seto-naikai)	34·20n	133·30e
Inubō-saki, *cape*	35·42n	140·53e
Iō-jima (Iwo *Jima), island*	30·48n	130·18e
Irb-zaki, cape	34·36n	138·51e
Ise-shima-*kokuritsu-kben, national park*	34·23n	136·48e
Ise-wan, bay	34·43n	136·43e
Ishikari, river	43·15n	141·32e
Ishikari-wan, bay	43·25n	141·01e
Iwaki, river	41·01n	140·22e
Iwaki-san, *mountain*	40·39n	140·18e
iwate-san, mountain	39·51n	141·00e
Iwo *Jima, see* Iō-jima		
Iyo-nada, *sea*	33·40n	132·20e
Izena-jima, *island*	26·56n	127·56e
Izu-hantō, *islands*	34·45n	139·00e
Izu-shotō, *islands*	34·00n	139·30e
Japan, Sea of	37·00n	132·00e
Jintb, river	36·46n	137·13e
Jōshin'etsu-kōgen-*kokuritsu-kben, national park*	36·46n	138·40e
Kagoshima-wan, bay	31·25n	130·38e
Kaimon-dake, *mountain*	31·11n	130·32e
Kammuri-yama, mountain	34·28n	132·05e
Kamui-misaki, *cape*	43·20n	*140.21e*
Kantō-*sammyaku, mountain range*	35·50n	138·50e
Kashima-nada, sea	36·15n	140·45e
Kasumiga-ura, lake	36·00n	140·25e
Kii-hantō, *oeninsuia*	34·00n	135·45e
Kii-sanchi, mountains	34·00n	135·50e
Kii-suidō, *strait*	33·55n	134·55e
Kikai-shima, island	28·19n	129·59e
Kinki-chihb, physicai region	35·00n	135·30e
Kino, river	34·13n	135·09e
Kinpoku-san, *mountain*	38·05n	138·22e
Kirishima-yaku-kokuritsu-kōen, *nationai park*	31·55n	130·51e
Kiso, river	35·02n	136·45e
Kiso-sammyaku, mountain range	35·43n	137·50e
Kitakami, river	38·25n	141·19e
Kitami-sanchi, mountain range	44·22n	142·43e
Ko-jima, *island*	41·22n	139·48e
Komaga-dake, mountain	42·04n	140·41e

MAP INDEX *(continued)*

Kompira-hana, point	33·40n	131·40e
Koshikijima-rettō, islands	31·45n	129·49e
Kō-zaki, cape	34·05n	129·13e
Kōzu-shima, island	34·13n	139·10e
Kuchinoerabu-jima, island	30·28n	130·12e
Kuchino-shima, island	29·57n	129·57e
Kuji, river	36·29n	140·37e
Kujū-san, mountain	33·05n	131·15e
Kuma, river	32·30n	130·34e
Kume-shima, island	26·20n	126·47e
Kuro-shima, island	30·50n	129·57e
Kutcharo-ko, lake	43·38n	144·21e
Kyōga-saki, cape	35·46n	135·13e
Kyūshū, island	33·00n	131·00e
Kyūshū-sanchi, mountains	32·35n	131·17e
La Perouse Strait	45·45n	142·00e
Me-akan-dake, mountain	43·23n	144·01e
Mimi, river	32·20n	131·37e
Minami-alps-kokuritsu-kben, national park	35·40n	138·13e
Mi-saki, cape	30·50n	131·04e
Mi-shima, island	34·46n	131·09e
Miyake-jima, island	34·05n	139·32e
Miyanoura-dake, mountain	30·20n	130·31e
Mogarni, river	38·55n	139·48e
Mu, river	42·33n	141·56e
Muroto-zaki, cape	33·15n	134·11e
Mutsu-wan, bay	41·05n	140·55e
Mybkb-zan, mountain	36·52n	138·07e
Nagoya Plain, see Nōbi-heiya		
Nakadbri-shima, island	32·57n	129·04e
Nakano-shima, island	29·49n	129·52e
Nakanoshima-suidb, sfrait	29·44n	129·49e
Nansei-shotō, see Ryukyu Islands		
Nantai-zan, mountain	36·46n	139·29e
Nasu-dake, mountain	37·07n	139·58e
Nemuro Strait	44·00n	145·20e
Nii-shima, island	34·22n	139·16e
Nikkō-kokuritsu-kden, national park	36·49n	139·33e
Ndbi-heiya (Nagoya Plain)	35·15n	136·45e
Nojima-zaki, cape	34·56n	139·53e
Noto-hantō, peninsula	37·20n	137·00e
Nyūdō-zaki, cape	40·00n	139·42e
O-akan-dake, mountain	43·27n	144·10e
Oga-hantb, peninsula	39·55n	139·50e
Ojika-hantō, peninsula	38·20n	141·30e
Okhotsk, Sea of	43·00n	144·00e
Oki-guntō, islands	36·15n	133·15e
Okinawa-jima, island	26·30n	128·00e
Okinoerabu-shima, island	27·22n	128·35e
Okushiri-tb, island	42·10n	139·27e
Omae-zaki, cape	34·36n	138·14e
Ōma-zaki, point	41·32n	140·55e
Ontake-san, mountain	35·53n	137·29e
Osaka-wan, bay	34·30n	135·18e
Ō-shima, island	34·43n	139·24e
Ō-shima, island	41·30n	139·22e
Oshima-hantb, peninsula	42·00n	140·30e
Ōsumi-hantō, peninsula	31·20n	130·55e
Ōsumi-kaikyō, sfrait	31·00n	131·00e
Ōsumi-shotō, islands	30·30n	130·00e
Otake, mountain	31·25n	130·29e
Ōu-sammyaku, mounfain range	38·45n	140·50e
Pacific Ocean	31·30n	134·00e
Philippine Sea	27·00n	129·00e
Rebun-jima, island	45·23n	141·02e
Rikuchūkaigan-kokuritsu-kben, national park	39·25n	141·57e
Rishiri-td, island	45·11n	141·15e
Rishiri-zan, mountain	45·11n	141·15e
Rydhaku-sanchi, mountains	36·09n	136·45e
Ryukyu Islands (Nansei-shotō)	27·15n	128·15e
Sada-misaki-hantb, peninsula	33·26n	132·13e
Sado, island	38·00n	138·25e
Sado-kaikyb, strait	37·50n	138·40e
Sagami-nada, gulf	35·00n	139·30e
Sai-kai-kokuritsu-kben, nafional park	33·12n	129·22e
Same-yarna, mountain	35·08n	132·37e
San'in-kaigan-kokuritsu-kben, national park	35·38n	134·38e
Sata-misaki, cape	30·59n	130·40e
Satsuma-hantb peninsula	31·25n	130·25e
Satsunan-shotb, islands	29·00n	130·00e
Sendai, river	31·51n	130·12e
Sendai-wan, bay	38·18n	141·18e
Sennokura-yama, mountain	36·49n	138·50e
Seto-naikai, see inland Sea		
Seto-naikai-kokuritsu-kben, national park	34·15n	133·28e
Seto uchi, see inland Sea		
Shakotan-hantb peninsula	43·20n	140·30e
Shichi-to, see Tokara-guntb		
Shikoku, island	33·45n	133·30e
Shikoku-sanchi, mountains	33·47n	133·30e
Shikotsu-tōya-kokuritsu-kben, national park	42·47n	141·00e
Shimando, river	32·56n	133·00e
Shimokita-hantd, peninsula	41·15n	141·00e
Shinano, river	37·57n	139·04e
Shioya-zaki, caoe	36·59n	140·59e
Shiretoko-kokuritsu-kben, national park	44·08n	145·10e
Shiretoko-misaki, cape	44·14n	145·17e
Shiribeshi-yama (Ezo-fuji), mountain	42·49n	140·49e
Shiriya-saki, cape	41·26n	141·28e
Shd, river	36·47n	137·04e
Shbdo-shima, island	34·30n	134.17e
Showa-Shin-Zan, volcano	42 30n	140·50e
Soya-misaki, cape	45·31n	141·56e
Suō-nada, sea	33·50n	131·30e
Suruga-wan, bav	34·51n	138·33e
Suwanose-jima, island	29·38n	129·43e
Suwanose-suidb, strait	29·32n	129·40e
Suzuka-sammyaku, mountain range	35·00n	136·25e
Suzu-misaki, cape	37·31n	137·21e
Takara-jirna, island	29·09n	129·13e
Tanba-kdchi, plateau	35·17n	135·30e
Tanega-shirna (Tane Island)	30·40n	131·00e
Tappi-zaki, cape	41·15n	140·21e
Tate-yama, mountain	36·35n	137·37e
Tenryū, river	34·39n	137·47e
Teshio, river	44·53n	141·44e
Teshio-sanchi. mountain range	44·15n	142·05e
Teuri-tō, island	44·25n	141·19e
Tobi-shima, island	39·12n	139·34e
Todo-saki, cape	39·33n	142·05e
Toi-misaki, cape	31·20n	131·22e
Tokachi, river	42·44n	143·42e
Tokachi-dake, mountain	43·25n	142·41e
Tbkai, historic region	35·00n	139·00e
Tō-kai, see East China Sea		
Tokara-kaikyb, strait	30·10n	130·10e
Tokara-guntb (Shichi-to), islands	29·37n	129·43e
Tokuno-shima, island	27·45n	128·58e
Tdkyb-wan, bay	35·25n	139·47e
Tonaki-shirna, island	26·21n	127·09e
Tone, river	35·44n	140·51e
Tbsan, historic region	36·00n	139·00e
Tosa-wan, bay	33·20n	133·40e
To-shima, island	34·31n	139·17e
Towada-hachimantai-kokuritsu-kben, nafional park	40·35n	140·53e
Towada-ko, lake	40·28n	140·55e
Tsugaru-hantō, peninsula	41·00n	140·30e
Tsugaru-kaikyb, strait	41·35n	141·00e
Tsukushi-sanchi, mountains	33·30n	130·30e
Tsuruga-wan, bay	35·43n	136·04e
Tsurugi-san, mountain	33·51n	134·06e
Tsushima, islands	34·20n	129·30e
Tsushima-kaikyb, channel	34·00n	129·00e
Uchiura-wan, bay	42·20n	140·40e
Uji-guntō, islands	31·1l n	129·27e
Unzen-amakusa-kokuritsu-kben, national oark	32·45n	130·17e
Unzen-dake, mounfains	32·45n	130·17e
Wakasa-wan, bay	35·45n	135·40e
Yakishiri-jima, island	44·26n	141·25e
Yaku-shima, island	30·20n	130·30e
Yatsushiro-wan, bay	32·20n	130·25e
Yokoate-shima, island	28·48n	129·00e
Yonana-dake, hill	26·43n	128·13e
Yoneshiro, river	40·13n	140·01e
Yoron-jima, island	27·02n	128·26e
Yoshino, river	34·05n	134·36e
Yoshino-kumano-kokuritsu-kben, national park	34·07n	135·55e
Yūbetsu, river	44·14n	143·37e
Yūfutsu, river	42·37n	141·44e
Yūwan-dake, mountain	28·18n	129·21e

was formed by warping. The Inner Zone is chiefly composed of granite, Paleozoic (between 225,000,000 and 570,000,000 years old), and volcanic rocks, which are arranged in complicated juxtaposition. The Outer Zone, consisting of the Akaishi, Kii, Shikoku, and Kyūshū mountain groups, in contrast, is characterized by a regular zonal arrangement from north to south of crystalline schists, Paleozoic, Mesozoic (between 65,000,000 and 225,000,000 years old) and Tertiary (between 2,500,000 and 65,000,000 years old) formations. The outstanding surface features of the Inner Zone present a highly complex mosaic of numerous fault blocks, while those of the Outer Zone are continuous except where the sea straits separate them into the four independent groups. The Inland Sea (Seto-naikai) is the region where the greater amount of depression has resulted in the invasion of sea waters. The northern edge of the Inner Zone is studded with the gigantic lava domes of the Dai-sen volcanic zone, which, together with volcanic Asb-zan, bury a considerable part of the western extension of the Inland Sea in central Kyūshū.

The Nansei Islands Region constitutes the main portion of the Ryukyn arc, which penetrates into Kyūshū Island as the Kirishima volcanic zone and terminates at Asb-zan. The influence of the arc is also seen in the trend of the many elongated islands off western Kyūshū, including Koshiki, Gotb, Hirado, and Tsushima islands.

The islands of the Shichito–Ogasawara region, far to the east of the Ryukyu arc, are comprised of volcanoes resting on the submarine ridge of the Shichito–Marina arc and the Ogasawara, or Bonin, Islands, which include Peel Island and Iwo Jima.

Drainage and soils. The increasing demand for freshwater because of the rice culture, industrialization, and urbanization is a serious problem. Difficulties of supply lie in the paucity of natural water reservoirs, the swift runoff of the rivers, and the engineering difficulties of constructing large-scale dams in the rugged mountains.

Major rivers of Japan

Japan's rivers are generally short and swift-running, and are supplied by small drainage basins. The most significant rivers are the Teshio and Ishikari rivers of Hokkaidō; the Kitakami, Tone, Shinano, Kiso, and Tenryii rivers of Honshii; and the Chikugo, the largest river of Kyūshū. Some of the rivers from the volcanic areas of northeastern Honshii are acidic and are useless for irrigation and other purposes.

Biwa-ko (Lake Biwa), the largest in Japan, covers 268 square miles of southern Honshii. All other lakes of considerable size are in the northeastern half of the country. Most of the coastal lakes, such as Kasumi-ga and Hamana on Honshii's east coast, are former drowned valleys the bay mouths of which have been dammed by sandbars. Inland lakes such as Biwa, Suwa, and Inawashiro of Honshii occupy tectonic depressions of recent fault origin. Lakes of volcanic origin, including Kutcharo of Hokkaidd and Towada and Ashino of Honshii, outnumber all other types.

The soils of Japan are customarily divided from northeast to southwest into a weak podzolic (soils with a thin organic mineral layer over a gray leached layer) zone, a brown-earth zone, and a red-earth zone. There are some local variations. The northern half of the Tbhoku area of northern Honshii is included in the brown forest soil area. The northern tip of Hokkaidb is classed as a subzone of the podzolic soils; the remainder of the island is included in the subzone of the acidic brown forest soils; and most of western Honshii is a transitional zone. Yellow-brown forest soils extend along the Pacific coast from southern Tdhoku to southern Kyiishii, while red and yellow soils are confined to the Seinan Islands.

Kuroboku soils (black soils rich in humus content) are found on terraces, hills, and gentle slopes throughout Japan, while gley (sticky, blue-gray compact) soils are found in the poorly drained lowlands. Peat soils occupy the moors in Hokkaidd and Tdhoku. Muck (dark soil, containing a high percentage of organic matter) and gley paddy soils are the products of years of rice culture. Polder soils (those reclaimed from the sea) are widely distributed, and immature volcanic ash soils are found on the uplands. The widespread reddish soils are generally regarded as the products of a former warm, humid climate.

Climate. Japan's present climate is influenced by the country's latitudinal extent, the surrounding oceans, and the neighbouring Asian landmass, whereas local climatic variations are the result of relief features. In winter, the high pressure zone over Eastern Siberia and the low pressure zone over the western Pacific result in an eastward flow of cold air (the winter monsoon) from late September to late March that picks up moisture over the Sea of Japan. The monsoon deposits its moisture as rain or snow on western Japan and brings dry, windy weather to the east. The pressure systems are reversed during the summer, and air movements from the east and south (the summer monsoon) from mid-April to early September bring warmer temperatures and rain. Cyclonic storms and frequent and destructive typhoons occur during the late summer and early fall, especially in the southwest.

The effects of ocean currents

The warm waters of the Kuroshio (Japan) Current, which corresponds to the Gulf Stream of the Atlantic, flow along Japan's eastern coast as far as latitude 35° N. The Tsushima Current branches from the Kuroshio Current off southern Kyiishii and washes off the west coast of Honshii and Hokkaidd; it is this current that lends moisture to the winter monsoon. The counterpart of the Labrador Current, the cold Oyashio (Chishima) Current, flows southeastward from the Bering Sea along the east coast of Hokkaidd and northeastern Honshū. Its waters meet those of the Kuroshio Current, causing dense sea fogs in summer, especially off Hokkaidb.

The chief physical feature to affect climate is the mountainous backbone of the islands. The ranges interrupt air flow from the west and east and cause the gloomy weather and heavy snows of winter along the Sea of Japan coast and the bright and windy winter weather along the Pacific. Temperatures and annual precipitation are about the same on both coasts, but they drop noticeably in the mountainous interior.

Temperatures are generally warmer in the south than in the north, and the transitional seasons of spring and fall are shorter in the north. At Asahikawa, on Hokkaidō, the minimum temperature in January, the coldest month, is 16° F (—9° C) and the maximum temperature in August, the hottest month, is 70" F (21° C), with an annual average temperature of 43" F (6° C). At Tokyo, the minimum temperature is 39° F (4° C), the maximum 81° F (27" C), and the annual average 59° F (15" C). Inland from Tokyo, Matsumoto is cooler, with an annual average temperature of 52° F (11° C); whereas the annual average of 57" F (14" C) occurs on the Sea of Japan at Kanazawa. The warmest temperatures occur on Kyūshū and the southern islands; at Kagoshima, the minimum annual temperature is 45° F (7° C), the maximum 81" F (27" C), and the average 63" F (17° C).

Precipitation in the form of rain and snow is plentiful throughout the islands. Maximum precipitation falls in the early summer, and the minimum occurs in winter except on the Sea of Japan coast, which receives the country's highest snowfall. The summer rainy season occurs through June and July; it is known as the ***baiu*** ("plum rain") because it begins when the plums ripen. Torrential rains accompany the typhoons.

Rainfall patterns vary with topography, but most of the country receives more than 40 inches of precipitation annually. The smallest amount of precipitation occurs on Hokkaidb, where only 37 inches fall annually at Obihiro, whereas the mountainous interior of the Kü-hantō (Kü Peninsula) of central Honshū receives more than 160 inches annually. Varying amounts of snow fall on Japan. From November to April snow blankets Hokkaidb, northern and interior Honshii, and the west coast.

Vegetation and animal lie. Much of the original vegetation has been replaced by agriculture or by the introduction of foreign species to the islands. Semitropical rain forest prevails in the Nansei and Ogasawara islands and contains various kinds of mulberries, camphor, oaks, and ferns; madder and lianas are found as undergrowth. In the Amami-ō-shima (Amami-o Islands) this type of plant life occurs only on lowlands, but it grows at higher altitudes southward.

Vegetational regions

The laurel forest zone of evergreen, broad-leaved trees extends from the southwestern islands northward to the lowlands of northern Honshii. Camphor, pasania, oak, tea, and holly are typical trees, and various kinds of ferns grow as undergrowth. In the Ryukyu Islands this type of vegetation also occurs as far north as Amami-6-shima, where it is mixed with the semitropical forest and grows to the top of the islands. In Kyūshū, the evergreen zone reaches to over 3,300 feet, but its altitudinal limit decreases northeastward across Honshii. In general, camphor dominates in the littoral lowlands; pasania, in sunny and well-drained sites; and oak, in the foggy and cloudy inlands. In the southwestern Hondo region (Honshii, Shikoku, and Kyiishii) are ficus, fanpalm, and treefern, as well as xerophytic plants (plants adapted to arid conditions). Stands of Cycas are found in the Gotbretto (Gotō Islands) off western Kyūshū, as well as in southern Kyiishii and the Ryukyu Islands. The coastal dunes are dominated by pine trees.

Deciduous broad-leaved forests develop in the higher and northward portions of the laurel forest zone. In Kyūshii, this type of forest occurs above 3,300 feet, but it gradually descends northward to meet the shoreline in northern Honshii. Its upper limit reaches 6,000 feet in Shikoku and 5,000 feet in central Honshii. The representative trees are beech, katsuratree, maple, oak, and birch; while various kinds of bamboo grasses grow as undergrowth. The trees have been occasionally replaced by larch, false cypress, false arborvitae, Japanese cedar, Japanese red pine, Japanese black pine, and other coniferous species. Natural stands of cedar occur above 2,300 feet on Yaku-shima. The deciduous zone extends into western Hokkaidb, where beeches terminate at the southwestern peninsula and, further northeastward, is replaced by basswood and maple. Some stands of conifers are mixed with the representative forests of this zone.

Coniferous trees are found in the north and eastern periphery of Hokkaidd up to 2,300 feet above sea level. Sakhalin spruce, Sakhalin fir, blue fir, and Yezo spruce are mixed with such deciduous trees as birch, oak, and maple and dense undergrowth of mosses and lichens. Coniferous trees are mixed with deciduous vegetation in southwestern Hokkaidb and occur in the higher portion of central Honshū and Shikoku.

High-altitude small shrubs, the creeping pine, and alpine plants grow in the high mountain knots of central Honshii above 8,000 feet. This zone gradually descends northward to Hakkōda-san, in northern Honshii, at 4,600 feet and to Daisetsu-zan, in central Hokkaidd, at about 3,600 feet.

The cherry tree, celebrated for its blossoms, is planted all about the country. The first day of the flowering of the cherry trees in Yoshino is regarded as the advent of spring.

Animal life

Except for bats, the land mammals of Japan are of species distinct from those of the neighbouring Asian continent. Bears, wild boars, badger, foxes, deer, and monkeys (including the Japanese macaque) were once abundant; but their numbers have been considerably reduced in the 20th century. Antelope, hares, weasels, and ptarmigan are found in the higher mountains.

Japanese waters are inhabited by whales, dolphins, porpoises, and fish such as salmon, sardines, sea bream, mackerel, tuna, trout, herring, grey mullets, smelts, and cod. Crustaceans include crabs, shrimp, prawns, and oysters and are important as food; mussels and oysters are raised commercially. The rivers and lakes abound in trout, freshwater crabs, and crayfish.

Reptiles include sea turtles, freshwater tortoises, sea snakes, and lizards. Most of the snakes, including the five-foot long Japanese rat snake, are harmless. Toads, frogs, and newts are common; and a giant salamander of Kyiishii and Honshii attains a length of five feet.

More than one-third of Japan's birds are water birds—shore birds, gulls, and auks (black and white diving seabirds). Other water birds include grebes, albatrosses, shearwaters, herons, storks, ibis, ducks, geese, swans, and cranes. The cormorant is sometimes trained to catch fish. There are about 150 species of songbirds, as well as vultures, hawks, falcons, pheasant, quail, owls, and woodpeckers.

THE LANDSCAPE UNDER HUMAN SETTLEMENT

Traditional regions. The concept of regions in Japan is inseparable from the historic development of administrative units. Instead of dividing the country into geographical units based on one physical feature, such as a drainage system or a lake region, care was taken to include various physical features in the larger administrative units so as to create a well-balanced geographic whole. Many of the ancient terms for administrative units have survived in the form of place names, and some are still relevant in the 20th century.

The origin of Japan's regions

The Taika reforms of AD 646 firmly established the concept of regions in Japan. The ri (roughly corresponding to the later village community) was established as the basic social and economic unit, and the more than 700 *gun* were the smallest political units to be governed by the central government. The *gun* were grouped to form 60 kuni, the largest political units, which were ruled by 'governors appointed by the central government. Each kuni was composed of maritime plains, interior basins, and mountains to constitute a more or less independent geographic entity. Several adjacent kuni that were linked by a trunk road or a convenient sea route were grouped into a *dō*; the term signified both the route and the region. The core region of the country was called the Kinai, or the land adjacent to the shifting Imperial capitals, which included Naniwa and Fujiwara.

During the Nara (710–784) and Heian (794–1185) periods, the region of Honsha to the east of the three great mountain barriers of Arachi, Fuwa, and Suzuka north, east, and southeast of Biwa-ko was called Kantō (*kan*, "barrier"; tii, east); and that to the west Kansai (kan, "barrier"; sai, "west"). As the empire's frontier shifted to the northeast, Kantb came to indicate the region to the east of Hakone barrier, and Kansai gradually came to include limited areas near the capital of Kybto as far as Osaka and present Kbbe. Northern areas that had not come under direct control of the central government were called Yezochi, "the land of non-subjugated people."

A third regional system was applied after the 10th century, in which kuni were amalgamated according to their distance from Kybto. The larger units were kingoku, or proximate kuni; Chiigoku, or intermediate kuni; and *en*goku, or remote *kuni.* Mutsu and Dewa in northeastern Honshū and islands such as Sado, Oki, Tsushima, and Iki were termed *henyō*, or peripheral, lands. The name of the present Chiigoku district is taken from this regional system.

After the Meiji Restoration in 1868 Hokkaidb was divided into eleven kuni, and Mutsu and Dewa were broken into seven kuni after the traditional manner of the 7th century. These divisions were only in recognition of geographic entities, and administrative power was not endowed to them.

In 1871 the feudal system was dissolved and the ken, or prefectural, system was established. At first the newly born prefectures were mostly fiefs for the formal feudal lords who were appointed as governors. There were three *fu* (urban prefectures) of Tokyo, Osaka, and Kybtb and more than 300 ken. During the next 20 years there were frequent changes in the *ken* pattern, both through amalgamation and partition. By 1888 there were 43 ken, three fu, and one *dō* (major region) of Hokkaidb. Okinawa was officially given the status of ken in 1879, and in 1893 the Santama area was transferred from Kanagawa-ken to Tokyo to insure a better water supply for the city. In 1943, Tokyo was given the status of to, or capital city, thus completing the present to-dd-fu-ken system.

Modern regional units

By the 20th century there was a need for larger regional divisions. By 1905 the Ministry of Education had established a system of eight chihii (major regions). The pattern is locational and is so arranged as to divide Japan from northeast to southwest. The chihii are Hokkaidb; Tbhoku, including Aomori, Iwate, Akita, Yamagata, Miyagi, and Fukushima prefectures of northern Honshii; Kantb, including Ibaraki, Tochigi, Gumma, Saitama, Chiba, and Kanagawa prefectures and Tokyo-to of the Kwanto Plain in eastern Honshii; Chiibu, including Niigata, Toyama, Ishikawa, Fukui, Yamanashi, Nagano, Gifu, Shizuoka, and Aichi prefectures of central Honshii; Kinki, including Shiga, Kyoto-fu, Osaka-fu, Nara, Hyōgo, Wakayama, and Mie prefectures of west central Honshii; Chiigoku, including Tottori, Shimane, Okayama, Hiroshima, and Yamaguchi prefectures of western Honshii; Shikoku, including Tokushima, Kagawa, Ehime, and Kbchi prefectures of Shikoku Island; and Kyūshū, including Fukuoka, Saga, Nagasaki, Kumamoto, Oita, Miyazaki, and Kagoshima prefectures of Kyiishii Island as well as the prefecture of Okinawa of the Ryukyu Islands, which, upon its return to Japan in 1972, began to appear in official publications as a ninth chihd.

Another system used by some governmental agencies is a modification of the chihd system. Chūbu-chihō is subdivided into Hokuriku (Niigata, Toyama, Ishikawa, and Fukui prefectures), Tōsan (Yamanashi, Nagano, and Gifu prefectures), and Tōkai (Shizuoka, Aichi, and Mie prefectures). This system is devised so as to group prefectures of similar geographic characters into one *chihō* and consequently is more effective for illustrating regional contrasts and comparing statistics.

Since 1950 there has been a tendency to adopt the Seto-uchi area (surrounding the Seto-naikai; Inland Sea) as a distinct entity because of its regional homogeneity and internal coherence. By this system, Chiigoku and Shikoku regions are amalgamated and then subdivided into San-in (Tottori and Shimane prefectures), Seto-uchi (Okayama, Hiroshima, Yamaguchi, Kagawa, and Ehime prefectures), and Nankai (Tokushima and Kōchi prefectures).

Since 1920 Japan's pattern of industrialization has been reflected in the terms *omote nihon* ("obverse Japan"), *ura nihon* ("reverse Japan"), and the Taihei-yb (Pacific

Ocean) belt. *Omote nihon* refers to the progressive and growing status of the Pacific coast, and *ura nihon* connotes the stagnant economic conditions of the Sea of Japan coast. The string of industrialized and urbanized areas along the seaboard lowlands of the Pacific between Kantb-chihb and northern Kyūshū comprise the Taihei-yb belt, which includes all the Japanese cities of more than 1,000,000 population except Sapporo in Hokkaidb.

Rural and urban settlement patterns. Since the late 19th century, economic and social changes have affected even the most remote rural villages, but many things Japanese have survived. In the villages, many features that are in common with those of other Asian villages are well preserved. Autonomous and cooperative systems of agricultural practices and rituals, as well as mutual assistance among the villagers, have been handed down to the present. An autonomous rural unit, generally known as a *mura*, consists of some 30 to 50 or more households. Now called an *aza*, this unit should not be confused with the administrative terms *mura* or *son* in use after 1888.

Rural settlement patterns

Most of the rural settlements are of age-old origin, and their histories are lost in time. Historically traceable settlements owe their origins largely to land reclamation after the 16th century. They are commonly called *shinden*, or "newly reclaimed field," but in terms of social structure they do not radically differ from the older settlements.

Considerable local difference is evident in the settlement pattern. Some villages are agglomerated, as are those of Kinki-chihb; some are dispersed, as in northeastern Shikoku; some are elongated, such as those on the rows of sand dunes in the Niigata-heiya (Niigata Lowlands) and on the natural levees of deltas; while others are scattered on the steeper mountain slopes. These differences are only superficial; without exception, the inhabitants are bound together to form a firm village community.

No village is regarded as purely rural. Those that are near to industrialized urban centres include commuters and industrial workers. The more remote settlements send out seasonal labourers during the winter months. The villages of Hokkaidb are based on commercial agriculture, and each household has direct contact with a nearby town.

Fishing villages were absent in Tbhoku-chihb until the beginning of the 17th century, when northward movement began. They were originally dependent upon nearby rice-producing villages; although some dried, salted, or smoked fish found more distant markets. The fishing villages are most numerous in the southwest, where an exchange economy has long been in practice. Mountain villages that depend solely on mountain products other than rice are exceedingly rare. Many of them were founded after the 17th century, when lumber, charcoal, and other such products found markets in the growing towns on the plains. There are some purely hunting villages *(matagi)* in the mountainous interior of western Tbhoku.

Japan's urban centres

Urban settlement is generally of recent origin. Except for the former capital cities of Nara, Kybto, and Kamakura, no sizable town of any significance appeared before the 16th century. Most of the provincial capitals, or *kokufu*, of ancient Japan were only administrative centres that contained official residences and were not developed towns. After the latter part of the 16th century, influential temples and feudal lords began to build towns by gathering merchants and craftsmen close to their headquarters. The power of the feudal lords stabilized when they built *joka-machi* (castle towns), which were located so as to command and control the main transporation routes and surrounding areas; the majority of Japan's important cities, including Tokyo, have developed from them.

Next in importance were the port towns, such as Hakata and Sakai, which have experienced more vicissitudes than the castle towns. In addition, some of the religious towns have grown to a considerable size, as in the case of Ise, Kompira, and Zenkōji. Under the Tokugawa shogunate regime (1603–1867), peaceful conditions fostered nationwide pilgrimages on a scale unknown in the preceding period, and temple and shrine towns such as Kyōto and Nara flourished.

Urbanization actually began in the late 19th century with the growth of the international ports of Kbbe, Yokohama, Niigata, Hakodate, and Shimoda and the naval bases of Yokosuka, Kure, and Sasebo. Industrialization ushered in the rapid growth of Japanese cities, and some of the industrial towns (*e.g.*, Yawata, Niihama, Kawasaki, and Amagasaki) were founded in response to economic development. Most of the former castle towns, and especially those along the Taihei-yb belt, have been expanded directly or indirectly by industrialization. In Hokkaidb and in southern Kyiishii raw materials and power resources have attracted a limited number of industrial plants, which alone are responsible for the existence of cities such as Tomakomai, Muroran, Nobeoka, and Minamata.

Japanese cities are bewildering mixtures of old and new, East and West. Oriental congestion exists side by side with the most modernized business centres and industrial establishments; and unsatisfactory sewage systems are a formidable menace in ever-expanding cities of skyscrapers, subways, and underground plazas. Other serious problems are the shortage of housing, land subsidence (sinking) on the seaboard lowlands, and the anticipated menace of earthquakes and floods.

II. People and population

GROUPS HISTORICALLY ASSOCIATED WITH THE CONTEMPORARY COUNTRY

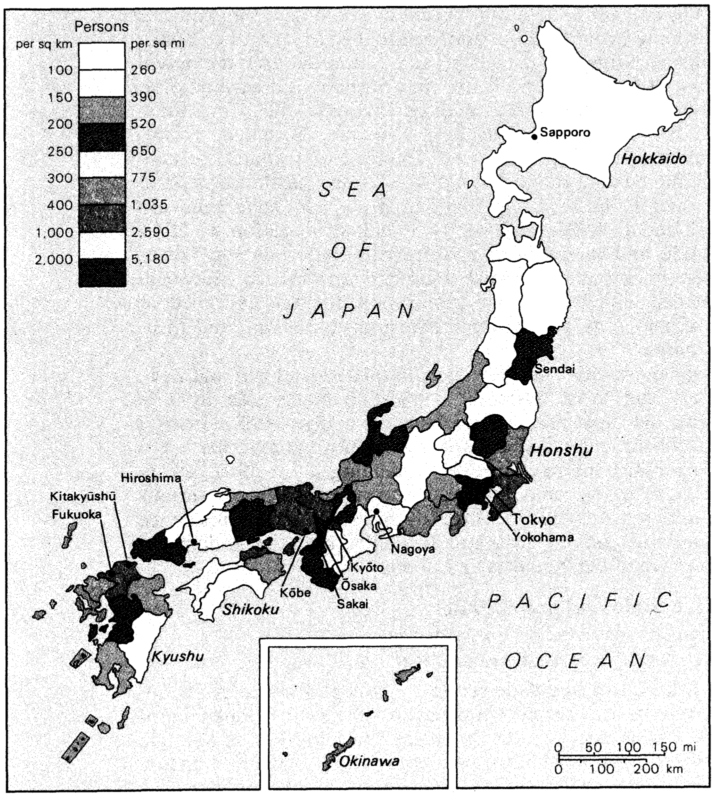
Ethnic and linguistic groups. Except for the Ainu, indigenous Caucasoid people who survive in limited numbers in Hokkaidb, the Japanese are regarded as a single ethnic group. The foreign community, which numbers less than 1,000,000, is comprised largely of Koreans and Chinese. Physically, the Japanese are classified as a branch of the Mongoloid race and are closely akin to the peoples of eastern Asia. Their general physical characteristics are black, straight hair; medium-coloured skin; dark-brown to black-brown eyes; scanty body hair; and height of about five feet two inches for adult males. In terms of blood types, type B is dominant all over Japan; regional differences include 0-type in northeastern Honshii and A-type in Chiigoku, Shikoku, and Kyiishii.

Japanese is the national language, and Ainu is almost extinct. The Japanese language is generally included in the Ural-Altaic linguistic group and is especially akin to Korean, although the vocabularies differ. The introduction of Chinese characters and literature in the 4th century AD greatly enriched the Japanese vocabulary. At first Chinese characters themselves were used to write Japanese, but in the 9th century a syllabary, known as *kana*, was developed from them. Since then, the *kana* has been used for written Japanese, together with some 3,000 to 5,000 Chinese characters. After World War II, the number of Chinese characters was reduced to less than 2,000, and spelling was standardized with pronunciation. Romanization was also introduced at the elementary school level.

The distribution of Japanese practically coincides with the territory of Japan; it is also spoken as a second language in Taiwan and Korea, and in Japanese communities outside of Japan. It is used by more than 100,000,000 people. Standard Japanese was established in the late 19th century through national education and increased communication. Although there are many local dialects, standard Japanese is understood throughout the country.

The dialects of the Japanese language

Japanese is broadly divided linguistically into the two major dialects of Hondo and Nantb. The Hondo dialect is used throughout Japan and may be divided into three major subdialects of the east, west, and Kyiishii. The Eastern, or *bandō*, subdialects were established in the 7th and 8th centuries as the *azuma* ("Eastern") language. After the 17th century there was a vigorous influx of the Kamigata (Kinai) dialect, which was the foundation of standard Japanese. The Eastern dialect extends westward to eastern Hokuriku, Tbsan, and Tbkai, and its boundary line with the Western dialects can be drawn approximately southward from Itoigawa on the Sea of Japan to Hamana-ko (Hamana Lake) on the Pacific coast.



Population density of Japan.

Among the Western dialects, Kinki dialect was long the standard language of Japan, although the present Kamigata dialect of Kydto–Osaka is of recent origin. The Izumo dialect along the Sea of Japan is presumed to be the original Japanese language.

The Kyiishii dialects have been placed outside the mainstream of linguistic change of the Western dialects, and retain the 16th-century forms of the latter. They extend as far south as Tane and Yaku islands.

The Nantb, or Okinawa, dialects are used by islanders from Amami-ō-Shima (Amami-ō Island) in Kagoshima Prefecture as far west as Yonaguni-jima (Yonaguni Island) by a total of more than 1,000,000 people. Long placed outside of the mainstream of linguistic change, they strongly retain their ancient forms.

Religious groups. In Japan, Shintbism—an indigenous folk religion—various sects of Buddhism, and Christianity exist together with new religions (Shinkd Shiikyb); no one of the religions is dominant, and each is affected by the others. For example, one person or family may believe in several Shintb gods and at the same time belong to a Buddhist sect. Intense religious feelings are generally lacking except among the adherents of some of the new religions.

Shintbism and Buddhism

Shintbism is a polytheistic religion in which men as well as natural objects are enshrined as gods. Some of the Hindu gods and Chinese spirits have also been introduced and Japanized. Each rural settlement has at least one shrine of its own, and distinguished men of each historical period are enshrined. Until the Meiji Restoration of the 19th century attempted to impose Shintdism on all of Japan, the religion coexisted comfortably with Buddhism. The amalgamation of these two religions gave birth to a special religion known as Shugen-dd.

Buddhism was officially introduced into the Imperial Court from Korea in AD 552. Direct contact with central China was maintained, and many sects were introduced. In the 8th century Buddhism was adopted as the national religion, and national and provincial temples, nunneries, and monasteries were built throughout the country. In the early 9th century, the Tendai and Shingon sects were introduced by Japanese monks; they still exert some influence in remote areas of Japan. Most of the Buddhist sects that have retained a strong influence are those that were modified in the 13th century by Japanese monks such as Shinran, who established the Jddo Shinshii sect, and Nichiren, who founded the Nichiren sect.

In the 1970s, Buddhism was classified into 12 major sects and almost 60 branches. Chief among them are the Jōdo Shinshii, Sōtō, Shingon, Jbdo, Rinzai, Nichiren, and Tendai sects.

Christianity was introduced into Japan in the 16th century by Spanish Roman Catholic missionaries and was well received both as a religion and as a symbol of European culture. Roman Catholicism was banned by the Tokugawa regime in 1637. Inaccessible and isolated islands and the peninsula of western Kyūshū contained "hiding Christian" villages until the ban was lifted by the Meiji government in 1873. Russian Orthodox, Roman Catholic, and Protestant missionaries came to Japan thereafter, and by the 1970s there were over 800,000 Christians.

The new religions were founded after the mid-1800s. Most of them have their roots in Shintōism, but their religious creeds have been influenced by Buddhism and Christianity. The Sbka-gakkai ("Value Creating Society") is a military nationalistic religion that claimed 10,000,000 adherents in 1962 and has elected members to the legislature. Other cults are the Tenri (Celestial Laws), Konko (Heavenly Light), and the House of Growth.

DEMOGRAPHY

Nothing is more misleading than the widespread belief of an overpopulated Japan. Although more than 105,000,000 people live in a land of 143,818 square miles with an average population density of about 744 persons per square mile, the increase of population since the 19th century has kept pace with economic development, and the standard of living has steadily risen. Japan's density is exceeded by only a few countries, such as South Korea, The Netherlands, and Belgium, despite the fact that more than 80 percent of the country is occupied by inhospitable mountainous areas. The congested concentration of population within the limited plains and lowlands is its distinguishing characteristic. The increased population of the 20th century has been absorbed into the ever-expanding urban areas, while the rural districts have remained constant or have declined. In the rural lowland areas, the density is generally less than 1,000 per square mile, lower than the densest portions of the Mekong Delta, some parts of Java, and the Ganges Delta.

Population growth trends

Japan has experienced spectacular growth since 1868, when the population numbered about 34,000,000. This increase is directly related to slow but steady urban growth; the development of Hokkaidō, Tbhoku, and southern Kyūshū; and the introduction of commercial agriculture. In 1897, when induatralization first began, the population numbered more than 42,000,000. From 1898 to 1918 growing industrial cities and mining towns absorbed a large population, as did Hokkaidd and the sericultural (silkworm raising) rural districts.

In 1920, when the first precise census was conducted, the population numbered 56,000,000. Between 1919 and 1945 Tokyo, Osaka, Nagoya, and northern Kyūshū developed as the nation's four major industrial districts. At the same time, some of the smaller cities that had failed to invite industrialization lost their ability to absorb a growing population, and some of them declined. By 1940 the population had grown to almost 72,000,000, or more than double that of 1868. During World War II, there was a'marked migration to the rural areas to avoid aerial bombing, and some cities such as Osaka were reduced to one-third their size. After 1945, the repatriated population of nearly 9,000,000 and the temporarily explosive increase in the birth rate caused abnormal growth.

The rapid rehabilitation of industry after 1950 has resulted in the continuing concentration of population in the Pacific coastal areas, which contained about 40 percent of the population in the early 1970s. The expansion of the Keihin area is not confined to Tokyo, Yokahama, and their adjacent suburbs, but extends to a much wider circle. The same is true of the Hanshin (Osaka-Kbbe) and Tyūkyō (Nagoya) areas, although northern Kyūshū remains somewhat stagnant. Rural areas outside of the direct in-

Table 1: Japan, Area and Population

	area*		population†‡	
	sq mi	sq km	1960 census	1970 census
Regions (*chihō*)				
Chiibu				
Prefectures ***(ken)***				
Aichi	1,963	5,085	4,206,000	5,386,000
Fukui	1,617	4,188	753,000	744,000
Gifu	4,015	10,399	1,638,000	1,759,000
Ishikawa	1,620	4,195	973,000	1,002,000
Nagano	5,071	13,133	1,982,000	1,957,000
Niigata	4,676	12,110	2,442,000	2,361,000
Shizuoka	2,827	7,322	2,756,000	3,090,000
Toyama	1,551	4,017	1,033,000	1,030,000
Yamanashi	1,642	4,254	782,000	762,000
Chiigoku				
Prefectures				
Hiroshima	3,262	8,448	2,184,000	2,436,000
Okayama	2,733	7,078	1,670,000	1,707,000
Shimane	2,558	6,626	889,000	774,000
Tottori	1,348	3,492	599,000	569,000
Yamaguchi	2,350	6,087	1,602,000	1,511,000
Hokkaidd				
Territory				
Hokkaidō	30,314	78,513	5,039,000	5,184,000
Kantd				
Metropolis *(to)*				
Tokyo	827	2,141	9,684,000	11,408,000
Prefectures				
Chiba	1,961	5,080	2,306,000	3,367,000
Gumma	2,454	6,356	1.578,000	1,659,000
Ibaraki	2,350	6,087	2,047,000	2,144,000
Kanagawa	922	2,387	3,443,000	5,472,000
Saitama	1,467	3,799	2,431,000	3,866,000
Tochigi	2,476	6,414	1,514,000	1,580,000
Kinki				
Prefectures				
Hydgo	3,226	8,355	3,906,000	4,668,000
Mie	2,229	5,774	1,485,000	1,543,000
Nara	1,425	3,692	781,000	930,000
Shiga	1,551	4,016	843,000	890,000
Wakayama	1,822	4,720	1,002,000	1,043,000
Urban prefectures (*fu*)				
Kyōto	1,781	4,612	1,993,000	2,250,000
Ōsaka	716	1,854	5,505,000	7,620,000
Kyūshū				
Prefectures				
Fukuoka	1,900	4,922	4,007,000	4,027,000
Kagoshima	3,532	9,149	1,963,000	1,729,000
Kumamoto	2,680	6,940	1,856,000	1,700,000
Miyazaki	2,779	7,197	1,135,000	1,051,000
Nagasaki	1,582	4,098	1,760,000	1,570,000
Oita	2,337	6,053	1,240,000	1,156,000
Saga	931	2,412	943,000	838,000
Ryukyu				
Prefecture				
Okinawa	922	2,389	883,000	945,000
Shikoku				
Prefectures				
Ehime	2,185	5,569	1,501,000	1,418,000
Kagawa	722	1,870	919,000	908,000
Kdchi	2,744	7,106	855,000	787,000
Tokushima	1,600	4,144	847,000	791,000
Tohoku				
Prefectures				
Akita	4,414	11,431	1,336,000	1,241,000
Aomori	3,474	8,998	1,427,000	1,428,000
Fukushima	5,321	13,781	2,051,000	1,946,000
Iwate	5,827	15,092	1,449,000	1,371,000
Miyagi	2,814	7,288	1,743,000	1,819,000
Yamagata	3,600	9,325	1,321,000	1,226,000
Total Japan	143,818	372,488	94,302,000	104,665,000

*Area figures for prefectures do not add to country total because disputed areas and some inland water areas are excluded from prefecture figures. †De jure. ‡Population figures do not reflect boundary changes subsequent to the 1970 census.
Source: Official government figures.

fluence of urbanization are subjected to a marked decline. Adult males migrate to the Pacific coast, and many of those who remain at home periodically leave as temporary labourers, creating a constant outflow of population from the mountainous areas and isolated islands. In many places, emigration is so marked that the remaining population cannot maintain a balanced communal society, and whole settlements are abandoned.

A striking demographic feature that has developed since World War II is the decline of birth and death rates in response to improved health conditions and birth control measures. In the 1920s, the birth rate was about 34 per thousand and the death rate was about 20 per thousand. The rates dropped rapidly after 1950; and in 1970, the birth rate was 18.7 per thousand. The death rate declined after 1947 to less than seven per thousand in 1970—one of the world's lowest. Life expectancy is 70.2 years for males and 75.6 years for females. Between 1970 and 1971 population grew almost 1.2 percent. (Ak.W.)

Table 2: Major Japanese Cities

	population (1970)
Tokyo, Tokyo	8,841,000
Ōsaka, Ōsaka	2,980,000
Yokohama, Kanagawa	2,238,000
Nagoya, Aichi	2,036,000
Kydto, Kydto	1,419,000
Kdbe, Hyōgo	1,289,000
Kitakyiishii, Fukuoka	1,042,000
Sapporo, Hokkaidō	1,010,000
Kawasaki, Kauagawa	973,000
Fukuoka, Fukuoka	853,000
Sakai, Ōsaka	594,000
Amagasaki, Hydgo	554,000
Sendai, Miyagi	545,000
Hiroshima, Hiroshima	542,000
Higashi-Osaka, Ōsaka	500,000

III. The national economy

Japan's
regional
and world
standing

Japan is remarkable for its extraordinarily rapid rate of economic growth in recent years. During the 1960s, the Japanese economy recorded an average annual growth of nearly 12 percent, which is greater than that of any other industrial country in the world. This spectacular progress was based on a rapid expansion of industrial production, the level of which was about 300 percent higher in 1971 than in 1962. As a result, Japan is the third largest economic power, ranking after the United States and the Soviet Union. It is the world's second or third largest producer of a large number of manufactured goods and the world's principal shipbuilder. Its exports have seen a fourfold increase during the 1960s: they now account for almost 7 percent of the world's total exports. By 1970, exports of manufactured products were more than 4.5 times higher than in 1960, accounting for about 8 percent of the total world exports of manufactures.

Despite its rapid economic progress, however, the country's standard of living is still relatively low when compared with other developed countries. In 1970, the per capita gross national product of U.S. $1,583 was about 15th highest in the world, well below that of other industrialized countries. The standard of living is expected to rise steadily, however; and if subsequent growth continues at its recent rate, Japan could become the world's most prosperous country by 1988.

THE EXTENT AND DISTRIBUTION OF RESOURCES

Mineral resources. Although Japan's mineral deposits are fairly diverse, with a few exceptions the reserves are small and production is inadequate to meet more than a small part of domestic requirements. The quality of the minerals mined is often poor, and since deposits are widely scattered the extractive industry is characterized by a large number of small and relatively inefficient mines that do not lend themselves to the application of modern, large-scale mining methods. Coal, copper, zinc, lead, and silver are among the most important minerals, and although a large number of others are mined on a minor scale, there is an almost complete lack of nickel, cobalt, bauxite, nitrates, rock salt, potash, phosphates, and oil.

Coal is the country's most important mineral. In 1970, output amounted to 39,700,000 tons and satisfied just under 45 percent of Japan's total requirements. Japanese coal is of relatively poor quality and difficult to mine. Production costs are high, and more than 70 percent of the country's requirements of coking coal are imported. Production is concentrated in Kyūshū and Hokkaidō, which account for more than 40 percent of total minable reserves. In 1969, these two fields were responsible for 90 percent of the total output. Despite its close proximity to the sea, the Kyiishii field has lost much of its importance because of the gradual exhaustion of the richer seams and the poor quality of the coal mined. During the 1960s, Hokkaidō increased its output by 17 percent. Although this field is some distance from the sea, its seams are thicker than those of Kyiishii, and the use of mechanized, large-scale mining methods have resulted in a relatively high output per man. Minor coalfields include Jōban in Tōhoku, which produces mainly lignite, and Ube in Chūgoku, which is an extension of the Kyiishii field. Total exploitable reserves of coal are estimated at about 3,200,000,000 tons.

Oil deposits are meagre, and estimated resources account for less than 0.01 percent of the world's total. Unless new fields are discovered, the reserves are likely to be exhausted by the 21st century. The oil-bearing belt extends from northern Honshii on the Sea of Japan to the Ishikari-Yiifutsu lowlands in Hokkaidd; virtually the whole of the country's output comes from Honshii, especially from Akita and Niigata prefectures. Between 1960 and 1971 output grew by 48 percent, accounting for less than 1 percent of total imports. Natural gas is produced in Honshii; the two most important fields are in south Kantd and Niigata prefectures. Output grew during the 1960s and early 1970s, but it accounts for only about 15 percent of the country's total gas supply.

Although Japan ranks as the world's third largest steelmaker, after the United States and the Soviet Union, domestic resources and production of iron ore are small. Output of iron ore fell during the 1960s and in 1970 amounted to less than 1 percent of total imports, which come mainly from the United States, Canada, Malaysia, the Philippines, and India. Japanese iron ore is of poor quality and is obtained from small mines, most of which are in Hokkaidō and northern Honshii. The most important producing area is in Iwate Prefecture in Honshii, which accounted for over half of total production in the late 1960s. Total workable deposits are estimated at 34,000,000 tons. Other domestic sources of iron include sand ore, deposits of which are widely distributed along the coast and are estimated at 153,000,000 tons, and iron pyrite sinter, which is a by-product of sulfuric acid plants.

Copper is Japan's most important metallic ore. It is produced in small mines; the larger mines, situated in northeastern Honshū and Shikoku, account for a significant part of total output. In 1970, the country produced copper concentrate with a metal content of almost 120,000 tons. This was far from enough to meet domestic demand, however, and almost 80 percent of the output of electrolytic copper was based on imported ores. Lead and zinc are often found in conjunction with copper; one of the largest producers of zinc is Kamioka, in Gifu Prefecture. During the early 1970s, Japan produced about 280,000 tons of zinc ore (metal content), which provided enough raw material for about half of the country's zinc output. Lead output in 1969 totalled 63,000 tons. Other metallic ores mined include silver, gold, tungsten, tin, antimony, molybdenum, mercury, titanium, vanadium, and manganese. Japan also has large sulfur deposits.

Agricultural resources. Because of the country's mountainous terrain, the supply of agricultural land is limited. Since the early 1960s, the cultivated land area has decreased slightly and accounts for about 15 percent of the total. The soil, largely infertile and immature, requires careful husbandry and fertilization. Soil erosion is a serious problem in many parts of the country. Timber resources, however, are extensive. In the late 1960s, forest land covered some 64,000,000 acres, of which more than 25,000,000 acres consisted of coniferous forest and 33,000,000 acres of broad-leaved forest. A sizable proportion of the forests is inaccessible or is suitable only for fuel wood.

About 67 percent of the total forest area is in private hands, and much of it is distributed among a large number of relatively small holders. The rest is publicly owned; it is in these areas that large-scale reafforestation has taken place. The output of logs fluctuated slightly during the 1960s, and in 1970 accounted for only about 6 percent more than in 1960.

Hydroelectric resources. As a result of its climate, Japan has considerable water resources. It also has an extensive network of rivers that can be used for irrigation, although flooding is a serious problem in many parts of the country. As a result of the mountainous terrain, the ample hydroelectric potential is distributed in an uneven fashion. Hydroelectric development is largely concentrated in central Honshii along the Shimanto, Tenryu, Tone, and Kiso rivers, in Tbhoku, and some parts of Kyiishii. This pattern of distribution ensures that the country's hydroelectric capabilities are well located in relation to the important industrial areas, which tend to be deficient in coal. In 1970, a total of 80,089,000,000 kilowatt-hours of hydroelectricity were generated, accounting for roughly 22 percent of the total supply. Between 1958 and 1970, output increased relatively slowly and further development is unlikely to be rapid. Although it is estimated that only about 40 percent of the total potential has been developed, the best sites are already utilized, and further additions to capacity could be increasingly expensive. Most plants cannot operate at full capacity for more than a few months of the year because of seasonal variations of rainfall and the difficulty of constructing adequate storage facilities.

SOURCES OF NATIONAL INCOME

Agriculture **and** fishing. Agricultural production, including forestry and fishing, accounted for about 8 percent of the net national income in 1970. Output in this sector grew considerably less rapidly than national output as a whole during the 1960s, and its share in the total experienced a spectacular decrease. Despite a rapid increase in yields, agricultural output per man is considerably less than in other sectors of the economy; although farming contributes only 8 percent of the nation's total output, it employs more than 17 percent of the total working population. Japanese agriculture is characterized by a large number of small and often inefficient farms. In 1970, the farmland area of nearly 15,000,000 acres was shared by 5,300,000 farm households. In the early 1970s nearly 3,000,000 rural households owned less than two acres, and many farmers had to rely on outside occupations for a substantial part of their income. Larger farms are generally found in Hokkaidō, where units of 25 to 50 acres are not uncommon. Approximately 48 percent of the cultivated area is devoted to the country's principal crop, rice. Other important farm products include wheat, barley, potatoes, soybeans, and tea.

The government's agricultural policy

The main objectives of the government's agricultural policy have been to encourage self-sufficiency in the more important commodities, to enlarge the size of the average holding, and to close the gap between rural and urban incomes. The central feature of this policy is an artificially high producer rice price, which is roughly twice as high as the world market price. This has succeeded in raising farm incomes and has led to a rapid rise in the production of rice. The boost in output, however, coincided with a drop in per capita consumption, and the government is faced with a growing surplus of expensive and often poor-quality rice. In spite of strong opposition from the farming lobby, the situation forced a re-examination of official policy and the producer rice price was not increased in 1969. Other measures announced or under consideration in the early 1970s included the provisions of incentives to cut rice acreage and to expand the production of those commodities such as fruits, vegetables, and tobacco that are in strong demand and that can be produced on a competitive basis. It is also planned to press for a further increase in the size of the average operating unit; the average holding will be about four acres by 1977, which is still small by advanced agricultural standards, however.

Japan has for some time been the second largest fishing nation in the world after Peru. In 1970, its total catch of 9,300,000 tons amounted to 28 percent more than that of its nearest rival, the Soviet Union. In spite of its dominant international position, the Japanese fishing industry faces a number of serious problems, which are partly the result of structural weaknesses within the industry and partly a matter of disputes with other marine nations over fishing rights and limits. The result of these and other adverse factors is that between 1960 and 1970 Japan's catch grew by only 50 percent as against an advance of about 72 percent for the world as a whole. Despite the efforts of the government to create larger and more efficient units, small and medium-sized enterprises and individual fishermen still account for the bulk of the total output. In 1967 there were 15 large fishing enterprises that accounted for 18 percent of total output. The middle layer of the industry was made up of some 10,000 smaller units that were responsible for just over 40 percent; and at the bottom of the scale there were about 250,000 fishing households that supplied the remainder. Since the early 1960s exports of fishery products grew erratically, while imports showed a steady increase. Much of the fishing fleet consists of small boats.

Mining and quarrying. Mining is a slow-growing and unimportant branch of the economy. In 1970 output was only 1 percent higher than in 1960, with the result that its share in the net national product declined from almost 2 percent to less than 1 percent. Well over half of the value of mining production is accounted for by coal, followed by copper, limestone, oil and natural gas, lead and zinc, sulfur, silver, and gold. The greatest problems are faced by the coal industry, which has been suffering from the competition of cheaper foreign coal and from a rapid shift to oil since World War II. Despite the closure of a large number of uneconomic pits and a fairly rapid growth in productivity, this competition has had an unfavourable effect on the financial results of most coal mining companies and the government has repeatedly found it necessary to aid the industry. In 1969 the government adopted a new coal rehabilitation program that rejected suggestions of nationalization and proposed financial assistance and the granting of interest-free loans. Other aspects of the program, which was planned to close in 1973, include government aid to meet producers' deficits that result from the cutting of the price of domestic coal to a level that is competitive with imports, rationalization of the industry and the distribution network, and greater cooperation among producers. In contrast, the situation in the second largest branch of the industry, copper mining, is relatively healthy, and most of the country's copper mines tend to be fairly profitable.

The faltering extraction industry

Manufacturing and power production. One of the most spectacular features of Japan's economic development since World War II has been the rapid advance of manufacturing. Between 1950 and 1960, the index of manufacturing production recorded a growth of about 180 percent, and by 1971 the figure was nearly three times higher than that of 1960. Apart from this remarkable quantitative growth, the 1960s also saw tremendous progress in terms of quality, variety, and efficiency. As a result, Japan is now one of the top three industrial nations in the world and is a greatly feared competitor whose products are in increasing demand. It has been the world's principal shipbuilder since 1956 and is a major producer of crude steel, synthetic rubber, aluminum, sulfuric acid, plastics, cement, passenger cars, refined copper, and cotton yarn. It has some of the world's largest and most advanced industrial plants; in 1968, for example, eight of the world's 11 largest blast furnaces were to be found in Japan.

Although all important branches of manufacturing recorded some advance since the early 1950s, the most spectacular progress was seen in the production of iron and steel, machinery, petrochemicals, and advanced electronic products. The index of production in the iron and steel industry was about three times higher in 1971 than in 1960, with crude steel output reaching 88,557,000 tons. The advance in machinery production was even faster, because of the grqwing demand for industrial equipment and the boom in sales of automobiles, television sets, and a large number of other durable consumer goods. Growth in the chemical industry was also impressive; and in 1971, output was about three times as large as in 1960—an advance that was mainly caused by the rapid development of the petrochemical industry. Some of the older industries, however, advanced relatively slowly. The lumber and wood industry, textiles, and foodstuffs failed to match the expansion in manufacturing as a whole.

Reasons for industrial growth

One of the principal reasons for Japan's highly satisfactory industrial performance has been the high level and rapid growth of capital investment. Between 1961 and 1970 the country devoted about one-third of its national income to investment, well over half of which was accounted for by capital equipment. The boom in equipment investment provided the iron and steel and machine-building industries with a rapidly growing home market, allowed for a spectacular increase in productive capacity and in the scale of operations, and led to a rapid replacement of old machinery. This, in turn, resulted in a considerable improvement in productivity throughout the economy and enabled industry to grow, despite the increasingly acute shortage of skilled labour that developed during the 1960s. Despite a rapid rise in wages, many sectors of Japanese manufacturing have a formidable advantage over their rivals, a fact that is well illustrated by the country's booming exports and healthy overseas payments position. According to official figures, labour productivity, or output per man, in manufacturing nearly doubled between 1963 and 1969. This was a very much better performance than that seen in other major industrial countries, and it is now estimated that in 1968 Japan's labour productivity was 35 percent higher than that of the United States, the United Kingdom, West Germany, France, and Italy. Another important reason for Japan's strong industrial position in the world is the extensive employment of the latest technological innovations. Official estimates show that about 44 percent of the increase in labour productivity in manufacturing between 1955 and 1965 can be attributed to technological advance.

During the 1960s industry was also characterized by a growing tendency toward tie-ups, mergers, and take-overs among the larger manufacturing and industrial concerns. These actions were made possible by the gradual relaxation and the increasingly flexible interpretation of the country's antimonopoly laws. The authorities have accepted the argument that greater concentration at the top is essential in order to improve efficiency, to make better use of the existing resources, and to increase or maintain international competitiveness. This argument was given additional force by the need to strengthen Japanese enterprises in the face of growing direct foreign investment, made possible under the government's capital decontrol program. In the 1950s, Japan's monopoly watchdog, the Fair Trade Commission, approved about 400 merger applications a year; but by the middle of the 1960s the number of approvals was about 900 a year. The merger of the Yawata and Fuji Steel companies into the Nippon Steel Corporation, which was approved in 1969, will create the world's second largest crude-steel producer after the United States Steel Corporation.

There is also a large number of small and relatively inefficient manufacturing enterprises that use a substantial amount of the scarce labour. In 1968 there were nearly 603,000 manufacturing establishments, of which 522,000 employed less than 20 workers. These occupied about 27 percent of the manufacturing labour force but contributed no more than 12 percent of total output. The larger enterprises, which employed more than 100 but less than 1,000 workers, were responsible for about 40 percent of production and one-third of employment, while the more than 700 giants, each with a labour force of over 1,000, occupied 17 percent of all workers in manufacturing and provided 29 percent of total output. By and large, small firms tend to be more prominent in the less dynamic industries, which produce food, textiles, wood, leather, and clothing.

Between 1960 and 1969, Japan's consumption of energy more than doubled. Despite this rapid increase, however, per capita energy consumption remains low when compared to that of other industrialized countries. The largest single source of energy is oil, which in 1969 accounted for about 66 percent of the nation's total energy supplies.

Since Japan has practically no indigenous oil resources, almost the entire demand is satisfied through imports, an important share of which comes from fields developed by Japanese companies. Crude oil imports grew nearly five and one-half times between 1960 and 1970, although a significant proportion was re-exported in processed form.

Because coal is relatively difficult and expensive to mine, its importance has steadily declined. In 1960 it accounted for roughly 41 percent of gross energy supplies, but by 1970 its share had dropped to 21 percent. During most of the 1960s, coal production was fairly static. Imports, especially of coking coal, are expected to grow; but the share of coal in total energy supplies is likely to continue to shrink.

In 1970, Japan generated almost 360,000,000,000 kilowatt-hours of electric power, which made it the third largest producer in the world. Because of the difficulties of hydroelectric development, more than three-fourths of the total is accounted for by thermal electricity. Although about half of the country's total coal consumption is accounted for by power stations, the late 1960s witnessed a rapid increase in the importance of oil-fired electricity plants. In 1971 there were two commercial nuclear generators in operation with a combined capacity of around 1,300,000 kilowatts, and several others were under construction. It is expected that capacity will rise to about 7,000,000 kilowatts by 1975, a level that could more than double by 1980. By 1985 nuclear stations could supply 30 percent of the nation's electricity needs. Most of the reactors in existence or under construction have been built by the United States, but there are plans to develop a Japanese reactor by the 1980s.

Gas production more than tripled during the 1960s; in 1970 it amounted to 50,899,000,000,000 kilocalories. The fastest growth has been seen in the output of oil gas, which expanded its share of the total from a mere 10 percent to about 50 percent in the early 1970s. In contrast, the output of coal gas showed a timid advance and its importance has been reduced from about 90 percent to 38 percent. In 1970, natural gas production was more than 13 times its 1960 level but provided only 17 percent of the country's total gas supplies. Future increase in gas consumption is expected to be relatively rapid—about 7 percent a year—because of the rapid growth of the urban population and the continuous extension of the gas network. There are over 200 gas companies, three of which account for around 80 percent of total production.

Financial services. Japan's financial system is complex and different from that of other developed countries in a number of important aspects. The Bank of Japan, established in 1882, is the sole bank of issue; it also plays an important role in the determination and enforcement of the government's economic and financial policies. The bulk of the domestic banking business is transacted through the 80-odd commercial banks, of which the 14 city banks (such as Fuji, Mitsubishi, and Sanwa) are the most important. There are also a number of long-term credit banks, some government financial institutions, including the Japan Development Bank and the Export-Import Bank, and several mutual savings and loan banks and credit associations.

A unique banking system

One of the most interesting features of the Japanese financial system is the high degree of interdependence between the central bank, the commercial banks, and industry. Traditionally, industry has relied on banks for a large part of its growing fund requirements, and although the importance of its own capital has increased, private and government financial institutions still account for a substantial part of the total. Since the commercial banks are responsible for most credit extended to industry, their influence on their client companies is considerable. Their active lending policy also means that their liquidity ratios tend to be low by Western standards and that they are forced to rely on call money (money that is readily available to banks as loans) and on large-scale borrowing from the Bank of Japan. The central bank is thereby in a strong position to influence bank operations and to bring about a quick adjustment in the volume of credit through credit ceilings, moral pressure, and other methods. Other sources of finance that are less susceptible to central bank influence include mutual savings and loan banks, credit associations, life insurance companies, and other non-bank financial institutions.

The bond market is relatively undeveloped because the government's low, long-term interest rate policy has made bonds relatively unattractive as compared with the comparatively high level of short-term rates. Individuals and institutional investors tend to buy discount debentures only, which account for less than 30 percent of total issues. Bond buying, therefore, is chiefly confined to banks and other financial institutions, which are expected to purchase government and government-guaranteed bonds according to an unofficial allocation quota. Bond issues have increased because of the high (but gradually decreasing) level of government issues since 1966. Although industrial debentures also have increased, they account for only about 10 percent of the total and their contribution to the supply of industrial funds is small. The secondary bond market has been open since the beginning of 1966, and although over-the-counter transactions have been rising rapidly, about half of the total business is accounted for by trading in financial debentures, mainly by the city banks and the Central Cooperative Bank of Agriculture and Forestry. It is generally accepted that improvement of the efficiency of the bond market is important, but significant progress can not be made without rectifying the imbalance between short and long-term interest rates.

There are nine stock exchanges in Japan; the two most important, Tokyo and Osaka, account for about 95 percent of the total business. Since the mid-1960s foreign investors have taken an increasing interest in Japanese stocks. Although foreign interest subsided a little in 1970, it is expected to be maintained at a relatively high level as long as rapid economic growth continues.

Foreign trade. An outstanding feature of Japan's economic development has been the rapid advance in its overseas sales. During the 1960s the country's exports grew at an average annual rate of about 16 percent, as against a growth of little more than 8 percent in world exports as a whole. The share of exports in the country's gross national product has not changed significantly; in 1970 the figure was some 10 percent, compared with an export ratio of about 30 percent for West Germany and 22 percent for the United Kingdom. From the point of view of individual industries and as a generator of growth, however, exports are much more important than their contribution to the national income would suggest. The growth in exports provided about 14 percent of the increase in the gross national product, and overseas sales accounted for around 20 percent of the country's total iron and steel and automobile production.

Reasons for this spectacular export performance are the growing variety of Japan's industrial output, the shift to products with a relatively high added value, more advanced sales promotion techniques, and Japan's export competitiveness. The rise in labour productivity has enabled industry to absorb much of the rapid rise in wage costs, with the result that on the whole Japan's export prices tend to be lower than those of its principal competitors. Export prices are estimated to be about 13 percent lower than those of the United States and the more important western European countries. Since indications are that the growth in labour productivity will continue to outstrip that of the developed world, the outlook for exports during the early and middle 1970s is highly encouraging.

The 1960s saw a significant change in the composition of exports. In 1958 nearly one-third of the total was accounted for by textile products, but 12 years later their importance had slumped to 12.5 percent. Exports of machinery grew tenfold, and their share of total exports increased to 46.3 percent. Other important export products include metal and metal manufactures and chemicals. Food exports have lost much of their earlier importance; despite a 100 percent increase in their value between 1963 and 1969, they accounted for only 3.4 percent of the total in 1970. The United States, Japan's largest single customer, accounts for nearly one-third of all exports. Southeast Asia is the second largest customer, followed by western Europe. Exports to the Middle East and some African countries have substantially decreased since 1960.

Growth of imports

Imports have grown at an average annual rate of around 15 percent. There is, however, some evidence of a gradual fall in the propensity to import, partly because of the growing ability of Japanese industry to satisfy domestic requirements of manufactured goods and partly because of a shift to products with a higher added value. There are some fears that the rapid rise in the standard of living and the gradual abolition of the remaining import restrictions could speed up the growth of overseas purchases. According to forecasts in the early 1970s, imports of goods and services are likely to record an annual average advance of 15 to 16 percent until 1975, which is comparable to the projected increase in the gross national product. As long as the increase in exports does not fall below expectations, however, such an import growth would continue to yield a trade surplus.

In view of Japan's meagre natural resources, it is not surprising that the bulk of its imports are raw materials, foodstuffs, and fuels. In 1970, metals and other basic raw materials accounted for about 30 percent, fuels almost 21 percent, and food about 14 percent. The largest single category of manufactured goods was machinery and allied products, followed by chemicals. Japan's largest supplier is the United States, which accounts for about 30 percent of its total imports, followed by Australia, Canada, Iran, Indonesia, West Germany, the Philippines, Saudi Arabia, and Malaysia.

MANAGEMENT OF THE ECONOMY

Private enterprise, the role of the government and taxation. Japan's system of economic management is probably without parallel in the world. Although the extent of direct state participation in economic activities is limited, the government's control and influence over business is stronger and more pervasive than in most other free-enterprise countries. This control is not exercised through legislation or administrative action but through constant—and to an outsider almost obsessive—consultation with business and through the authorities' deep indirect involvement in banking. Consultation is mainly by means of joint committees and groups that keep under review, monitor performance of, and set targets for just about every branch and sector of the economy. In addition, there are several agencies and government departments that concern themselves with such aspects of the economy as exports, imports, investment, and prices, as well as overall economic growth. These are staffed by experts, who are not only in constant touch with business but are also close to the minister concerned; they form an integral part of a system that is quick to collate and interpret the latest economic indicators and to respond to changes in the situation. The most important of these agencies is the Economic Planning Agency. Like the Bureau of Statistics, it forms part of the Prime Minister's Office and, apart from monitoring the daily running of the economy, it is also responsible for long-term planning.

The system works well, without any major crises in government-business relations, because of the unusual self-discipline of Japanese businessmen in their dealings with the authorities and the government's deep understanding of the role, needs, and problems of business. The need for large-scale government participation in economic activities is thereby obviated and, unlike many governments in the free-enterprise world, the state appears to be positively reluctant to extend its direct role. The government has monopoly of tobacco and salt, which is operated through the Japan Monopoly Corporation. It also runs the telephone and telegraph services through the Nippon Telegraph and Telephone Public Corporation, while another publicly owned company, the Japan National Railways, is responsible for about two-thirds of rail transport. The government has an interest in air travel, radio, and television. It plays no part in gas production or—except for providing electricity in economically undeveloped areas—in electricity generation.

The government role in banking

The government's economic influence is supplemented by its substantial role in banking. The state owns a number of financial institutions, such as the Japan Development Bank, the Export-Import Bank, the Small Business Finance Corporation, and the Housing Loan Corporation, whose principal objective is to provide finance to private enterprise in areas in which it is considered particularly desirable. The Ministry of Finance and the Bank of Japan have considerable influence over business investment decisions because of the close interdependence of business, the commercial banks, and the central bank. Industry looks to the banks for a substantial part of its investment funds, which are in turn dependent on central-bank financial support.

Tax revenues account for about 85 percent of the government's total income. Japanese taxes can be divided into three main categories: taxes on individuals, on business, and miscellaneous levies. Individuals are subject to progressive income, prefectural, and municipal inhabitant taxes. Income tax is divided into withholding tax, which is deducted at the source of income, and assessment tax, which is payable annually. The local taxes consist of a per capita levy and an income levy. An enterprise tax is sometimes levied on individuals who carry on specified business activities. Businesses are subject to local inhabitant taxes, which consist of a per capita and a corporation levy, enterprise tax, and corporation tax. Other taxes include gift and inheritance taxes, excise taxes levied on a number of consumer goods, and liquor, gasoline, travel, entertainment, securities-exchange, and automobile-acquisition taxes. Income and corporation taxes each bring in about one-third of total tax revenue; the third most important levy, accounting for about 10 percent of the total, is liquor tax.

Trade unions and employers' associations. Japanese trade unions have a relatively short history. Although there were several labour organizations before World War II, trade unions became important only after the United States occupation forces introduced legislation that gave workers the right to organize, to bargain with employers, and to strike. In the early 1970s trade unions had a combined membership of about 11,431,000. Since Japanese trade unions are generally organized on a plant or enterprise basis, their number is relatively large. In 1970, there were a total of nearly 61,000 unions, and in a number of cases there were different organizations for different plants of the same company. The great majority of the enterprise unions are affiliated to federations that are loosely organized on craft lines, such as the Federation of Iron and Steel Workers Unions and Federation of Textile Workers Unions, and to one of the four national labour organizations. They retain much of their independence, however, in dealing with employers. While the craft and national federations formulate general policy, discuss and advise on strategy, and coordinate wage offensives, serious negotiations are usually conducted by individual unions and the employees. One result of Japan's industrial, as opposed to craft, unionism is that demarcation disputes and interunion rivalry for members are relatively rare. Furthermore, if judged in terms of working days lost, Japanese labour relations have been noticeably better than those in other developed countries, such as the United Kingdom, Italy, and the United States.

Four national labour organizations

The national labour organizations are the left-wing and highly political Sōhyō (General Council of Trade Unions), the more moderate and less political Dōmei Kaigi (General Council of Japan Labour Organizations), the Shinsambetsu (National Federation of Industrial Trade Unions), and the Chūritsurōren (Liaison Council of Neutral Trade Unions). Sōhyō is the largest of the four, its membership accounting for some 40 percent of all unionists. Dōmei Kaigi has 17 percent of the total; Chūritsurōren, 10 percent; and Shinsambetsu, only 1 percent of all union members. Sōhyō and Dōmei Kaigi are the two principal rivals; Chūritsurōren often associates itself with Sōhyō, especially when planning the annual "spring labour offensive."

Apart from the political demands of the left-wing unions—such as the ending of the Japan–United States security treaty—labour organizations have been mainly concerned with such questions as wages, prices, and working conditions. Problems of the evolution of a comprehensive industrial policy, of greater centralization, and of the union of the rival national organizations became important during the 1960s mainly because of the trend toward growing concentration in industry and greater cooperation among the various employers' organizations. There is also a growing feeling that in this age of rapid technological progress and change, the ideology-ridden policies of Sōhyb and Dōmei Kaigi are no longer adequate. Although all of the interested parties pay lip service to unification, the pressure for greater cooperation, pragmatism, and professionalism comes mainly from the independent unions in private industries. Many of these seem to feel that the national federations are incapable of self-induced reform and that the initiative for a more modern and united trade union movement must come from those unions that are not fully committed to either side.

Contemporary economic policies and problems. During the 1960s the principal aims of the government's economic policy were rapid growth and the attainment of strong external payments. This policy had three main supports—a high and rapidly increasing level of private equipment investments, a spectacular advance in exports, and growth in private consumption. Subsidiary themes that were regarded as essential for the fulfillment of the two main objectives included a steady shift of agricultural labour to industry, a rapid improvement in labour productivity, rationalization and concentration in industry, and the large-scale import and application of foreign technology. Also part of the program was the relaxation of the strict antimonopoly laws, as well as strong protection of industry against foreign capital intrusions and, in some cases, competition by imports. Judged in terms of its objectives, this policy has been enormously successful. There is, however, less reason for satisfaction over the trend of prices. Between 1960 and 1971, retail prices rose by nearly 100 percent, and there is little doubt that this rapid increase was one of the major reasons for inflationary wage claims and settlements. Another less encouraging aspect of the period was slow growth in public investment. In order to minimize the tax burden, the government's own expenditure has been kept relatively low, with the result that the growth of social capital, which contributes to public as opposed to private affluence, has been far from adequate.

The problem of inflation

As the standard of living rises, the pressure for better social services, education, roads, housing, and transport facilities is likely to grow. It will be necessary to reconcile higher spending in these areas with a rapid growth in private investment, without placing undue strain on the country's resources. It is widely felt that while the fulfillment of these and other social objectives could have some effect on private consumption and investment, it need not significantly affect the rate of overall economic growth. Inflation is the most serious threat to rapid and balanced economic progress. During the late 1960s, the growth in wages tended to outstrip improvements in productivity, a trend that could seriously affect Japanese prices and competitiveness. Because of the growing shortage of skilled labour, the bargaining powers of the unions are likely to strengthen; and if the price spiral is not slowed down, the pressure for a faster increase in wages than in productivity may be very difficult to resist.

Other problems arise out of the impending liberalization of trade and foreign investment. Until the late 1960s, the economy received strong protection against foreign competition through restrictions on imports, nontariff barriers to trade, and strict control of foreign investment. Some of these measures have been abolished or relaxed, but many still remain in force, and, in order to prevent a serious deterioration in trade relations with the developed world, liberalization must be completed by 1980. Although liberalization could lead to a faster increase in imports and a significant rise in the influence of foreign-affiliated enterprises, it must be emphasized that greater direct foreign investment and the accompanying inflow of technology could have a highly beneficial effect on the economy. Another problem that will require close attention is the growing labour shortage; in order to mitigate its effects it will be necessary to shift labour from the agricultural and service sectors into industry, to promote retraining, to encourage labour mobility, and to rationalize industry so as to release labour from the large number of fairly inefficient small enterprises. **(E.I.U.)**

IV. Transportation

OVERALL TRANSPORTATION PATTERNS

Until the latter part of the 19th century, the majority of Japanese people travelled on foot. There were no vehicular means of transportation, except for small wagons, carts, or palanquins (kago) carried by men or animals. The first railway was built between Tokyo and Yokohama in 1872, and others soon followed, though the rugged terrain required the construction of many tunnels and bridges. Iron ships were built at about the same time, and modern ports were constructed. Road construction, however, tended to lag behind the development of other means of transport, resulting in the present congestion.

Japan's great cities attract large numbers of passengers and vast quantities of goods. Tokyo, especially, with a metropolitan population of some 22,000,000 people, is an incomparably large focus for transportation; it is followed by the Osaka metropolitan area, in which the three cities of Osaka, Kōbe, and Kydto together have a metropolitan population of almost 15,000,000. The third largest focus of transportation is Nagoya, with nearly 6,600,000 people living within a 30-mile radius of the city. All of these large urban agglomerations are served by large and internationally known ports. Other cities, such as Kitakyiishu, Fukuoka, Sapporo, Sendai, and Hiroshima, also function as hubs of the transportation network.

Transport networks

The largest volume of intercity or inter-regional transport, in both passengers and goods, moves between the two largest metropolitan regions—by rail, road, coastal waterway, and air. Kyūshū, the third largest island. is connected with Honshū, the largest island, by the world's first undersea railway tunnel (built in 1941) and by an undersea double-decked road tunnel (built in 1958); a huge suspension bridge linking the two is also under construction. A railway tunnel 38 miles long, of which 14 miles runs under the sea, is being built to connect Hokkaidō, the second largest island, with Honshū, and is expected to be completed in 1978. Three large bridges are planned to connect Honshii and Shikoku, the fourth largest island. Many smaller islands near the main islands have also been connected to them by bridges. As a result Japan step-by-step is slowly losing its character as an archipelago.

The Japanese network of telecommunications and of postal services is among the best in the world. Its hundreds of islands, as well as its remotest villages deep in the mountains, are effectively interlinked by these services. In 1971 Japan had about 26,300,000 telephones; 92 percent of the telephone service was automated.

COMPONENT SYSTEMS

Road networks. The development of the road network is retarded in comparison with Japan's general economic progress and in view of its large number of cars. In 1970 there were more than 600,000 miles of public road, but some 200,000 miles were unusable by cars. By the early 1970s, however, some expressways had been built, including the Tōmei Expressway (linking Tokyo and Nagoya) and the Meishin Expressway (linking Nagoya and Kbbe). The metropolitan regions of Tokyo and Ōsaka have a limited-access highway network within their respective built-up areas. In the early 1970s about 2 percent of the gross national product was being devoted to road construction, but the construction of a nationwide network of expressways was expected to require many years. The fact that many roads, both in built-up and in rural areas, are narrow and winding causes additional planning problems. Until the 1920s most Japanese roads were used exclusively for pedestrians, and even horse-drawn vehicles rarely used them. This fact, as well as the limited area of land in proportion to population, affected the pattern of road development considerably. Even in the early 1970s a great number of narrow roads were still being constructed, although a minimum width of at least 13 feet is recommended in built-up areas, except in special cases. In rural areas, footpaths are still in general use, despite the existence of newly built or newly widened motorable roads. Extremely dense networks of footpaths in many mountain areas are used for farming, lumbering, or recreation.

Japanese city street patterns are manifold. Cities such as Kyōto and Nara still preserve the gridiron street pattern of the ancient Chinese city plan, though with modifications in built-up inner parts of the cities. In many rural areas as well, the ancient pattern of land division and the resultant road pattern take the rectangular gridiron form, which is similar to the U.S. township layout in concept and pattern, although it was evolved much earlier and is smaller in size. Feudal towns, especially fortified (castle) towns, have somewhat similar street patterns, though with many modifications for defense purposes. Contemporary Japanese cities and towns still tend to follow an approximate gridiron pattern. Despite this, there are many irregular roads, as well as many modern geometrically curved roads.

Increases in vehicular traffic

The number of four-wheeled motor vehicles and three-wheeled trucks has increased at a phenomenal rate. Numbering only 125,000 in 1935, the number rose to more than 10,000,000 in 1967 and to about 19,472,000 in 1971, a number exceeded only in the United States. Among the world's major countries, Japan has the highest density of motor vehicles per unit area in the plains and in other inhabited areas. Trucks represent a much higher proportion of vehicular traffic than in other major motorized countries; only in 1970 did passenger cars outnumber trucks for the first time. The quantities of freight transported by trucks in the early 1970s surpassed those carried by rail, and the role of the trucking industry was expected to increase still further; the use of large trucks and of containers was also increasing. The use of passenger cars was no longer regarded as a luxury for the majority of people, and commuting to work by car had become an increasingly common practice, resulting in congestion in the big cities and in industrial areas. Railways, however, still played the major role in carrying commuters. No practical solution to the problem of how to reduce the number of cars on the roads has been discovered, though the increase in poisonous exhaust gases and in the noise of the traffic had become topics of daily conversation. The necessity of preserving natural or historical sites or of keeping areas free from traffic noise often proved an obstacle to road construction. At the same time, various technological devices to reduce pollution from automobiles and other problems, as well as to increase efficiency and safety, were under study.

Railways. Despite the competition from road transport, railways play an extremely important role in transporting passengers. In 1970 the total route length of Japanese railroads was 17,000 miles, of which about 7,100 were electrified. Railroads continue to give way, however, to competition from road and air transport.

The first railroad in Japan, which ran from Tokyo to the port of Yokohama, was British financed and was built in 1872 by British engineers. There was strong opposition to its construction, because nationalists feared the expansion of foreign economic and political influence. This opposition was silenced after the completion of the first railroad. After another railroad was built connecting the Osaka area with the port of Kobe (1874) and with Kyōto (1877), Japanese engineers began to build railroads at a rapid rate. The first streetcar line was constructed in Kyōto in 1891, using the electricity from the nation's first power station. In subsequent years, Japan, unlike most other Asian countries, developed quite extensive intra-urban and suburban railroad systems; the period between the two world wars, in particular, was characterized by the construction of many suburban railroad lines to the suburbs to serve the needs of growing numbers of middle-income people. In 1927, the first subway was built in Tokyo's downtown district. Construction of new railroads continued until the outbreak of World War II, but the Japanese defeat and its aftermath prevented further construction for some time. From about 1955 onward, however, railroad construction was resumed. Subway construction, in particular, progressed, and by 1972 Tokyo, Osaka, Nagoya, Sapporo, and Yokohama had subway systems; among these, the Tokyo system was the most extensive, totalling more than 80 miles, of which about 66 miles were underground.

The New Tōkaidō Line

In 1964 the renowned New Tbkaidb Line, a part of Japan National Railways, began operations, providing fully automatic electrified service, for passengers only, on a double-track line between Tokyo and Osaka, at a maximum speed of 130 miles per hour. In 1971, 38 super-expresses, stopping only at Nagoya and Kyōto on the 320-mile route, and 67 limited expresses, which stopped at 11 stations, ran daily between the two cities. 'The railroad had proved highly profitable and had experienced no serious accidents.

In the early 1970s plans were made to introduce freight traffic on the New Tbkaidb Line, as well as to build a nationwide network connecting all major cities, to be called the New Trunk Railway system, based on its method of operation. It was planned to extend the Tokyo to Osaka line to Okayama by 1972 and further westward to Fukuoka in Kyiishii by 1975. Three lines radiating outward from Tokyo were also under construction—to Morioka (and ultimately to Sapporo), to Niigata, and to Narita, respectively. The maximum operating speed was raised to about 155 miles an hour in 1972.

The most serious traffic problem is caused by congestion on railroad transport in the large cities. Most commuter trains are very crowded during rush hours, with some trains carrying twice—and (rarely) three times—the number of passengers for which they were designed. Services are being expanded to cope with the growing demand, with trains running at intervals of only a few minutes and sometimes of only one minute. In Tokyo, a monorail (completed in 1964) operates over a distance of about eight miles between downtown Tokyo and the international airport at Haneda. Monorails are numerous in Japan, but most are primarily used for recreational purposes, as are the cable cars that operate in the mountains. The length of commuting travel time is a serious traffic problem; in rare instances it consumes two hours each way in a working day. In the early 1970s most Japanese railroads, especially local lines, ran at a deficit.

Port facilities. Japan is one of the principal seagoing nations, having engaged in seafaring since early times. In about 1600 the port of Sakai, just south of present-day Osaka, prospered from its trade with China and Southeast Asia. Soon afterward, however, the feudal regime greatly restricted foreign trade by imposing a policy of isolation from the rest of the world—a policy that was to last for about 250 years. As a consequence, no ports engaged in foreign trade with the exception of Nagasaki, which was opened to a limited number of Chinese and Dutch merchants. After Japan reopened its doors to foreign trade in 1858, it was some time before large modern trading ports were developed; Yokohama and Kōbe became the leading trade ports of the new Japan, the former being the outport of Tokyo and the latter the outport for Osaka and Kyoto. Many other modern ports subsequently came into existence.

In 1971 Japan had the second largest merchant fleet in the world; with a total gross tonnage of more than 30,500,000 tons, it was exceeded in size only by the ships registered in Liberia. The Japanese merchant fleet consisted of about 8,850 ships of more than 100 gross tons, of which about 1,400 were oil tankers, including some that were over 250,000 tons in size. In 1969 Yokohama was the leading import port, and Kōbe the leading export port. They were followed by Tokyo, Nagoya, Osaka, Kawasaki, Chiba, Moji, Yokkaichi, and Shimizu (in descending order of importance). Tbkyb-wan (Tokyo Bay), with its four large ports of Yokohama, Tokyo, Kawasaki, and Chiba, handled more cargoes than either

Europoort or the New York Port Authority; for the moment, however, the four ports are under separate management, even though they maintain close mutual cooperation.

Air transport. Both domestic and international air transportation are rapidly increasing in importance. Before World War II, air transportation was considerably restricted, but, since the foundation of Japan Air Lines (JAL), a government-supported private company, in 1953, international flights have proved profitable. In the early 1970s Japan was connected with most major countries in the world both by JAL and by foreign air companies. Domestic flights were also experiencing a boom.

Tokyo is by far the nation's largest single focus of air transport, followed by Osaka. Other major airports are in Nagoya, Sapporo, and Fukuoka. All other metropolitan areas are also connected by air routes. Generally speaking, southwestern Japan is covered by a denser network of air transport than other regions, primarily because of the presence of many islands.

V. Administrationand social conditions

THE STRUCTURE OF GOVERNMENT

The role of the emperor

Constitution **and** government. Japan's constitution was promulgated in 1946 and came into force in 1947; it superseded the 1889 constitution, from which it differed in the following points: first, the emperor is the symbol of the state and of the unity of the people, while sovereign power now rests with the people; second, Japan renounces war as a sovereign right; and third, fundamental human rights are guaranteed as eternal and inviolable. Furthermore, the government is now based on a constitution that aims at maintaining Japan as a peaceful and democratic country in perpetuity.

The emperor has no powers related to government. His major role as emperor consists in such formalities as appointing the prime minister and the chief justice of the supreme court, convoking sessions of congress, promulgating laws and treaties, and awarding of state honours —all with the advice and approval of the Cabinet. Before being appointed by the emperor, the prime minister is first designated by the Diet, the members of which are elected by the people.

Legislative powers are vested in the Diet, which consists of two houses. The House of Representatives, or lower house (Shiigiin), ultimately takes precedence over the House of Councillors, or upper house (Sangiin). The House of Representatives has 491 seats, representing all the 47 prefectures of Japan (including Okinawa); the House of Councillors has 252 seats, representing all prefectures as well as the nation as a whole. The House of Representatives controls the budget and approves treaties with foreign powers. Executive power is vested in the Cabinet, which is organized by the prime minister. If the House of Representatives passes a resolution of no-confidence or refuses to pass a vote of confidence in the government, the Cabinet must resign, unless the House of Representatives is dissolved within ten days. In 1972 there were 12 governmental ministries and five agencies in addition to the Prime Minister's Office. All central-government offices are located in and around the Kasumigaseki district in central Tokyo. An independent constitutional body called the Board of Audit is responsible for the annual auditing of the accounts of the state.

Local government. Japan is divided into 47 prefectures, 43 of which are ken (prefectures proper) but one of which (Tokyo) is a to (metropolis), one (Hokkaidd) a do (territory), and two (Ōsaka and Kyiito) are fu (urban prefectures). (Okinawa, which was under United States administration from the end of World War II until 1972, then became a ken.) Prefectures, which are administered by governors and assemblies, vary considerably both in area and in population. The largest prefecture is Hokkaidd, with an area of 30,314 square miles, while the smallest is Osaka, with 716 square miles. The most populous prefecture in 1970 was Tokyo, which had a population of 11,408,000; the least populated was Tottori, with 569,000. A prefecture is further subdivided into minor civil divisions; these include the city (shi), town (*machi* or *chō*), and village (*mura* or son). All these local government units have their own mayors, or chiefs, and assemblies. Before World War II, there were also counties (gun), consisting of towns and villages but excluding cities within a prefecture. In the early 1970s, however, the county system only survived in the form of statistical units.

An intermediate level of governmental services is formed betweeen the central and prefecture levels: the branch offices of the government that are located in certain cities, which—as regional centres—generally administer several prefectures together. Designated cities (*shi*tei-toshi), which must have populations of at least 500,000 each, are divided into wards (ku). In 1972, there were ten cities in this category—Tokyo, Osaka, Yokohama, Nagoya, Kyōto, Kdbe, Kitakyūshū, Sapporo, Kawasaki, and Fukuoka. A ward has a chief and an assembly, the former being nominated by the mayor and the latter elected by the residents.

Terms of office; constituencies

The political process. Elections. Members of both the House of Representatives and the House of Councillors are chosen by general elections. Members of the House of Representatives serve for a four-year term, but this term may be terminated earlier if the house is dissolved; the 491 seats are contested in 124 different constituencies located throughout Japan. Members of the House of Councillors are elected for a six-year term, with half of the members being elected every three years. The electoral procedure for the House of Councillors differs from that for the House of Representatives in that 100 members out of the total 252 are elected from a national constituency, requiring the participation of voters throughout Japan; the remaining 152 members are elected from 47 prefectural constituencies. The number of seats for each constituency was determined largely by the population density in each area in 1947, with some modifications resulting from the population increase in urban constituencies. Heads of local governmental units, such as prefectures, cities, towns, and villages, are elected by local residents. Universal adult suffrage is available to all men or women who are 20 years old or older.

Political parties. Freedom to organize political parties is guaranteed by the constitution. Since the enactment of the 1947 constitution, many political parties have been organized, have merged, or have been dissolved. In 1971, the Liberal-Democratic Party (Jiyū-Minshutō) was the party in power, representing somewhat conservative elements. Such conservative parties have held power since 1948.

The Japan Socialist Party (Nihon Shakaitii) has long been the largest opposition party, drawing much of its support from labour unions and inhabitants of the large cities. It has a neutralist policy, urging the establishment of a peaceful security system covering Japan and East Asia by means of a treaty between Japan, the United States, the Soviet Union, and China. It opposes the Treaty of Mutual Cooperation and Security concluded between Japan and the United States, which has been upheld by the ruling Liberal-Democratic Party since 1960.

In 1971, the Kdmei Party (Kōmeitō) meaning fair or fair play party, was the second largest opposition party, drawing its main support from the Sōka-gakkai, a religious group belonging to the Nichiren Buddhist sect. Although the party was formed in 1964 under the influence of the Sdka-gakkai, it has renounced any formal ties with it. The Kōmei Party's major policy aim is the establishment of a welfare system and the promotion of "human Socialism." The Democratic Socialist Party (Minshatii) was formed in 1960 by a group from the Socialist Party that was opposed to extreme ideologies. It aims at creating a Socialist society through democratic processes and draws most of its support from the same sources as the Socialist Party itself. The Japan Communist Party (Nihon Kyōsantō) is one of the minor parties, although its influence on political and intellectual developments in society is strong.

The participation of the citizen. Since the late 1960s, the role of the citizen in politics has been often discussed both in journalism and in daily conversation, especially at election time. Many citizens today wish to participate in solving such problems as traffic congestion, waste disposal, air and water pollution, the shortage of parks and playgrounds, and noise control. In many cities there are numbers of commercial and residential streets that have been closed to automobile traffic.

Justice. The judiciary is completely independent of the executive and legislative branches of the government. The judiciary system consists of the Supreme Court, eight high courts, a district court in each prefecture, with the exception of Hokkaidō, which has four, and many summary (informal) courts. Family courts also abound.

The Supreme Court is by any standard the most important one in Japan; it consists of one chief justice and 14 other justices. The chief justice is appointed by the emperor upon designation by the Cabinet, while the other 14 justices are appointed by the Cabinet. The appointment of the justices of the Supreme Court is subject to review in a national referendum, first at the time of the general election following their appointment and then at the general election following the elapsing of a ten-year period. An impeachment system also exists; the court of impeachment consists of members of the House of Representatives and of the House of Councillors. The Supreme Court determines questions of the constitutionality of any law, order, regulation, or official act. Lower court judges are appointed by the Cabinet from a list of persons nominated by the Supreme Court. The appointment term is for ten years, and re-appointment is allowed. All judges of lower courts must retire at the age of 70, according to law.

The Self-Defense Forces

The armed forces. Under its present constitution, Japan cannot maintain armed forces for purposes of aggression; in consequence, national security is maintained by the Self-Defense Forces (Jietai), as well as by the collective security system in which the United States participates. The Self-Defense Forces play a role that is entirely limited to defense and internal security.

Between 1945 and 1950, Japan had no armed forces except for police; after the outbreak of the Korean War, however, the government, at the suggestion of the Allied occupation forces, established a National Police Reserve of 75,000 men and increased the strength of the Maritime Safety Agency to 18,000 men. In the early 1970s, the Self-Defense Forces consisted of ground, maritime, and air forces, under the civilian-controlled National Defense Council; in 1971 they had a total combined strength of 259,000 uniformed personnel. The Ground Self-Defense Force alone had 179,000 uniformed personnel, divided into five armies and 13 divisions; it also had 350 aircraft and more than 100 missile launchers. The Maritime Self-Defense Force had about 38,300 personnel, as well as 210 ships and other vessels totalling 133,000 tons, more than 150 antisubmarine aircraft, and 100 other aircraft. The Air Self-Defense Force had 41,700 personnel, with 970 aircraft and 75 missile launchers. The majority of weapons were supplied by the United States.

The constitutionality of the Self-Defense Forces has often been disputed. The Supreme Court ruled in 1959 that it did not violate the constitution, because its purpose was defensive. United States military bases operated in many parts of Japan under the Treaty of Mutual Cooperation and Security, concluded between Japan and the United States in 1960. The treaty may be terminated one year after either signatory so notifies.

ADMINISTRATIVE SERVICES

Education. A great many educational institutions existed even in the feudal period preceding the Meiji Restoration of 1868, a number of which had been subjected to Chinese cultural influences since ancient times. Numerous private schools (terakoya), mostly in towns, functioned as elementary schools; reading, writing, and arithmetic were taught by unemployed warriors, monks, or others. Provincial lords (daimyd) also established special schools for children of the warrior class. Yet another type of school was operated primarily for the children of wealthier merchants and farmers.

The modern educational system was introduced immediately after the Meiji Restoration. The government set up elementary and secondary schools throughout Japan in 1872, and in 1886 a system providing three to four years of education was inaugurated. The introduction of modern education did not encounter many problems, primarily because it was possible to utilize the educational system already functioning. Free compulsory education was introduced in 1900 and in 1908 was extended to a period of six years. Since 1947, compulsory education has been for a nine-year period, beginning at the age of six.

In the early 1970s the educational system of Japan was organized as follows: kindergarten lasted from one to three years but was not compulsory, being attended by about two-thirds of the children at the age of five; compulsory elementary school lasted six years, compulsory middle school three years, and high school (not compulsory) another three years. Higher education consisted of from two to three years, for junior colleges, and four years, for ordinary colleges and universities; a master's degree could be obtained in two years after earning a bachelor's degree, and a doctor's degree in three years after earning a master's degree. In addition, there are five-year technological colleges (combining high school and junior college education). Public elementary and middle schools are free; a modest fee is required for public high schools and public higher educational schools. About 5 percent of Japan's gross national product is spent on education.

Higher education

Japan is one of the few countries in the world that provide a complete and thorough education for almost all the people. Although neither kindergartens nor high schools are compulsory, attendance at both is fast becoming virtually universal; higher education is also becoming popular. In 1970, there were more than 380 colleges and universities with more than 1,400,000 students and about 480 junior colleges with more than 260,000 students; nearly 30 percent of the attendance in both categories was female. The percentage of freshman students pursuing higher education represented about 20 percent of all students. In higher education, about 75 percent of the students attend private colleges and universities, and 90 percent attend private junior colleges. Slightly more than half of the students in higher education studied the humanities and social sciences. In 1969 Tokyo prefecture had more than 100 out of the national total of 379 institutions of higher education.

The Tokyo Metropolitan Area, including Yokohama and many other satellite cities, had nearly 1,000,000 students of higher education. Such a concentration led to congestion as well as to an active intellectual life. A large project called Academic City, located about 30 to 40 miles northeast of Tokyo near Tsukuba-san (Mt. Tsukuba), is to comprise a large new national university and many national and private research institutes; it is to be completed by 1980.

Educational administration is decentralized, with the Ministry of Education playing a coordinating role. Responsibilities for the budget, curriculum, teacher appointments, and the supervision of elementary and middle schools are in the hands of local educational boards.

Health and welfare services. In the early 1970s, social-welfare services were not yet satisfactorily established, although they had expanded considerably since the end of World War II. In the 1970 governmental budget, about 15 percent of the total was allocated to social welfare, including social insurance, services for the aged and the physically and mentally handicapped; and care for disadvantaged children. Social insurance itself consisted of health insurance, pension insurance, unemployment insurance, and workmen's accident compensation insurance. After 1961 the health-insurance system covered all Japanese people, although the scale of payments varied and in some cases no payments were required. The government had established more than 800 health centres throughout Japan, aiming primarily at improving environmental sanitation and at preventing communicable diseases in their early stages. Out of a total of about 8,000 hospitals in 1970, about 450 were operated by the national government, more than 1,000 by local governments, and the majority by unions, associations, or individuals. Aged people received many services, including medical examinations, home help services, recreational services, and institutional care, as well as varying amounts of financial aid. Local governments are obliged to provide welfare services for the physically handicapped and mentally retarded. Various children's welfare programs also exist—for example, medical-care services are free to expectant mothers and to young children from low-income families. Many voluntary and private associations also provide supplementary services.

Wartime destruction of houses

Housing. The housing shortage has posed major political and social problems. It is due to the following: during World War II, 70 percent of the houses in 70 percent of the major cities in Japan were ruined; constantly rising land prices, especially in and around major cities; the general use of lumber as a building material, requiring earlier replacement than brick or stone; the frequent occurrence of earthquakes, typhoons, and heavy rains bringing floods; the government's inclination to encourage economic growth rather than house construction; rapidly rising standards of living, which create a demand for larger and better houses; and an increase in the number of families, resulting from the breaking up of families into smaller units.

The penchant for living in single family homes may be attributed to the Japanese love of nature, to the influence of the garden-city movement, and to the fashion for imitating the tree-shaded residences of the feudal warrior class. The government and some private construction companies have encouraged construction of multistory apartment houses.

To cope with the housing shortage, a semigovernmental agency, the Housing Loan Corporation, was established in 1950 to finance house construction at low interest rates. In 1955, another semigovernmental agency, the Japan Housing Corporation, was organized; by 1970 it had built almost 700,000 units. Local governments have also built a number of units, mostly of the apartment-house type and primarily for low-income families. Between 1950 and 1970, 13,000,000 units were built; in the early 1970s the number was increasing at an annual rate of more than 10 percent. In 1969, less than 20 percent of the houses built were subsidized by governmental or semigovernmental funds. Generally, the size of housing units is increasing as per capita national income rises.

Police services. Japan's police services are under the supervision and control of the government's police agency, which in 1970 employed about 8,000 persons; the nationwide police services themselves had a strength of about 170,000 persons, all of whom were employed by local governments. Police services operate relatively smoothly; many problems that plague other countries are absent because of Japan's insularity and its uniform ethnic composition. The ratio of arrests of suspects is usually very high; the low ratio of extremely cruel crimes stands in contrast to that of most advanced countries.

SOCIAL CONDITIONS

Wages and cost of living. With Japan's very rapid economic growth, wages have been rising at a rapid rate. In the five years from 1965 to 1970, real wages rose by 50 percent and nominal wages by about 75 percent or more; this was largely due to increases in efficiency of production. In prewar Japan, per capita national income reached a peak of about U.S. $200 just before the outbreak of World War II but fell sharply to less than U.S. $100 in 1946. Recovery was somewhat slow during the first postwar years, but, after about 1950, economic growth was fairly consistent and was accompanied by a rapid increase in wages. In the early 1960s it was debatable whether Japan had joined the ranks of the developed countries; by 1964, for example, gross national product per capita had only reached U.S. $660. By the early 1970s, however, very few contended that Japan was still only a developing country rather than a developed one; gross national product per capita rose to U.S. $1,430 in 1969 and to more than U.S. $1,795 in 1970.

Wage trends

As of 1969, the monthly average income for all employed persons in Japan was 64,300 yen (U.S. $179); male income averaged 75,900 yen, while female income was only 36,800 yen (360 yen = U.S. $1.00; 864 yen = £1 sterling, on December 30, 1970). Wages paid by large enterprises are significantly larger than those paid by small ones. In the case of manufacturing industries, in 1968 workers in large factories with more than 1,000 employees obtained nearly double the wages of those in factories with ten to 19 employes, the ratio being 100 to 58. At the same time, the productivity of workers in the large factories was greater than that of workers in small factories, the ratio being 100 to 37; thus, workers in small factories were better paid in relation to their productivity. Workers in large enterprises are usually provided with relatively good housing and with recreational facilities in return for very low payments.

The cost of living in the early 1970s was generally rising at a rapid rate; this was especially true in large cities, where residents were obliged to pay dearly for their metropolitan amenities. Generally speaking, manufactured goods, such as clothes, machines, and cars, were low priced, and public-transportation fares were low, while food and raw materials were comparatively expensive, as were such utilities as electricity. Land prices were extremely high, primarily because of high population density in both urban and rural areas, though some critics attributed them to the government's indifference. A private house built in an urban or suburban area often costs less than the site. Land prices in some areas in downtown Tokyo are said to be higher than those in any other city in the world. In general, high land prices tend to encourage the prevalence of small apartment houses.

Health conditions. Sanitary conditions had improved considerably by the early 1970s. The average life span of the Japanese people was 70.2 years for men and 75.6 for women (1970). Higher living standards, including better nutrition and better living conditions, as well as progress in medical care, contributed much to the increase in the life span. Numerous hospitals, clinics, and health centres throughout the nation, as well as health education in schools and among the public at large, had virtually eliminated such diseases as typhus, diptheria, and scarlet fever. Tuberculosis and dysentery were much less prevalent than they once were. Smallpox, cholera, leprosy, and rabies had long been practically nonexistent, except for rare cases among tourists. Increases in the so-called "diseases of civilization" had, however, become a serious problem. Apoplexy, high blood pressure, heart ailments, mental disorders, and similar diseases had become principal causes of death, as had traffic accidents. The number of hospital beds available for patients in 1970 was more than 1,300,000, and the number of persons per medical doctor was about 900—making Japan one of the most favoured nations in the world in this respect. Japanese medical practice is usually of the Western type, but classical Chinese techniques are also used.

Changing dietary conditions

In the early 1970s, the Japanese people were obtaining increasingly better food. Although calorie consumption—which averaged about 2,250 daily—was much lower than that of Europeans or Americans, overnutrition causing excess weight had nevertheless become a serious problem. The traditional Japanese food was being replaced partly by Western types of food and partly by Chinese food—to such a degree that the average Japanese no longer regards Western or Chinese food as alien.

Social and economic divisions. During the feudal period, there was a social division of commoners into four classes (warrior, farmer, craftsman, and merchant), with a peer class above and an outcast class (variously called *shinheimin* or *eta*) below. With the possible exception of the outcast class, this social-class system has almost disappeared. In 1959, for example, Crown Prince Akihito married a commoner's daughter. Insofar as a social-class system does persist it has no ethnic basis because all residents in Japan are Japanese, who regard themselves as belonging to one and the same ethnic group. The few exceptions include aliens (non-Japanese citizens) living in Japan—particularly Koreans who came to Japan as labourers before and during World War II—and Japanese citizens of Ainu origin, who are scattered over the island of Hokkaidō. Before World War II there was a tendency to distinguish the people of Okinawa from other Japanese because many of them exhibited minor differences in physiognomy and cultural life, but this tendency has disappeared. Okinawan culture, including its dialect and religion, is now recognized as sharing many traits with Japanese culture and, in some respects, represents a prototype. The ethnic unity of the Japanese is reflected in the fact that, of all the peoples of the world, the Japanese have been the least inclined to intermarry with foreigners. A vernacular word gaijin, which literally means a foreigner or foreigners, is often used in daily conversation; the term implies that gaijin are fundamentally different from Japanese and, as such, cannot understand Japanese culture.

Vast discrepancies between the condition of the rich and poor have been reduced since World War II largely as a result of the agricultural reforms of 1946 to 1950 and of the application of a graduated income tax. Although differences exist in income and property, more than 95 percent of the Japanese regard themselves as in the middle-income group. Most of those in the upper-middle income group own their own homes, usually houses with several rooms surrounded by a garden; those in the lower middle-income group usually live in a one-to-four room house or (in urban areas) in an apartment house. The social attitudes of the Japanese and the absence of strict zoning in urban areas have contributed to the mixed land uses characteristic of Japan's cities; thus, functionally different establishments, such as shops, factories, or houses, are found adjoining one another, so that mixed rather than exclusive social patterns result.

VI. Cultural life and institutions

THE CULTURAL MILIEU

Japan's long history has produced a cultural milieu that differs significantly from that of other countries. In general, this milieu is characterized by an inseparable mixture of traditional Japanese culture with introduced Chinese and Western cultural forms.

Ancient Chinese influences

Prehistoric Japanese culture was subjected to ancient Chinese cultural influences that were introduced some 1,500 years ago. One consequence was the imposition of the gridiron system of land division, which long endured; even today, it is possible to trace the ancient place-names and field division lines of this system. Chinese characters and many other Chinese developments were also introduced. The Buddhist religion, which originated in India and subsequently underwent modification in Central Asia, China, and Korea, also exerted a profound influence on the Japanese cultural life, but in the course of history, the process of Japanization continued. The Japanization of the introduced Chinese culture was greatly accelerated during the 250-year period of isolation that ended in 1868. The climate of Japan, for example, which was much more humid than that of China, led to such cultural adaptations as the use of lumber for building in place of the mud and brick used in China. Similarly, the Chinese characters had only a limited use because they did not fit the Japanese language, which belongs to a completely different family. Another trait of the Japanese life differing from the Chinese is that indoors the Japanese sit on the floor rather than using chairs, and they generally sleep on mattresses and quilts spread on the floor, whereas the Chinese normally use chairs and beds.

After the Meiji Restoration of 1868, Japan began to modernize and to industrialize on the European and U.S. pattern. In the period since then, the United States has exerted a more conspicuous influence on Japanese cultural and social life than has Europe, as a whole. Western cultural traits have been introduced on a large scale through the schools and the mass-communication media, such as the radio, newspapers, and television. Western scientific and technical terms have been widely diffused in translation and have even been re-exported to China and Korea. Today, U.S., English, French, German, or Soviet influences on Japanese culture are in evidence in literature, the visual arts, music, education, science, recreation, and ideology.

Effects of Western influences

Modernization has often been accompanied by cultural changes. Rationalism and Socialism based on Christianity, as well as Marxism, have become inseparably related to everyday Japanese life. Western or Westernized music seems to be preferred to traditional Japanese music at most social levels. Although Japanese Christians form slightly less than 1 percent of the population, Christmas is enjoyed, if not celebrated, quite widely, almost as a folk event. The use of Western dress among the Japanese, in place of the kimono, is widespread, although the kimono tends to be used by young women at celebrations and by a considerable number of male adults and older women for home wear. House construction has also been changed considerably by the introduction of Western architectural forms and functions. In shape, in colour, and in building materials, many Japanese houses today are significantly different from the traditional ones; they now have more modernistic shapes, use more colours, and are more often made of concrete and stucco.

Everyday use of modern transport and of modern communication media has brought Japanese urban life close to that of the West. Japanese forms of recreation are similar to those in other developed countries, although there are some notable differences. Outdoor recreational activities include hiking, mountaineering, skiing (there were more than 10,000,000 Japanese skiers by 1971), skating, golf, swimming, boating, fishing, baseball, tennis, and football (soccer). Indoor recreations include shdgi (a kind of chess), go (a strategy game also similar to chess), Japanese and Western card games, basketball, volleyball, pingpong, bowling, wrestling, *jūdō,* kendd (Japanese fencing), karate, and gymnastics. *Sumō* wrestling is also practiced, or watched, both indoors and out.

ARTS, FOLK TRADITIONS, AND POPULAR CULTURE

The Japanese cultural tradition includes many forms of the fine arts and folk arts. Local variations are found throughout Japan's mountainous archipelago, where most river basins, valleys, or islands have their own specific folklores.

The highly refined traditional arts of Japan include flower arranging (ikebana), the tea ceremony (*cha-no-*yu), painting, calligraphy, dance, music, theatrical plays (Including such forms of drama as Kabuki, a highly stylized form of drama characterized by singing and dancing; *bunraku*, the puppet theatre; N6, the classic form of dance-drama), and gagaku (court music), gardening, and architecture. Delicacy and exquisiteness of form, together with simplicity, characterize traditional Japanese artistic taste. The Japanese tend to view the traditional Chinese arts generally as being too grandiose or showy. The newly introduced Western arts are also felt to suffer from the same flaws, though in a different fashion.

With the advance of modernization, many folk traditions and forms of folklore are rapidly disappearing. The widespread use of common Japanese (the Tokyo dialect) has accelerated this trend, since local cultures are directly related to dialects, which are so diverse that many of them cannot be generally understood. Folk songs, for example, are generally no longer commonly sung except in some remote areas in northern and southwestern Japan. Folk music and dance are related to local life and are often significantly concerned with the local religion (whether animistic, shrine Shintō, or Buddhist), agriculture, or human relations (including the theme of love). Some, however, still enjoy a great popularity, which has been increased through the mass media. On informal social occasions, even in the large cities, folk and popular songs are often sung. Such traditional arts as ikebana, cha-no-yu, and calligraphy are studied and practiced by a great many Japanese; *ikebana* and cha-no-yu, in particular, are popular among young unmarried women, since these are regarded as cultural or aesthetic attributes for future housewives. Traditional Japanese painting, dance, and music have, however, lost much of their traditional popularity, though the poetic forms of haiku and waka continue to flourish.

Changes in social customs

In social life, the arranged marriage (miai-kekkon) is being replaced by the love match, though nearly half of all marriages are still arranged or initiated by parents or other older persons or, sometimes, by friends. Modern (usually Western) popular culture has gained a strong foothold in Japan. Jazz, rock and roll, and the blues are enjoyed by the younger generation, along with half-Westernized or half-Japanized folk and popular songs. Many more or less Japanese songs are sung to the accompaniment of Western musical instruments; at the same time, many more or less Western subjects are treated in Japanese-style drama or song.

In 1971, Japan had 12 national holidays. New Year's Day was traditionally regarded as the most important of these holidays, with millions of people engaging in a kind of pilgrimage that leads to shrines and temples starting at midnight of December 31; several shrines and temples each draw more than 1,000,000 people during the three day New Year holiday season. In addition to the 12 national holidays, there are also such other nationwide festivities as Girls' Day (March 3), which is comparable to Boys' Day (May 5), now celebrated as Children's Day (a national holiday). May Day (May 1) is celebrated by many workers. Many temples and shrines celebrate their own specific festivals, attracting large numbers of people. City, town, and village authorities, as well as local communal bodies, often organize local festivals.

CULTURAL INSTITUTIONS

In addition to its cultural institutions such as libraries, museums, art galleries, theatres, parks, gardens, and schools of various kinds, Japanese department stores also play a role in the dissemination of culture by offering free or low-cost exhibitions.

As of 1970, the National Diet Library in Tokyo (which also included branch libraries), with almost 6,000,000 books, was the single largest library in Japan. In 1968, higher educational institutions, including universities, colleges, junior colleges, and technical colleges, had more than 1,200 libraries, with a total of some 60,000,000 books. Secondary and elementary schools are also equipped with libraries as a matter of course. In addition, in 1968 there were about 725 city, town, or village libraries, some equipped with mobile facilities. The overwhelming majority of library books are in the Japanese language.

Museums

There were nearly 340 museums of all kinds in 1968; these included 55 general museums, as well as 43 science, 83 historical, 77 art, and 4 outdoor museums, as well as 22 zoos, 18 botanical gardens, and 31 aquariums. Museums of all kinds have been increasing in number, as well as in the quantity of their exhibits and in their attendance.

Local governments provide youth educational services, offering more than 5,000 classes on various topics in 1968. Adult education is also conducted by local governments, as well as by private institutions. In 1967 alone more than 4,000,000 adults enrolled in classes in general education, vocational training, technology, homemaking, home economics, arts, physical education, and recreation. Many institutions also help to promote nature studies and recreation; in 1971, these included about 500 public and private youth hostels, about 260 national lodging houses, 21 national vacation villages, 23 national parks, 44 quasi-national parks, and a great many prefectural natural parks.

Miscellaneous schools also function as agencies of cultural dissemination; in 1969 these numbered almost 8,000, plus about 60 branch schools, and had a total enrollment of 1,400,000 students. These schools, which are recognized by the local authorities, offer courses in such subjects as dressmaking, handicrafts, cooking, abacus calculation, foreign languages, driving, and nursing.

PRESS AND BROADCASTING

The role of press and broadcasting has long exerted an important influence in an ever changing Japan.

The press. Japan ranks as one of the major book publishing countries in the world. In 1971 nearly 27,000 different books were published, of which about one-quarter consisted of literary works. Books on the sciences and social sciences are gaining popularity. In the same year, the total number of copies of these books was about 510,000,000. About 2,500 different magazines were published, totalling about 1,700,000,000 copies, with more than half of these copies consisting of weeklies. About half of the book-publishing companies are located in Tokyo, where more than 90 percent of all books are published.

The role of newspapers is of great importance. Major newspapers print both morning and evening daily editions. Their total daily circulation in 1970 was about 53,000,000—an average of two newspapers a day to each Japanese household, more than in almost any other country. In 1970, there were 117 newspaper companies. At least five of them could be regarded as operating on the national level; these five were *Asahi* (with a morning circulation of about 5,700,000); *Yomiuri* (5,400,000); Mainichi (4,600,000); Sankei (1,900,000); and Nikkei (*Nihon Keizai* [1,160,000]). All of these papers also published evening editions, and all had their main headquarters in Tokyo. Some local newspapers also have large circulations—for example, *Chūnichi*, with headquarters in Nagoya, with a circulation of almost 1,900,000. Japan's three largest newspapers rank first, second, and third in the world in circulation, even if morning and evening editions are counted separately. All of these large newspapers are generally considered to maintain high editorial standards.

Radio and Television. Radio and television are used in Japan far more extensively than in any other Asian country and, indeed, than in most other countries in the world. Radio broadcasting began in 1925 with the establishment of Nippon (Nihon) Hōsō Kyōkai, or NHK (the Japan Broadcasting Corporation)—a public corporation financed by license fees that according to law must be paid by television-set owners. NHK broadcasts many quality programs, both on radio and on television; no commercial advertisements are permitted. In 1970 NHK was operating more than 300 AM radio stations on two networks and about 250 FM radio stations, offering stereophonic and other broadcasts.

The first television broadcast was made by NHK in 1952. In 1970, there were nearly 2,000 NHK television stations, broadcasting to practically all parts of mainland Japan, as well as to many isolated islands. NHK has been broadcasting overseas programs such as "Radio Japan" since 1953 and in the early 1970s conducted broadcasts in 23 different languages. Private commercial broadcasting began in 1951 and may have become more popular than NHK; in 1970, there were 13 commercial radio companies, 43 commercial television companies, and 35 combined radio and television companies. The total number of commercial television stations was almost 900, and there were about 150 commercial radio stations. In 1970, television closely followed newspapers as the most important media of commercial advertising; magazine and radio advertising were less significant.

Since World War II, Japanese political trends have always taken a democratic direction, for the most part under strong conservative influence. As a result of Japan's very rapid economic growth, all phases of the society and culture have experienced fundamental changes. In the early 1970s, Japan's gross national product was exceeded only by those of the United States and the Soviet Union. Income and wages were expected to parallel those of the highly developed northwestern European countries before the end of the 1970s, if the rate of economic growth was maintained.

Since prehistoric times, the Japanese nation has often remained secluded from the rest of the world for long periods of time, so that the Japanese have developed their own peculiar culture, though Chinese and Euro-U.S. influences cannot be ignored. As a result, the Japanese became one of the least internationally minded of nations. Under the influence of the mass media and of international exchanges, however, Japan's culture is now undergoing revolutionary change. Regional variations, once extremely conspicuous, are now disappearing. A new and very modern nation seems to be emerging—one that will probably retain many of its traditional cultural traits.

(Y.M.)

BIBLIOGRAPHY. Two good gazetteers of Japan are *Nihon chimei jiten*, compiled by AKIRA WATANABE, 3 vol. (1954–56); and *Nihort chimei daijiten*, compiled by AKIRA WATANABE *et al.,* 7 vol. (1967–68). The *Nihon kokusei-zue,* an industrial digest issued annually, contains physical and economic statistical data with explanations. Recent representative geographical works in English include *A Report on Japanese Natural Resources Prepared in General Headquarters in 1948* (1948), an official document comprised of studies on physical and cultural aspects by Japanese and American scholars; *Regional Geography of Japan,* 2 vol. (1957), the proceedings of a conference of the International Geographical Union; *Japan: Guide Book for the Congress Tours* (the XIth Pacific Science Congress, 1966); and ROBERT B. IIALL, JR., *Japan: Industrial Power of Asia* (1963), which contains a brief analysis of postwar industrial development. Geomorphology is covered by AKIRA WATANABE, "Landform Divisions of Japan," *Bull. Geogr. Surv. Inst., Tokyo,* 2:81–94 (1950–51); climatology is dealt with by EIICHIRO FUKUI, "Climate and Agricultural Land Use in Japan," in *Festschrift Leopold G. Scheidl* (1967); see also Fukui's work, "Climatic Division of Japan," in *Geogrl..Rev., Japan,* vol. 9 (1933); TAKESHI MATSUI, "General Characteristics of the Soil Geography of Japan," *Pedorojisto/Pedologist,* 12:25–36 (1968), contains recent information; YOSHIKATSU OGASAWARA, "Land Use of Japan," *Bull. Geogr. Surv. Inst., Tokyo,* 2:95–119 (1950–51); and *Japartese Cities: A Geographical Approach* (Association of Japanese Geographers Special Publication, no. 2, 1970), discuss many subjects pertaining to these respective fields. The following two government pamphlets (both in English) contain official information: "Statistical Handbook of Japan" (Bureau of Statistics, annual); and "The Japan of Today," compiled by the Ministry of Foreign Affairs (annual). The BUREAU OF STATISTICS, *Japan Statistical Yearbook* (annual), provides the best readily available information source on Japan as a whole. The MINISTRY OF FOREIGN AFFAIRS OF JAPAN, *Japan in Transition: One Hundred Years of Modernization* (1968), gives general descriptions and is illustrated. HISAO AONO and SHOHEI BIRUKAWA (eds.), *Nihon chishi* (1967–), is a comprehensive series dealing with Japan's regional geography by prefectures. EDWARD A. ACKERMAN, *Japan's Natural Resources and Their Relation to Japan's Economic Future* (1953), is a basic source for understanding the economic development. For general background on the economy, see also GEORGE C. ALLEN, *Japan's Economic Recovery* (1958); ROBERT J. BALLON (ed.), *Doing Business in Japan* (1967); RICHARD K. BEARDSLEY (ed.), *Studies on Economic Life in Japan* (1964); ALICE H. COOK, *An Introduction to Japanese Trade Unionism* (1966); WILLIAM W. LOCKWOOD (ed.), *The State and Economic Enterprise in Japan* (1965); and YUTAKA MATSUMURA, *Japan's Economic Growth, 1945–1960* (1961). Detailed current information on various aspects of the economy may be found in publications of the Japanese government, such as the BANK OF JAPAN, *Economic Statistics of Japan* (annual); and the MINISTRY OF AGRICULTURE AND FORESTRY, *Abstract of Statistics on Agriculture, Forestry, and Fisheries* (annual). RYUZIRO ISIDA, *Geography of Japan,* rev, ed. (1969), is concerned with major aspects of the physical, economic, and cultural environment. GLENN T. TREWARTHA, *Japan: A Geography* (1965), also deals with geographical aspects in detail. A detailed field survey of village life from geographical, historical, and social viewpoints is made in RICHARD K. BEARDSLEY, JOHN W. HALL, and ROBERT E. WARD, *Village Japan* (1959). Japan's cities are treated in detail in RONALD P. DORE, *City Life in Japan: A Study of a Tokyo Ward* (1958). The JAPAN CENTER FOR AREA DEVELOPMENT RESEARCH, *Nihon rettd no shōraizō* (1969), is a comprehensive atlas covering geographical, economic, social, and cultural aspects, with explanations in both Japanese and English.

(Ak.W./Y.M.)

Japan, History of

This article is divided into the following sections:

- I. Ancient and medieval Japan to *c.* 1550
  - Prehistoric Japan
  - The ancient period (*c.* AD 250–710)
  - The Imperial state
  - The Kamakura *bakufu* (1192–1333)
  - The Muromachi (or Ashikaga) shogunate (1338–1573)
  - The period of the "warring country"
- II The early modern history of Japan (1550–1850)
  - Unification
  - The *bakuhan* system
  - The weakening of the *bakuhan* system
  - The Tempō reform and the collapse of the *bakuhan* system
- III. Japan since 1850
  - The Meiji Restoration
  - Imperial Japan
  - After World War II

I. Ancient and medieval Japan to c. 1550

PREHISTORIC JAPAN

Pre-Ceramic culture. It is not known when man first settled on the Japanese archipelago. It was long believed that there was no Paleolithic occupation in Japan, but since World War II several Paleolithic tools have been uncovered. These include both core tools, made by chipping away the surface of a stone, and flake tools, made by working with a stone flake broken off from a larger piece of stone.

It seems likely that the men who used these implements moved to Japan from the Asian continent. At one stage, land connections via what are now the straits of Korea dnd Tsushima made possible immigration from the Korean peninsula, while another connection, via what are now the Sōya and Tsugaru straits, allowed men to come in from northern Asia.

The Paleolithic Age in Japan is variously dated from 30,000 to 10,000 years ago. Nothing certain is known of the culture of the period, though it seems likely that men lived by hunting and gathering, used fire, and made their homes either in pit-type dwellings or in caves. No bone or horn artifacts of the kind associated with this period in other areas of the world have yet been found in Japan. There was no knowledge whatsoever of pottery; hence, the period is referred to in Japan as the Pre-Ceramic or Pre-Pottery era.

Jōmon culture (5th or 4th millennium to c. 250 BC). The Pre-Ceramic period is followed by two Neolithic cultures, the Jbmon and the Yayoi. The former takes its name from the *jōmon* ("cord marks") pottery found throughout the archipelago. A convincing theory dates the period during which Jbmon pottery was used from about 10,000 years ago until the 2nd or 3rd century BC. Of the features common to Neolithic cultures all over the world—progress from chipped tools to polished tools, the manufacture of pottery, the beginnings of agriculture and pasturage, the development of weaving, and the erection of monuments using massive stones—the first two are prominent features of the Jōmon period, but the remaining three do not appear until the succeeding Yayoi period. The manufacture of pottery, however, was highly developed, and the work of Jbmon culture has a diversity and complexity of form and an exuberance of artistic decoration. It is customary to take changes in the type of pottery used as a basis for subdividing the age and to distinguish very early, early, middle, late, and very late periods. It must be remembered, however, that since Jōmon culture spread over the whole of the archipelago, it also developed regional differences, and this combination of both chronological and regional variations gives the evolution of Jbmon pottery a high degree of complexity.

The Japanese Neolithic

The pottery of the very early period includes many deep, urnlike vessels with tapered bases. In the early period, the vessels of eastern Japan become roughly cylindrical in shape, with flat bases, and the walls contain an admixture of vegetable fibre. The middle period saw rapid strides in pottery techniques; the pots produced during this time in the central mountain areas are generally considered to be the finest of the whole Jbmon era. The surface of these generally cylindrical vessels is covered with complex patterns of raised lines, and powerfully decorative projections rise from the rim to form handles. From the middle period onward there is increasing variety in the types of vessels, and a clear distinction developed between high-quality ware using elaborate techniques and simpler pots made for purely practical use. The amount of the latter increases steadily, preparing the way for the transition to Yayoi pottery.

Jōmon dwelling sites have been found in various parts of the country. They can be classified into two types: one, the pit-type dwelling, consisted of a shallow pit with a floor of trodden earth and a roof; the other was made by laying a circular or oval floor of clay or stones on the surface of the ground and covering it with a roof. Remains of such dwellings have been found in groups ranging from five or six to several dozen, apparently representing the size of human settlements at the time. Most of these settlements form a horseshoe shape, with a space in the centre that seems to have been used for communal purposes. Nothing certain is known, however, concerning social or political organization at this period. It can be deduced that each household was made up of several family members and that the settlement made up of such households was led by a headman or magician.

The men of the Jbmon period lived mainly by hunting and fishing and by gathering edible nuts and roots. The appearance of large settlements from the middle period onward has been interpreted by some scholars as implying the cultivation of certain types of crop—a hypothesis supported by the fact that the chipped stone axes of this period are not sharp but seem to have been used for digging soil. Weaving was still unknown, and clothes were probably made of skins or bark. Jewelry included bracelets made of seashells, earrings of stone or clay, and necklaces and hair ornaments of stone or bone and horn. From the latter part of the period, the custom also spread over the whole country of extracting or pointing certain teeth, probably performed as a rite marking the attainment of adulthood.

No especially elaborate rites of burial were evolved, and the dead were buried in a small pit dug near the dwelling. Sometimes the body was buried with its knees drawn up or with a stone clasped to its chest, a procedure that probably had some religious or magical significance. A large number of clay figurines have been found, many representing female forms that were probably magical objects associated with primitive fertility cults.

The Ainu

For years certain scholars have claimed that the men responsible for the Jōmon culture were not of Japanese stock but were ancestors of the Ainu, an aboriginal Caucasian people now found in northern Japan. Scientific investigation of the bones of Jbmon men carried out since the beginning of the 20th century has disproved this theory. The Jbmon people were a particular people who might be called proto-Japanese, and they were spread all over the country. Despite certain variations in character arising from differences in period or place, they seem to have constituted a single stock with more or less consistent characteristics. The present Japanese people were produced by an admixture of certain strains from the Asian continent and from the South Pacific, together with adaptations made in accordance with environmental changes. Linguistic evidence suggests that a people speaking a language belonging to the primitive Ural-Altaic family moved eastward across Siberia and entered Japan via Sakhalin and Hokkaidō. Nothing can yet be proved concerning their relationship with the people of the Pre-Ceramic period, but it cannot be asserted that they were entirely unrelated.

Yayoi culture (c. 250 BC to c. AD *250).* The new Yayoi culture that arose in Kyushu (the southernmost of the four Japanese islands), while the Jbmon culture was still undergoing development elsewhere, spread gradually eastward, overwhelming the Jiimon culture as it went, until it reached the northern districts of Honshu, the largest island of Japan. The name Yayoi derives from the name of the district in Tokyo where, in 1884, the unearthing of pottery of this type first drew the attention of scholars. Yayoi pottery was fired at higher temperatures than Jbmon pottery and was wheel turned. It is distinguished partly by this marked advance in technique and partly by an absence of the proliferating decoration that characterized Jiimon pottery. It developed, in short, as pottery for practical use. It is accompanied by metal objects and is associated with the wet cultivation of rice. Culturally, it represents a notable advance over the Jōmon period and is believed to have lasted for some five or six centuries, from the 3rd or 2nd century BC to the 2nd or 3rd century AD.

In China, the 3rd and 2nd centuries BC corresponded with the period of the unified empire under the Ch'in (221–206 BC) and Han (206 BC–AD 220) dynasties, which had already entered the Iron Age. In 108 BC, Han Wu Ti occupied the Korean peninsula and established Lo-lang and three other colonies. They provided a base for a strong influx of Chinese culture into Korea, which, in turn, spread to Japan. The fact that Yayoi culture had iron implements from the outset, and bronze implements somewhat later, probably indicates borrowings from Han culture. Iron objects rust easily, and comparatively few have been found, but they seem to have been widespread at the time. These include axes, knives, sickles and hoes, arrowheads, and swords. The bronze objects are also varied, including halberds, swords, spears, ***taku*** (small bell-shaped devotional objects from China), and mirrors. The halberds, swords, and spears seem not to have been used in Japan for the practical purposes for which they were evolved in China but to have been prized as precious objects.

Wet cultivation of rice

The wet cultivation of rice, possibly borrowed from southern China, was one of the most important features of Yayoi culture. The earliest Yayoi pottery and sites, discovered in northern Kyushu, have yielded marks of rice husks as well as carbonized grains of rice, which suggests that rice growing was carried on in Japan from the earliest days of the culture. Traces of paddy fields, their divisions marked with wooden piles, have been found close to sites of settlements in various districts, along with irrigation channels equipped with dams and underdrains, showing that techniques of making and maintaining paddy fields were quite advanced.

Generally speaking, the settlements of this period were built on lowlying alluvial land to facilitate the irrigation of the paddies, but at one stage they were built in the hills or on high ground instead. It is not clear whether this was dictated by the needs of defense or whether dry cultivation was being practiced. Much as in the Jbmon period, there were two types of dwelling, the pit type and the type built on the surface; but in addition to these, raised-floor structures now appeared that were used for storing grain.

With the acquisition of a knowledge of textiles, clothing made great strides compared with the Jbmon period. The cloth was woven on primitive looms using vegetable fibres.

The dead were buried in either large clay urns or heavy stone coffins. Both were common in northern Kyushu and neighbouring areas, and similar urns and coffins are also found in Korea, where they probably originated. The graves were usually marked by mounds of earth or circles of stones, but a special type employed a dolmen (a large slab of stone supported over the grave by a number of smaller stones). Since the erection of dolmens was widely practiced in Manchuria and Korea, these, too, are believed to be a sign of an influx of continental culture. Normally, graves occur in clusters, but occasionally one is found apart, surrounded by a ditch and with swords, beads, and mirrors buried along with the dead. Such special graves suggest that society was already divided into classes.

It is natural to suppose that these new cultural elements represent a migration to Japan from Korea or China. Yet it is also certain that the migration was not of an order to change the character of the men who had inhabited the islands from Jbmon times. Although Yayoi culture undoubtedly represents an admixture of new sanguineous elements, it seems likely that the chief strain of proto-Japanese found all over the country during the Jōmon period was not disrupted but was carried over into later ages. This point of view is supported by the accounts of the "men of Wo," found in the Chinese history ***Wei chih.***

Chinese chronicles. Japan first appears in Chinese chronicles under the name of “Wo” (in Japanese, Wa). The Han histories relate that "in the seas off Lo-lang lie the men of Wo, who are divided into more than 100 states, and who bring tribute at fixed intervals." Lo-lang was one of two Han colonies established in the Korean

The "state of Nu in Wo"

peninsula in 108 BC, and it is beyond doubt that the country of Wo was Japan. A Later Han (23–220) history records that in AD 57 the "state of Nu in Wo" sent emissaries to the Later Han court and that the Emperor gave them a gold seal. The "state of Nu," located around what is now Hakata Bay, in Kyushu, was one of more than 100 states into which Wo was divided. This account was confirmed by a gold seal, apparently the identical seal awarded by the Chinese emperor, unearthed on the island of Shikano, at the mouth of Hakata Bay, in 1748. Later, in the latter half of the 2nd century, there was civil war in the state of Wo; a woman called Pimiko (Himiko in Japanese) used her religious authority to pacify the land, and there came into being a union of more than 30 states, which opened communications with the Wei dynasty (AD 220–264) in China. Wei, too, sent emissaries to Wo, and friendly relations between the two sides continued during the first half of the 3rd century. The *Wei chih* contains a detailed account of the route from Lo-lang to the court of the Wo queen in "Yamatai." Scholars are divided as to whether Yamatai was located in northern Kyushu or in the Kinai district. If it was in northern Kyushu, then the union of states was a purely local government, unrelated to the Yamato court of later times, but if it was in the Kinai district, then it would be natural to see it as the ancestor of that court. This would suggest, in turn, that Japan had already achieved a considerable degree of political unification. But it seems most likely Yamatai was a local centre of power in Kyushu and that unification did not take place until a century later.

According to the *Wei chih*, the people of Wo had already reached a fairly high degree of civilization. Society had clearcut divisions of rank, and the people paid taxes. There were impressive raised-floor buildings. The various provinces held fairs where goods were bartered. Since there were exchanges of letters with Wo, it seems, too, that there were already some who could read and write.

THE ANCIENT PERIOD (*c.* AD 250–710)

Date of unification

The Yamato court and the unification of the nation. The question of how the unification of Japan was first achieved and of how the Yamato court, with the *tennō* ("emperor of heaven") at its centre, came into being has inspired many hypotheses, none of which has so far proved entirely convincing. Thanks to Chinese and Korean records, however, it is possible to get at least an approximate idea of the date when unification occurred. The relations that Yamatai had begun with Wei were continued with Chin (AD 265–317), the dynasty that replaced Wei; but following the dispatch of a mission in 266, all records of exchanges cease, and it is not until 147 years later, in 413 during the Eastern Chin dynasty (AD 317–419) in China, that the name of Wo again appears in Chinese documents. It is most likely that the blank period resulted from conditions within Japan that made exchanges with other countries impossible. This period probably saw the collapse of Yamatai and the birth pangs of the united nation that took its place.

It is possible to push the date of unification of the nation back a few decades earlier than 413 because a memorial erected in 414 commemorating the achievements of King Kwanggaet'o (Japanese Hotae) of Koguryd (a Korean state, 37 BC–AD 668), describing the fighting between Wo and Koguryb that took place on the Korean peninsula from the end of the 4th century into the beginning of the 5th century, makes special mention of a great army sent to the peninsula in 391 by Wo that succeeded in subjugating the kingdoms of Paekche, Kaya (or Karak), and Silla. Such military success presupposes a long period of preparation and the prior establishment of a Wo foothold on the peninsula. The 8th-century *Nihon shoki* ("Chronicles of Japan"), one of Japan's two oldest histories, mentions the dispatch of troops by Japan in 369. Displays of strength of this kind would hardly have been possible unless Japan were already unified, and the date of the unification of the country may therefore be set at the middle of the 4th century at the latest.

The rise and decline of the Yamato court. At the time of unification, Japan already seems to have been an extremely powerful nation, as attested by the fact that it took on Koguryb, whch dominated Korea, and established a base for its own power in southern Korea. Paekche, in the west of southern Korea, was a friendly state that paid tribute to Japan, while Kaya (Japanese, Mimana), at the southern extremity of the peninsula, was under direct Japanese jurisdiction. Tributes from these states lined the coffers of the Yamato court and encouraged a marked rise in standards of living. Weavers, smiths, and irrigation experts migrated to Japan from these areas, and the Chinese ideographic script came into Japan at this time, together with Confucian works written in that script.

The Korean rebellion

The Yamato court reached its peak in the eary 5th century and thereafter went into a rapid decline. The main reason was that the states of the Korean peninsula, as a result of shifts in international relationships, broke away from Japan, so that the latter was no longer able to rely on tribute from them. Japan therefore conceived the idea of borrowing the authority of the Chinese court in achieving the subjugation of the Korean kingdoms. Beginning in 421, it sent envoys to the Liu-Sung dynasty (420–479) to ask that the Japanese emperor be granted the title of generalissimo, with military control over the states of Korea. Such envoys were sent on a number of occasions during the Liu-Sung dynasty and continued until 502, during the Liang dynasty (502–557). These attempts proved futile and are themselves evidence of the decline of Japan's military power.

Japan's difficulties abroad were paralleled by an impasse in domestic affairs. The Yamato court was headed by a hereditary emperor, while its members were drawn from the group of powerful *muraji* (clan leader) families, which had been vassals of the emperor from the start, and another group of powerful families, the *omi* (chieftain), which had sworn allegiance during the process of national unification. The highest officers of government were the *ō-muraji* and the *6-omi*, the heads and representatives of those two groups. In time, however, some members of these families began to cool in their allegiance to the emperor or even to plot with the states of Korea. In addition, there were ceaseless struggles involving succession to the throne within the Imperial family itself. As a result, Mimana, Japan's domain in Korea, was captured in 562 by the kingdom of Silla, depriving Japan of a powerful foothold on the peninsula. By the end of the 6th century, Japan had reached a low point in both foreign and domestic affairs.

Introduction of Buddhism

During the declining years of the Yamato court, however, there was one event of the utmost cultural importance: the introduction of Buddhism from the Korean Paekche kingdom. The date of its introduction is traditionally set at either 538 or 552, but it seems likely that Buddhist beliefs had begun spreading among ordinary Japanese at a much earlier date. Buddhism at first was an object of wonder and admiration, a rare item of foreign culture symbolized by its beautiful statuary, its imposing religious paraphernalia, and its majestic temples. The Buddhism that first spread among the Japanese was almost certainly a simple reliance on the magical powers of the religion in seeking various benefits in the present world. True understanding of its doctrines did not come until the time of Shōtoku Taishi (Prince Shōtoku).

The period from the latter half of the 3rd century until the beginning of the 7th century is known to archaeologists as the age of the *Kofun* ("tumulus") culture, for burial mounds were then erected over a wide area, and their shapes and contents (the objects buried with the dead) give a good idea of the material aspects of the everyday life of the time. These include the well-known *haniwa* tomb sculptures.

The idealized government of Shōtoku Taishi. The Yamato court that fell into such desperate straits toward the end of the 6th century was to be resuscitated by efforts made within the Imperial family itself, efforts that in the course of a century reformed the government of the country and set it moving toward formation of a centralized state more suited to the new age. The movement was touched off by the theories of ideal government expounded by Prince Shbtoku, who, as regent for his aunt, the empress Suiko, took charge of the nation in these difficult times. Prince Shbtoku took the Buddhist spirit of peace and salvation for all beings as the ideal underlying his government. He made no move, even, to charge the known murderer of the previous emperor but worked to convince him gradually, through the ideas of Buddhism, of the wrong he had done.

The constitution of 604

The Prince's most striking achievement in the field of domestic government was his establishment of a system of 12 court ranks in 603 and the "Seventeen Article Constitution" in 604. The former, which made clear the relative stations of persons working at the court by giving them caps of different colours, aimed to encourage efficient use of persons of ability and give the court a proper organization and etiquette of its own. The constitution consists of 17 simple articles setting forth the ideals of the state and rules for human conduct. It distinguishes the ruler, his ministers, and the people as the three human elements making up the state and clearly lays down the duties and rights of each; it thus set the pattern of a centralized state presided over by a single ruler, and it provided a kind of basic law of the nation.

Shbtoku's chief achievement in foreign relations was the opening of relations with Sui dynasty (518–618) China. The exchanges between Japan and the Southern dynasties (418–589) in the 5th century had placed Japan in the position of a tributary state. Prince Shdtoku opened relations with Sui on an equal basis, and envoys were exchanged by the two countries. He also sent Japanese students to China to learn directly from Chinese culture, which had hitherto reached Japan via the states of Korea.

Shōtoku was a profound student of Buddhism who gave lectures on the scriptures and himself wrote commentaries. His commentary on the *Lotus Sūtra,* four volumes of which survive in the original draft written by the Prince himself, may be called the oldest written work of known authorship in Japan.

As Buddhism gained ground, imposing temples were built in the Chinese style. The astonishment aroused by these great buildings—often with more than one story and with massive tiled roofs—that were built where there had been only low, thatched houses may well be imagined. A new civilization descended on Japan almost overnight. Of the temples built at the time, all that has survived of most of them are the foundation stones, but the Hbryii-ji, founded in 607 at Ikaruga in present Nara Prefecture, still preserves its ancient wooden structures; indeed, it is the oldest wooden building in the world.

The Taika reforms. The death of Prince Shbtoku, in 629, prevented his ideals of government from bearing full fruit. The Soga family, regaining its former powers, exterminated Prince Shōtoku's son Yamashiro be in 643 and all his family. At the same time, however, the students whom Shbtoku had sent to China were returning to Japan with accounts of the power and efficiency of the T'ang dynasty (618–907), which had overthrown the Sui dynasty and unified China. These accounts impressed on educated men the need to reform the government, strengthen the power of the state, and take every step to prepare against possible pressure from outside.

Abolition of private land-ownership

In 645 Prince Nakano be and Nakatomi Kamatari engineered a coup d'etat within the palace, killing the Soga family and wiping out all forces opposed to the Imperial family. They then set about establishing a system of centralized government with the emperor as absolute monarch at its head. An edict issued in 646 abolished private ownership of land and men by the wealthy families. The land thus taken over by the state was to be parcelled out among all who had attained a certain age, with the right to cultivate, in exchange for which the tenants were to pay a fixed tax. Provisions were also made for a governmental system embracing a capital city and local administration and for defense and communications facilities. A system was also established whereby a kind of "complaints box" was installed at court to give people a chance to appeal directly to the emperor. The main outlines of the reforms were drawn up in about five years. They are given the name Taika reforms after the *nengō* ("year period")—the first such in Japanese history—that was given to the era at that time. In the countries of the Far East, era names are a symbol of an independent nation, a sign that the sovereign's authority is effective.

Not long after the Taika reforms, Japan became involved in a dispute that led it to send troops to Korea. Paekche, whose royal castle had been captured in 660 by the combined forces of T'ang (China) and Silla (another Korean kingdom), called on Japan for help. Japan, which had traditionally been friendly with Paekche, sent a large army; it was crushed, however, in 663, by a combined T'ang and Silla army at the mouth of the Pak River. Japan withdrew entirely and gave up any further intervention on the peninsula. The Japanese ruler of the time, the empress Saimei, went to northern Kyushu and directed operations personally, even though she was already 67 at the time. The Empress was succeeded by Prince Nakano be, who, as the emperor Tenchi, directed his attention to domestic affairs. He built fortifications in Kyushu to prepare for an expected T'ang and Silla invasion and amended the system established by the Taika reforms so as to make it more suitable to the practical needs of the state. His younger brother, Emperor Temmu, similarly devoted his energies to domestic government; he had the Taika reforms set forth in written codes, which comprised *ritsu-ryd* political structure.

The ritsu-ryō system. The *ritsu-ryd* can be divided between *ritsu,* the criminal code, and *ryō,* the administrative and civil codes. A similar system had long been in force in China, and the Japanese *ritsu-ryō* was an imitation of the *lü-ling* of T'ang China and incorporated some of its articles just as they stood. Where different local conditions called for amendment, however, amendments were made without hesitation; it is a good early example of the skill of the Japanese in importing foreign culture.

The Imperial government

The Japanese emperor, for example, was in some respects an absolute monarch who ruled over the whole country as the head of a bureaucracy in the same manner as the emperor of China. Yet at the same time he was also the traditional high priest who maintained peace for the land and people by paying tribute to the gods and sounding out their will. Thus the central government was headed by twin agencies—the Dajōkan (Council of State), which combined within its functions the various practical aspects of administration, and the Jingikan (Office of Deities), which was in charge of the worship of the gods. Prospective bureaucrats were required to study at a central college and to pass prescribed examinations; during their term of office their achievement was subjected to scrutiny once a year, and their rank and position were adjusted in accordance with the results. This was based on the highly developed bureaucratic system of China, yet the *ritsu-ryō* system was not too bound by its provisions to provide special favours for men of high rank and good family. This, too, was a compromise between the new principles of the *ritsu-ryd* system and the old spirit of respect for birth. The provinces were divided into three types of administrative division: the *kuni,* or *koku* (province), the *kōri,* or *gun* (county), and the *sato,* or *ri* (village), to be administered by officials known as *kokushi, gunji,* and *richō,* respectively, The posts of *kokushi* were filled by members of the central bureaucracy in turn, but the posts of *gunji* and *rich6* were filled by members of prominent local families.

The people were divided into two main classes, freemen and slaves. The slaves were the possession of the government, the aristocracy, and the shrines and temples; as such they were obliged to provide unlimited labour, but their total number accounted for less than one-tenth of the population. The majority of the free population were farmers. At the age of six, each male child was apportioned paddy fields that remained his to cultivate for life. A tax was levied on the produce of the paddies, and a head tax was levied on adult males. The paddy field tax was low (about 3 percent of the crop), but the head tax, payable in handicrafts such as silk and hemp, imposed a heavy burden. Moreover, the transport of the goods from the provinces to the capital was the responsibility of the taxed, which involved an enormous labour for those living in distant parts. Adult males were also obliged to give military service and to provide labour for public works at the command of the local *kokushi,* amounting to not more than 60 days per year. Since the government's finances depended on such tribute from the common people, whenever the latter found the burden too much and fled from their registered homes to avoid paying taxes the government felt the pinch immediately.

The lowest ranking freemen were the groups of smiths, tanners, and others engaged in manufacturing. They were mostly the descendants of immigrants who inherited their trades and paid their taxes in the form of manufactured goods or by working for fixed periods in the government workshops.

All land was in principle the property of the state. Most of the land was distributed equally among the people, but, apart from this, land of a certain annual yield was given to bureaucrats and other high-ranking persons as stipends and to Shinto shrines and Buddhist temples as sources of revenue. Land other than paddy fields was left to the individual to use as he pleased. There was a need to open up new paddy fields as a means of providing for a growing population,, but the *ritsu-ryd* system made inadequate provision for this process. In time, the government began to encourage the opening up of new land, and in 743 the system was changed to give permanent private possession of such land to the person who had first put it under cultivation. As a result, the aristocrats and the shrines and temples frantically set about putting land under cultivation in order to increase their own privately owned territories. The principle of public ownership of land provided for in the *ritsu-ryd* system began to crumble, and as it did so the whole system of government grew increasingly shaky.

Shrine and temple aristocracy

THE IMPERIAL STATE

Government and culture in the Nara period (AD 710–784). In 710 the Imperial capital was shifted from Fujiwara to Nara. For the next 75 years, with minor gaps, Nara was the seat of government, and the old custom of changing the capital with each successive emperor was finally discarded. During this period, the centralized government provided for under the *ritsu-ryd* structure worked well, but a still conspicuous feature is the brilliant flowering of culture, especially Buddhist culture. The leaders in its promotion were the emperor Shōmu and his consort, Kōmyō. Immediately on his accession, Shōmu, who from childhood had been given a thorough schooling as future emperor, showed an eager concern to promote the stable livelihood of the people. Convinced that the Buddhist faith was a means to ensuring both the happiness of the individual and peace for the country as a whole, he introduced strong doses of Buddhism into his government.

Official encouragement of Buddhism

One of the measures he took was the founding of the temples known as *kokubun-ji.* Each province was to build a monastery known as *kokubun-ji* and a nunnery known as *kokubun-niji,* each with a seven-story pagoda and each housing a statue of the Sakyamuni Buddha. Each monastery was to have 20 monks, each nunnery 10 nuns, whose constant task would be to recite the scriptures and offer up prayers for the welfare of the nation. Just as the temporal world had its *kokushi* (governors) in each province to attend to its administrative and juridical matters, so the spiritual world would have officially appointed monks and nuns, distributed evenly among the provinces, to attend to the spiritual needs of the people.

The second measure taken by Shōmu was the construction of the Tōdai-ji as *kokubun-ji* of the capital and the installation within it of a huge bronze figure of the Vairocana Buddha as supreme guardian deity of the nation. The casting of the Daibutsu (Great Buddha) was a tremendously difficult task. The Emperor, however, called on the people at large to contribute to the project, in however humble a way, and thereby partake of the grace of the Buddha. The great image that was produced as a result, though damaged in later ages, still stands in the Tddai-ji and is famous the world over as the Great Buddha of Nara.

The marriage of Buddhism and politics that was Shōmu's ideal was to cause trouble in the following era. The temples gradually amassed vast wealth, and the monks acquired high political positions and began to interfere in secular affairs. A movement to counter such abuses arose among the aristocracy, the leaders of the movement being the Fujiwara family, descendants of the Nakatomi Kamatari, who had played such an important role in the Taika reforms. Kamatari and his son Fuhito (later given the surname Fujiwara) had supervised compilation of the codes that comprised the *ritsu-ryb* system and had become prominent figures at court as a new type of bureaucrat-noble. The subsequent progress of the family's fortunes was not always smooth. In particular, the emphasis on Buddhism in government had obliged them to lie low. At the end of the 8th century, however, when it seemed to them that the evils of Buddhistic government were threatening the future of the nation, they set on the throne a new emperor, Kōnin, who had no leanings toward Buddhism. Kōnin's son Kammu, who was of a similar mind, shifted the capital to Heian (present Kyōto) to sever connections with the temples of Nara and re-established government in accordance with the *ritsu-ryb* system.

The cultural flowering centring on Buddhism was an outcome of lively exchanges with other nations. Four times within 70 years the government sent official missions to the court of T'ang, and each time they were accompanied by a large number of students who went to study in China. By this time T'ang had formed a great empire that controlled not only the central plains of China but parts of Mongolia and Siberia to the north and of Central Asia to the west.

Japanese culture, borrowing from the T'ang, whose capital, Ch'ang-an, was a great international city, thus showed in the Nara period marked international flavour. The ceremony of consecration of the Great Buddha of the Tōdai-ji, for example, was conducted by a Brahmin high priest born in India, while the music was played by musicians from all over the Far East. But despite this strongly international flavour, respect was also shown for traditional Japanese ways and outlooks. An outstanding example of this respect is the collection of Japanese verse known as *Manyō-shū* (c. 8th century AD), an anthology of 4,500 poems both ancient and contemporary. The poets range over all classes of society, from the emperor and members of the Imperial family through the aristocracy and the priesthood to farmers, soldiers, and prostitutes, and the scenery celebrated in the verse represents districts all over the country. The poems deal directly and powerfully with basic human themes, such as love between men and women or between parents and children, and are deeply imbued with the traditional spirit of Japan, scarcely influenced at all by Buddhist or Confucian ideas. The anthology had immense influence on all subsequent Japanese culture.

The *Manyō-shū* anthology

The compilation of Japan's two most ancient histories, the *Koji-ki* and *Nihon shoki,* also took place at the beginning of the 8th century. Both works are extremely important, for they draw on oral or written traditions handed down from much earlier times.

Changes in ritsu-ryō government. In 794, as noted above, the emperor Kammu shifted his capital to Heian, cut the ties between government and Buddhism, and revived government in accordance with the *ritsu-ryb.* Commanding that the provisions of the *ritsu-ryb* system be enforced, he also amended those articles that were no longer relevant to the age. Since it was difficult in practice to carry out the allocation of rice fields once every six years, this was amended to once in 12 years. A tighter watch was imposed on corruption among local officials. The original system of raising conscript troops was abolished, and troops were thenceforth selected from among the sons of local government officers and persons of rank. An alien tribe known as the Emishi in the northern districts of Honshu was brought under government control. Those Emishi who submitted to government forces were resettled throughout the empire and quickly assimilated to the existing population.

Buddhism was forbidden to interfere in affairs of state.

but as a religion it was encouraged to fulfill its proper functions. Two brilliant monks, Saichō and Kūkai, were sent to China to study. Each of them, on his return to Japan, established a new sect of Japanese Buddhism: the Tendai sect, founded by Saichō, and the Shingon sect, founded by Kūkai. In the Nara period, Buddhism had been no more than a transplantation of the Buddhism of T'ang China, but the two new sects, though basically derived from China, were reworked in a characteristically Japanese fashion. As headquarters of their new sects, Saichō and Kūkai founded, respectively, the Enryaku-ji (also known as the Hieisan-ji) on Mt. Hiei and the Kongōbu-ji on Mt. Kōya *(ji,* "temple"). The two sects were thenceforth to form the twin mainstreams of Japanese Buddhism.

Decline of *ritsu-ryd* government

After Kammu, successive emperors carried on his policies. and society enjoyed some 150 years of peace. The formal aspects of government, at least, were carefully observed, and the supplementing of the legal codes, the compilation of histories, and the minting of coins all took place frequently in accordance with precedent. The social reality, however, became increasingly chaotic, so that form and actuality were soon travelling along quite different courses. The very foundations of *ritsu-ryb* government had begun to crumble because of the difficulty of finding enough rice fields to distribute to the people and the decline in government revenue resulting from the impoverishment of the masses.

A good example of the split between form and reality is the fact that while, on the surface, appointments to official posts at court were made just as they had always been, real power shifted to other posts that were newly created as the occasion demanded. Typical of such new posts were those of *kurōdo,* a kind of secretary and archivist to the emperor, and the *kebiislzi,* who had total control over the police and the judicature. The two supreme examples of such posts were those of *sesshō* (regent) and *kampaku* (chief councillor). The original role of the *sesshd* was to attend to affairs of state during the minority of the emperor, while the *kampaku's* role was to attend to state matters for the emperor even after he had come of age. Neither post had been foreseen by the *ritsu-ryō* system, which was rooted in the principle of direct rule by the emperor.

The rule of Imperial succession

Around the middle of the 9th century, however, when the emperor Seiwa ascended to the throne at the age of nine, his maternal grandfather, Fujiwara Yoshifusa, became *sesshō.* Yoshifusa's son Mototsune became *sessho* in the time of the emperor Yōzei, then in the reign of the emperor Uda he became *kampaku* in turn. It thus became the established custom that a member of the Fujiwara family should serve as *sesshō* and *kampaku.* In order to become *sesshō* or *kampaku,* it was necessary that the person concerned should marry his daughter into the Imperial family, then establish the resulting offspring as emperor. In other words, an important qualification was that one should be the emperor's maternal grandfather or father-in-law. As a result of this complex system, there were constant struggles at court involving the expulsion of members of other families by the Fujiwaras or wrangling among the Fujiwaras themselves.

One of the most celebrated affairs involving the expulsion of a member of another family by the Fujiwaras was the removal of Sugawara Michizane from his post as minister and his exile to Kyushu. Born into a family of scholars, Michizane was himself an outstanding scholar whose ability in writing Chinese verse and prose was said to rival that of the Chinese themselves. Recognizing his talent, the emperor Uda singled him out for an attempt to break the authority of the Fujiwaras. As part of his plan, he simultaneously appointed Fujiwara Tokihira as his minister of the left and Sugawara Michizane as minister of the right. The Fujiwaras objected strenuously. Tokihira falsely reported to the Emperor that Michizane was plotting treason; the matter was taken up officially, and Michizane was stripped of his ministerial post and his family was sent into exile.

At the end of the 9th century, at Michizane's suggestion, Japan cut off formal relations with T'ang China, perhaps because of the expense involved in sending regular envoys. The practical result, however, was the stimulation of a more purely Japanese cultural tradition.

The culture of the 9th century was a continuation of that of the 8th, insofar as its foundations were predominantly Chinese. The writing of Chinese prose and verse was popular among scholars, and great respect for Chinese customs was shown in the daily lives of the aristocracy. Many Buddhist monks went to China to bring back as yet unknown scriptures and iconographic pictures. Buddhist sculpture and paintings produced in Japan were done in the T'ang style. Japanese touches were, of course, gradually added to these basically T'ang styles, and a new culture was slowly coming into being, but it was not until the 10th century and later that this tendency became a strong current.

Aristocratic government at its peak. From the 10th century and through the 11th, successive generations of the Fujiwara family continued to control the nation's government by monopolizing the posts of *sesshō* and *kampaku,* and the wealth that poured into their coffers enabled them to lead lives of the greatest brilliance. The high-water mark was reached in the time of Fujiwara Michinaga. Four of his daughters became consorts of four successive emperors, and three of their sons became, respectively, the emperors Ichijō II, Suzaku II, and Reizei II. Government during this period was based mostly on precedent, and the court had become no more than a centre for ceremonies. Court ministers were content to perform prescribed rites on prescribed days and were utterly unfitted to deal with any sudden social crisis that might confront them.

The growth of private estates

The *ritsu-ryd* system of public ownership of land and men survived in name alone; land passed into private hands, and men became private citizens. Typical of the new privately owned lands were the *shōen* ("manors"), which developed on the basis of rice fields under cultivation since the adoption of the *ritsu-ryd* system. Since the government encouraged the opening up of new land during the Nara period, the temples and aristocrats with resources at their disposal hastened to develop new areas, and vast private lands accrued to them. The owners of the new lands used one pretext or another to obtain special exemption from taxes, so that the *shōen* gradually became nontaxpaying estates. The increase in such *shden* thus came to pose a serious threat to the government, which accordingly issued frequent edicts intended to check the formation of new estates. This merely served, however, to establish more firmly the position of those already existing and failed to halt the tendency for such land to increase. Since the owners of the *shden* were the same aristocracy and high officials that made up the government, it was extremely difficult to change the situation.

Although the aristocracy and temples around the capital enjoyed exemption from taxes on their *shden,* the same privileges were not available to powerful families in the provinces. These, accordingly, presented their *shden* to members of the Imperial family or the aristocracy, concluding agreements with them that the latter should become owners in name while the former themselves retained rights as actual administrators of the property. Thanks to such agreements the estates of the aristocracy went on increasing steadily and their incomes swelled proportionately. The *shden* of the Fujiwara family in particular reached such vast proportions that it was said that among them they owned the whole country.

While the aristocracy was leading a life of luxury on the proceeds from its estates, the provinces were witnessing the first stirrings of a new power in the land: the warrior, or samurai, class. Younger members of the Imperial family and lower ranking aristocrats who were dissatisfied with the Fujiwara monopoly of high posts would take up posts as local officials in the provinces, where they settled permanently, acquired lands of their own, and established their own power. In order to protect their territories, they began to press the local inhabitants into service and to give them arms, thus building up armed forces of their own. This meant that wherever, in addition, there happened to emerge men of true martial ability, a situation developed where the slightest incident was liable to provoke armed conflict. The risings of Taira Masakado (died 940) in the Kantō district and of Fujiwara Sumitomo (died 941) in western Japan had an enormous effect in lowering the government's prestige and encouraging the desolation of the provinces.

The development of the kana syllabary

The 10th century witnessed the development of a truly Japanese culture, one of the most important contributing factors being the emergence of an indigenous script, the kana syllabary. Until then, Japan had no writing of its own; Chinese ideographs were used partly for their sense and partly for their pronunciation in order to represent the Japanese language, which is entirely different from Chinese. In the natural course of events, however, the educated men and women of the day gradually evolved a system of writing that used a purely phonetic, syllabic script formed by abbreviating a certain number of the Chinese characters. This script, known as kana, made it possible to write the national language with complete freedom, and its invention was an epoch-making event in the history of the expression of ideas in Japan. Thanks to the kana, a great amount of verse and prose in Japanese was to be produced.

Particularly noteworthy in this respect were the daughters of the Fujiwaras, who, under the aristocratic government of the day, became consorts of successive emperors and surrounded themselves with talented women who vied with each other in learning and the ability to produce fine writing. The kana script provided such women with an opportunity to create works of literature. Among such works, the *Genji monogatavi* (*The Tale of* **Gengi)**, a novel by Murasaki Shikibu; and the *Makura-no-sōshi* ("Pillow Book"), a collection of vivid scenes and incidents of court life by Sei Shōnagon, a lady in waiting to the empress Sadako, are masterpieces that hold a place in world, and not merely Japanese, literature. The waka, the native Japanese verse form, was an indispensable part of the daily lives of the aristocracy, and proficiency in verse-making was counted an essential accomplishment for an intellectual. Such circumstances led to the compilation in 905 of the *Kokin-shu,* the first of a series of anthologies of verse made at Imperial command.

The same trend toward the development of purely Japanese qualities became strongly marked in Buddhism also. Both the Tendai and Shingon sects produced a succession of gifted monks and continued, as sects, to flourish. But, being closely connected with the court and aristocracy, they tended to pursue worldly wealth and riches at the expense of purely religious goals, and it was left to the Jbdo ("Pure Land) sect of Buddhism to preach a religion that sought to arouse a desire for salvation in ordinary men. This "Pure Land" Buddhism, which expounds the glories of the paradise of Amida (Amitabha, or Buddha of Infinite Light)—the world after death—and urges all men to renounce the defilements of the present world for the sake of rebirth in that paradise, seems to have offered an ideal hope of salvation for men in the midst of the collapse of the old order. It was a very approachable religion in that it eschewed difficult theories and ascetic practices, teaching that in order to achieve rebirth it was only necessary to invoke the name of Amida and dwell on the marks of his divinity. This same teaching also inspired artists to produce an astonishing number of representations of Amida in both sculpture and painting. The mildness of his countenance and the softly curving folds of his robe contrasted strongly with the grotesque Buddhist sculpture in the preceding age, and represented a much more truly Japanese taste.

The signs of the growing independence of Japanese culture, apparent in every field, were an indication that by now, two centuries after the first busy ingestion of continental culture, the process of absorption was nearing completion.

Government by cloistered emperors (insei) and the rise of the samurai. The powerful authority wielded by the Fujiwaras as *sesshō* and *kampaku* was maintained by their blood relationship on the maternal side to successive emperors; once such a relationship disappeared, their power was bound to weaken. This, in fact, is what happened; the emperor Sanjb II ascended the throne even though he was not born of a daughter of the Fujiwaras, while Michinaga's sons Yorimichi and Norimichi both gave their daughters to be Imperial consorts without obtaining the desired birth of an Imperial prince. As a result of these and other circumstances, the latter half of the 11th century saw real political power pass from the *sesshō* and *kampaku* to the "cloistered emperors"; in other words, to emperors who had-already abdicated and taken Buddhist vows, thus nominally renouncing the world, yet who wielded a very real power behind the scenes. This system, known as *insei* ("cloistered rule"), was perhaps a more natural arrangement insofar as it represented a shift from government by matrilinear relatives of the emperor to government by patrilinear relatives, yet in continuing to treat the actually reigning emperor as a pure figurehead it was no whit better than the old *sesshō–kampaku* system. The cloistered emperor system continued for a long period, the retired emperors Shirakawa, Toba I, and Shirakawa II holding absolute behind-the-scenes power for 43 years, 28 years, and 34 years, respectively.

The power of retired emperors

The system of government by cloistered emperors was inspired by no particular ideal and conformed to no particular rules. The one common feature of each reign was that the emperor became a Buddhist priest and governed in a way that theoretically respected the teachings of Buddhism. In practice, this "Buddhism" was more preoccupied with construction of ostentatious temples than with true belief. Other signs of the same trend were the frequent journeys of retired emperors to worship at distant temples and their edicts strictly prohibiting the killing of living creatures; it did not concern them that such edicts deprived a large number of their subjects of their occupations.

The nominal respect for Buddhism spurred on the secularization of the religion. Properly speaking, the world of Buddhism should have been one in which factionalism could have no part, a world in which nothing counted but wisdom, virtue, and experience. But at this time large numbers of aristocrats were taking holy vows and going to live in the temples, which thus became centres of factionalism and intrigue. Most of the higher positions in the religious world were occupied by members of the Imperial family and former aristocrats. This effectively closed advancement to commoners, and the lower ranking monks in the temples harboured a grudge against their superiors on this account. Whenever some particularly serious grievance arose, they would march in a body on the capital and try to force acceptance of their demands by a direct appeal to the court. Some idea of the nuisance they constituted can be gained from the fact that even the retired emperor Shirakawa—the most powerful of the cloistered emperors—ranked them with the waters of the Kamo River and the dice in games of chance as one of three superhuman forces that he was powerless to control. Nor did the monks hesitate to resort to armed force; it was an age in which a priesthood ostensibly committed to compassion and respect for life in all its forms could openly bear arms and engage in slaughter.

The rise of the warrior class

Another feature of the age was the rise of the warrior class. With the development of government by the cloistered emperors, the more powerful of the samurai, who, as noted above, first established their power in the provinces, gradually gathered in or near the capital, where they acted as military police. Associating with the court nobles and aristocracy, they gradually established a foothold at court. Outstanding among these samurai were the Minamoto family, descendants of the emperor Seiwa, and the Taira family, descendants of the emperor Kammu. The Taira had at first settled in the Kantō district, where they extended their influence over a wide area; but they had suffered a setback in the Taira Masakado rising and had finally lost their hold on the Kantb district as the result of another, later rising by Masakado's descendant Tadatsune. The Minamotos, favourites of the Fujiwaras, had been a prominent family in the capital from the start, but their fame as a warrior clan was greatly heightened in the mid-11th century when after 12 years of hard fighting they quelled a rising by the Abe family in the Tohoku district. Minamoto Yoshiie, who played an important part in the fighting, became the nation's most celebrated warrior, and many powerful clans made voluntary vows of allegiance to him and presented him with land in return for his protection. Yoshiie, however, had no son to match him in military prowess, and the Tairas took advantage of this relative decline to advance their own fortunes again. Taking advantage of the *insei* system as a means to their own political advancement, they curried favour with the retired emperors. Taira Masamori and his son Tadamori served as governors in provinces in western Japan, building up their own power in the area, and aided the retired emperors' temple-building programs by erecting a large number of new temples and pagodas. Tadamori also tried his hand at trade with Sung dynasty China as a means of amassing wealth. In such ways, the social position of the Tairas rose steadily, so that Tadamori's son Kiyomori could take his place alongside the aristocracy.

The Heiji War

Discord between retired emperors and reigning emperors combined with internal differences within the Fujiwara family to split the Imperial family and nobility into two parties, which enlisted the Minamoto family and the Taira family, respectively, on their own sides. The two sides eventually clashed openly in Kyōto in what is known as the Hbgen War (July 1156). The war was on a small scale, the outcome determined by a single night's fighting, yet it was highly significant as showing that the power of the samurai was sufficient to sway the nation's government. In the Heiji War (1159) that followed, the Minamotos were thoroughly defeated, and Taira Kiyomori emerged as the chief power in the land. Although he was a samurai by birth, Kiyomori shared certain of the Fujiwaras' aristocratic tendencies, and the 20-odd years of Taira rule that followed had a special character of their own. He himself became grand minister of state (*dajō-daijn*) at the court, and more than 50 other official posts were filled by members of his family. His daughter became the consort of the emperor Takakura, and the prince born of the union ascended to the throne in his infancy—a return to government by matrilinear relatives of the emperor. Kiyomori's rule also had its more drastic, soldierlike aspects; thus in a single move he swept 42 court officials from their posts and into exile, and he razed to the ground troublesome temples such as the Tōdai-ji and Kōfuku-ji. His repairing of the Inland Sea route and his encouragement of trade with Sung China were measures that would never have occurred to a Fujiwara government.

While the Tairas thrived in the capital, the descendants of the original Minamotos were quietly building up their strength in the provinces. Finally Yoritomo, a descendant in the direct line of the Minamoto family, who grew up in exile at Izu, rallied the Minamotos and sent his younger brothers Yoshitsune and Noriyori to attack Kybto. The final rout of the fleeing Taira forces on the sea off the island of Shikoku put a more or less decisive end to the swing of fortune between Minamoto and Taira.

It also marked an important turning point in Japanese history, since Yoritomo's establishment of a military government, or shogunate, in Kamakura may be seen as the beginning of rule by a samurai class backed up by a feudal system and the end of the ancient monarchical period in which power was in the hands of the court and the aristocracy. The shogunate (or bakufu, from the headquarters of the government) was to hold effective political control in Japan until the restoration of Imperial power in 1867. (T.Sa.)

THE KAMAKURA BAKUFU (1192–1333)

The establishment of military government. The establishment of the shogunate by Minamoto Yoritomo at the end of the 12th century marks the beginning of a new and independent government by the warrior class opposed to the political authority of the civil aristocracy. Yoritomo established his supreme headquarters (bakufu) in Kamakura and entrusted the suppression of the powerful Taira family to his younger brothers Noriyori and Yoshitsune. Meanwhile, he gathered a following in eastern Japan as a foundation for a new military government. As a first step he set up in 1180 the Samurai-dokoro (Board of Retainers), a disciplinary board to control his military vassals. General administration was handled by a secretariat opened four years later and known as the Kumonjo (later renamed the Mandokoro). In addition, a judicial board, the Monchiijo, was set up to handle lawsuits and appeals. Under these institutions, the organization of the bakufu gradually took shape.

***Bakufu* institutions**

In 1185, after the destruction of the Taira family, Yoritomo appointed military governors (shugo) in all the provinces and military stewards (jitd) in both public and private landed estates. It was the job of the shugo to recruit metropolitan guards and keep strict control over subversives and criminals. The jitd collected taxes, supervised the management of landed estates, and maintained public order.

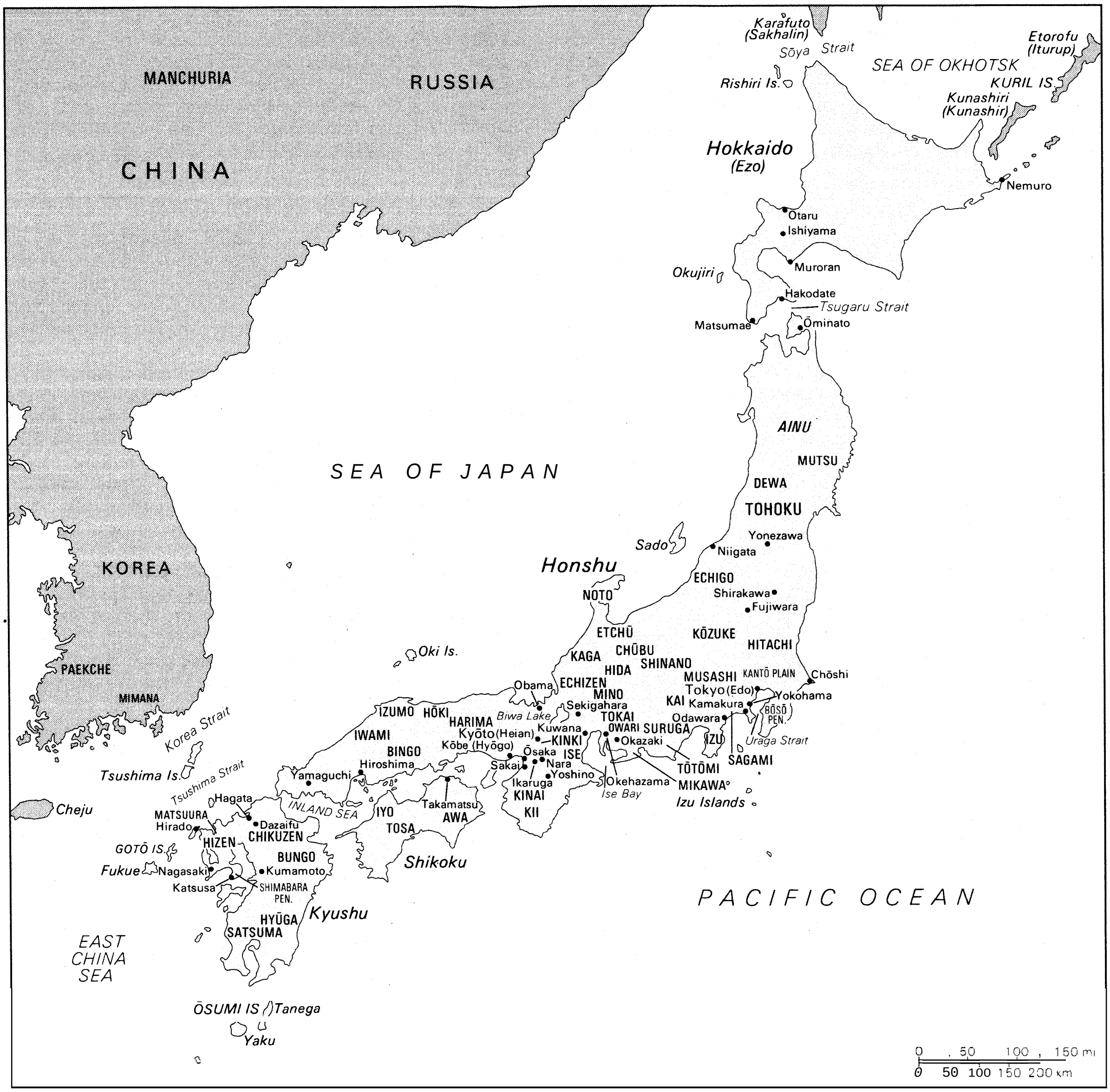
In 1189 Yoritomo finally destroyed the great Fujiwara family of Mutsu Province, which had sheltered his rebellious brother Yoshitsune. Three years later Yoritomo went to Kyōto and was appointed shogun, an abbreviation of seii *taishōgun*) ("barbarian-quelling generalissimo"), the highest honour that could be accorded a warrior. At first the chief base of the *bakufu* lay in the landed estates seized from the Taira family and in the limited administrative revenues from public estates in provinces granted to Yoritomo by the Imperial court. But later the *bakufu* was able to expand its influence over those public estates that were still controlled by the civil provincial governors, as well as the private estates of the civil aristocracy and the temples and shrines.

The regency government. After the death of Yoritomo in 1199, real power in the bakufu passed into the hands of the Hbjb family, from which Yoritomo's wife, Masako, had come. In 1203 Hōjō Tokimasa, Masako's father, assumed the position of regent (shikken) for the shogun, an office that was held until 1333 by nine successive members of the Hbjb family. Taking advantage of disputes among Yoritomo's generals, the Hōjō overthrew their rivals, and after three generations the direct line of descent from Yoritomo had become extinct. Though wielding actual power, the Hōjō family was of low social rank, and its leaders did not aspire to become shoguns themselves. Kujō Yoritsune, a distant relative of Yoritomo, was appointed shogun, while Tokimasa's son Hōjō Yoshitoki (shikken 1205–24) handled most government business. Thereafter, the appointment and dismissal of the shogun followed the wishes of the Hōjō family.

The Jōkyū War

This increasing political power of the military led to a conflict with the aristocracy. Hence, the emperor Toba II, seeing in the demise of the Minamoto family a good opportunity to restore his political power, in 1221 issued a mandate to the country for the overthrow of Yoshitoki, but few warriors responded to his call. A bakufu army occupied Kybto and Toba II was arrested and banished to the island of Oki in an incident known as the Jōkyū War, after the era name Jōkyū (1219–22). The bakufu now set up its headquarters in Kyōto to supervise the Imperial court and to control the legal and administrative business of the western provinces. The estates of the civil aristocrats and warriors who had joined Toba II were confiscated and distributed as rewards among the shogun's vassals. The political power of the bakufu now extended over the whole country.

Meanwhile, the regent Hōjō Yasutoki, to strengthen the base of his political power, reorganized the council of leading retainers into an advisory council known as the Hyōjō-shū. In 1232 the council drew up a legal code known as the Jdei Shikimoku (Jdei Formulary). Its 51 articles set down in writing for the first time the legal precedents of the bakufu. Its purpose was simpler than that of the ritsu-ryd, the old legal and political system of the Heian and Nara civil aristocracy. In essence it was a body of pragmatic law laid down for the proper conduct of the warrior way of life. Distinctive features of the formulary included a strong emphasis upon the lord–vassal relationship and paternal power and a recognition of female inheritance of land; its sway was gradually



Important Japanese historical sites.

extended over the whole country. In 1249 the regent Hōjō Tokiyori also set up a supreme court, the Hikitsuke-shū, to secure greater impartiality and promptness in legal decisions.

The Mongol invasions. The establishment of the regency government coincided with the rise of the Mongols under Genghis Khan in Central Asia. In the space of barely half a century they had established an empire extending from the Korean peninsula in the east and as far west as Russia and Poland. In 1260 Genghis Khan's successor, Kublai, became Great Khan in China and fixed his capital at present-day Peking. In 1271 Kublai adopted the dynastic title of Yüan; shortly thereafter the Mongols began preparations for an invasion of Japan. In the autumn of 1274 a Mongol and Korean army of 25,000 men set out from present-day South Korea. On landing in Kyūshū it occupied the Matsura district of Hizen Province (part of present-day Sago Prefecture) and advanced to Chikuzen. The *bakufu* appointed Shōni Sukeyoshi as military commander, and the Kyūshū military vassals were mobilized for defense. A Mongol army landed in Hakata Bay, forcing the Japanese defenders to retreat to Dazaifu; but a typhoon suddenly arose, destroying more than 200 ships of the invaders, and the survivors returned to South Korea.

The *bakufu* took measures against a renewed invasion. Coastal defenses were strengthened, and a stone wall was constructed extending for several miles around Hakata Bay to thwart the powerful Mongol cavalry. Apportioned among the Kyūshū military vassals, these public works took five years to complete and required considerable expenditure. Meanwhile, the Mongols made plans for a second expedition. In 1281 two separate armies were arrayed: an eastern army consisting of about 40,000 Mongol, North Chinese, and Korean troops set out from South Korea; and a second army of about 100,000 South China troops under the command of the Chinese general Han Bun-ko. The two armies met at Hirado and in a combined assault took Hakata Bay and pressed on to Dazaifu. But again a fierce typhoon destroyed nearly all of the invading fleet, forcing Han Bun-ko to retreat precipitately. The remnants of the invading army were captured by the Japanese; it is said that of 140,000 invaders, fewer than one in five made good their escape. The defeat of the Mongol invasions was of crucial importance in Japanese history. The military expenditure undermined the economic stability of the Kamakura government and led to the insolvency of many of the military vassals. It led to another prolonged period of isolation from China that was to last until the 14th century. Moreover the victory gave a great impetus to a feeling of national pride, and the *kamikaze,* or "divine wind," that destroyed the invading hosts gave the Japanese the belief that they were a divinely protected people.

Samurai groups and farming villages. The Japanese feudal system began to take shape under the Kamakura bakufu. At its inception in the Kamakura (1192–1333) and Muromachi (1338–1573) periods, Japanese feudalism presents a different appearance from that of later eras. In the earlier period feudal warrior landlords lived in farming villages and carried on agriculture themselves, while the central civil aristocracy and the temples and shrines held huge public and private estates in various provinces and wielded power comparable to that of the bakufu. These estates were, in reality, managed by influential resident landlords who had become warriors. They were often the original developers of their districts, becoming officials of the provincial government, agents of the private estates, and military stewards appointed by the bakufu. As leaders of a large number of villagers, they laboured to develop the rice fields and irrigation works in the areas under their jurisdiction, and they and other influential landlords constructed spacious homes for themselves in the villages and hamlets where they lived.

Among these were some who were military vassals of the shogun and others who were connected to the aristocracy or the temples and shrines. The military vassals owed their loyalty to the shogun, for whom they performed public services such as guard duty in Kyōto, where the emperor lived with his court. In return, the shogun not only guaranteed these men the traditional landholdings but rewarded them with new lands. This connection between lord and vassal, on which grants of landownership or management were based, gave Japanese society a distinctly feudal character.

The samurai served on the battlefield and in times of peace engaged in hunting and training in the military arts, nourishing a lugged and practical character. The *kyūba*-no-michi ("the way of the bow and horse"), the samurai ideal of chivalry, grew out of this daily training and the experience of actual warfare. Pride of family name was especially valued, and loyal service to one's overlord became the fundamental morality. This was the origin of the more highly developed code of Bushidd, or Code of Warriors, of later ages.

The status of women in the samurai families was comparatively high; they were allowed to inherit a portion of the estates, a practice that gradually came to be restricted.

Economic changes in the Karnakura period

After the middle of the Kamakura period, the farming villages in which the warriors resided underwent changes as agricultural practices advanced, and other aspects of society were changing as well. Artisans were usually attached to the proprietors of the private estates and progressively became more specialized along with the growth of consumer demand. Centres for metal casting and metalworking, paper manufacture, and other skills appeared in the localities. The exchange of agricultural products, manufactured goods, and other products thrived; local markets, held on three fixed days a month, became common. Copper coins from Sung (960–1279) China circulated in these markets, while itinerant merchants increased their activity. Bills of exchange were also used for payments to distant localities. In the large ports, specialized wholesale merchants appeared who were called *to*-imaru. They served as contractors who stored, transported, and sold goods. Further, it was common for many merchants and artisans to form guilds, known as *zu*, organized under the temples, shrines, or civil aristocrats. Such guilds, which resembled European trade guilds in the Middle Ages, gained special monopoly privileges and exemptions from customs duties.

Introduction of Zen sects

Kamakura culture: the new Buddhism and its influence. During the Kamakura period the newly arisen samurai class came to dominate the ancient civil aristocracy, which continued to maintain the classical culture. Vigorous overseas trade fostered the transmission of Zen Buddhism (in Chinese, Ch'an) and Neo-Confucianism from Sung China. Chinese influences were seen in monochrome painting style (sumi-e), forms of architecture, certain skills in pottery manufacture, and the custom of tea drinking—all of which assisted the formation of the samurai culture and exerted an enormous influence on everyday life.

In matters of religion, the great social changes that took place between the end of the Heian period (794–857) and the early Kamakura period caused the people to demand a simple standard of faith, in place of the complicated teachings and ceremonies of the ancient Buddhism. The warriors of the farming villages, in particular, demanded a religion that would suit their personal experience. Several new Buddhist sects sprang up which eschewed difficult ascetic practices and recondite scholarship. Among these may be included the Jōdo, or Pure Land sect, and its offshoots, the Ikkō and Rinzai sects (see BUDDHISM). The Zen sect sought to open the way to insight by self-effort; hence, it met with a ready response, satisfying the demands of the samurai. At the same time, scholarship and the arts were still deeply linked with esoteric Buddhism, which was a vigorous influence even in Shinto circles.

In scholarly and literary circles, the civil aristocrats of Kyoto confined themselves to the annotation and interpretation of the ancient classics and to the study of precedents and ceremonies. But at the beginning of the Kamakura period, a brilliant circle of waka (poems of 31 syllables) poets gathered around the retired emperor Toba II, and an Imperial selection of poems entitled the *Shin-kokin waka-shū* was compiled. The waka of this period is characterized by the term *yūgen*, which may be described as a mood both peaceful and profound.

Impact of military on the arts

Just before the Jōkyū War the monk Jien (posthumous name Jichin) completed his Gukan-shd ("Jottings of a Fool"). This is the first work of historical philosophy in Japan, and it provides a comprehensive picture of the rise and fall of political powers from the Buddhist viewpoint. Meanwhile, with the rise of the samurai style of life, warriors with a love of scholarship and a delight in waka poetry appeared. One was Hōjō Sanetoki, who collected Japanese and Chinese books and founded a famous library, the Kanazawa Bunko, in the Shōmyō-ji (at what is now Kanazawa). Military epics became popular. The most famous that has come down to us is the anonymously written Heike monogatari ("Tale of Heike"). They were recited throughout the country by Buddhist troubadours called biwa hdshi. After the middle Kamakura period, as Buddhist pessimism grew fainter, various kinds of instruction manuals and family injunctions were composed, while collections of essays such as Yoshida Kenkō's *Tsurezure-gusa* also made their appearance. The new nationalistic fervour found expression in Koken Shiren's *Genkō shakusho* (1332), a 30-volume history of Buddhism in Japan.

In the visual arts the carving of wooden images of famous monks flourished, and, after the middle of the Kamakura period, Chinese styles of the Sung dynasty also began to enter Kamakura wood carving. In painting, in addition to Buddhist themes, picture scrolls (*emaki*-mono) became popular, taking as their themes the history of temples and shrines, the biographies of founders of the sects, and military epics.

Decline of Kamakura society. The later Kamakura period witnessed the breakdown of family solidarity among the samurai (see below).

During the troubled state of society at the end of the Kamakura period, feudal landlords, out of the necessity of defending their own lands, seem to have devoted themselves chiefly to the military arts and to have entrusted the running of agriculture to their household dependents. Moreover, lands were no longer divided among younger sons but were now kept entirely in the hands of the eldest son. Power thus became concentrated in the head of the household, to whom other family members were subordinated in a lord–vassal relationship.

At the same time, major economic changes began to undermine the position of the bakufu vassals. Yet, despite the social crises among the landholders, trade was flourishing. Coins came into increasing circulation, and city styles of living began to be imitated in the country. But the landowners were often unable to meet their expenditures from the income of their limited holdings. Therefore, they borrowed money at high rates of interest from rich moneylenders, and many were forced to surrender their holdings when unable to repay their loans. Thus the gap between rich and poor became marked among the shogun's vassals. In particular, the local military governors, who had the right to raise troops, progressively gained control over the resident landlords, establishing a lord–vassal relationship with them. Moreover, deputies sent out by the heads of families to oversee their distant landholdings often broke with the main family and became vassals of the military governors. This powerful new class of local magnates was called daimyo (domain lords) and soon began to challenge the authority of the Hōjō regents in the *bakufu*.

The Ashikaga, Sasaki, Shōni, and Shimazu families were among the most powerful of the new class. The *bakufu* began to totter, shaken by the disputes between the Hōjō family and the rival military governors. When the And6 family raised a revolt in Mutsu Province at the end of the Kamakura period, the government found it difficult to suppress, partly because of the remoteness of the site of the uprising. In addition, regional unions of small landlords developed in the *kinai* (the five home provinces centred around Kyoto).

Taking advantage of the accumulating weaknesses of the *bakufu*, a movement arose among the civil aristocracy to regain political power from the military. The occasion was provided by the question of the Imperial succession. In the mid-13th century there emerged two competing lines for the succession—the senior line centred on the Jimyō-in monastery in the north and the junior line centred on the Daikaku-ji monastery in the south. In the last half of the century, each side sought to win the support of the *bakufu*. In 1317 the *bakufu* proposed that the two lines serve by turns as emperor, but the dispute did not cease. Finally, in 1318 Prince Takanaru, aged 31, acceded to the throne as Emperor Daigo II.

Dispute over the succession

THE MUROMACHI (OR ASHIKAGA) SHOGUNATE (1338-1573)

The "Kemmu restoration" and the dual dynasties. On the accession of Daigo 11, the retired emperor Uda 11 broke the long-established custom and dissolved the "cloistered" Imperial government. As a result, the entire authority of the Imperial government was concentrated in the hands of a single emperor, Daigo II. A party of young reforming court nobles gathered around the Emperor, who revived the Kirokusho (Records Office) and strove to renovate the government. But to realize his ideal of a true Imperial restoration, it was necessary for Daigo 11 to rid himself of the interference of the *bakufu*. His plans for its overthrow were discovered, however, and he was arrested and exiled to Oki Island. But in the *kinai* area, local leaders, supported by militant Buddhist monks, raised an army to overthrow the *bakufu*. The Imperial forces were led by Prince Morinaga and Kusunoki Masashige, but the decisive victory was brought about by the powerful Kantō families of Ashikaga Takauji and Nitta Yoshisada, discontented vassals of the Hōjō family. In 1333 Takauji successfully attacked the Hōjō headquarters and forced the *bakufu* leaders to commit suicide. Yoshisada meanwhile conquered the *bakufu* in Kamakura. Thus, after 140 years' rule, the *bakufu* government was brought to an end.

Overthrow of the Kamakura *bakufu*

The return of Emperor Daigo II to Kyoto in 1333 is known as the Kemmu restoration. The Emperor immediately set about to restore direct Imperial rule. He abolished the powerful office of *kampaku* (chief councillor) and set up a central bureaucracy. He established the Kirokusho to settle lawsuits in the provinces; a *zasso ketsudansho* (court of miscellaneous claims) to handle minor suits; and a *musha-dokoro* (guard station) to keep order among the warriors in Kyoto. He placed Prince Morinaga in charge of his military forces and set up members of the Imperial family as provincial leaders in the north and east.

Those local warriors, however, who had joined the Imperial forces in the overthrow of the *bakufu* were disappointed in the division of the spoils. A rebel army formed under the leadership of Ashikaga Takauji, and in 1336 it drove the Emperor from Kybto. Takauji set up an emperor from another branch of the Imperial line (the Jimyō-in), while Daigo II and his followers set up a rival court in the mountainous country of Yoshino. For the next 60 years political power was divided between the Daikaku-ji, or southern, court in Yoshino and the Jimyō-in, or northern, court in Kyōto. It remained for Takauji's grandson Yoshimitsu to establish peace (1392) between the two courts; thereafter, the Imperial succession remained with the descendants of the northern court. Throughout the long dispute, however, local warriors attached themselves to the military governors, who increasingly asserted their independence from the central authority, whose power greatly declined.

Division of northern and southern courts

The establishment of the Muromachi bakufu. After the withdrawal of Daigo II to Yoshino, Ashikaga Takauji set up a *bakufu* at Nijō Takakura in Kyōto. But in 1378 Takauji's grandson, the shogun Yoshimitsu, moved the *bakufu* to the Muromachi district in Kybto, where it remained and took final shape. Yoshimitsu, assisted by the successive *kanrei* (deputy shogun) Hosokawa Yoriyuki and Shiba Yoshimasa, gradually overcame the great military governors. He destroyed the Yamana family in 1391, and in 1399, with his power further enhanced by his success in uniting the northern and southern courts, he attacked and destroyed the great military governor Ōuchi Yoshihiro, thus gaining control of the Inland Sea. Yoshimitsu was now raised to the highest office of grand minister of state, or *dajō-daijin*. He constructed the famed Golden Pavilion (destroyed by an arsonist in 1950; rebuilt in 1955) at his country seat in Kitayama, taking great pride in its luxurious display; he also carried on trade and diplomacy with Ming dynasty China under the title of king of Japan.

The Muromachi *bakufu* inherited almost unchanged the structure of its Kamakura predecessor (see above), setting up a Mandokoro, Monchiijo, and Samurai-dokoro. But after the appointment of Hosokawa Yoriyuki as *kanrei* (deputy shogun), this post became the most important in the *bakufu* government. The official business of the Mandokoro was to control the finances of the *bakufu*, and later the Ise family, who were hereditary retainers of the Ashikaga, came to inherit this office. The Samurai-dokoro, besides handling legal judgments, was entrusted with the control of the capital. Leading officials called *shoshi* who held the additional post of military governor of Yamashiro Province (Kyoto Prefecture) were next in importance to the *kanrei*. In local administration, a special administrator was set up in Kamakura to control the 10 provinces of the Kantō area. This office came to be held by heads of the Ashikaga Motouji family. The 11 provinces of Kyushu were controlled by an office known as the Kyushu *tandai*.

Structure of the Muromachi *bakufu*

The finances of the Muromachi *bakufu* could not be met simply from its receipts from the lands under its direct control. So according to their needs, the military governors and stewards of each province were ordered to levy a money tax, either on every unit of land or every household, but this was not fully effective. As a result, taxes were levied from such dealers as pawnbrokers, and sake brewers, who were among the wealthiest merchants of the time. Financial deficiencies were also supplemented by engaging in trade with China. But the foundations of this *bakufu* began to be shaken by the increasing power of the military governors and by the frequent uprisings of local samurai and farmers.

In the Kamakura period, the authority of the military governors was limited to security matters. In the latter half of the northern and southern courts period, their executive power over the areas under their control was increased. As disturbances increased, they gained wide powers of military command. Sometimes the private estates were made depots for military supplies on the pretext of protecting them from the depredations of the samurai, and half their yearly taxes were given to the warriors. This was called the equal tax division, or *hanzei*. Many military governors succeeded to their domains by inheritance, and in cases such as that of the Yamana family a single military governor sometimes held a number of provinces. Thus arose a new class of official known as the *shugo* daimyo.

From the outset, the controlling power of the Ashikaga *bakufu* was weak, and, especially after the death of Yoshimitsu, the tendency for powerful military governors to defect became marked. Hence, as time passed the office of shogun became increasingly impotent.

The rise of commerce and manufacturing

In the villages around Kyoto, the status of farmers rose markedly as agriculture became more highly developed, and commerce and small-scale manufacturing prospered. Also, confederations of the middle and small landlords, or *mydshu,* proceeded apace and often resulted in uprisings. Such confederations appeared where farming by the greater *myōshu* had dissolved and middle and small *myōshu* had established themselves on a wide scale. These smaller landlords endeavoured to defend themselves against the ravages of local warfare, and they formed unions to manage the forests in common and to maintain irrigation works. In such confederations, a leader called the elder, or *otona,* would be selected to carry on village government. Assemblies were held regularly among its members at the village shrine or temple, and regulations were drawn up for the maintenance of the community life.

As self-government became strong in the communities, the resistance of farmers became fierce. After the northern and southern courts dispute, armed uprisings arose among the farming villages, demanding reductions in yearly taxes against the old proprietors and a moratorium on debts owed to the moneylenders. A large-scale uprising of this kind took place in 1428 in the last years of Yoshimitsu's reign. In 1429 an uprising broke out in Harima Province (present Hyōgō Prefecture) aimed at the expulsion of the warriors from the province. In 1441 farmers living around Kydto attacked the pawnbrokers and demanded a moratorium on debts from the *bakufu.* Thereafter, uprisings occurred on a greater or lesser scale almost yearly—testimony to the fading power of the *bakufu.*

Trade between China and Japan. Trade with Ming dynasty China began after the suppression of Japanese piracy. Ashikaga Takauji had sent ships of the Tenryū-ji to trade with the Yüan dynasty. But trade then ceased on account of the internal disturbances, and pirates from the maritime districts of west Japan raided the Korean peninsula and the continental mainland. When Korea came under the control of the Yi dynasty and when the Ming dynasty emerged in China, it requested the *bakufu* to open formal trade relations with the aim of suppressing piracy. Yoshimitsu, both in response to the desires of the merchants and in order to supplement the finances of the *bakufu,* began formal trade relations with Ming China and Korea, repatriating a large number of Chinese who had been taken captive by the pirates. In response, the Ming also began to trade, but under the form of tribute from Yoshimitsu, "King of Japan," to the Emperor of China. In this trade, in order to distinguish between pirate ships and trading ships, seals received from the Ming called *kangōfu* were used. Hence the use of the term *kangd* trade.

The *kangd* trade

The profits of this trade were important to the *bakufu,* but later the control of this trade came into the hands of the Hosokawa and Ōuchi families, under whose protection trading merchants became active in Hagata, Hyōgo, and Sakai. After the Ōnin War (see below), the Ōuchis controlled this trade, but on their destruction the *kangō* trade ceased and piracy again became rife. Trade with Yi dynasty Korea was carried on through the agency of the Sō family of Tsushima, and various domain lords and the merchants of Hagata were actively involved in it, importing cotton and other goods. Japanese traders resided in Pusan and elsewhere in Korea. Also included in the trade with China and Korea were goods imported by Japanese merchants from the Ryukyu Islands, lying between Japan and Taiwan, and dye materials, pepper, and other special products from the South Seas.

The Ōnin War (1467–77). In the reign of the shogun Ashikaga Yoshimasa a general civil war broke out in the area around Kydto, caused by econoniic distress and by a dispute over the Imperial succession. Indeed, severe famines engendered rebellion nearly every autumn, and it is said that during his term as shogun Yoshimasa issued 13 edicts for the cancellation of debts known as *tokuseirei,* or "acts of grace." Lacking children of his own, Yoshimasa at first proposed that his younger brother should succeed him. But when he later fathered a child a quarrel arose over the succession for control of the family. The two chief administrators, Shiba and Hatakeyama, and the great military governors also took sides in the power dispute, with Kosokawa Katsumoto and Yamana Sōzen (also known as Yamana Mochitoyo) at the head. In 1467, the first year of the Ōnin era, fighting broke out between the eastern army of the Hosokawa party and the western army of the Yamana party. The eastern army had the advantage of the support of both the Emperor and the shogun, but the western army, assisted by the Ōuchi family, recovered its power, and fighting continued for 11 years, centred on Kyōto. Destruction around Kyōto was severe, and many large temples and residences were burned. After 11 years of warfare, the fighting spread to the provinces. As a result, the farming villages held conferences and quite frequently mounted armed uprisings in self-defense. The leaders of these armed uprisings were local samurai with roots in the farming villages. Such men frequently established themselves as domain lords (daimyo) during the war disturbances. They formed associations and often mounted uprisings that extended over a whole province and challenged the great military governors. In the autumn of 1485, 36 representatives of the local warriors of southern Yamashiro Province met in the Byōdōin Temple at Uji and successfully demanded the withdrawal of the two Hatakeyama armies. As a result, the southern Yamashiro area became self-governing for more than eight years.

Edicts for cancellation of debts

In these wars, the civil aristocracy and priests lost the income from their private estates. Many of them left the capital, moving to Sakai or Nara or even taking up residence in the castle towns under the protection of local domain lords. This migration assisted in the diffusion of the central culture to the localities. Old traditions were destroyed, but from the ashes a new culture was born.

The shogun Ashikaga Yoshimasa, for example, finally turned his back on a troubled world and set up a country residence in the Higashiyama ward of Kyōto, where he lived in elegance and refinement, paying no attention to matters of government. The political power of the *bakufu* thus became virtually extinct, and real power came into the hands of the chief administrators of the Hosokawa (1490–1558) family. In the 16th century this power then came into the hands of their retainers, the Miyoshi (1558–65) family, until it was finally usurped by their own retainers, the Matsunaga (1565–68) family.

THE PERIOD OF THE "WARRING COUNTRY"

The emergence of new forces. After the Ōnin War, the power of local leaders became increasingly strong, and there were many instances in which the deputies of military governors usurped the domains of their superiors, retainers overthrew their overlords, and branch families seized power from main families. Because of this tendency for "inferiors to overcome superiors" (*gekokujō*), the previous military governors almost completely disappeared from Kyoto and the surrounding provinces, and a new type of domain lord took their place. They are called *sengoku* daimyo ("civil war barons") or feudal lords of the period of warring country.

Until the first half of the 16th century, domain lords in the various localities were thus building up strong military bases. During this period, the provinces held by the domain lords were almost completely free of *bakufu* control. The domain lords included the local leaders among their retainers, taking away their independence by enforcing land surveys and directly controlling the farming villages. Domain lords such as the Imagawa, Date, and Ouchi issued their own laws called *bunkoku-hd.* These provincial laws, while drawing on the samurai laws of the Jōei Formulary and thereafter, also included regulations for farmers, and they applied strict controls over retainers. It was made a principle that inheritance by retainers should be restricted to the main heir alone, and the overlord's permission was necessary for his vassals to inherit property or to marry. In the farming villages the domain lords, in addition to carrying out land surveys, built irrigation dikes and opened new rice fields. In order to concentrate their power they also readjusted the disposition of local fortified strongholds, gathered their retainers into castles, and reorganized roads and post stations to centre on their castle towns.

Growth of cities and trade

Commerce and towns made marked development at this time. Markets also came to be held six times a month and were set up all over the country. Despite the obstructions of the customs barriers, the products of all the districts were available in these markets. In large cities such as Kyoto, commodity exchange markets were set up to handle huge quantities of rice, salt, fish, and other goods; wholesalers, or toiya, specialized in dealings with distant areas. The circulation of coined money also became vigorous, but in addition to the various kinds of copper coin imported from China of the Sung, Yiian, and Ming dynasties, privately minted coins also circulated within the country, giving rise to confusion of exchange rates. The bakufu and domain lords issued laws to prohibit people from hoarding good coins but without effective results. The guilds now showed a strong monopolistic tendency *in* order to protect themselves against new-style merchants who emerged while new guilds were set up in the castle towns (*jōka-machi*) under the direct control of the domain lords. Among the cities of the time, next to Kyōto and Nara, Uji and Yamada outside the gates of Ise Jingū Shrines flourished. Besides these, towns grew up around the castles of the domain lords, such as Navetsu of the Uesugi family, Yamaguchi of the Ōuchi family, Ichijō-dani of the Asakura family, and Odawara of the later Hōjō. As the castles of the domain lords were moved from mountain fortresses to strongholds in the plains, markets were opened outside the castle walls, and merchants and artisans gathered there to live. Harbour towns (*minato-machi*) such as Sakai, Hyōgo, and Onomichi on the Inland Sea, Suruga and Obama on the Japan Sea, and Kuwana and Ōminato on Ise Bay also flourished as exchange centres. Sake brewers, brokers, and wholesale merchants were leading townsmen (machi-shu), and town elders, called otona, were chosen to carry on local government through assemblies. In the trading port of Sakai, for example, an assembly of 36 men drawn from the wholesale guilds carried on the city government. They maintained soldiers and constructed moats and other defenses, and while profiting from the confrontation of the domain lords, they resisted their domination. The Jesuit missionaries (see below) compared Sakai to the free cities of Europe in the Middle Ages and described its flourishing condition in their reports.

The arrival of the Europeans. As the warring domain lords carved out their territories, the central authority ceased to maintain control over overseas trade. Further, Japanese marauders in association with Chinese pirates again became active. At this time, the Spanish and Portuguese made their appearance in the archipelago. In 1543 (according to Japanese sources), the first Portuguese were shipwrecked on the island of Tanega, off southern Kyushu. These were the first Europeans to arrive in Japan, and the art of musket construction they passed on at this time immediately spread to Sakai and other places.

The Jesuit missionaries

In 1549 the Jesuit missionary Francis Xavier arrived in Kagoshima. After missionary work for two years and three months, he left Japan, and thereafter Jesuit missionaries arrived continuously. The missionaries made use of the trade from the Portuguese ships to propagate Christianity, and there were cases in which merchant ships would not enter the ports of domain lords who did not show good will toward missionary activity. Thus, domain lords, seeking the profits of foreign trade and the acquisition of military equipment and supplies, progressively protected Christianity, until finally some domain lords even became Christian converts. Three Kyushu Christian lords—Ōtomo Sōrin, Arima Harunobu, and Ōmura Sumitada—sent an embassy to Rome. Farmers also increasingly became converts under the influence of the social relief work and medical aid that accompanied missionary activity.

The establishment of military culture. While absorbing the traditional culture of the civil aristocracy, the military families that established themselves in Kyoto also introduced the continental culture of the Sung, Yiian, and Ming dynasties, especially the culture associated with Zen Buddhism, thus fashioning a new military family culture. This began with the golden age of the shogun Ashikaga Yoshimitsu at the end of the 14th century. In this period, scholarship and the arts flourished in the five Zen monasteries of Kyoto under the patronage of the shogun. Renga (linked verse) and Nō drama became vigorous. The essence of this culture finds concentrated expression in the Golden Pavilion (Kinkaku-ju) in Yoshimitsu's country estate at Kitayama, now a part of the Rokuon-ji (see above The establishment of the *Muromachi* bakufu). Facing a garden of refined elegance, the Golden Pavilion is built in the Japanese *shinden* style (an ecclesiastical style with Zen influence) in its first and second stories, while its upper story is in the *kara* ("Chinese") style of the Zen sect. This proves that Kitayama culture, while absorbing Zen influences from China, also shows many influences of the native aristocratic culture. In the time of the shogun Yoshimasa, the samurai culture, in addition to even deeper Zen taste, shows a refined appreciation of simplicity and an air of quiet profundity. The Silver Pavilion (Ginkaku-ji) and its garden built by Yoshimasa on his country estate at Higashi-yama (Mt. Higashi; now part of the Jishō-ji) displays the essence of this polished Higashi-yama culture. While adopted by the local domain lords, Higashi-yama culture also gave rise to a new culture centred on the townsmen of Kyoto and Sakai, and is the forerunner of the famed Azuchi–Momoyama and Edo cultures.

Increasing temporal role of the monasteries

In Buddhism, the ancient great temples like the Enryaku-ji became mere shadows of their former greatness with the diminution of their landed estates. Since the Kamakura period, the Rinzai sect had been linked to the upper military families. The Muromachi shogunal family (the Ashikaga) gave special protection to the group of the priest Musō Soseki of this sect, which flourished in the Gozan monasteries (the five most important Zen monasteries) in Kyoto. The monks of the Gozan became advisers to the bakufu in government, diplomacy, and culture, studying the Neo-Confucian philosophy of Chu Hsi that came from China along with Zen, publishing books, and writing poetry and prose in the Chinese style. But the five monasteries became vulgarized because of their excessive links with the political world, and they ceased to prosper as the *bakufu* declined. In contrast, the Myōshin-ji and Daitoku-ji groups arose, and the latter is famous for the monk Ikkyū, who propagated his own teaching. At this time Rennyo (1415–99) of the Ikkō sect came forth from the Kyōto Hongan-ji, teaching his principles in simple phrases and spreading the faith by organizing groups called ka. However, he came under persecution from the Enryaku-ji and fled to Echizen, establishing a school of instruction at Yoshizaki. He then moved to Settsu Province, where at Ishiyama (now a part of Osaka) the Hongan-ji achieved its golden age. While also persecuted, the Hokke (Lotus) sect progressively gained adherents among the warriors and merchants. At this time, the custom of pilgrimages to the holy places of the Buddhist deity Kannon, to the Ise Shrine of Shintō, and to the summit of Mt. Fuji also became popular. Within this trend, a worldly Shinto belief arose, and in the 15th century the scholar Yoshida Kanetomo, while proclaiming Shinto principles, also took the occasion to free Shintō shrines from Buddhist control; he believed that only a deep religious faith could cure the people of their despondency. In the arts the Nō drama developed in the Kamakura period under the influence of agricultural festival dances, and guilds (za) were formed to serve at the ceremonies of temples and shrines and at funeral services From among the four guilds attached to the Kōfuku-ji and the Kasuga Shrine of Yamato Province (present Nara Prefecture), the father and son Kan-ami and Zeami appeared, who under the patronage of the shogun Yoshimitsu laid the foundations for a flourishing Nō drama, bequeathing many masterpieces of chant, which forms the foundation of the art. Kydgen (dialogue play with dance), which developed from the comic elements of *sarugaku*, was performed in the intervals of Nō drama, and, taking its topics from everyday life of the common people, was widely appreciated by them. Traditional Japanese waka verse was also composed, but renga (linked verse) became ever more popular and was enjoyed by the warriors and the common people. After a time, however, renga became overly formal, as the waka did, and lost its freshness; hence, in the "warring country" period the free-style verse called haikai was born.

Developments in architecture and the visual arts

Along with the prosperity of Zen, the *shoin* style of building closely connected with this sect was widely adopted by the military families and civil aristocrats in the construction of their residences, becoming the foundation of present-day domestic architecture. The *shoin* was originally a room in which monks read the Buddhist scriptures. In constructing the shoin, an entrance called *genkan* is built, while within the room straw mats called tatami are laid out, paper-covered sliding partitions are used, and an alcove and shelves at different levels are set up. The custom of hanging a monochrome painting in the alcove and placing flowers or an incense bowl before it also arose at this time. In gardens, a delight was first taken in adding the Zen mood of retreat from the world to the *shinden* style, making symbolic use of streams, flowers, and bushes. Later, even more symbolic gardens were constructed using arrangements only of stones, raked sand, and gravel. The carving of images of the Buddha and the Buddhist paintings that had flourished in the Kamakura period now declined, as did the ancient sects themselves, and new ones arose. Yamato-e painting also declined, and the picture-scrolls lost their freshness. On the other hand, with the increase of interest in Zen, monochrome painting in the Sung and Yüan style was begun by the monks of the Gozan. In the time of Yoshimasa, the great painter Sesshū broke away from imitation of Chinese models and opened new frontiers in monochrome paintings. The father and son Kanb Masanobu and Hisanobu introduced the gentle models of Yamato-e to monochrome painting and became the founders of the succeeding Kanb school. Tea drinking, introduced from Sung China by the Buddhist priest Eisai in the Kamakura period, spread to the warriors and the common people in the northern and southern courts period. In particular, in the time of the shogun Yoshimasa, Jukb came from Nara and began the wabi-cha form of tea ceremony by bringing together the *cha-no-yu* of the civil aristocracy and the cha yoriai of the common people. This new form spread among the warriors and great merchants and was further purified in the "warring country" period by the Sakai merchant Jōō. Together with the development of the tea ceremony, new forms were brought about in the construction of tearooms, flower arrangement, pottery, and Japanese cakes. In the "warring country" period, Higashiyama culture became further diffused among the common people, and as the livelihood and education of the merchants and artisans of the cities reached higher levels they enjoyed Nō and *kyōgen* dramas, the tea ceremony, and renga (linked verse). Fairy stories were also widely enjoyed, being easy to read, and included stories that had been related among the people since ancient times. These became popular not only among the children of the civil aristocracy and warriors but also among those of the townsmen who were educated in temples and shrines. The local domain lords also promoted culture within their domains, enhancing their dignity as lords by building temples and shrines in their castle towns and actively introducing the culture of Kyōto.

Thus, while warfare was rife in the Kamakura and Muromachi periods, these were, nevertheless, eras that gave Japan some of its most distinctive cultural institutions. (T.T.)

II. The early modern history of Japan (1550–1850)

UNIFICATION

The Oda regime. In the 1550–60 period the sengoku ("civil war") daimyo domain lords, who had survived the wars of the previous 100 years, moved into an even fiercer stage of mutual conflict. These powerful daimyo were harassed not only by their mutual conflicts but also by the social development of the common people within their domains. The lords sought to resolve these contradictions by acquiring land and people to widen their domains and, finally, by trying to grasp the leadership of the whole country. To accomplish this purpose, they considered it necessary to control Kybto, the political centre since ancient times. In the midst of these struggles, one such *sengoku* daimyo, Oda Nobunaga of Owari Province in modern Aichi Prefecture, succeeded in entering the capital as the first feudal unifier.

The first feudal unifier

On the emergence of Oda's regime, the feudal disintegration of the previous century began to show a clear tendency toward unification. Oda's bold wars of suppression, which entitled him to be called a military genius, led to a great redrawing of the political map, previously split up among the domain lords of the whole country. In the Kinai district, on which Oda's conquered territory was centred, however, he established control by dividing his new domain among his commanders. Rather than completely abrogating the long-established privileges of the temples, shrines, and local landlords (kokujin), he at first recognized these, regarding them as an important adjunct to the strengthening of his military power and using them as followers in the unification wars. The land surveys aimed at strengthening feudal landownership were at this stage carried out not so much to gain control over the complicated landholding and taxation system of the farmers as to define the size of fiefs (chigyo) of his retainers in order to confirm the extent of their military services and obligations.

The unification policy of the Oda regime was upheld by the separation of the warriors from the farmers, but it could not be fully achieved because of resistance from old political forces. Unification became more clearly established later, during Toyotomi Hideyoshi's regime.

The Hideyoshi regime. Oda was the son of an Owari domain lord, whereas Hideyoshi was the son of a farmer from the same province. Entering Oda's service, Hideyoshi was greatly esteemed for his brilliant talents, and before Oda's death Hideyoshi had become one of his most powerful commanders. After Oda's suicide during the Honnō-ji (a monastery in Kybto) Incident of 1582—when his retainer and commander, Akechi Mitsuhide, rebelled against him—Hideyoshi eliminated many rivals by his superb political judgment and shrewd actions, thus firmly establishing himself as successor. In the footsteps of Oda, Hideyoshi proceeded to unify the whole country at a rapid pace, and, in 1590, all Japan—from Kyūshū in the west to Tohoku in the east—had come under his control. On the pretext of giving rewards for distinguished service, Hideyoshi gave the Kantō domain, formerly of the Hōjō family, to Tokugawa Ieyasu, causing him to move to Edo; this was, in fact, a stratagem to remove the Tokugawa family from the Tōkaidō and Chūbu districts, where its power had been nourished.

The keynote of the Hideyoshi unification policy was its firm establishment in principle of the separation of the warriors and farmers. Thus, the clear-cut contrast between feudal landowners and feudal small farmers (or serfs) became the basic model of the system.

The Taikō land survey

The Taikō land survey played an important part in this process. Taikd was a traditional title for former regents, and it was assumed by Hideyoshi in 1591. The Taikō land survey was carried out over the whole country from 1583 to 1598, just before Hideyoshi's death. As a result of this survey, the complicated connections of rights to landownership that had developed since the Kamakura period were set in order. Landowning relations were now based on *kokudaka—i.e.*, on the actual product of the land. Moreover, this actual product now came within the landlord's grasp in every village, and land taxes were now levied on the village as a unit. This is called the mura-uke, or village responsibility, system. In addition to this definition of the rights held by the farming population, the kokudaka system also applied to the landholdings of domain lords (the chigyd) to be distributed among their retainers. In place of previous land taxes (*nengu*) assessed in money as so many hundred or ten thousand kan of silver, an assessment of kokudaka was made as so many hundred or ten thousand koku of rice, the amount being also used as a standard on which military services were levied in proportion.

During the Taikb land survey, a land-survey register was drawn up in every village, and farmers so registered were recognized in their rights as cultivators to the extent of the land thus registered; in return they were bound to pay land taxes in rice and forbidden to neglect the cultivation of their fields or to move elsewhere. Farmers were thus reduced to rural serfs, tied to the land, and exploited. The promulgation of an order of social-status control in 1591 prohibited people who neither cultivated the land nor performed military service (*i.e.*, the artisans and merchants) from residing in the villages, showing a further advance from the separation of the warriors and farmers to a feudal social class system of warriors, farmers, artisans, and merchants. The order of "sword hunt" (katana-gari), which took away arms from the farmers, was also an important prerequisite for this policy. By establishing the kokudaka system, the Taikō land survey delivered the final blow to the shden system, a system of manors in medieval times, which had already greatly declined. The feudal chigyd system, based on the kokudaka system, was established throughout the country, the domain lords all submitted to the despotic control of the Hideyoshi regime, and the alliance-like relationship between Oda and the former sengoku daimyo changed to a clear lord–vassal relationship.

The political structure of the Hideyoshi regime was not yet fully equipped, however, to be the unified governing authority controlling the whole country. For example, the kurairechi, or lands under its direct control, which were the immediate financial base of the regime, amounted to about 2,000,000 koku. But setting aside those of the metropolitan and surrounding provinces, these lands were in many cases divided among the distant, independent domain lords (tozama daimyo), and the management of these lands was entrusted to them. Such lands were thus not firmly in the grasp of the regime. By contrast, the lands under the direct control of the Tokugawa shogunate amounted to more than 4,000,000 koku, or better than double those of the Hideyoshi regime, and 80 percent of them were managed by officials known as gundai and daikan, who were direct retainers of the shogunate, with only 20 percent deposited with domain lords. This inner contradiction in the political structure of the Hideyoshi regime gave rise to internal power struggles and finally drove Hideyoshi to reckless action such as the Korean expedition (aggressions against Korea in 1592 and in 1597). The Hideyoshi regime collapsed on the failure of that expedition and as the direct result of Hideyoshi's subsequent death. Tokugawa Ieyasu was the strongest candidate to form the next regime.

THE BAKUHAN SYSTEM

The Matsudaira family

The establishment of the system. The ancestors of Tokugawa Ieyasu, the founder of the Edo bakufu, were the Matsudaira, a sengoku daimyo family from the mountainous region of Mikawa Province (in present Aichi Prefecture). The Matsudairas had built up their base as domain lords by advancing into the plains of Mikawa. But when they were attacked by the powerful domain lords of the Oda and Imagawa families from the west and east, Ieyasu's father, Hirotada, suffered the bitter misfortune of losing his domain. Ieyasu became a hostage of the Oda family at a young age, but he was then captured by the Imagawa family and spent 12 years under detention. In 1560 Imagawa Yoshimoto was killed by Oda Nobunaga in the Battle of Okehazama, in which Nobunaga destroyed the Imagama family and confirmed his course of unification, and Ieyasu was finally released. He returned to Okazaki in Mikawa and brought this province under his control. As Oda's ally, he guarded the rear for the advance on Kyōto, and he thereafter fought his own military campaigns, advancing eastward. By 1582 he was a powerful daimyo, possessing, in addition to his home province of Mikawa, the four provinces of Suruga and Tōtōmi (modern Shizuoka Prefecture), Kai (Yamanashi Prefecture), and southern Shinano (Nagano Prefecture).

When Hideyoshi seized the ruling power, Ieyasu at first opposed him. But he then submitted, and, rising to be the most powerful domain lord among Hideyoshi's vassals, he became chief of the five *tairō*, the highest officers of the Hideyoshi regime. After Hideyoshi's death Ieyasu won the Battle of Sekigahara in 1600, in which all daimyo in the country took part, establishing his national supremacy. In 1603 Ieyasu set up the Edo bakufu (more commonly known as the Tokugawa shogunate [1603–1867]) to legalize this position. Control over the domain lords was firmly exercised at this time. On the pretext of allotting rewards after Sekigahara, he dispossessed, reduced, or transferred a large number of opposing daimyo and gave the confiscated lands either to relatives and retainers of the Tokugawa family to establish them as domain lords and to increase their holdings or reserved the lands as the family's domains. In addition, Hideyoshi's son Hideyori was reduced to the position of a domain lord of the Kinki district. Two years after the establishment of the bakufu, Ieyasu relinquished the office of shogun to his son Hidetada, retiring to Sumpu (modem city of Shizuoka) to devote himself to strengthening the foundations of the bakufu. In 1615 Ieyasu stormed and captured Osaka Castle, destroying the Toyotomi family. Immediately afterward, the Buke Sho-hatto (Laws for the Military Houses) and the Kinchū narabi-ni Kuge Sho-hatto (Laws for the Imperial and Court Officials) were promulgated as the legal basis for bakufu control of the domain lords and the Imperial court. In 1616 Ieyasu died.

Completion of the *bakuhan* system

Under the second and third shoguns, Hidetada and his successor, Iemitsu, the bakufu control policy advanced further until the bakuhan system—the government system of the Tokugawa shogunate; literally a combination of bakufu and han (daimyo domain)—reached its completion. By reorganizations in 1633–42 the executive of the bakufu government was almost completed, as represented by the offices of senior councillors (*rojū*), junior councillors (waka-doshiyori), and three commissioners (bugynin) for the temples and shrines of the country, the shogun's capital, and the treasury of the bakufu. Confiscations and reductions were continuously made against the domains, and wide-scale transfers also took place, distributing the strategic districts of Kantō, Kinki, and Tōkaidō among the relatives and retainers of the bakufu, thus keeping the "outside" (tozama) domain lords in check. Along with the rearrangement of the daimyo, the lands under the direct control of the bakufu were also increased at key points throughout the country. The most important cities and mines were also placed under direct bakufu control and used to control commerce, industry, and trade.

The bakufu also revised the Laws for the Military Houses and systematized the *sankin* kdtai (alternative attendance), by which the domain lords were required to pay ceremonial visits to Edo every other year, while their wives and children resided permanently in Edo as hostages. This sankin-kdtai system was unique to Japanese feudalism and never appeared in European feudalism. In addition, the daimyo were forced to assist in such public works as the construction of castles in the bakufu domains, thus being driven into financial difficulties. The bakufu domains now amounted to more than 7,000,000 koku—about one-quarter of the whole country. Of these lands, more than 4,000,000 koku were under its direct control, and 3,000,000 koku were distributed among the hatamoto and *gokenin*, the liege vassals to the bakufu. In addition, because the bakufu had a monopoly of foreign trade and alone had the right to issue currency, it had considerably greater financial resources than did the domain lords. In military strength, it was also far more powerful than any individual daimyo.

In step with the structural organization of the bakufu as the supreme power, the domain governments of the daimyo also progressively took firm shape. The relationship between the shogun and the daimyo was linked with the lord–vassal relationship, based on the feudal *chigyō* system. The land of the whole country belonged to the shogun, who divided this among the domain lords as a special favour, or *go-on*. In return, the domain lords had a duty to provide military and other services to the shogun. This same connection existed between the domain lords and their retainers, and in order to concentrate and strengthen the ruling power of the domain lords, it was necessary for them to tighten this connection. Applying restrictions to the traditional right of the domain warriors to *chigyd*, or subdomains, retainers were supplied with rice stipends *(kuratnai)* in place of *chigyd*, thus increasing their dependence on the domain lord. At the same time, this increased the lands under the direct control of the daimyo, strengthening the economic base of the domain. These were the same methods employed by the *bakufu*. In this way, a hierarchical, feudal *chigyō* system was established by means of the *koku* system, which extended from the shogun and the domain lords to their retainers. This was the fundamental condition that made possible the concentration of ruling power in the hands of superiors, the principal distinguishing feature of the Japanese feudal system.

Control of the villages

Having been inherited from the Oda–Hideyoshi regimes, the control of the farmers, which was the main object of feudal rule, was now further strengthened. The Taikō land survey had recognized the rights of the lower class farmers as the actual cultivators of the land and made them bear the responsibility for the payment of land taxes. While also carried out on this basis, the land surveys of the *bakufu* and the daimyo domains were much more detailed and precise in standards of survey, adopting the policy of extracting the greatest possible tax yield, limited only by the necessity for the farmers to continue to produce. The Tokugawa villages thus differed from those of the preceding ages, which had been controlled by patriarchal landlords, or *myōshu*. The Tokugawa villages were composed of a main core of feudalistic petty farmers, who went under the general title of *hombyakushō* ("farmers proper"). At the same time, the village became an administrative unit of the new feudal regime; and to carry out its functions, a system of village officers was organized in three grades—*nanushi* (or *shdya), kumigashira*, and *hyakushd-dai*. The inhabitants of the towns and villages of the whole country were required to form *gonin-gumi* (five-household groups), or neighbourhood associations, to foster joint responsibility for tax payment, to prevent offenses against the laws of their overlords, to provide one another with mutual assistance, and to keep a general watch on one another. The so-called outer economic controls over farmers were further strengthened: farmers were strictly prohibited from buying and selling land, from absconding from their land, or from changing their occupation; minute restrictions were also applied to keep their clothes, food, and houses as simple as possible. The Keian no O-furegaki (Ordinance of the Keian era) was promulgated by the *bakufu* in 1649, a compendium of *bakufu* policies toward rural administration.

The enforcement of national seclusion. The 1630s also marked an important dividing line in foreign relations with the enforcement of *sakoku*, or national seclusion. The path toward seclusion had been prepared during the formation process of the *bakuhan* system. The seeds of this policy had been sown in trade control and in measures against Christianity by the Oda–Hideyoshi regimes. Driven on by his consciousness of Japan as a "divine country" in his position as feudal overlord, Hideyoshi, though strongly attracted to trade as a source of national wealth and military strength, had issued an order for the exclusion of the missionaries. Ieyasu, even more strongly attracted by profits, made efforts to trade not only with the Portuguese Catholics but also with Protestant Holland and England, protecting trade with the southern regions by granting special licenses, or *shuin-jō* ("red seal license"), to oceangoing merchant ships. But Ieyasu's encouragement of trade was above all aimed at establishing a *bakufu* trade monopoly. In 1604 a special system for the purchase of silk was established, and the Chinese silk imported to Japan by Portuguese ships was sold at fixed prices to the powerful merchants of Kyōto, Sakai, and Nagasaki, who formed a guild and then distributed this silk to the domestic retail merchants. Moreover, under the name of "emperor's silk" (*go-yō ito)*, Ieyasu enjoyed a preferential purchase of a part of the imported silk prior to the said guild's treatment and reaped a huge profit on releasing this to the domestic markets.

As a result of his eagerness for trade, it was natural for Ieyasu to be generous to the propagation of Christianity. But Ieyasu feared that the Christians would link up with the Toyotomi family to resist the *bakufu*, and he took steps to prohibit Christianity before his destruction of the Toyotomi family. An order prohibiting Christianity was promulgated in 1612, and the persecution of its adherents began immediately thereafter. This persecution became much more severe during the reigns of Hidetada and Iemitsu, until finally it became official policy to stamp out Christianity even at the sacrifice of trade. This policy became manifest with the seclusion orders of the 1630s. Thus, in 1635 Japanese were forbidden to make overseas voyages or to return to Japan from overseas, which was a severe blow to traders.

The Shimabara Rebellion

In 1637, in resistance to heavy taxes and the prohibition of Christianity, an uprising took place in the Shimabara Peninsula of Kyūshū, consisting of farmers led by masterless Christian samurai. For five months they put up a fierce fight against the *bakufu* army. Having suffered this bitter experience, known as the Shimabara Rebellion, the *bakufu* thereafter stepped up its strict controls on Christians and attempted to root them out by such means as *fumi-e*, in which one was required to trample on an image of Christ or the Virgin Mary. The system of registration at Buddhist temples, or *terauke-seido*, was begun, and all the inhabitants of the country had to belong as parishioners to a parent Buddhist temple, called a *danna-dera* ("family temple"), which every year had to guarantee that the parishioner was not a Christian. In 1639, when Portuguese ships were completely forbidden to visit Japan, the seclusion orders were fully effective both in name and reality. Only the Dutch and the Chinese were allowed to trade as before, but this trade was also extremely restricted and was confined to the island of Dejima at Nagasaki.

Various views exist as to the influence of national seclusion on the Japanese, but the depth of its impact was probably without parallel in Japanese history. It is an undeniable fact that the vigorous desire of the Japanese to expand overseas prior to the seclusion policy was thenceforth diverted into a negative national character known as *shima-guni konjō*, or insularity. While the seclusion policy was useful in enabling the Tokugawa shogunate to exercise an enduring and stable political dominance for nearly 300 years, this simply resulted in the long continuation of a rigid feudal system to an extent unknown elsewhere in the world. Industry developed and gave rise to a unique popular culture, but it was a limited Japanese feudal culture with no international characteristics.

Status distinction system. Thus, the *bakuhan* system was firmly established in the second half of the 17th century, as determined by domestic and foreign conditions. This establishment of a feudal class structure of warriors, farmers, artisans, and merchants was truly the final consummation of the *bakuhari* system. Distinctions between the statuses of warriors, farmers, artisans, and merchants were strictly enforced, but the distinction between the warriors and the other three statuses was especially strict. Forming barely 7 percent of a total population estimated at 30,000,000, the warriors levied taxes on the farmers, who formed over 80 percent of the population and who thus provided the economic foundation of the system. To symbolize this society, the warriors wore swords in everyday life because the system was maintained by their great military power. While peace lasted for 300 years, it was in fact no more than an armed peace. Another special feature of this society was that, even in family relationships, absolute obedience was demanded from members of the family toward the house head (*ka-chd*). Among the family members, the status of women was especially low, and the idea of *danson-johi* (respect for the male, contempt for the female) was prevalent.

Relations between superiors and inferiors

The establishment of the *bakuhan* system created a need for a new world view and a system of ethics to support it. In these circumstances, the Shintō and Buddhist ideologies of early medieval society were useless. Only the morals of Confucianism, especially the Chu Hsi studies, could provide an intellectual rationalization for the status-oriented social structure of the *bakuhan* system, with its stratified feudal landowning structure headed by the shogun and its attribution of superior status to warriors over the other social classes. The feudal rulers, in particular, demanded Shushigaku ("Chu Hsi studies") because, among the various schools of Confucianism, it was the most systematic as a doctrinal system. The Chu Hsi scholar Hayashi Razan was employed by the *bakufu* and was well suited to the *bakufu* foundation period. He is said to have had a hand in the drafting of all *bakufu* official documents and in the formulation of *bakufu* laws. His political ideas, as seen in his *Honchd hennen-roku* ("Chronological History of Japan") and in the *Honchō tsugan* ("Survey History of Japan"), compiled by his son Gahō, provided a historical justification for the establishment of the Tokugawa shogunate. The role of Chu Hsi studies was to repudiate the revolutionary idea of *gekokujd*, or the overthrow of superiors by inferiors, and to emphasize the idea of *kenshin*, linking this to Confucian moral conceptions. The central moral conceptions of Confucianism were *cha*, or "loyalty," and *kd*, or "filial piety." But in contrast to China, more emphasis was placed on *chū* as a support for feudal lord–vassal relations than on *kb*, which was a family ethic. Chu Hsi studies opposed the world view and logic of Christianity, which gave more importance to God than to the ruler–subject relationship and which also bitterly criticized Buddhism, which had been the ideology of the Middle Ages.

Wang Yang-ming studies (Oybmeigaku) also held a special place in Confucian circles in the early Edo period. These studies were characterized by a strong subjective idealism, but in other aspects they had practical elements. Nakae Tbju, said to be the father of Japanese Wang Yang-ming studies, was so earnest in performing virtuous acts that he was called the sage of Ōmi. One of his followers, Kumazawa Banzan, transformed Wang Yang-ming studies from a means for individual spiritual training to a method for providing the energy for political reformation.

Industries, cities, and culture. With the establishment of the *bakuhan* system, the anarchy of the earlier period ended, bringing in its wake a tendency to domestic peace.

Commercial economy. As a result, industry was promoted and cities developed. This trend took place in the latter half of the 17th century, centred in the Kinki district, where productive power was the most advanced in the whole country. Feudal petty farmers, who went under the general name of *hombyakusho*, now became widespread, and, while paying heavy taxes and performing various kinds of labour services, they sought to enjoy a better standard of living, however slight. In addition to their efforts as cultivators, they reclaimed new lands and produced various commercial crops and handicraft goods for sale in the city markets. Representative of such commercial crops were cotton and rapeseed oil in the Kinki district and silk in east Japan. Communications and transportation also developed for the circulation of such goods. But this circulation of commercial goods among the people was centred on the side roads and waterways rather than on the five great highways constructed by the *bakufu*. As a result of the development of industry and communications, such new-style merchants as wholesalers and brokers came to the fore, and powerful financiers also appeared.

Commercial goods production

As for the cities, the castle towns of the domain lords were the most numerous, but these gradually came to acquire the character of commercial cities. Purely commercial cities and post towns (towns along highways) also arose throughout the country. The cities of Edo, Osaka, and Kyōto, under the direct control of the *bakufu*, were especially developed. When its warrior inhabitants are included, Edo in the early years of the 18th century had a population of more than 1,000,000 and thus became the largest city in the world.

Cultural activities. In this period remarkable artists and masterpieces appeared in the fine arts and crafts. Great artists representative of the culture include Ihara Saikaku in *ukiyo-zdshi* ("tales of the floating world") genre novels, Chikamatsu Monzaemon in *jōruri* (puppet play) drama, and Matsuo Bash6 in haikai (humorous unconventional verse). Saikaku was a townsman who spent all his life in Ōsaka. He first aspired to write haikai, and for more than 30 years he was active as a haikai composer. But his forte was in *yakazu-haikai*, which was a competition to compose as many haiku (the poetic form of haikai) as possible within a fixed period of time. Saikaku set a new record by composing 23,500 haiku in a single day and night—one haiku every four seconds. In 1685 Saikaku gave up haikai and next set out to write *ukiyo-zōshi*, producing about 20 masterpieces in succession, beginning with *Kōshoku ichidai otoko* (1682; *The Life of an Amorous Man*, 1964). Between *1624* and 1643 the *ukiyo-zōshi* novel had transformed the *kana-zōshi* (storybooks written in kana) into an even more thoroughly townsman form of literature after the latter had replaced the previous *otogi-zōshi* ("fairy-tale books") in popularity. The unique spirit of the age can be seen in the world *ukiyo*, which had meant "sad world" in medieval times. Written with a different calligraphic character in *bakuhan* times, it now meant "floating world" and implied cheerfulness. The consistent aim of Saikaku's works was to depict, as accurately as possible, human desire for love and gain, taking the realistic viewpoint that "human beings are bundles of desire equipped with arms and legs." His viewpoint also contained a keen criticism of the warriors as men so bound by social status and moral principles that they could not live a free life.

Matsuo Bash6 remained closely attached to haikai and fashioned it into up-to-date popular poetry. Bashō came from a warrior family, but after becoming a masterless samurai, or *rōnin*, he devoted himself to haikai, while suffering various hardships in his means of livelihood. Bash6 found the existing haikai styles unsatisfying, and he developed a new style called *shōfū* or "Bash6 style." Bash6 proclaimed what he called *makotono* ("true") haikai, seeking the spirit of this poetic form in sincerity and truthfulness. He also brought a new beauty to haikai by the use of small and simple words. He brought about an artistic flowering of the highbrow conceptions of medieval poetry by grafting onto them the commonplace feelings of the Tokugawa people. Rather than repudiating tradition, Bashō's haikai brought it to completion.

During the period 1592–1614 the *Jōrurihime monogatari* (a romantic ballad), drawing on the traditions of the medieval narrative story, was for the first time arranged in the form of dramatic literature accompanied by puppetry and the samisen (a lutelike musical instrument). It continued to develop until the three masters—Takemoto Gidayū as narrator, Chikamatsu Monzaemon as composer, and Tatsumatsu Hachirobei as puppeteer—made *jōruri* the representative form of Tokugawa performing arts.

Gidayū was the son of a farmer, while Chikamatsu, like Bashō, came from a warrior family. Chikamatsu wrote more than 80 *jidaimono*, or historical dramas, and 20 *sewamono*, or domestic dramas of contemporary townsman society, both for *jdruri*. He also wrote more than 30 Kabuki plays. The chief theme running through Chikamatsu's works is the idea of *giri* (or "duty"), which is to be understood not so much as feudal morality enforced from above but rather as the traditional consciousness of honour and dignity in one's motives and position and of social consciousness in human relations. The compositions of Chikamatsu's later years seek the motif of tragedy in the fact that this *giri*, while proof of men's humanity, cannot be thoroughly achieved because of their immorality and lack of principle. Beginning with his *Shinjū*

The idea of *giri*

ten no Amijima (1720; The Love Suicide of Amijima, 1953), the leading male and female characters in his sewamono dramas are unable to keep the postulates of giri in this world and so die by *shinjū* (a suicide pact between lovers) in order to realize their love in a future life. Buddhist elements can be seen in this conclusion, as well as the unresolvable contradictions that faced the townsman in Genroku period society (1688–1703).

Kabuki also established its foundations as Tokugawa drama in this period. The okuni Kabuki, named for the maiden dancer Izumo Okuni, which had been popular at the turn of the 17th century, afterward developed from the ynjo Kabuki, or courtesan style, to the *wakashū* Kabuki, or young-man style. The *wakashū* Kabuki was prohibited by the bakufu moral censors, however, because of widespread homosexuality among its performers, and it developed into the *yarō* Kabuki style, played by adult males distinguished from the *wakashū* by shaven forelocks. At the same time, Kabuki now developed from its previous dancing-act form into a regular drama centred on a dramatic plot with realistic acting. The Kabuki was centred in the cities of Kyoto, Osaka, and Edo. In Kyōto and Osaka Kabuki, the speciality was the wagoto, a drama with a pronounced comical element, centred on love, whereas the popular form of Edo Kabuki was the *aragoto*, which had developed from the *kinpira jōruri*, based on the theme of the heroic tale of Kinpira—son of Sakata Kintoki and one of the leading followers of Minamoto Raiko—and adopted the manners of the *kabukimono*—the rowdy bucks of the age.

The traditional arts of Nō drama, the tea ceremony, and flower arrangement also reached new stages of development in the period. The tea ceremony (*cha-no-yu*)—in particular, became popular and was practiced not only by the shogun and domain lords but also by the newly risen merchants, who used their wealth to become eager collectors of tea-ceremony utensils with historical associations. As the tea ceremony became popular, many schools came into existence, including that of Sen-ke (Sen house), the school of Sen Rikyū, fostering the tendency for the art of the tea ceremony to be monopolized by the house heads of the various schools and for the profession of tea master to develop. This "house head" (*iemoto*) system gradually also spread to flower arrangement and to other arts. One result of this was to inhibit further development of these artistic forms.

Distinctive tendencies also arose in the fine arts and crafts. One representative of this is Ogata Kōrin, who brought decorative painting to completion and bequeathed to posterity many splendid masterpieces in gold lacquer (*maki-e*) and other work. The technique of dyeing and weaving was also improved, and, in Kyōto, Miyazaki Yiizen brought the splendid *yūzen-zome* (rice-paste batik method of dyeing) to completion, while the kimono became even more colourful. In Edo, Ukiyo-e drawing in traditional styles was further developed by Hishikawa Moronobu, who depicted not only the usual courtesans and actors but also vividly portrayed various aspects of the lives of ordinary people. Besides his original drawings, he used the Chinese technique of wood-block printing to satisfy popular demand. Famous centres of pottery production also arose at various places throughout the country.

Changes in ceremonies and daily customs

The old ceremonies of the Imperial court and the various forms of deportment developed in the bakufu were extended to the common people and shaped the manners of the cities and farming regions. Japanese customs in dress, food, and housing became established at this time. The custom changed from eating twice a day to three times a day; in the cities rice became the normal food, and a rich variety of cakes and sweets was consumed.

THE WEAKENING OF THE BAKUHAN SYSTEM

On entering the 18th century, the inner contradictions of the bakuhan system came to the surface and began to show signs of weakness. The finance of the bakufu and domain lords was based on a rice-using economy, in which executives endeavoured to levy taxes—to be paid in kind, mostly in rice, centred on the yearly crop. Rice and other crops were then transported to the great central cities like Edo and Osaka and exchanged into money. This forced the farmer-producers to subsist at a low standard of living and to be as self-sufficient as possible, being able to purchase only iron tools, salt, medicines, and other such goods that they could not produce for themselves. Taxes, however, came to be paid in money to a fair extent, and, because farmers also wanted to widen the scale of their activities and to enjoy a more comfortable life, it was unavoidable that they would show a strong interest in commercial-goods production. But whereas rich farmers profited from commercial-goods production, the vast majority of poor farmers and peasants become more deeply impoverished as a result of their involvement in such production. Squeezed by the merchant intermediaries, who forced them to sell their products cheaply, many were forced to part with their lands.

Rural and urban unrest

Thus, as the commercial economy extended into the farming villages, social divisions arose among the farmers. Tax-paying connections became unstable, and the financial difficulties of the warriors, who existed on the taxes paid by the farmers, were naturally aggravated. Because the level of agricultural technology in the feudal period was generally backward, hundreds of thousands of people starved or left their villages during periodic crop failures and famines, and the abandonment of cultivated land also became conspicuous. The samurai class had long since taken up normal residence in the cities. With the development of the urban way of life, they now suffered from increased expenditures. The bakufu and the domains tried to suppress commercial-goods production, which destroyed the self-sufficient economy of the farmers. But when all such attempts failed, they encouraged such production, seeking to supplement their finances by monopolizing the farmers' commercial goods and selling them themselves. This policy was the monopoly system, a phenomenon that may be termed the merchantization of the feudal lords. Thus, on top of increased taxes, the farmers were deprived of the profits of their commercial goods. Unable to restrain themselves, they violated the strict legal prohibitions with repeated farmer uprisings (*hyakushō* ikki). Meanwhile, the city poor, who were driven to the border line of starvation by the rising price of rice and other commodities, often rioted, plundering and destroying rice shops and pawnshops; these urban riots were called *uchi-kowashi* ("destruction").

The "rice shogun"

Political reform in the bakufu and the domains. Continual political reforms made by the overlords in response to this crisis in the feudal system characterize the latter half of the Tokugawa period. Such reforms began with the eighth shogun, Tokugawa Yoshimune (ruled 1716–45). Yoshimune appointed civilians to official posts in finance and rural administration in order to increase government efficiency. As an emergency policy, he ordered the domain lords to make rice contributions (agemai), which he allotted to the hatamoto to supplement their stipends. More characteristic was his effort to increase tax yields by opening new lands to cultivation and revising the method of taxation. He also sought to bring fresh air into a stuffy, despotic government by such means as setting up complaint boxes, or meyasu-bako, to hear the views of the people and drawing up a part of a legal code, the Kuji-kata O-sadamegaki, which mitigated criminal punishments. He made efforts to control the falling price of rice, earning him the name of "the rice shogun." But when the price of rice rose sharply in a great famine in the 1730s, the common people of Edo attacked the wholesale rice dealers who had cornered the market. This was the first such riot in Edo. Yoshimune's reforms took many twists and turns; but in 1744, the year before his retirement, the receipts of the bakufu both in total land taxes and in tax receipts reached their highest level for the entire Edo period. For this reason Yoshimune was called the "great ruler who restored the bakufu." His success, however, was possibly due to the fact that the disturbances had not yet become very grave, whei-eas the political power of the bakufu was still quite strong.

When Yoshimune's son Ieshige became the ninth shogun, government by the personal attendants of the shogun, which Yoshimune's personal rule had prevented, was revived. Chamberlains, or *soba-ydnin,* who handled communications with the senior councillors, gained strong powers of authority as his spokesmen when they won the Shogun's confidence. One such man was Tanuma Okitsugu, who rose from chamberlain to be senior councillor. Okitsugu delighted in bribery, and he was criticized by an opposition group for corrupting the *bakufu* government. But he was an active reformer who further developed some of Yoshimune's programs. Instead of suppressing the activities of the big-city merchants, Okitsugu used them to promote production; and while advancing the development of the commercial-goods economy, he sought to control it. His ready recognition of the commercial and industrial guilds, or *kabu-nakarna,* seems to have been aimed not so much at gaining contributions (*myōga-kin*) for the treasury as for establishing a controlled commercial-goods circulation, linking the city guilds with the village producers. Okitsugu was criticized by the people for issuing large amounts of debased coinage that caused a rise in prices, but he may have been trying to increase the amount of currency to match the developing commercial-goods circulation.

Okitsugu's progressive political attitude is best revealed in his development of Ezo (present-day Hokkaido) as a precaution against the southward advance of the Russians; he even considered trading with Russia. Various natural disasters occurred in his time, however: a great eruption of Mt. Asama (Asama-yama) in 1783 was followed by a great famine in 1783–87, in which large numbers of people starved to death. In addition, derelict land deserted by its cultivators increased, and the custom of *mabiki,* or infanticide, became common among parents unable to rear their children. Some people became landlords by collecting together the lands of poor farmers; others became powerful merchants. The anti-feudal struggle of the farmers also rose to an unprecedented pitch, until, in 1787, a large-scale riot lasted for three days and threatened to reduce Edo to anarchy. Okitsugu had already been dismissed as senior councillor in the previous year, and Matsudaira Sadanobu, grandson of Yoshimune and the lord of Shirakawa domain (in modern Fukushima Prefecture), was indicated as his successor. But Okitsugu's supporters in the *bakufu* made every effort to prevent Sadanobu's appointment, and for more than six months the political situation remained a complete vacuum. The farmers' riot suddenly changed the situation, however, and Sadanobu was appointed senior councillor.

Natural disasters

Sadanobu is renowned as the director of the Kansei reforms (1789–1801). He rejected Okitsugu's free and easy administration and instituted a policy of retrenchment. He set out to reduce the high prices in the great cities and had a fund established in Edo under the name of *shichibu-tsumikin* (70 percent reserve fund); knowing that land and house rents were high in the shogun's capital because of the heavy *machinyūyō* tax levied on its landlords, he reduced this tax and set aside 70 percent of it as a fund for the relief of the poor. To relieve the hardships of the *bakufu* retainers, he took emergency measures to cancel the debts of the *hatamoto* to the *fudasashi,* the Edo merchants who handled the exchange of their stipends. The farming villages, which were the foundation of the *bakuhan* system, had been devastated in the great famine of the 1780s, so while encouraging the *daikan* (a head official managing the *kuraireichi)* to bring land back into cultivation and to increase the population of the villages, Sadanobu also issued orders that annual rice taxes were to be increased as much as possible.

Sadanobu was a firm admirer of Chu Hsi studies, and he believed that government must be conducted on the basis of Confucian benevolent rule. He established an examination system for promotions among *bakufu* officials and also prohibited all teachings except those of the Chu Hsi school at the Shōheikō, the *bakufu* official college.

Sadanobu's reforms give the impression of an overly severe reaction to Okitsugu's administration; and whereas people at first welcomed them, antipathy gradually increased. Within the *bakufu* also, the O-Ōku, or shogunate domestic quarters, composed of ladies, disliked Sadanobu's reforms and forced him to resign by covert stratagems (the Shogun's favourite mistresses tempted the Shogun to expel him). While Sadanobu was senior councillor, a Russian envoy, Adam Laxman, landed at Nemuro in 1792 and requested trade relations, but the *bakufu* did not give its assent. Sadanobu ordered that plans be drawn up immediately for a coastal defense system centred on Edo Bay (now known as Tokyo Bay), while he himself inspected the coastline of Izu, Sagami, and Bōsō. Because of Sadonobu's resignation (1793), these plans were not carried out; but the *bakufu* councillors of this time were the first to react to the footsteps of the foreign nations advancing on Asia, which now came to be heard through the wall of national seclusion.

In conjunction with the *bakuju* programs, reforms were carried out within the various daimyo domains. A distinctive feature of domain reforms at this time, however, was that they tried to apply stronger regulations and control over the commercial-goods economy of the farmers.

Sadanobu was senior councillor during the reign of Tokugawa Ienari, the 11th shogun. Ienari was restrained by Sadanobu's strict political reforms, but when the latter left the *bakufu* council, the Shogun was able to relax. Even so, Ienari was not completely free while the councillors who had supported Sadanobu's reforms were still alive. During the period 1804–31 these men died one after another, and the *bakufu* government became even more lax than in the time of Okitsugu. Mizuno Tadaakira, a senior councillor in this period, had risen as a personal attendant to Ienari and had business ability. But he welcomed bribery, and the other officials followed his lead, greatly increasing the corruption of the *bakufu.* On the surface, things seemed peaceful, but underneath the stagnation of the feudal system became even more grave. Especially in the farming villages of Kantō, right in the lap of the *bakufu,* homeless ruffians and gamblers continually caused disturbances. The *bakufu* therefore set up an office called the Kantō Torishimari-deyaku (Supervisors of the Kantō District) to strengthen police control of the area, and it ordered the villages of Kantō to form associations to assist this office.

The growth of the northern problem. In the early 1800s foreign relations became a fairly pressing problem, and the situation in Ezo fell into confusion. In 1804 another Russian envoy, N.P. Rezanov, came to Nagasaki to request commercial relations. When the *bakufu* refused this request, his men attacked Etorofu Island. Prior to this, the *bakufu* had relieved the Matsumae domain of eastern Ezo and placed it under its direct control; and in 1807 the *bakufu* also took direct control of both eastern and western Ezo for defensive purposes. In 1808 the English warship "Phaeton" made an incursion on Nagasaki, and three years later the Russian naval lieutenant V.M. Golovnin landed on Kunashiri Island and was arrested by the Japanese. When these incidents were resolved, peace continued for a time in the northern regions, and the *bakufu* relaxed its precautions, returning all Ezo to the Matsumae domain in 1821. In the south, English ships often appeared in Japanese waters after the "Phaeton" incident, and the *bakufu,* in 1825, cancelled the order for the provision of firewood and water (*shinsui-kyūyo-rei*) and promulgated an order for the driving away of foreign ships (*ikokusenuchiharai-rei*). While attempting to preserve the iron law of seclusion to the bitter end, *bakufu* policy had no consistency (to foreign ships it was at times offensive and at times not), and it was utterly powerless when it received the full weight of foreign pressure in about the 1840s.

The Russian attack on Etorofu

New learning and thought. As this weakening of the *bakuhan* system grew more serious, new movements took place in scholarship and culture. In the latter half of the 17th century the Kogaku (Study of Antiquity) school arose; scholars criticized Chu Hsi studies and advocated a return to the original ideals of Confucianism. This view was developed by It6 Jinsai and Ogyū Sorai. Especially in his work *Seidan,* Sorai insisted that the main reason for the financial distress of the warrior class in both the bakufu and the domains was that there was no "system" in things, and that when the warriors left the villages and moved to the cities, they had to buy everything, whereas before they had been able to live self-sufficiently. Various other schools of Confucianism arose, such as Setchū-gaku (Eclectic Studies) and Kōshōgaku (Positivistic Studies). Conflict between the various schools became fierce, and the academic world grew confused. The authority of Chu Hsi studies grew weak, and, at the time of the Kansei reforms the bakufu attempted to reinvigorate them. It issued an order prohibiting all other schools of Confucianism in the college of the *bakufu*, but it had no marked results. At the same time, studies like Confucianism spread remarkably in the provinces, as can be seen from the establishment of domain schools (*hankō*), principally in the later Edo period, for the education of the domain samurai. Thus, learning and culture arose in the domains, accompanied by a growth of scholarship with local colouring. Among such schools, the Kaitoku-db in Osaka was famous as the "townsmen's university." This school was founded by cooperation between Confucian scholars and rich merchants, and both samurai and merchants sat together to hear lectures. It is not surprising that the unique thinker and scholar Yamagata Bantō should have been produced from this rationalized way of study.

New movements also appeared in Shintō, which, with Confucianism and Buddhism, acted as the ideology of popular education. The Confucian scholar Yamazaki Ansai formulated Shintō from the standpoint of Confucianism and proclaimed the Suika form of Shintō. But in the later Tokugawa period popular interest in Shinto grew progressively stronger, centred on the Ise faith. This tendency was spurred on by lectures that explained Shintō in terms easily understood by the common people. Thereafter, Shingaku (Heart Learning) arose, which explained Confucian, Shinto, and Buddhist teachings to the townsmen and farmers. Its founder was Ishida Baigan. Kokugaku (National Learning) was also established against the same social background. Kamo Mabuchi studied the Manyō-shū, the most ancient collection of poetry in Japan, and other ancient writings, urging a return to old ways that had not been defiled by foreign ideas, such as Confucianism and Buddhism. By studying the ancient language of Japan's oldest classic, the *Koji-ki* ("Records of Ancient Matters"), his pupil Moto-ori Norinaga clarified Japan's ancient system of morality, called *kannagara-no-michi* ("way of the gods"). Inheriting Norinaga's explanation of Shinto, called Fukkō Shinto, Hirata Atsutane regarded Japan as the centre of the world; and fiercely upholding the conception of Japan as a divine country (shinkoku), he strongly advocated reverence for the Imperial house. Hirata's thought, along with Mitogaku (school of Mito), provided the ideological foundation for the "Revere the Emperor! Drive out the Barbarians!" (*sonnō-jōi*) movement of the late Tokugawa period.

Western studies

Studies of Euroyean modern science also arose, termed *yōgaku* ("Western learning") or *rangaku* ("Dutch learning"). A great stimulus to the concrete development of Western studies was provided by the publication, in 1774, of the Kaitai *shinsho*, a translation by Sugita Gempaku and others of an anatomical book imported from the Low Countries. Thus, Western studies became progressively more vigorous, centred on medicine. But as the crisis in the feudal system grew more serious, many scholars of Western studies criticized the seclusion policy, making the bakufu nervous. The persecution of Watanabe Kazan, Takano Chōei, and other representative scholars of Western studies in the *bansha-no-goku* incident (1839), which resulted from a conservative plot within the bakufu, seriously undermined Western studies in Japan. Thereafter, as consciousness of the foreign threat grew stronger, Western studies came to place heavy emphasis on the field of military technology.

Other philosophers also appeared who repudiated feudal society. And6 Shōeki, in his *Shizen shin-eidō*, portrayed an ideal society in which all people equally engaged in farming, without social distinctions or exploitations. This shows the extent to which the common people were troubled by the contradiction of feudalism. Even if Shōeki is considered exceptional, other men successively appeared with an anti-feudal world view directly or indirectly influenced by empirical science and Western studies. Miura Baien of Kyushu called his learning Rational Studies (Jōrigaku); it contained a dialectical method of thought. Hiraga Gennai, the son of a foot soldier of the Takamatsu domain, disliked the restricted life of the warrior; he became a masterless samurai (*rōnin*) and thought and acted freely. He advocated that Japan prevent the outflow of gold and silver by promoting domestic production and exchanging these products for foreign goods. Because this view agreed with Tanuma Okitsugu's desire for a production development policy, Hiraga was employed by Okitsugu and sent to Nagasaki. While experimenting with such things as functional dynamos and thermometers, Gennai gave full play to his genius by cultivating sugar and carrots, producing Dutch-style pottery, and surveying and developing mines in various provinces of the country. He also produced masterpieces as a dramatist. An attitude of naïve materialism grew in his thought.

The work of Shiba Kōkan and Yamagata Bantō also appeared at this time. Kōkan is known as the pioneer of etching in Japan; hut he also displayed an evolutionistic attitude, repudiating the feudal status system on the ground that both the emperor and the beggar were similar human beings, thus insisting on human equality. Bantō was chief manager to a rich Osaka merchant and had financial ability. In his work *Yume-no-shirō*, he reconstructed the Japanese history in the age of gods from the world view of natural science. He, too, had a strong feeling for human equality, saying "both in ancient times and at the present, there is no upper or lower among human beings."

Growth of popular knowledge

The common people of the Tokugawa period, by their production and labour, were progressively reared in empirical knowledge, and their self-awareness as human beings became stronger. At the outset of the period only a handful of upper class farmers, such as the shdya or *nanushi*, were literate; by the end of this period, with the exception of the very lowest class, farmers were all at least partly literate. This is also connected with the diffusion of the *terakoya* (temple schools), the educational organs of the common people. But in any case, it reflected the growth in popular knowledge extending over 300 years. This is also shown by the fact that "village conflicts" (*murakata sōdō*) became more fierce in the later part of this period, as the farmers sought to censure the improper acts of village officials and to make the village more democratic. The fact that the leadership in these conflicts was taken by the middle and lower class farmers was a natural phenomenon. In comparison with the present, however, it is undeniable that, in the Tokugawa cities and villages, backward social tendencies were still firmly rooted. The cause of this backwardness may be attributed to the collective abuses brought about by the feudal system, but the role played by such religions as Shintō and Buddhism cannot be overlooked.

Buddhism especially had strong powers of regulation over the lives of the common people. Among the various Buddhist sects in this period, the Jōdo, Jōdo Shin, Zen, and Nichiren made striking advances because their temples were guaranteed in their privileged status by the implementation of the terauke ("temple certificate") system of the bakufu. Besides their previous role in conducting funeral rites and masses, they now had charge of registration and census; thus grafted onto the livelihood of the people, their operation was generally stabilized. The tendency also became strong for the people to seek genze *riyaku—i.e.*, to pray for happiness during their lifetime, such as for commercial prosperity or restoration of health—and not to wait for happiness after their death. In response to this, various ceremonies were conducted and efforts made to swell the financial income of the temples. Among such ceremonies, the most important were *kaichō* and *tomitsuki*. *Kaichō* consisted of allowing the people to worship a Buddhist image that was normally kept concealed and not generally displayed. Gradually this came to be performed by moving the image out to other cities and villages. Tomitsuki was an officially authorized lottery, and in Edo the raffles of the Yanaka Tennō-ji, the Yushima Tenjin, and the Meguro Fudō (better known as the Ryusen-ji) temples were famous. Among the Buddhist priests who profited from the trend of the times, there were some who led loose private lives, providing the Confucian scholars with examples for demanding that Buddhism be stamped out.

Various sorts of popular faiths flourished in the cities and villages. Shugen-do had, as its special characteristic, prayers by itinerant monks (yainabushi) for curing illness or bringing happiness, but in its teachings, while centred on Buddhism, it also contained beliefs drawn from Shintō and elsewhere to meet the religious feelings of the people. A new faith in healing spirits arose from the view that human suffering could be cured only by men who had suffered the same hardships themselves, and in the late Tokugawa period a development took place toward faith in living gods (ikigami) who could respond to the various requests of the common people and who became revered as sect founders (*kyōso*). These included the Kurozumi-kyō sect, founded by Kurozumi Munetada, the Konkō-kyō sect of Kawate Bunjirō, and the Tenri-kyō sect of Nakayama Miki. People like Nakayama Miki reflected the confused social conditions of the late Tokugawa period, and their advocacy of *yo-naoshi*, or relief of the world by social reform, had political undertones. Influenced by the popularity of the cult of Shintō shrines, periodic pilgrimages to Ise, called *okage-mairi* or *nuke-mairi*, became popular. Pilgrimages were also made to Ise by groups of several hundreds of thousands of common people. Among such pilgrims, there were those who had political hopes for *yo-naoshi* in common with the faiths of the sect founders.

Literary styles of the Edo period

The maturity of Edo culture. The early 19th century brought the city culture that had arisen in Edo to full maturity in learning and craftsmanship. This Edo culture was supported by rich townsmen and warriors, and it was also widespread among the townsmen generally. Literary styles took various forms; representative authors are Santō Kyōden in the *sharebon*, or genre novel, Jippensha Ikku in the kokkeibon, or comic novel, and Takizawa Bakin in the yomihon, or regular novel. They examined in detail such things as the townsman's way of life, customs, conceptions of beauty, and ways of thinking. In the world of art, Ukiyo-e came to completion both in form and content and established its position as popular art. Nishiki-e, wood-block printing in many colours, was invented by Suzuki Harunobu and entered its golden age with the prints of sumo wrestlers by Tōshūsai Sharaku and of courtesans by Kitagawa Utamaro. In the last years of the Edo period, the masters of wood-block landscape prints, And6 Hiroshige and Katsushika Hokusai, opened new boundaries and even exerted an influence on foreign art. As a result of links with the city culture, scholarship arose even in local towns and villages, where crafts and products with distinctive local colouring were supported by landlords and merchants. A national culture emerged from these city and local cultures and became the foundation of a modern culture that developed during and after the Meiji period.

But signs of stagnation and corruption also appeared in some aspects of Edo culture—a reflection of the crisis in the *bakuhan* system. The crisis had reached new levels by the third decade of the 19th century. A great famine, lasting several years, dealt a savage blow to the impoverished villages, and farmer uprisings and city riots reached unprecedented peaks. In 1836 an uprising took place in the Gunnai district of Kai Province (Yamanashi Prefecture), then under direct *bakufu* control, and spread westward to the Kuninaka plain across the Sasago pass, by this time numbering more than 50,000 adherents. Dividing into two groups, insurgents attacked the residences of nearly 500 prominent men and for a time reduced the centre of Kai to anarchy. How deeply the bakufu was shocked by this can be seen in the sentencing of 562 persons to crucifixion for their part in the uprising. A year later in the city of Osaka, also under direct *bakufu* control, Oshio Heihachirō, a former city official, raised a revolt to overthrow the officials and rich merchants and to relieve the poor; although the uprising was speedily suppressed, the bakufu was greatly astonished that a former faithful official would lead a revolt. In addition to these domestic disturbances, the European powers began to press more heavily upon Japan. The Opium War (1839–42) arose between the Ch'ing dynasty of neighbouring China and the British, and foreign gains following this war filled the bakufu authorities with a sense of crisis. Tokugawa Nariaki, the lord of the Mito domain, urged the bakufu to make definite political reforms; he called the domestic resistance of the common people the domestic anxiety and the pressure of the foreign nations the foreign anxiety.

Internal and external threats

THE TEMPO REFORM AND THE COLLAPSE OF THE BAKUHAN SYSTEM

Faced by this threatening situation in both domestic and foreign affairs, the chief senior councillor, Mizuno Tadakuni, instituted a reform program; this Temp6 (the name of the year period from 1830 to 1844) reform lasted only from 1841 to 1843, however. Tadakuni set in order the regulations for the government officials and encouraged them to practice frugality and the literary and military arts. He also aimed to restore the farming villages devastated by the great famine. Tadakuni was stricter than earlier reformers in that he planned to force temporary residents in Edo to return to their home villages, and to restrict the commercial-goods production of the farmers to make them produce rice. In order to lower the price of commodities in the cities, he applied detailed regulations to the townsmen. He further ordered that the kabu-nakama, or merchant and artisan guilds, be dissolved because he regarded them as the cause of rising commodity prices. Tadakuni planned to reclaim by military means the Imba swamp (Chiba Prefecture) so that food supplies could be conveyed to Edo from the provinces of Hitachi and Kazusa (Ibaraki and Chiba prefectures) if Edo Bay were blockaded by foreign ships. Plans for the defense of Edo Bay took concrete form and included the seven islands of Izu. Tadakuni also promulgated a land-requisition (*agechi*) order to bring daimyo and hatamoto domains surrounding Edo and Osaka under direct bakufu control; the main object of this was the defense of Edo, but it was also a plan to supplement the finances of the bakufu. The *agechi* order was finally withdrawn, however, in the face of fierce opposition from the daimyo, hatamoto, and people of the domains concerned, and, as a direct result of this failure, Tadakuni was driven from his seat among the senior councillors in 1845. Tadakuni predicted that because of his reforms, the Tokugawa shogunate would keep its government for another 30 years, and it was in fact 30 years after his reforms that the bakufu's downfall took place (1867). In the same Temp6 period, administrative reforms were carried out in many of the domains, and the reforms of the powerful southwestern domains, such as Chōshū and Satsuma, are noteworthy. These reforms adopted the policy of "wealth accumulation and military strength" (*fukoku-kyōhei*); they improved the domain finances, and the resulting surpluses were used to modernize the military armaments of the domains. In the powerful domains of southwestern Japan, middle and lower class samurai came forward as reformers, replacing the previous conservative officials. The way was thus gradually being prepared for the emergence of the leaders of the Meiji Restoration (1868) and of modern Japan.

In 1845, when Abe Masahiro replaced Mizuno Tadakuni as chief senior councillor, various movements appeared that can be called reactions to the Temp6 reform. One of these involved the commissioners for the shogun's capital, with close connections with the Edo merchants, and tried to restore the kabu-nakama guilds. In 1851 an order was finally issued for the revival of the nakama. But the guilds included those (sometimes called *kumiai*) that had arisen since the dissolution of the Temp6 era *kabu-nakama*, and their controlling power over city markets was thus extremely restricted. The confrontation of the city merchants with the village producers and local merchants over monopoly of commercial-goods circulation routes had grown more fierce, and the former had been forced to yield further.

Foreign rivalries in the opening of Japan

Thus, the domestic reaction to the Temp6 reform was comparatively calm, and the major stumbling block facing the bakufu was the foreign problem. The Netherlands, the only European power trading with Japan, saw that if Britain succeeded in forcing Japan to open the country, it would lose its monopoly; so the Dutch now planned to seize the initiative in opening Japan and to thus turn the situation to their own advantage. In 1844 the king of The Netherlands, William II, sent a diplomatic mission urging the bakufu to open the country, but Abe and the bakufu rulers refused this suggestion. Visits by foreign ships, however, increased progressively. In 1844, 1845, and 1846, British and French warships visited the Ryukyu Islands and Nagasaki to request commercial relations. In response, the bakufu, in 1845, established the new office of Kaibo-gakari for coastal defense and various diplomatic posts. The defense system of Edo Bay was also revived, the number of domains on guard duty was increased, and new gun emplacements were built. In 1848 it was decided not to revive the order to drive away foreign ships, which had been rescinded during the Temp6 reform, but that extensive military preparations should be made.

Rumours had long circulated among the various foreign countries that the United States government would send an expeditionary fleet to Japan. In 1846 Comdr. James Biddle of the American East Indian fleet appeared with two warships in Uraga Harbour (Uraga-kō) and held consultations on the question of commercial relations. When refused by the bakufu, he left empty-handed. The United States, however, eagerly desired ports for fuel and provisions for its Pacific merchant and whaling ships and would not give up trying to open Japan. But the bakufu had for so many years kept its place as overlord of the political regime by strictly maintaining the ancestral law of seclusion that it could not muster up the resolution to step forward and open the country. The opening of Japan was thus postponed until the last possible moment as a result of the vacillation of its rulers, and the opening had to be effected unilaterally by foreign pressure, backed by massive naval strength. (K.Ma.)

III. Japan since 1850

THE MEIJI RESTORATION

The term restoration is commonly applied to the political changes that returned power to the throne during the reign of Mutsuhito, who took the reign name Meiji ("enlightened rule," 1868–1912). The slogan return to antiquity (*fukkō*) made it possible to interpret sweeping changes as traditional in motivation. Actually, the Meiji changes constituted a social and political revolution that began before 1868, and political innovations ended only with the promulgation of a constitution in 1889.

Fall of the Tokugawa. The arrival of the foreigners in the 1850s provided a new issue for domestic politics and a new measure for the effectiveness of the feudal administration. When it became clear that the Shogun was unable to protect Japan from the barbarians and that his concessions to them were made in spite of their known repugnance to the Imperial court in Kyōto, the two shogunal boasts of loyalty to and protection for the court proved spurious. The slogan *sonnō-jōi* ("Revere the Emperor! Drive out the Barbarians!") was first raised by men who sought to influence shogunal policy and then taken up by others who wanted to embarrass the Tokugawa. The Shogun's ratification of the Harris Treaty and of others that followed was carried out in the face of strong opposition from the Kydto court, and it brought to the surface antagonisms that had developed during the long years of peace and study. They centred in the Tokugawa house of Mito, which had done much to sponsor Confucian scholarship. The Mito daimyo made vigorous attempts to involve the Kyōto court in affairs of the shogunate with a view to establishing a nationwide program of preparedness. For this he was punished by the head of the Edo council of elders, Ii Naosuke. In 1860 Ii was assassinated by men from Mito and Satsuma, an act that inaugurated years of violence. Many of those who took part were young samurai from Edo; their swords availed little against the foreigners' guns, but they took a heavy toll of political enemies.

Reaction against foreign intervention

Years of extremism followed. The Tokugawa shogunate, anxious to rally support among its feudatories and to help them to prepare their defenses, relaxed its controls and regulations. In many fiefs young enthusiasts tried to push their feudal superiors into a less cautious and more strongly antiforeign position. It soon became obvious to most men that expelling the foreigners by force was impossible. Antiforeign acts provoked stern countermeasures and diplomatic indemnities, which tightened the foreign hold on the country. After the bombardment of Kagoshima and Shimonoseki, there could be no doubt of the foreigners' military superiority. Thereafter, the slogans of antiforeignism and exclusion continued to be used chiefly as a means of obstructing and embarrassing the shogunate. The Edo policy makers were forced to make surface concessions to the antiforeign elements, which aroused the hostility and distrust of the treaty powers. After the arrival of the British minister Sir Harry Parkes in 1865, Great Britain, in particular, began to tire of negotiating with a shogunate that stood between it and the Kyōto court and began to consider ways of dealing directly with the latter. It gradually became clear that ultimate authority lay in Kyōto.

In some fiefs the young extremists found themselves unable to budge their superiors from their conservative positions. From Chōshū (now part of Yamaguchi Prefecture), foreign shipping in the straits of Shimonoseki was shelled in 1863. This drew a foreign bombardment the following year. Samurai opinion grew so vehement that after the fief authorities submitted to Tokugawa discipline in 1864, a swift military coup brought to power, as the daimyo's counsellors, a group of men who had led the radical antiforeign movement. But they were no longer blindly antiforeign; several had secretly travelled to England. Their aims had become national—to overthrow the shogunate and create a new government headed by the Emperor. The same men developed new militia units based on Western training methods and arms. Chōshū became the centre for discontented young samurai from other fiefs who were impatient with their leaders' caution. In 1866 Chōshū allied itself with the great fief of Satsuma in expectation of a Tokugawa attempt to crush all tozama daimyo opponents and erect a centralized despotism with French help.

Anti-Tokugawa alliance

The Tokugawa armies were successfully repulsed at Chōshū in 1866, causing the shogunate to lose power and prestige. The death of Shogun Iemochi in 1866 brought to power as the last shogun Hitotsubashi Keiki (or Yoshinobu), who was aware of the pressing needs for national unity. He spurned suggestions that he seek French help to put down his enemies. When he was urged by a lord of Tosa to resign his powers, he did so rather than risk a full-scale assault by Satsuma and Chōshū, confident that as lord of eastern Japan he would emerge as an important figure in whatever new political organization should develop. But the young Meiji emperor, who had succeeded to the throne in 1867, was guided by several nobles in close touch with leaders of Satsuma and Chōshū, and the last shogun was manoeuvred into a choice between giving up his land, which would risk revolt from his vassals, or appearing disobedient, which would justify punitive measures. Keiki's armies advanced on Kyōto, only to be defeated. Satsuma, Chōshū, and Tosa units, now the Imperial army, advanced on Edo, which was surrendered without a battle; fighting continued to the north until the summer of 1869, but the Tokugawa cause was doomed. In January 1868 the principal lords were summoned to Kybto to learn of the restoration of Imperial rule. The next year the capital was moved to Edo, from this time on Tokyo, and the building of the modern state began.

From feudal to modern state. The Meiji government was dominated by the Satsuma, Chōshū, and court figures who had outmanoeuvred the Shogun. They were convinced that Japan would need a unified national government in order to achieve military and material equality with the Western powers. Most of them, like Kido Kōin and It6 Hirobumi of Chōshū and Saigō Takamori and Okubo Toshimichi of Satsuma, were young samurai of modest rank, but they did not represent in any sense a class interest. Indeed, their measures destroyed that class. In order to gain backing for their policies, they enlisted leaders of fiefs with which they had worked—from Tosa, Saga, Echizen—and maintained their cooperation with such court nobles as Iwakura Tomomi and Sanjō Sanetomi.

Missions to the West

The cooperation of the impressionable young emperor was essential. It was taken for granted that Western strength depended on constitutionalism, which produced national unity; on industrialization, which produced material strength; and on a well-trained military. The new slogan of the day became *fukoku-kyōhei* ("rich country, strong arms"). Knowledge was to be sought in the West, the goodwill of which was essential if the unequal treaties were to be revised. Therefore, a number of missions to the West were organized. In 1871 Iwakura Tomomi led a large number of his fellow government leaders to visit Europe and the United States. The experience gained abroad strengthened convictions already formed as to measures of modernization that would be required.

Abolition of feudalism. The Meiji leaders began with measures to lessen the feudal decentralization on which they blamed much of Japan's weakness. In 1869 the Satsuma, Chōshū, Tosa, and Saga leaders persuaded their daimyo to return their lands to the throne; other lords hastened to follow suit. The court took steps to regularize and make uniform administration in the fiefs, but it appointed the former lords as new governors. In 1871 the governor daimyo were summoned to Tokyo, and feudalism was declared abolished. The approximately 300 fiefs became 72 prefectures and three metropolitan districts; this number was later reduced by one-third. For the most part, the daimyo lost contact with administration, and, although they were rewarded with titles in a new European-style peerage, set up in 1884, their political importance was slight.

It was necessary to end the complex system of social stratification that had existed under feudalism; yet, it was difficult to make arrangements for the samurai, who numbered, with dependents, almost 2,000,000. In 1869 the old hierarchy was replaced with a new and simpler division whereby court nobility and feudal lords were termed aristocracy (kazoku); upper and middle samurai, shizoku; other samurai, *sotsuzoku* (a rank soon abolished); and all others, commoners (heimin), including the previously unlisted pariah groups. The samurai were given pensions equal to a part of their old income. When the regime found these pensions too heavy for its treasury to carry, the pensions were changed to interest-bearing but nonconvertible bonds. During the same years, the samurais' special hairdo was discouraged; the wearing of swords, the former badge of class, was later banned.

Samurai revolts

Many of the bonds were soon squandered, because few warriors had had occasion to develop commercial aptitude, and the inflation that accompanied government expenditures lessened their value greatly. In 1873, moreover, a nationwide conscription was instituted, depriving the samurai of their traditional monopoly of military service. There were a number of samurai revolts, the most serious in the southwest, which had led in the restoration movement and where warriors previously had reason to expect the greatest rewards. Some revolts, as in Chōshū, were expressions of discontent against administrative measures that deprived samurai of their importance, while in Saga the dissidents championed a proposed foreign war to employ samurai.

The last and greatest revolt came in Satsuma (1877), led by the restoration hero Saigō Takamori. The new conscript levies were hard pressed to defeat Saigō, and the government had to enlist former samurai and empty its military academies in order to put down the revolt. But the revolts merely expressed regional discontents and were never coordinated. Even in the case of the Satsuma war, the loyalties of most of the Satsuma men in the central government remained with the Imperial cause.

Fiscal and economic policies. In 1873 land surveys were begun to determine the amount and value of land on the basis of average yield in recent years, and a tax in money of 3 percent of the value was then set as the land tax. Out of the same surveys came certificates of ownership of land for farmers, who were also released from feudal controls. The land measures involved basic changes, and there was widespread confusion and uncertainty among the farmers, frequently expressed by short-lived revolts and demonstrations. The establishment of private ownership, along with measures to promote new technology, fertilizers, and seeds, soon produced a rise in recorded agricultural output. The land tax, supplemented by printed money, was the principal source of the government's income for several decades.

Although hard pressed for money, the government also began a program of industrialization, seen as essential for national strength. Aside from military industries and strategic communications, it was carried out in private hands, although the government set up pilot plants to provide encouragement. Trade and manufacturing benefitted from the new national market and legal security, although unequal treaties made it impossible to protect industries with tariffs until 1911.

Rise of the zaibatsu

In the 1880s fear of excessive inflation resulted in a decision to sell most of the new plants to private investors—usually persons who had close relations with government officials. A small number of individuals came to dominate many enterprises; they were known as the zaibatsu, or financial cliques. With tremendous opportunities and few competitors, the same firms appealed in enterprise after enterprise. Their aims were close to those of the government leaders, and there were often close friendships between them. The House of Mitsui, for instance, had close relations with Meiji leaders, while that of Mitsubishi was founded by a colleague of the restoration leaders.

National loyalties. Equally important for building a modern state was the development of national loyalties. True national unity required the propagation of new loyalties among the masses, previously inarticulate and powerless. The early restoration government, influenced by a Shintō revival, elevated a bureau of Shinto, the state cult, to the highest position in the new political hierarchy and strove to replace Buddhism with a strong cult of the national deities. Christianity was legalized in 1873, with great reluctance, at the urging of the Iwakura mission, and thereafter it seemed important to bolster traditional outlooks without risking foreign condemnation by forcing a state religion upon the Japanese. The education system proved an ideal vehicle for ideological orientation. A system of universal education was announced in 1872. For a time its organization and philosophy were Western inspired; but during the 1880s, as the government leaders saw their countrymen turning to Western ideas and learned of a new nationalist orientation of schooling in Europe, the Japanese system was altered to include emphasis on "ethics," and in 1890 the Imperial Rescript on Education laid out the lines of Confucian and Shintō ideology, which constituted the moral content of later Japanese education. Thus, loyalty to the emperor, who was hedged about with Confucian teaching and Shintō reverence, became the centre of a citizen's ideology. Meanwhile, to avoid charges of indoctrination, the state distinguished between this secular cult and actual religion; in this way, the leaders could permit "religious freedom" while requiring a form of worship as the patriotic duty of all Japanese subjects. This uniform system of mass education was also utilized to project into the nation at large the ideal of samurai loyalty that had been the heritage of the ruling class.

Constitutional movement. It was widely believed that constitutions provided much of the unity that gave Western countries their strength, and Japanese leaders were

Charter oath of 1868

eager to bring themselves abreast of the world in this respect. A government plan (1868) experimented with a two-chamber house, but it proved unworkable because the government leaders preferred to have their own way. The Emperor's charter oath (April 1868), however, committed the government to seek knowledge and wisdom throughout the world, to abandon customs of the past, to allow all subjects to fulfill their proper aspirations, and to allow popular opinion to influence their decisions.

Creation of political parties. To these statements of intent were added protests from below. A democratic movement grew out of a split in the leadership group over government policy. Itagaki Taisuke and other leaders of the Tosa faction combined with members of the Saga fief in 1873. Their demands for a punitive expedition against Korea, the obscurantist government of which had insulted Japanese envoys, had been refused because domestic reforms were to come first, and they resigned their positions. Instead of championing the old order, however, Itagaki and his friends called for a popular assembly so that future decisions would reflect the will of the people (by which they initially meant their fellow samurai) and thus preserve unity. Itagaki and his Tosa followers developed discussion and mutual-help groups and, gradually growing in political confidence and ability, organized themselves on a national basis as the Liberal Party (Jiyū-t6) in 1881. It should be noted that the movement had only a narrow social and regional base at this time and that its purposes were to promote effective national unity rather than tolerance of diversity and dissent.

When the remaining Meiji leaders were asked to submit their opinions on constitutional problems in 1881, Ōkuma Shigenobu, a Saga leader, revealed a relatively liberal draft instead of first submitting it for the scrutiny of his colleagues. He also revealed sensational evidence of corruption in the disposal of government assets on the island of Hokkaido. Ōkuma was forced out of the government and he organized the Progressive Party (Kaishintō) in 1882. Itagaki's Liberal Party had a predominantly rural backing of former samurai and village leaders, many of whom objected to government taxation policies; Ōkuma's party had an urban base and attracted support in the business and journalistic worlds.

Economic distress of the late 19th century

The Meiji contribution. The Emperor promised that a constitution would be instituted in 1889; the parties were urged to await the Imperial decisions quietly. The constitution was prepared behind the scenes by a commission headed by It6 Hirobumi. The period of constitution writing coincided with one of intense economic distress as the government sought to stem the inflation caused by the spending of the 1870s. But deflationary measures caused hardship in the countryside and provided a situation in which party agitation could easily kindle direct action. Several instances of this and severe government repression in the form of police and press controls forced the parties to dissolve temporarily in 1884. Itagaki travelled to Europe and returned more than ever convinced of the need for national unity in the face of Western condescension.

It6 Hirobumi also travelled to Europe for help in preparation of the new constitution. In Germany he found what seemed an appropriate balance of imperial power and constitutional forms that seemed to offer modernity without sacrificing effective control. As a balance to a popularly elected house, It6 first organized a new European-style peerage in 1884. The government leaders, military commanders, and former daimyo were given titles and readied for future seats in a house of peers. A Cabinet system was installed in 1885, and a privy council, designed to judge and safeguard the constitution, was set up in 1888. It6 resigned as premier to head the council.

The constitution was completed by 1889, and elections for the lower house were held to prepare for the initial diet, which met in 1890. The constitution took the form of a gracious grant by the Emperor, and it could be amended only upon Imperial initiative. Its provisions were couched in general terms. Rights and liberties were granted "except as regulated by law." If the diet refused to approve a budget, the previous year's could be followed. The emperor was "sacred and inviolable"; he commanded the armies, made war and peace, and dissolved the lower house at will. Effective power thus lay with the executive, which could claim to represent the Imperial will. The education rescript of 1890 was to guarantee that future generations accept the Imperial will and authority without question. In spite of its antidemocratic features, the constitution provided a much greater area for dissent than had previously existed. The lower house could initiate legislation. Private property was inviolate, and freedoms, even when subject to legislation, were greater than none at all. The budgetary arrangements meant that increased support for the military could be had only with Diet approval. Initially, a tax qualification of 15 yen limited the electorate to about 500,000; this was lowered in 1900 and 1920, and in 1925 universal manhood suffrage came into effect. The government leaders had difficulty controlling and manipulating the lower house, despite their power of dissolution and their resources for intimidation and bribery, thus illustrating that the constitution had altered the political picture. And the party leaders' cooperation with their erstwhile enemies when given a reasonable amount of prestige and patronage showed their large areas of agreement with the Meiji oligarchs.

End of the Restoration

The constitution ended the Meiji Restoration and revolution. The government leaders soon retired behind the scenes to influence the political world as elder statesmen (*genro*) and acted to maintain and conserve the balance of ideological and political institutions they had worked out.

IMPERIAL JAPAN

Foreign affairs. Achieving equality with the Western powers had been one of the major goals since the beginning of the Meiji period. Treaty reform, designed to end the foreigners' judicial and economic privileges provided by extraterritoriality and fixed custom rates, had been attempted as early as the Iwakura mission of 1871; but the Western powers refused to consider it until Japanese legal institutions had been brought into line with those of the West. Japanese attempts at compromise arrangements in the 1880s were denounced by the press and opposition groups in Japan. The treaty provisions for extraterritoriality were formally changed in 1894, after the completion of the Meiji institutional reforms; tariff autonomy came into effect in 1911.

Relations with China. Asian matters took second place to internal problems during most of the Meiji period. In 1874 a punitive expedition was launched against Formosa to chastise the aborigines for murdering Ryukyuan fishermen. This lent support to the Japanese claim to the Ryukyus, which had been under Satsuma influence in Tokugawa times; the islands were incorporated into Japan in 1879 despite Chinese protests. Adventures in Korea, however, although espoused by nationalists and, on occasion, by liberals, were avoided by the government, which was conscious of its need for internal reform and foreign approval. The matter was complicated by a growing Chinese readiness to resist Japanese interference in the affairs of Korea, China's most important tributary state. The Chinese were alert to the danger of Japanese gains. Incidents in 1882 and 1884 that might have led to war with China and Korea were instead settled by compromise. In 1885 China and Japan agreed that neither would send troops to Korea without first informing the other.

Sino-Japanese War (1894–95)

By the early 1890s Chinese influence in Korea was clearly becoming predominant. In 1894 Korea requested Chinese assistance in putting down a rebellion. When the Chinese informed Tokyo of this, Japan quickly rushed troops to Korea and, after the rebellion was crushed, showed no inclination to withdraw. Hostilities between Chinese and Japanese forces broke out first at sea and then in Korea in July–August 1894. The Japanese navy sank or captured much of the northern Chinese fleet, and a peace treaty was negotiated at Shimonoseki between Japan and China on April 17, 1895. Both powers recognized the independence of Korea; China ceded Formosa, the Pescadores Islands, and the Liaotung Peninsula, granted Japan all rights enjoyed by European powers, and made significant new economic concessions; new treaty ports were opened, and Japan received an indemnity of 200,000,000 taels in gold in two installments. A subsidiary treaty of commerce (1896) gave Japan freedom to engage in trade, manufacture, and industry in China's treaty ports and provided for tax exemption within China for all goods so manufactured. Japan thus marked its own emancipation from unequal treaties by imposing even harsher terms on its neighbour. But the European powers were not yet prepared to welcome Japan as a full equal in the imperialist scramble in China. Germany, France, and Russia forced Japan to return the Liaotung Peninsula to China. In 1898 Russia forced China to grant it the lease of that peninsula with its important naval base at Port Arthur. The war thus demonstrated that the Japanese could not maintain Asian military victories without Western sufferance. Nevertheless, the war proved a tremendous source of prestige for Japan and brought the Tokyo government much internal support; it also strengthened the hand of the army in national affairs.

Relations with the West. Instead of accepting Japanese leadership, Korea sought the help of the Russians as a counterweight. During the Boxer Rebellion in China (1900), Japanese troops took a major part in the allied expedition that rescued foreign nationals in Peking, but Russia occupied south Manchuria, thereby strengthening communications with Korea. Realizing the need of protection against a possible combination of European enemies, the Japanese government began talks that led to an Anglo-Japanese Alliance (1902). Each signatory agreed to aid the other in the event of an attack by two or more powers, while remaining neutral if the other was at war with a single power. The Tokyo government was thus prepared to take a firmer line with respect to Russian advances in Manchuria and Korea. In 1904 Japanese ships attacked the Russian fleet at Port Arthur without the formality of a declaration of war. Japanese arms were everywhere successful; the most spectacular victory was in Tsushima Strait, where the ships of Adm. Tog6 Heihachirō destroyed the Russian Baltic fleet. But Japanese armies were strained to their utmost, and it was with relief that Japan accepted the United States president Theodore Roosevelt's offer of good offices for the negotiations that led to peace, signed at Portsmouth, New Hampshire, September 5, 1905. Japanese primacy in Korea was recognized, and Russia surrendered to Japan its economic and political interests in south Manchuria (including the Liaotung Peninsula) as well as the southern half of the island of Sakhalin. The victory over Russia' altered the balance of power in Asia, and Japan's ability to cope with a great European power accelerated the development of nationalist movements in Asia. Within Japan, however, the failure to secure a Russian indemnity to cover the costs of the war made the treaty unpopular.

Russo-Japanese War (1904–05)

Japan as a major military ***power.*** After the conclusion of the war, Japanese leaders now had a free hand to guide the course of reform in Korea, and Korean resistance was met with force. It6 Hirobumi, sent to Korea as resident general, forced through treaties that gave Korea little more than protectorate status and forced the abdication of the Korean king. In 1909 It6 was assassinated, and the following year Korea was formally annexed to Japan. Korean liberties and resistance were crushed under military rule. By the end of the Meiji period, Japan had thus achieved equality with the West and had, in fact, become the strongest military and imperialist power in Asia.

Japan had abundant opportunity to use its new power in the years that followed. World War I found the Western powers fully occupied in Europe. Japan took part in the war in compliance with the Anglo-Japanese Alliance, but generally it limited its participation to the seizure of the German Pacific Islands and the German holdings on the Shantung Peninsula. When China pressed for return of these, the Japanese government presented the so-called Twenty-one Demands in January 1915. China reluctantly agreed to extend the duration of the Manchurian leases and to joint control of steelworks and ironworks in central China. The German Shantung holdings were to be settled by agreement between Japan and Germany at the time of the peace treaty; subsequently, Japan agreed to hand back the territory in return for further commercial privileges. China promised not to alienate harbours in Fukien province to any other power without Japanese approval. But the Chinese resisted group V of the Twenty-one Demands, which would have reduced China to the status of a Japanese ward. Japan had gained abundant opportunity for the exploitation of Manchuria, but the ill feeling aroused by the negotiations, together with Chinese chagrin at failure to recover its losses in the Treaty of Versailles, cost Japan any hope of Chinese friendship. Subsequent Japanese sponsorship of corrupt warlord regimes in Manchuria and North China helped to confirm the anti-Japanese nature of modern Chinese nationalism.

The Twenty-one Demands

Japanese behaviour when the Allies intervened in Siberia in 1918 after the Bolshevik Revolution furthered the impression of Japanese rapacity. One of the principal reasons for a disarmament conference in Washington, D.C. (1922), was an attempt to lessen Japanese influence. A network of treaties was worked out that placed restraints on Japanese ambitions while guaranteeing Japanese security. Japan, Great Britain, the United States, and France concluded a four-power pact that replaced the Anglo-Japanese alliance; a five-power pact (with Italy) for disarmament set limitations for capital ship construction on a ratio of five for Great Britain and the United States to three for Japan. Parallel guarantees against fortifying advanced bases assured Japan of safety in Pacific waters. A nine-power pact would, it was hoped, protect China from further unilateral demands. Japan subsequently agreed to retire from Shantung, and, shortly afterward, Japanese armies withdrew from Siberia and northern Sakhalin. In 1925 a treaty with the Soviet Union extended recognition and ended active hostilities.

The mid-1920s thus saw the end of Japan's great surge forward in the Pacific and brought hope that a new quality of moderation and reasonableness, based on the absence of irritating reminders of inferiority and weakness, might characterize Japanese policy.

Constitutional government. The inauguration of constitutional government in 1890 saw a vigorous and often obstreperous opposition in the lower house of the Diet, and it was probably general determination to prove that parliamentary institutions could work in Japan that forced the party and government leaders to cooperate sufficiently to make the system work. The first Cabinets, led by Yamagata Aritomo, Matsukata Masayoshi, and Itō, attempted to maintain the principle that the government, which in their view represented the emperor, should be aloof from parties and that it was the duty of the lower house to approve government requests. This policy failed because the parties desired to increase their power and patronage and therefore sought Cabinets responsible to the lower house. It was only the Sino-Japanese War that produced the kind of unity the constitution makers had envisaged. In the years that followed, the oligarchs formed alliances with the two parties, usually exchanging a Cabinet seat or two for support in the lower house. These arrangements proved unsatisfactory as party leaders soon raised their sights. In 1898 Itagaki and Ōkuma combined forces to form a single party, the Kenseitō, and, because this ruled out successful administration by a nonparty Cabinet, they were allowed to form a government. But their alliance was of short duration, as long-standing animosities and jealousies enabled antiparty forces among the bureaucracy and oligarchy to force their resignation within a few months.

Lower-house opposition

A discernible division now developed among the dwindling group of Meiji leaders. Yamagata Aritomo dominated the army and much of the bureaucracy. During the two years he held power after the fall of the Kenseitō Cabinet, he strengthened legal and institutional safeguards against political-party rule and secured an Imperial ordinance that service ministers should be career officers on the active list; this gave the army or navy power to break a Cabinet. Partly in reaction, It6 Hirobumi. also of

Chōshū, formed his own political party in 1900, the Rikken Seiyiikai, enlisting most of the former followers of Itagaki. Thereafter, practical political goals of power and patronage softened the hostility between oligarchs and politicians.

After 1901 both It6 and Yamagata retired from active participation in politics; until 1913 Cabinets were led by their protégés Saionji Kimmochi and Katsura Tarō. Basic decisions of politics and policy, however, continued to be made by the core group of elder statesmen, who advised the Emperor on all important decisions and selected prime ministers by rotating power between the two principal factions. Saionji was the last to be recruited into this extraconstitutional body.

Changing political patterns

With the death or enfeeblement of the first generation of leaders, the pattern of political manipulation changed. No subsequent group could match the prestige the Meiji leaders had enjoyed. The Meiji emperor died in 1912 and was succeeded by a son who took the reign name Taishō ("great righteousness," reigned 1912–26); but mental illness prevented him from approximating his father's fame. The growth in prestige and power of businessmen found expression in their control of the political parties and resulted in an increasing role for professional party politicians. The genro's last attempt to seat Katsura in 1912 ended in failure, while his successor, Adm. Yamamoto Gombe, was discredited by scandals in naval procurement. Ōkuma Shigenobu emerged from retirement to head a Cabinet during World War I and was succeeded by a military Cabinet under Gen. Terauchi Masatake. In 1918, however, discontent with Terauchi's reactionary posture and administrative incompetence combined with the rising power of the party professionals to bring about the appointment of Hara Kei (Hara Takashi) as prime minister. Hara was the first nontitled person to hold that office, and his appointment marked the first party Cabinet. His assassination in 1921 cut short his cautious efforts to reduce the power of the military and the bureaucracy and to extend the franchise. After several short-lived Cabinets, a successful party Cabinet was organized in 1924 by Katō Takaaki. The army was reduced in size; moderate social legislation was enacted; and universal manhood suffrage extended the franchise to 14,000,000 voters. Meanwhile, Japan avoided stronger steps in China's civil war and pursued a conciliatory course with Russia, despite demands from nationalists, who utilized alleged outrages in China and a discriminatory United States Immigration Act of 1924 to warn of the futility of appeasing or cooperating with other powers.

Reaction to social movements

But, as the parties grew in power, they tended to look to bureaucrats for leadership. The businessmen who supported the parties and the bureaucrats who led them shared a fear of the social movements that followed industrialization and the importation of foreign ideas. A growing labour movement had already been checked by a special police law introduced in 1900. This was strengthened under Katō in 1925, as conservatives generally began to fear subversion in labour and tenant movements. A small Communist party was organized by a group of intellectuals in 1922, and a general interest in Marxist thought contributed to more fears of subversion. Under the Meiji constitution, party governments had to make their peace with the military, with the house of peers, and with the conservatives close to the throne; whatever ideas for reform they had therefore had to be worked out with the utmost caution. Frequently, the Diet found itself virtually powerless, and this encouraged corruption and disorders in the chamber, which did little to win popular respect for the machinery of representative government. There were no institutional changes that enabled a government to be firmly based on popular support. The Meiji Constitution was so ambiguous in its provisions for the executive that the party prime ministers could achieve little unless they secured, through compromise. the cooperation of forces antagonistic to democratic government.

Social change. Changes in the social and intellectual scene outstripped those in the political. Many of them were related to the development of industry. After the Treaty of Shimonoseki the government utilized the Chinese indemnity to subsidize the Yawata Iron and Steel Works, which were established in 1897 and began production in 1901. Yawata depended on China for its ores. After 1900 Japan's population exceeded the capabilities of domestic food production so that there was need for import of food as well. Growing textile and other consumer-goods industries expanded to meet Japanese needs and to earn credits required for the import of raw materials. Heavy industry was encouraged by government-controlled banks, which provided needed capital. Strategic industries, notably steel and the principal rail trunk lines, were in government hands, but most new growth was in the private sector, albeit somewhat concentrated in the zaibatsu financial and industrial giants.

Social protest

The enlarged urban population produced movements of social inquiry and protest. In 1895 the industrial labour force numbered about 400,000. Several efforts to organize socialist movements speedily met with police repression. Peace preservation laws were passed in 1900 and 1925, and in 1928 it became a capital crime to agitate against private property or Japanese state policy (*kokutai*). In 1903 a small group organized the *Heimin shimbun* ("Commoner's Newspaper"); it published *The Communist Manifesto* and opposed the Russo-Japanese War in the name of the workers of Russia and Japan before being forced to cease publication. The labour and Socialist movements gained strength after World War I, but leadership was usually theoretical and doctrinaire, with little real contact with the workers. Police repression and the difficulties of organizing a labour force of diverse industrial empires such as those of Mitsui and Mitsubishi also retarded the labour movement. Meanwhile, the increasing confidence and power of management came to influence and at times control the political parties. The Katō Cabinet of 1924–26 was sometimes referred to as a Mitsubishi Cabinet.

In the countryside the principal reflection of the new trade patterns was an additional impetus to silkworm production to augment the farmers' income. Farm villages also provided the bulk of the labourers for the new industries, and farm daughters were found in many textile plants. The early 20th century was not a time of agricultural prosperity. Farmers were handicapped by growing fragmentation of holdings and increasing tenantry. The rising number of tenants resulted in the growth of tenant organizations, especially during and after World War I. Government efforts to encourage voluntary reform brought only a law for mediation of disputes in 1924. But a financial panic in 1927 aggravated rural conditions and indebtedness even before the collapse of the U.S. silk market in 1929 spelled disaster for the farmers and workers alike.

The most lasting social changes were found in the great metropolitan centres, where a growing labour force and new middle-income groups were concentrated. The Tokyo–Yokohama area was devastated by the great Kantō earthquake in 1923, and its reconstruction as a modern metropolis symbolized the growth and orientation of the urban society. The currents of enthusiasm during and after World War I were uniformly international and largely U.S. in inspiration. Western music, dancing, and sports became popular, and rising standards of living and expectation produced the need for more and better higher education. The participation of women in office work in the new enterprises and the rise of a feminist movement, however unsuccessful, marked the beginning of changes in the family system.

The educated class grew in numbers and in vigour. Currents of thought included Western-style democracy and the new radicalism of the Soviet Union; the Marxist influence went far beyond the ranks of the struggling Communist Party—which was, in any event, soon crushed by the police. Political liberalism was championed by the University of Tokyo figure Yoshino Sakuzii, who formed student and intellectual groups the title of which—shinjinkai (New Peoples Association)—symbolized the self-conscious break with tradition. Minobe Tatsukichi, a distinguished constitutional theorist, introduced the idea that the emperor was an organ of the state and not the sole source of sovereignty. Such men faced sharp criticism and had, in time, to resign their positions, but they had great influence and symbolized and stimulated advanced currents of thinking.

Attempts to break with tradition

The base for these new currents was precarious. Politically and institutionally, no advances—beyond the universal manhood suffrage of 1925—were scored, while, under the peace-preservation laws of 1928, a special police corps was established to seek out "dangerous thoughts." Economically, the urban classes were dependent upon the continuance of the favourable trade patterns of the 1920s. When the Great Depression at the end of the decade wrecked Japan's foreign markets and removed the possibility of the villagers' augmenting rice income with that of silk and when the irresponsibility and occasional corruption of Diet representatives contrasted with poverty elsewhere in Japanese society, many were prepared to listen to charges that the political-party government, dominated by selfish zaibatsu interests, had neglected Japan's markets in China, imperilled morality and decency at home, and allowed subversive trends to flourish, while the politicians reaped personal gains.

The rise of the militarists. The notion that expansion through military conquest would solve Japan's economic problems gained currency during the Great Depression of 1929. A key argument advanced to support it was that Japan's population had grown from 30,000,000 at the time of the Meiji Restoration to almost 65,000,000 in 1930; each year the problem grew worse, and the imports of needed foodstuffs increased. It was also argued that emigration to many areas was cut off because of discrimination against Oriental peoples. Efforts made by Japan and China to secure a racial-equality clause in the League of Nations covenant had been frustrated by Western statesmen who feared the anger of their constituents. So the argument ran that no recourse could be expected without resort to force.

Military distrust of party government

To these economic and racial arguments was added the military's distrust of party government. The Washington conference had allowed a smaller ratio of naval strength than the navy had desired, and the government of Prime Minister Hamaguchi Osachi in 1930 accepted and gained approval of the London Naval Conference limitations of cruiser strength over military objections. The Katō government had cut the army strength. Many service leaders had also bridled under Japan's moderation during the Chinese Kuomintang northern expedition in 1926 and 1927, and they would have preferred a much stronger stand. The Seiytikai Cabinet under Prime Minister Tanaka Giichi reversed that policy by intervening in Shantung in 1927 and 1928. Tanaka was forced out in 1929 and replaced by Hamaguchi, under whom the policy of moderation returned. It seemed to many that such vacillation earned Japan ill will and expensive boycotts in China without gaining any advantage.

Anti-government thought. Many military leaders resented the restrictions that civilian governments had placed upon them, and their power was considerable. It would be wrong to attribute such views to all or even most of the high command, but enough army officers in particular held this position to furnish a possible focus for dissatisfaction among other groups in Japanese society. The idea of the frugal, selfless samurai was peculiarly useful as a contrast to the stock characterization of the selfish party politician.

These economic pressures and political misgivings were exploited by civilian ultranationalists who opposed parliamentary government as "un-Japanese." Since Meiji times a number of rightist organizations had formed, dedicated to the theme of internal "purity" and external expansion. They sought to preserve what they thought unique in the Japanese spirit and fought against excessive Western influences. Some originated in the Meiji period, when nationalists had felt obliged to work for a "fundamental settlement" of differences with Russia; the Kokuryūkai (Black Dragon or, more accurately, Amur Society) was one such, while others, such as the Seisantō (Productivity Society), were keyed to labour and social problems. The Kokusūikai, or National Purity Society, worked to preserve national purity, while the Ketsumeidan, or League of Blood, was terrorist. Their leaders were against political parties, big business, acculturation, and Westernization, and, by allying with other rightists, they alternately terrorized and intimidated their presumed opponents. A number of business leaders and political figures lost their lives, and the assassins' success in publicizing and dramatizing the virtues they claimed to embody had a considerable importance in the ethos of the troubled 1930s. It is clear, however, that the terrorists never had as much influence as they claimed or as the West believed.

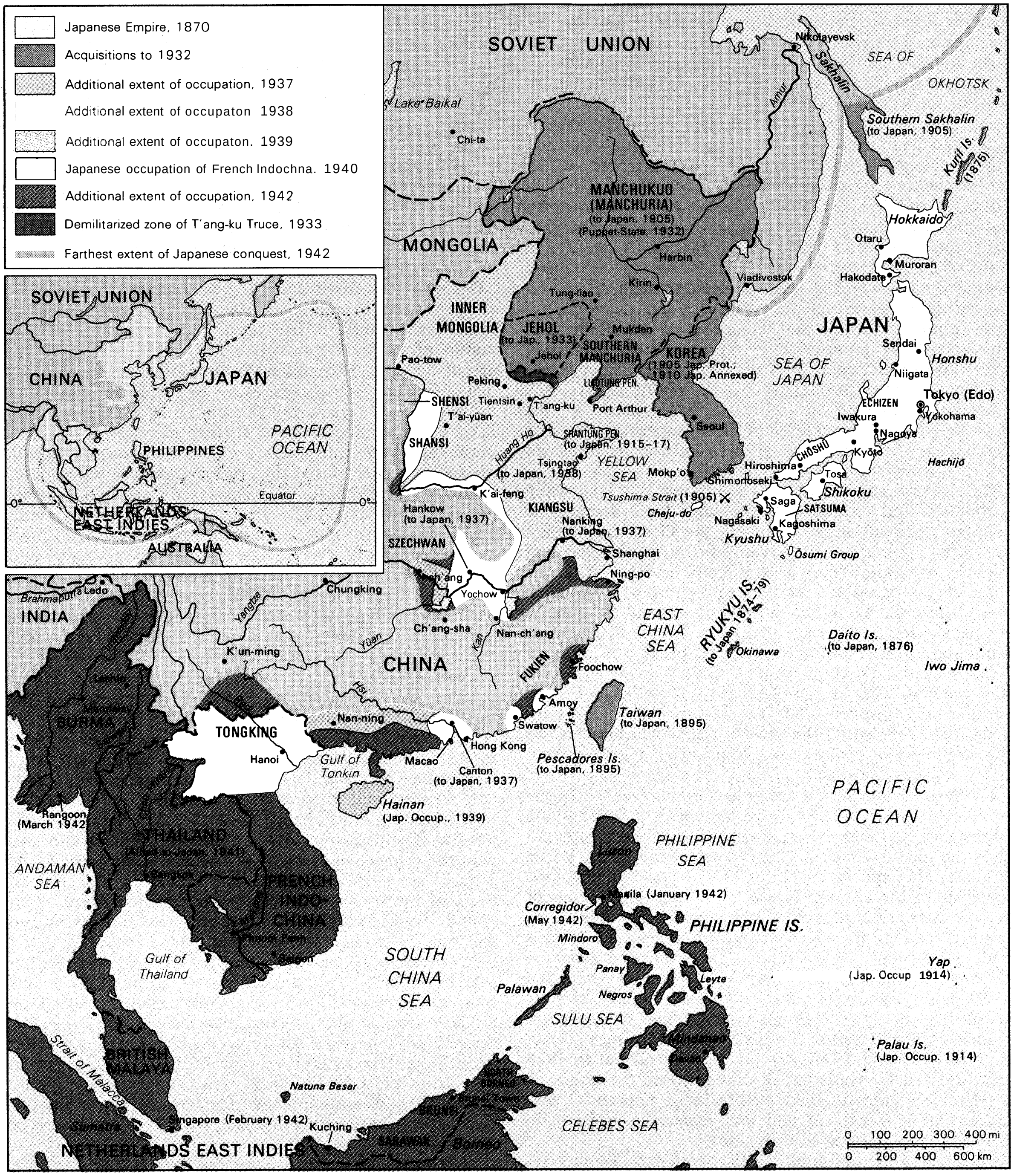
The principal force against parliamentary government was provided by junior military officers. Largely from rural backgrounds, distrustful of their senior leaders, ignorant of political economy, and contemptuous of the urban luxuries of politicians, the officers were ready marks for rightist theorists. Many of them were animated by goals that were national-socialist in character. Kita Tkki, a former Socialist and former member of the Kokuryiikai, wrote in his outline plan for the reconstruction of Japan that the Meiji Constitution should be set aside in favour of a revolutionary regime advised by "national patriots" and headed initially by a military government, which should nationalize major forms of property, limit wealth, end party-government and peerage systems, and prepare to grasp the leadership of a revolutionary Asia. Kita helped persuade a number of young officers to take part in the violence of the 1930s, and in large measure their plots were designed to create a disorder so great that military government would follow.

Aggression in Manchuria

Anti-government acts. The Kwantung Army, which invested the Kwantung (Liaotung) Peninsula and patrolled the South Manchurian Railway zone, provided a rich harvest of officers keenly aware of Japan's continental interests and prepared to take steps to further them. They hoped to place the civilian government in an untenable position and to force its hand. The Tokyo terrorists similarly sought to change foreign as well as domestic policies. The pattern of direct action in Manchuria began with the murder in 1928 of Marshal Chang Tso-lin, the warlord ruler of Manchuria. The action, though not authorized by the Tanaka government, helped bring about its fall. Tanaka's Cabinet, however, dared not investigate and punish those responsible, and this convinced extremist officers that their lofty motives would make retribution impossible. The succeeding government of Prime Minister Hamaguchi showed intentions of restraining military activists and powers, however, and the next plots centred around plans for replacing civilian government altogether; Harnaguchi was mortally wounded by an assassin in 1930. In March 1931 a coup involving highly placed army generals planned to terrorize civilian politicians into a grant of martial law, was abandoned because of disagreement among the principals.

Terrorism and revolts

On September 18, 1931, came the Manchurian incident, which launched aggression in East Asia. A Kwantung Army charge that Chinese soldiers had tried to bomb a South Manchurian Railway train (which arrived at its destination safely) resulted in speedy and unauthorized capture of Mukden, followed by the occupation of all Manchuria. The civilian government in Tokyo could not stop the army, and even army headquarters was not always in full control of the field commanders. Prime Minister Wakatsuki Reijirō gave way, in December 1931, to Inukai Tsuyoshi. Inukai's plans to stop the armies by Imperial intervention were frustrated. In 1932 naval officers took the lead in extremism; a terrorist attack in Tokyo in May took the life of Inukai; the terrorists failed to secure a proclamation of martial law. The army, however, now announced that it would accept no party Cabinet. To forestall its desires for power, the last genro, Saionji, suggested retired Adm. Saitō Makoto as prime minister. Plotting continued, culminating in a revolt of a regiment about to leave for Manchuria. In February 1936 several outstanding statesmen (including Saitō) were murdered; Prime Minister Okada Keisuke escaped when the assassins mistakenly shot his brother-in-law. For sev-



Japanese expansion in the late 19th and 20th centuries.

eral days the rebel unit held much of downtown Tokyo. When the revolt was put down on February 29, the ringleaders were quickly arrested and executed. Therewith, the influence of the young extremists, often referred to as the Imperial Way faction (Kōdō-ha), gave way before that of the more cautious Control faction (Tōsei-ha), which had less sweeping plans for internal reform but shared many of the foreign-policy goals of the young fanatics.

The only possible source of prestige sufficient to thwart the military lay with the throne. The senior statesmen, however, were cautious lest they imperil the Imperial institution itself. The young emperor Hirohito had succeeded to the rule in 1926, taking as his reign title Shōwa. His outlook was more progressive than that of his predecessors; he had travelled in the West, and his interests lay in marine biology (of which the ultranationalists disapproved in one whose role it was to embody the Japanese mystique). The palace advisers feared that a strong stand by the Emperor would only widen the search for victims and might lead to dethronement of the monarch. As international criticism of Japan's aggression grew, many Japanese rallied to the support of their soldiers.

The road to World War II. Each advance by the military extremists gained them a new compromise concession by more moderate elements in the government and brought greater foreign hostility and distrust. Rather than attempt to thwart the military, the government agreed to reconstitute Manchuria as the "independent" state of Manchukuo. The last Manchu emperor of China, Hsiian-t'ung, was first declared regent and then enthroned as emperor in 1934. Actual control lay with the Kwantung Army, however; all key positions were held by Japanese, with surface authority for cooperative Chinese and Manchus. A League of Nations committee recommended in October 1932 that Japanese troops be withdrawn, Chinese sovereignty in Manchuria recognized, and a large measure of autonomy granted to Manchuria. The League called upon member states to withhold recognition from the new puppet state. In March 1933 Japan

Creation of Manchukuo

formally withdrew from the world body. Thereafter, Japan poured technicians and capital into Manchukuo, exploiting its rich resources to establish the base for the heavy-industry complex that was to undergird the new order in East Asia.

Consolidation and expansion. In north China, boundary areas were consolidated in order to enlarge Japan's economic sphere. In early 1932 the Japanese navy precipitated an incident at Shanghai in order to end a boycott of Japanese goods there; but Japan was not yet prepared to challenge other powers for control of central China, and a League of Nations commission arranged terms for a withdrawal in May 1932. Frustrated naval officers returned to Tokyo to carry out the violence that killed Inukai on May 15. A move southward from Manchuria into Jehol in January 1933 led to the T'ang-ku Truce in May, whereby a demilitarized zone was set up between Peking and the Great Wall. This brought the fighting to a temporary close. In 1934, Japan made it clear that it would brook no interference in its China policy and that Chinese attempts to procure technical or military 'assistance elsewhere would bring Japanese opposition.

Further external ambitions, however, had to wait for the resolution of domestic crises. The military revolt in Tokyo in February 1936 marked the high point of the extremist faction and the consolidation of power by the Control faction within the army. Finance minister Takahashi Korekiyo, whose policies had brought Japan out of its depression slump, was killed, and his opposition to further inflationary spending was thus stilled. When further efforts by the palace advisers to defer full power for the military failed, the leadership went to the popular but ineffective Konoe Fumimaro, scion of an ancient court family (June 1937). The same period saw the kidnapping of Chiang Kai-shek by Chinese border armies at Sian in December 1936 and his agreement to consolidate Nationalist and Communist efforts into an anti-Japanese front. To this was added evidence that the Japanese people were not yet prepared to renounce their parliamentary system. In the spring of 1937, general elections showed a startling strength for a new Social Mass Party, which received 36 seats out of 466, and a heavy majority for the two parties (the Seiyūkai and its rival the Minseitō), which had combined forces against the government and its policies. The time seemed ready for new efforts by civilian leaders, but the field armies anticipated them.

The "China incident"

In July 1937 Japanese troops opened fire on Chinese units near Peking; thus began the "China incident." Japanese armies took Nanking, Hankow, and Canton despite vigorous Chinese resistance; to the north, Inner Mongolia and the provinces of Shansi and Shensi were invaded but not fully invested. On discovering that the Nationalist government, which had retired to Chungking in Szechwan, refused to compromise, the Japanese installed a more cooperative regime at Nanking in 1940.

Alliances. Meanwhile, Japan had signed the Anti-Comintern Pact with Germany in November 1936 and later with Italy. This was replaced by the tripartite pact in September 1940, by which Japan was recognized as the leader of a new order for Asia, and the three signatories agreed to assist each other if any one was attacked by a power not then at war. This was directed against the United States, since the Soviets and Nazis were then allied; the Soviet Union was, indeed, invited to join in the tripartite pact later in 1940.

Japanese relations with the Soviet Union were considerably less cordial than those with Germany. The Soviets consented, however, to sell their Chinese Eastern Railway holdings to the South Manchurian Railway in 1935, thereby strengthening Manchukuo. In 1937 the Soviet Union signed a nonaggression pact with China, and in 1938 and 1939 Russian and Japanese armies tested each other in two full-scale battles along the border of Manchukuo. The Soviet–Nazi pact of August 1939, however, was followed by a neutrality pact between the Soviet Union and Japan in April 1941.

The German–Japanese tie was never a close or effective one. Both parties were limited in their cooperation by distance, distrust, and claims of racial superiority. The Japanese were uninformed about Nazi plans for attacking the Soviet Union, and the Germans were not told of Japan's plans to attack Pearl Harbor. Nor, despite formal statements of rapport, did Japan's state structure approach the totalitarianism of the Nazis. A national-mobilization law (1938) gave the Konoe government sweeping economic and political powers, and in 1940, under the second Konoe Cabinet, the Imperial Rule Assistance Association was established to merge the political parties into one central organization; yet, the institutional structure of the Meiji Constitution was never altered, and the wartime governments never achieved full control over interservice competition. The Imperial Rule Assistance Association never succeeded in mobilizing all segments of national life around a leader. The emperor remained but a symbol, albeit an increasingly military one, and no *Fuhrer* could compete without endangering the national polity. Wartime social and economic thought retained important vestiges of an agrarianism and familism that were in essence premodern rather than totalitarian.

Relations with the democratic powers

War in Europe. Japan's relations with the democratic powers deteriorated steadily. The United States and Great Britain did what they could to assist the Chinese Nationalist cause. The Burma Road permitted the transport of minimal supplies to Nationalist forces. Constant Japanese efforts to close this route were successful briefly in 1940, when the British felt they could not risk a second war. But anti-Japanese feeling had strengthened in the United States, especially after the sinking of a U.S. gunboat, the "Panay," in the Yangtze River in 1937. In 1939 U.S. secretary of state Cordell Hull denounced the 1911 treaty of commerce with Japan, and thus embargoes became possible in 1940. Franklin D. Roosevelt's efforts to rally public opinion against aggressors included efforts to stop Japan, but, even after the outbreak of war in Europe in 1939, public opinion in the United States was averse to courting war by stronger measures.

The European war presented the Japanese with tempting opportunities. After the Nazi attack on Russia (1941), the Japanese were torn between German urgings to join the war against the Soviets and their natural inclination to seek richer prizes from the colonial powers to the south. In 1940 Japan had occupied northern Indochina in an attempt to block access to supplies of the Chinese Nationalists, and in July 1941 it announced a joint protectorate with Vichy France over the whole colony. The way was prepared for further moves in Southeast Asia.

The United States reacted to the occupation of Indochina by freezing Japanese assets and declaring an embargo on oil to Japan. The government was faced with the alternatives of withdrawing from at least Indochina and possibly China or seizing the sources of oil production in the Netherlands East Indies. Negotiations with Washington were carried on under the second Konoe Cabinet. Konoe was willing to withdraw from Indochina, and he sought a personal meeting with Roosevelt, hopeful of some U.S. concessions or favour with which he might convince his military leaders. But the U.S. State Department refused to agree to a meeting without prior Japanese concessions. Pressed by his war minister, Gen. Tōjō Hideki, Konoe resigned in October 1941 to be succeeded by Tōjō. Secretary of State Hull refused to agree to Japan's "final offer": Japan would withdraw from Indochina after China had come to terms in return for U.S. promises to resume oil shipments, cease aid to China, and unfreeze Japanese assets. With Japan's decision for war made, the negotiators received instructions to continue to negotiate. Preparations for the opening strike against the U.S. fleet at Pearl Harbor were already in motion. The Japanese military elected to try to establish, through a "new order in East Asia," a co-prosperity sphere in which Japan, as the centre of an industrial bloc comprising Manchuria, Korea, and North China, would draw from the rich colonies of Southeast Asia the raw materials it needed, while inspiring them to friendship and alliance by destruction of their previous masters. But, in practice, "East Asia for the Asiatics," Japan's slogan, turned out to mean "East Asia for Japan."

Preparations for war

World War II and defeat. The attack on Pearl Harbor (December 7, 1941) achieved complete surprise and success. It also unified U.S. opinion and determination to see the war through to a successful finish. The Japanese had expected that, once they fortified their new holdings, a reconquest would be so expensive in lives and treasure that it would discourage the "soft" democracies. Instead, the U.S. fleet was rebuilt with astonishing speed, and the chain of defenses was breached before these riches could be effectively tapped by Japan.

Early successes

The first years of the war brought Japan great success. Japanese troops occupied Manila in January 1942, although Corregidor held out until May; Singapore fell in February, the Netherlands Indies and Rangoon in early March. The Allies had difficulty maintaining communication lines to Australia, and the loss of the British battleships "Repulse" and "Prince of Wales," added to the U.S. Pacific fleet disaster, seemed to promise the Japanese Navy freedom of action. Tōjō grew in confidence and popularity and began to style himself somewhat in the manner of a Fascist leader. But the U.S. Navy had not been permanently driven from the South Pacific. The Battle of Midway in June 1942 cost the Japanese fleet aircraft carrier strength it could ill afford to lose, and the battle for Guadalcanal in the Solomons ended with Japanese withdrawal in February 1943.

After Midway, Japanese naval leaders came secretly to the conclusion that Japan's outlook for victory was poor. When the fall of Saipan in July 1944 brought U.S. bombers within range of Tokyo, the Tōjō Cabinet was replaced by that of Koiso Kuniaki. Koiso formed a supreme war direction council designed as a link between the Cabinet and the high command. It was becoming evident that Japan was losing the war, but no group had a program acceptable to the military leaders. There were also grave problems about breaking the news to the Japanese people, who had been told only of victories. Great fire-bombing raids in 1945 brought destruction to every major city except the old capital of Kyoto; but the generals were still determined to continue the war, confident that a major victory or a protracted battle would be the best way of gaining honourable terms. The Allied talk of unconditional surrender provided a good excuse for continuing the fight.

In February 1945 the Emperor met with a group of senior statesmen to discuss steps that might be taken. When U.S. landings were made on Okinawa in April, the Koiso government fell. The problem of the new premier, Adm. Suzuki Kantarb, was not whether to end the war but how best to do so. The first plan advanced was to ask the Soviet Union, with which Japan was still at peace, to intercede with the Allies. The Soviet government, however, was planning to enter the Pacific war, and reply was delayed while Soviet leaders took part in the Potsdam Conference in July. The Potsdam Declaration of July 26 offered the first ray of light with its statement that Japan would not be "enslaved as a race nor destroyed as a nation."

The atomic bombs

On August 6 and 9 the atomic bombs took their toll of life in Hiroshima and Nagasaki. On August 8 the Soviet Union declared war and on the 9th marched into Manchuria, where the Kwantung Army could offer only slight resistance. The Japanese government attempted to gain as its sole condition for surrender a qualification concerning the maintenance of the Imperial institution; after the Allies agreed to respect the will of the Japanese people, the Emperor insisted on surrender. The Pacific war came to an end on August 14. The formal surrender was signed on September 2 in Tokyo Bay aboard the USS "Missouri."

Military extremists made an unsuccessful attempt to prevent the radio broadcast of the Emperor's announcement to the nation. There were a number of suicides among the military officers and nationalists who felt themselves dishonoured, but the Emperor's prestige and personal will, once expressed, sufficed to bring an orderly transition. To increase the appearance of direct rule, the Suzuki Cabinet was replaced by that of Prince Higashikuni Naruhiko.

Investigators concluded that neither atomic bomb nor Soviet entry was central to the decision to surrender, although they probably helped to advance the date. It was decided that submarine blockade of the Japanese islands had brought economic defeat by preventing exploitation of Japan's new colonies, sinking merchant tonnage, and convincing Japanese leaders of the hopelessness of the war. Bombing brought the consciousness of defeat to the people. Destruction of the Japanese Navy and Air Force jeopardized the home islands. Japan's largest armies, however, were never defeated, and this was responsible for the army's eagerness to fight on. Occupation found Japan's cities destroyed, its stockpiles exhausted, and its plants gutted. The government stood without prestige or respect. An alarming shortage of food and rising inflation threatened what remained of national strength. The time was ripe for changes.

AFTER WORLD WAR II

SCAP and its objectives. Gen. Douglas MacArthur, Supreme Commander, Allied Powers (SCAP), received his orders for the occupation of Japan through U.S. military channels; a Far Eastern Commission made up of Pacific war Allies was to make policy in Washington and provide consultation through an Allied Council for Japan, which sat in Tokyo. In fact the occupation became an American affair, and SCAP grew into a large headquarters, SCAP worked through the Japanese government. In the early years it provided direct instructions frequently, but with time suggestions were made more discreetly.

Purposes of the occupation

Occupation purposes had been held out in general terms in the Potsdam Declaration, with its promises of freedoms and statements of intent to remove undemocratic tendencies; those purposes were defined more precisely in a document that was worked out by the U.S. departments of state, war, and navy. Its emphases were on demilitarization, so that Japan would not again become a danger to peace; on democracy, so that (although the U.S. was not to impose any particular form of government) a responsible Japanese government would guard individual rights; and on encouragement of the Japanese to develop an economy that would be adequate for peacetime needs.

MacArthur responded enthusiastically to the idea of a demilitarized and democratic Japan and utilized the complex pattern of authority under which he functioned to ward off interference from Washington or from the Allies. He rushed constitutional reform to anticipate outside suggestions and first ignored and then delayed moves for partial Japanese rearmament after the Cold War changed U.S. priorities. The occupation measures created an open historical situation in which new forces could and did rise; SCAP measures proved lasting in cases where they coincided with trends already present within Japanese society, and those measures were vital to Japan's recovery as a free society and economy.

The early months of the occupation saw SCAP move swiftly to remove the principal supports of the militarist state. The armed forces were demobilized; State Shinto was disestablished; nationalist organizations were abolished and their members removed from important posts. Also removed from active roles were all persons prominent in wartime organizations and politics, including commissioned officers of the armed services and all high executives of the principal industrial firms. In Tokyo an international tribunal tried General Tōjō and other war leaders, sentencing seven to death, 16 to life imprisonment, and two to shorter terms. Millions of Japanese were repatriated from the former colonies and from Southeast Asia. The Home Ministry, which had controlled wartime Japan through its appointive governors and national police, was abolished, and the Education Ministry was deprived of its sweeping powers to control compulsory education. Because central control and military influence were being attacked by a military government that needed centralized powers in order to be effective, the occupation's role was often contradictory. Geography and economic rationality reinforced the logic of centralization, and many of the moves toward decentralization were modified or reversed a half decade later.

Political reform. SCAP informed leading Japanese citizens that constitutional reform should receive first attention. Between October 1945 and February 1946 a Cabinet committee headed by Matsumoto Jōji prepared revisions of the Meiji Constitution, but the changes were few and superficial. MacArthur's government section rushed a new draft and submitted it to the Japanese government as a basis for further deliberations. Despite the misgivings of conservative statesmen, it was approved by the Emperor and submitted for amendment to the first postwar Diet, which had been elected in April 1946 (in these elections women had voted for the first time). The constitution, slightly modified, was promulgated on November 3, 1946, and went into effect on May 3, 1947. Its preface stated the intention of the Japanese people to ensure peaceful cooperation with all nations and the blessings of liberty for themselves and their descendants. The constitution included a 31-article bill of rights, and Article 9 renounced war as a "sovereign right of the nation" and pledged that "land, sea, and air forces, as well as other war potential, will never be maintained." The Emperor was described as the "symbol of the state and of the unity of the people, deriving his position from the will of the people with whom resides sovereign power." Earlier, on January 1, 1946, the Emperor had renounced claim to divinity. The constitution provided for a bicameral Diet, with the greatest power for a House of Representatives, the members serving four-year terms. The old peerage was dissolved and the House of Peers replaced by a House of Councillors, the members serving six-year terms. The prime minister was to be chosen by the Diet from its members, and an independent judiciary had the right of judicial review.

New constitution

The new constitution thus reversed the Meiji pattern and contributed to responsible government by specifying the locus of executive authority. Despite its hasty preparation and foreign inspiration, it gained wide public support. Although the ruling conservatives desired to revise it after Japan regained its sovereignty in 1952, and an official commission favoured changes in 1964, the decreasing likelihood of mobilizing the two-thirds majority of the Diet necessary to secure approval for changes gradually made the possibility moot. By then, Article 9, with its renunciation of war, had in any case been partially evaded with the ruling that self-defense (hence self-defense forces) was an inherent right.

By a peace treaty that went into effect in 1952, moreover, elements of the political pattern had already changed, and subsequent governments showed their ability to modify by administrative actions a constitution that remained unchanged. Decentralization in some fields had proved expensive and inefficient. The police, for instance, while less centralized than in the days of the Home Ministry, had returned to substantially national organization. Despite the announced goals of local decentralization, changing patterns of communications and administration had shown the logic of incorporating many small units of administration into larger units, a trend particularly marked in the countryside, where villages and towns merged to form a more rational tax structure. Article 9 had been compromised by a decision taken by SCAP to form the National Police Reserve of 75,000 men in 1950, during the Korean War. The force, later (1954) renamed the Self-Defense Forces, came to number 240,000 in the late 1970s.

Nevertheless, the basic principles of the constitution of 1947 enjoyed support among all factions in Japanese politics. Executive leadership was a chief asset of the new institutions. With the abolition of the competing forces that beset the premiers of the 1930s, the postwar prime ministers found themselves in charge of the administration and, with rearmament, of the armed forces as well. Thus, responsible leadership gradually replaced the ambiguous claims of Imperial rule of earlier days.

Economic and social changes. SCAP'S political democratization was reinforced by economic and social changes that were designed to create interest groups prepared to use their new rights to safeguard and protect the new political and economic structure. Changes in the countryside, in industry, and in social legislation all had the same purpose of breaking or weakening the old pattern of hierarchic control that had distinguished the "family-state" ideal of the Meiji leaders.

Agriculture. In agriculture the occupation established a program of land reform to convert tenants into owners. Tenancy had risen after World War I, and only ineffectual measures had been taken against it by the prewar Japanese government, perhaps partly because the political parties were based largely on a rural electorate dominated by landowners. Japan's wartime governments, however, made important changes in land relationships. In their attempts to achieve national unity and equal sacrifice, they had created agricultural associations to collect all rice. Absentee landlords had received a lower rate of payment, and the tenant's relations with his landlord had become much less important. Moreover, peasant sons in the armed services had been able to send home part of their pay, while the shortage of labour had made it possible for many members of farm families to secure gainful employment in factories. Thus, important preliminaries in rural well-being, not least among them the opportunities of the black market, had taken place prior to SCAP's instructions to the Japanese government to prepare a land-reform plan.

SCAP land reform

The government plan proved inadequate in the opinion of occupation authorities, and in the spring of 1946 a SCAP plan was drawn up; it became law in October. By its terms village and prefectural land commissions were elected with tenant, owner-farmer, and landlord representation to select land for purchase and eligible purchasers from among tenants. The government then bought the land at pre-inflation prices and sold it to the tenant. Four years later, the reform had changed the ownership of more than two-thirds of Japan's cultivated acreage, and advantageous tax and price arrangements had enabled the majority of the new owners to pay for their land. The average family holding remained about 2% acres (one hectare), and, in view of the larger population and the change in laws covering primogeniture, there was an increase of the fragmentation of land. Nevertheless, the reform helped produce a striking rise in rural prosperity.

Business. Although initial Allied plans contemplated exacting heavy reparations from Japan, the unsettled state of other Asian countries that were to have been recipients brought reconsideration. With the exception of Japanese assets overseas and a small number of war plants, reparations were very nearly limited to agreements worked out between Japan and each of its Asian victims after the peace treaty signed in San Francisco in 1951.

SCAP and the *zaibatsu*

Similar moderation marked the course of planning for deconcentration of the great ***zaibatsu*** firms. At first, they were considered Japan's chief potential war makers, but they later came to be seen as essential elements in economic recovery. Of 1,200 concerns marked for investigation and possible dissolution in 1948, only 28 were broken up by SCAP, though the major units of the ***zaibatsu*** empires—holding companies—were dissolved and their securities made available for public purchase. New legislation sought to enforce fair trading and to guard against return to monopolies. Taxes on the profits of war wiped out many large fortunes and affected all large concentrations, while capital levy, inheritance, and graduated income taxes were designed to equalize the tax burden. The extension to the business world of the removal of wartime leaders prevented any action by the senior executives of 250 concerns during most of the occupation years. By 1950 extensive changes, although far short of those initially proposed, had taken place in the industrial world. The large banks, however, had not been broken up, and they proved to be the centres for a measure of reconsolidation in the years after the occupation ended.

Labour. The balance of economic power was also affected by measures that produced a strong and vigorous labour movement, which contested with management for political and economic primacy. After the Home and Welfare ministries were dissolved, a new Labour Ministry was established in 1947. The release of all political prisoners, carried out in the early months of the occupation, freed the core of the Japan Communist Party, and most of those released turned their attention to organizing the labour movement, hoping to use it as a path to power. Laws on trade unions and labour relations, modelled on New Deal legislation in the United States, were passed, and soon a strong union movement appeared, led by men with political ambitions. When a general strike was announced for February 1947, with the avowed purpose of overthrowing the government, SCAP issued an injunction against it. Thereafter, occupation policy was concerned with reconstruction and no longer exclusively with liberation, and steps against inflation, political radicalism, and Communist control of labour unions followed. Under the Socialist Cabinet of Prime Minister Katayama Tetsu there was emphasis on labour education, and in July 1948 SCAP ordered the government to take steps to deprive government workers—including members of the communications unions—of the right to strike. A new labour organization, the General Council of Japanese Trade Unions (Sōhyō), was sponsored as a counterweight and gradual replacement for the Congress of Industrial Labour Organizations (Sambetsu), which had become dominated by the left. After 1951, however, Sōhyō too became increasingly anti-government and anti-American. Although some of the occupation measures deprived labour of useful weapons for fighting its way to power, the strength of organized labour, expressed through the Socialist Party, remained significantly different politically from what it had been before the war. The government's failure to carry through a police law designed to curb labour radicalism and sabotage in 1958 demonstrated the powerful support that labour could command in the Diet, in the press, and in public opinion generally.

Labour unions

Social reforms. The social legislation of the postwar era saw energies and hopes long repressed by the Japanese government spring to full flower. The civil code, which had reinforced the power of the male head of the family with numerous legal supports, was rewritten to allow for equality between the sexes and joint inheritance rights. Women were given the right to vote and to sit in the Diet. The abolition of the peerage, which had been created in Meiji days, symbolized the modernization of society.

The years after the peace treaty, which became effective on April 28, 1952, saw a number of changes in the pattern of occupation reforms. The land system remained unaltered, and rural well-being was higher than it had ever been. Industrial reconsolidation reunited many of the earlier financial empires, which required maximum efficiency in order to compete on international markets. Proportionately, however, the old firms had not regained their former power. Social legislation, moreover, had created such strong interest groups that prospects of substantial reversals of the postwar changes seemed slight. For many of the newly liberated generation, the new order was symbolized by the marriage of Crown Prince Akihito and a commoner, Shoda Michiko, in 1959.

International relations. Japan's return to international relations at the end of the occupation found it stripped of its conquests and even deprived of some of its own territory. The Republic of China on Taiwan, the People's Republic of China on the Chinese mainland, the Republic of Korea (South Korea), and the Democratic People's Republic of Korea (North Korea) all possessed military establishments far larger than Japan's Self-Defense Forces. International relations were not destined to be conducted on the pacifist lines envisioned by Article 9 of the constitution of 1947. The United States maintained its occupancy of Okinawa and the Ryukyus, while the Soviet Union occupied the entire Kuril chain in addition to reclaiming southern Sakhalin. The Korean War, which broke out in June 1950, increased the urgency of a peace treaty. Arrangements were worked out between the principal non-Communist allies before and during the command of Gen. Matthew B. Ridgway, who succeeded MacArthur as supreme commander in April 1951.

The peace treaty. The San Francisco Conference that convened in September 1951 to sign the Japanese peace treaty ratified arrangements that had been worked out earlier by John Foster Dulles under the direction of Secretary of State Dean Acheson. Japan recognized the independence of Korea and renounced all rights to Taiwan and the Pescadores, the Kurils, and southern Sakhalin and gave up its rights in the Pacific islands to which it had held mandate under the League of Nations. The Soviet Union attended the San Francisco Conference, but it failed to make its objections to the treaty heard and consequently did not become a signatory. This enabled Japan to retain the hope of regaining at least the Kuril islands closest to Hokkaido—territory that it had not seized in war—through diplomatic efforts.

The San Francisco peace treaty recognized Japan's "right of individual and collective self-defense," which was exercised through the enactment of a security pact with the United States whereby U.S. forces would remain in Japan until Japan could "assume responsibility for its own defense." Japan agreed not to grant similar rights to a third power without U.S. approval. U.S. assistance was extended to the Japanese defense forces, while U.S. units, with the exception of air detachments and naval bases, were gradually removed to Okinawa.

Post-treaty relations. The peace treaty went into effect in April 1952. It made no arrangement for reparations for Japan's Pacific war victims but provided that Japan should negotiate subsequently with the countries concerned. Consequently, effective resumption of relations with the nations of Asia came only after treaties covering reparations had been worked out. These were signed with Burma in 1954, with the Philippines in 1956, and with Indonesia in 1958. In 1956 Japan also restored diplomatic relations with the Soviet Union but without working out a formal treaty of peace. In December of that year, with the Soviet Union no longer invoking a veto, Japan became a member of the United Nations. Subsequently, Japan took an active part in UN meetings and specialized agencies. It also became a contributing member of the Colombo Plan group of nations for economic development in South and Southeast Asia, of the General Agreement on Tariffs and Trade (GATT), and of the Organization for Economic Co-operation and Development (OECD). For many Japanese, their country's return to international status and eminence was symbolized by its acting as host country for the Olympic Games in 1964 and an international fair (Expo 70) at Osaka in 1970. Japan also played a leading role in the creation of the Asian Development Bank in 1965–66.

Reparations treaties

China policy. At the time of the San Francisco treaty, Prime Minister Yoshida Shigeru had intended to delay committing Japan to either of the two Chinas, and the absence of both governments from San Francisco made this seem possible. But John Foster Dulles convinced Yoshida that the treaty would meet opposition in the United States Senate unless some assurance was given that Japan would recognize the Republic of China on Taiwan; thus, Tokyo soon negotiated a peace treaty with that regime but a treaty that did not prejudice possible subsequent negotiations with Peking. A lively trade developed between Japan and Taiwan, and Japanese contributions to the economy of Taiwan were considerable. The treaty also encouraged the development within Japan's Liberal-Democratic Party of a so-called Taiwan lobby. Because Japan's relations with Peking remained tenuous, Chiang Kai-shek was for a time able to hold the Japanese government to its commitments by threatening to cut off Taiwan trade if Tokyo considered developmental loans to the mainland.

Chinese-Japanese trade

Mainland trade relationships developed slowly in the absence of political ties. In 1953 an unofficial trade pact was signed between private Japanese groups and authorities of the People's Republic, the first of a number so negotiated. Together with the industrial countries of western Europe, Japan shortened the list of goods that were under embargo for mainland trade; as late as 1972, however, 167 items remained on that list. In addition to unofficial agreements with Japanese firms designated as friendly by Peking, the 1960s saw the development of an informal, semi-official "memorandum" trade that became increasingly important. But the Peking government made skillful use of trade for political purposes, in the hope of embarrassing or weakening Japan's conservative governments, and intervals of ideological tension and political instability on the mainland were soon reflected in falling levels of trade with Japan. In 1958 China's Great Leap Forward campaign resulted in a temporary closure of all trade with Japan, and in the mid-1960s the Great Proletarian Cultural Revolution also resulted in a severe decline. Nevertheless, Japan gradually became China's most important trading partner.

In 1971 Pres. Richard M. Nixon's announcement of a forthcoming visit to Peking produced a rapid growth in Japanese willingness to compromise ties with Taiwan in favour of closer relations with Peking. The Chinese premier, Chou En-lai, moving with new confidence, now found it possible to attach the sort of conditions that Chiang Kai-shek had earlier worked for in the 1950s by warning that China would not trade with firms engaged in development activities on Taiwan or in South Korea (with which Japan had in 1965 negotiated a treaty establishing diplomatic relations and providing for reparations and assistance). While refusing to deal with Prime Minister Satō, Peking indicated new interest in formal relations with Japan, subject to the revocation of Japan's treaty with Taiwan. Japanese leaders showed themselves increasingly prepared to meet these conditions. The People's Republic was admitted to the United Nations in 1971; in September 1972 Prime Minister Tanaka Kakuei reached agreement with Peking on steps to normalize relations, and simultaneously Japan severed its ties with Taiwan, replacing its embassy with a nonofficial office. Subsequently Japan pursued trade opportunities with the People's Republic vigorously, and in August 1978 the two countries concluded a Treaty of Peace and Friendship that bound both to "perpetual peace and friendship" and pledged them to oppose "hegemony" from whatever source and to foster economic and cultural relations.

Japanese–Chinese Treaty of Peace and Friendship

Postwar politics. After the surrender in 1945, Japanese politics at first returned to the pattern that had been interrupted by the militarist domination of national life. Extremists of the right were discredited by their identification with the lost war. Their major figures were removed from office or arrested, and until 1952, when all but those convicted by the international tribunal were permitted to resume their careers, little rightist organization was possible. Thereafter, some figures of the 1930s re-emerged, but the rightists lacked unity and could offer no program of leadership in Asia. They were handicapped by a decline of influence in the military and business sectors that had formerly supported their activities. Most important, rightist ideology found few listeners among the postwar generation accustomed to new freedoms. Except for a few spectacular incidents, such as the murder of the Socialist leader Asanuma Inajirō in 1960, rightist activities were limited to efforts to revive national holidays, such as February 11 (Foundation Day, for Emperor Jimmu, a campaign that succeeded in 1966), and demonstrations against the Soviet Union and China.

The left fared better for a time. With the release of political prisoners after the war, and with the repeal of the peace-preservation laws that had hampered political organization in prewar days, prominent Communist Party leaders returned to action. Although land reform deprived them of an issue they had used elsewhere in Asia, the postwar years, with their confusion and economic hardship, provided a favourable climate for Communists. A high point of Communist influence at the polls came in the general election of 1949, when Communists placed 35 candidates in the House of Representatives and received nearly 10 percent of the vote.

On the outbreak of the Korean War in 1950, SCAP ordered the removal of Communist leaders from politics. Most chose to go underground, reappearing after the occupation ended. By that time popular sympathy with the Communist cause had declined markedly. The steady rise of living standards, the uncooperative attitude the Soviet Union had shown in negotiating over the Kuril Islands and in fishing treaty discussions, a popular distaste for Communist opposition to the Imperial institution, and widespread dislike of the extremist tactics shown by leftist labour unions—all combined to create an unpromising climate for Communist politicians.

While not reversed, these trends were importantly modified in the late 1960s. The spectacular economic growth of that decade produced great urban migrations that provided promising settings for mass organization and politics, and both conservative-religious (*e.g.*, Komeitō, or Clean Government Party, the political arm of the Sbka-gakkai movement) and radical political movements grew in strength. The conservative government's policy of giving priority to export development over social-welfare measures, though partly justified by a rising standard of living in which all groups shared, did little to alleviate the difficult conditions of life for many of the new recruits to the urban labour force. Communist leaders exploited their possibilities skillfully. In the mid-1960s they broke publicly with Peking to establish an autonomous and somewhat nationalist image, and their student organizations followed a relatively moderate line during anarchic disruptions at the universities in the late 1960s. These policies helped produce large pluralities in many urban elections, as in contests for the House of Councillors, where voter constituencies were large. Nevertheless, the Japan Communist Party remained far from power and found it difficult to establish satisfactory coalition arrangements with the more strongly pro-Peking Japan Socialist Party (Nihon Shakaitb) and other "reformist" elements.

The vicissitudes of right and left made it natural for the prewar moderates to dominate postwar politics. Career diplomats and bureaucrats possessed the ability and command of English to enable them to work with SCAP authorities, and, because they had been out of action since the 1930s, they had not become liable to the removal of militarists from office. Thus, figures of the 1920s and '30s re-emerged, as did also the remnants of the party organizations of those years. The liaison agency, staffed largely by former diplomats, assumed immediate importance. The Cabinet that emerged shortly after the arrival of U.S. forces was headed by Shidehara Kijūrō, who was replaced in May 1946 by Yoshida Shigeru; both were diplomats. In 1947 and 1948 there was an interval of rule under Katayama Tetsu, a Socialist who headed a coalition Cabinet but who was unable to carry out a Socialist program. In 1948 Ashida Hitoshi held office for five months, after which Yoshida returned as prime minister and remained until December 1954, setting a record for modern Japanese prime ministers. Yoshida negotiated the peace treaty and the security pact in 1951 and set Japan's postsurrender course of close cooperation with the United States.

Dominance of Yoshida

Hatoyama Ichirō became a candidate for Yoshida's position, and the Liberal Party was split between their respective followers as a result. The San Francisco treaty and security pact split the Socialist Party into two factions also: the left opposing both the treaty (because it did not include the Communist countries) and the security pact with the United States, the right wing favouring the treaty while opposing the security pact.

Formation of Liberal-Democratic Party

In 1955 the Liberals and Democrats united to form the Liberal-Democratic Party (Jiyū-Minshutō), which thereafter was the dominant party.

After independence. The Korean War marked the turn from depression to prosperity for Japan. As the staging area for the UN effort in South Korea, the country profited from the many services it provided.

Economic growth. The return of independence in 1952 thus found the Japanese economy in the process of growth and change, a process that continued without interruption into the 1960s. Sustained prosperity and consistently high growth rates changed all sectors of life in Japan. The countryside, where farmers had benefitted from the land reform, began to feel the effects of small-scale mechanization and a consistent migration to industrial centres. Agricultural yields rose as improved strains of crops and modem technology were introduced, as household appliances appeared in remote villages, and as the changing, more diversified patterns of food consumption of the cities provided a market for more cash crops, truck (market) garden fruits and vegetables, and meat products. Population control slowed the birth rate, and steady industrial growth brought full employment and even a labour shortage.

Particularly in the 1960s, the structure of the Japanese economy changed in order to concentrate on products of highly advanced technology, which emphasized Japan's need for stable, advanced trading partners instead of the Asian markets for inexpensive textiles that it had earlier developed. Improvements in transportation—*e.g.*, cargo-handling methods and bulk transport by large ore carriers—were removing the disadvantage of the greater distances over which Japan's new materials were moving. Most important of all, a large and growing domestic market was rendering invalid many earlier generalizations about Japan's need for cheap labour and captive Asian markets for inexpensive exports. By the late 1960s Japan had experienced the highest and most sustained growth rates of any country in the world for more than a decade. Japan led the world in shipbuilding, was second in the production of motor vehicles, and was third in steel production.

International policies. Japanese leaders attempting to raise the national income felt their options in international affairs severely restricted by the alliance with the United States, Taiwan, and South Korea, which prevented closer ties with the Soviet Union, China, and North Korea. Prosperity was a universal goal, but international politics proved sharply divisive. Public-opinion polls showed firm agreement against military power and continued horror of atomic or nuclear developments; and it was agreed that the Self-Defense Forces could not be utilized for international or UN causes.

U.S.–Japanese treaty revision

As noted above, restoration of relations with the Soviet Union and membership in the United Nations, both in 1956, were the principal efforts and achievements of Hatoyama Ichirb, who succeeded Yoshida in 1954. Hatoyama was followed by Ishibashi Tanzan in December 1956 and by Kishi Nobusuke in 1957. Kishi, who had been named, though not tried, as a war criminal because of his membership in the Tōjō Cabinet, continued the policies of cooperation with the United States that Yoshida had initiated. But, because of his war record, he was the target of much criticism. In 1958 the Peking government, occupied with its Great Leap Forward, closed all trade contacts with Japan. At the same time, revisions in the U.S.-Japan treaty of mutual cooperation and security were being discussed. A proposed treaty revision, which was to be in force for 10 years, alarmed many Japanese who had felt only slightly involved in the original agreement negotiated at the time of independence. Issues were further complicated by plans for a state visit by Pres. Dwight D. Eisenhower. Originally planned as a follow-up to a visit to Moscow, the visit changed drastically after the Soviet Union shot down a U.S. U-2 reconnaissance plane on May 1, 1960. Eisenhower's trip abroad had begun as a symbol of peace and coexistence; but it now became a strategic tour of U.S. Pacific allies and bases, and its critics charged that it was designed to sustain the falling popularity of the Kishi government. After the Kishi Cabinet used its majority to force the treaty revisions through the Diet, opposition to the Prime Minister, the treaty, and the Eisenhower visit increased steadily. Gigantic student demonstrations shook Tokyo day after day. The treaty survived, but Eisenhower's visit was cancelled, and Kishi resigned in July 1960. In 1970, when the treaty had run its course, both governments were reluctant to see a repetition of the events of 1960; they agreed to invoke its provisions indefinitely, subject to one year's revocation by either party.

Ikeda, Sato, **and their successors.** Kishi was followed by Ikeda Hayato in 1960. A specialist in economic policy, Ikeda set for his country the goal of doubling national income in 10 years. The goal was more than met, as Japan's economy grew at rates of more than 10 percent annually, the highest in the industrialized world. The administration of Pres. John F. Kennedy in the U.S. caught the imagination of many Japanese, and Kennedy's designation of the popular scholar Edwin O. Reischauer as American ambassador further improved Japanese–American relations. By the late 1960s the unpopularity of the war in Vietnam and the turbulence of the world student movements, in which Japanese students played a full role, threatened to disturb American relations again.

The era of Sat0

Satb Eisaku, who became prime minister when Ikeda resigned because of illness in 1964, continued Ikeda's policies and proved an able and resourceful figure until he left office in 1972. In 1965 Japan entered treaty relations with South Korea. At home, Satō worked for continued economic growth and did his best to free Japan from the reminders of defeat that remained. The Bonin (Ogasawara in Japanese) and Ryukyu islands had been left under American occupation by the San Francisco treaty of peace. U.S. military bases on Okinawa (in the Ryukyus) seemed essential to American commitments to South Korea, and they were an important link to American forces in Vietnam: in the late 1960s, Japanese opposition to the Vietnam war made them highly objectionable. Satb's government secured return of the Bonin Islands in 1967, and the retrocession of the Ryukyus became effective in 1972. Satb hailed this as "the end of the postwar era." The U.S. retained bases on Okinawa, but they were now subject to the restrictions that affected other U.S. bases in Japan. Satō resigned in 1972 and two years later was awarded the Nobel Prize for Peace for his role in maintaining Japan's policy against nuclear weapons.

Tanaka scandals

Tanaka Kakuei, Satō's successor, seemed to promise a new stage of Japanese strength. One of his first acts was to take a trip to Peking, on the heels of President Nixon, to reverse Japan's China policy by recognizing the People's Republic. Tanaka reacted to increasing public concern with problems of pollution and overcrowding by calling for the redistribution of industry throughout the Japanese islands. Soon he was being charged with worsening inflation as prices of land throughout Japan rose rapidly. More serious was the effect of the petroleum crisis of 1973 on a country completely dependent on imported oil. Outbreaks of panic buying by consumers brought much reflection on the essential fragility of Japan's economic position; the rapid rise in the price of oil seemed to indicate the end of an era of relatively cheap and abundant resources. Japan next experienced the world recession of the 1970s, and its recovery seemed slower because of the decade or more of exuberant growth that had gone before. The Tanaka era ended in disaster in 1974 with a scandal based on irregularities in the accumulation of his private fortune. Shortly afterward, and worse, it was charged that Tanaka was implicated in improper use of official influence to bring about the selection by Japanese airlines of airplanes manufactured by Lockheed Aircraft Corp.

Tanaka was succeeded by Miki Takeo, the leader of a small faction in the governing Liberal-Democratic Party who won favour while larger factions were in disarray because of the scandals. Miki was determined to pursue full disclosure of the Lockheed affair, and a lengthy investigation and prosecution of Tanaka and his associates was begun by the government; this culminated in the arrest of Tanaka in July 1976 and indictment on a charge of bribery. The economy continued slow. Elections to the Diet in December 1976 brought an end to the Liberal-Democrats' absolute majority, and Miki resigned. The new prime minister was Fukuda Takeo, who had rich experience in many branches of government and was considered a specialist in economic policy. The problems of the Japanese economic turndown proved difficult, however, and combined with party factional differences to bring about Fukuda's defeat in a newly designed party presidential "primary" in November 1978. He was succeeded by Ohira Masayoshi, who announced his intention to continue his predecessors' foreign policies.

Trade policies seemed among the most difficult problems. Domestic consumption was down, inventories were high, and industries were committed to their work forces, with the result that production continued high. There followed a vigorous effort to increase exports, which resulted in Japanese manufacturers' taking larger shares of all foreign markets, including that of the United States. The yen grew stronger at the expense of other currencies, and particularly the dollar, in which most purchases of raw materials were figured. Consequently Japan again saw its import prices diminish, although its manufactures, now more expensive in foreign markets, were rendered less competitive in price. The 1970s produced particularly large Japanese surpluses in Japanese–American trade. Japan was now the largest or second largest trading partner of virtually every country with which it traded. It was investing heavily in other countries, and its manufacturers were building plants throughout the world.

Trade agreement with China

The People's Republic of China was particularly important to Japan's long-range economy, and the Chinese turn from ideological rigidity to pragmatism in the 1970s, after the death of Mao Tse-tung, encouraged increasing numbers of Japanese to seek Chinese markets. In 1978 an eight-year agreement for a total of $20,000,000,000 in industrial contracts was signed with China, and in August Japan and China signed the Treaty of Peace and Friendship referred to above.

Within Japanese politics and opinion, the long-standing polarization over Japan's treaty with the United States was made obsolete by these developments. Although opponents of the treaty had long argued that the American tie and the Japanese Self-Defense Forces threatened to cut Japan off from the People's Republic, Peking now saw both that treaty and those forces as constraints against its Soviet rival and no longer criticized them. The Socialist and Communist ideology of class warfare also lost ground in the atmosphere of general affluence, and the overwhelming majority of Japanese showed that they considered themselves to be members of the middle class. Meanwhile the Liberal-Democrats' dominance had been eroded by movements of population that made Japan increasingly urban. As a result, that government's inclination to attempt revision of the constitution of 1947 had only academic interest in a situation in which it required allies in the Diet to maintain its hairline majority.

The late 1970s thus found the Japanese aware that the shibboleths of the postsurrender decades were obsolete. Japan was not poor but wealthy, not weak but a power in the international economy, not isolated but the largest trading partner of almost every country in the world. This awareness was reflected in a surge of introspection in which writers discussed the role for their country and the nature of their society. What remained was a growing consensus around general principles, summed up by Prime Minister Fukuda in 1978: Japan should adhere to its decision not to become a major military power, and should promote friendly cooperation with all nations and work to accept growing responsibility within the international community. Japan remained unusually dependent upon the stability of the world economy, but that stability in turn was more dependent upon the quality of Japanese participation than it had ever been before.

(M.B.J.)

BIBLIOGRAPHY

Introductory works: Among various critical guides in Western languages, JOHN WHITNEY HALL, *Japanese History: New Dimensions of Approach and Understanding,* 2nd ed. (1966), is the most convenient and skilled presentation of scholarly work in this field. Additions to published materials will be found, though not annotated, in the annual bibliographical issue of the *Journal of Asian Studies.* Generally speaking, Japanese scholars in the field of Japanese history are far more productive than their foreign colleagues, and an idea of how much has been produced is available in JOHN WHITNEY HALL, *Japanese History: A Guide to Japanese Reference and Research Materials* (1954); and HERSCHEL WEBB, *Research in Japanese Sources: A Guide* (1965). Besides general surveys and monographs, the following three compendiums by groups of recognized scholars are suggested for advanced students who are accustomed to reading Japanese: SHOGAKUKAN (ed.), *Zusetsu Nihon bunkashi taikei* ("Illustrated Compendium of Japanese Cultural History"), 2nd ed., 14 vol. (1965–68), a collection of essays arranged in chronological order and with numerous illustrations for the general reading public, the last volume devoted to bibliography and related materials; SABURO IENAGA (ed.), *Iwanami Koza: Nihon rekishi* ("Iwanami's Lectures on Japanese History"), 22 vol. (1962–64), a collection of articles contributed by many authors and arranged in chronological order by periods, well documented and directed to somewhat advanced readers; and KAWADE SHOBO (ed.), *Nihon no rekishi* ("A History of Japan"; 1965–67), composed of 27 volumes written by different authors with 4 volumes devoted to illustrations, in chronological order.

Surveys: Authoritative textbooks in English include EDWIN O. REISCHAUER and JOHN K. FAIRBANK, *East Asia: The Great Tradition* (1960); and JOHN K. FAIRBANK, EDWIN O. REISCHAUER, and ALBERT CRAIG, *East Asia: The Modern Transformation* (1965). Though dated, JAMES MURDOCH, *A History of Japan,* 3 vol. (1903–26, reprinted 1964), is a pioneer work that gives a detailed political history. A later survey is GEORGE B. SANSOM, *A History of Japan,* 3 vol. (1958–63). Shorter excellent interpretive works include SANSOM'S *Japan: A Short Cultural History,* rev. ed. (1962), and *The Western World and Japan* (1950, reprinted 1965); EDWIN O. REISCHAUER, *Japan: The Story of a Nation* (1970) and *The Japanese* (1977); SABURO IENAGA, *History of Japan,* 5th ed. (1961); MIKISO HANE, *Japan: A Historical Survey* (1972); BRADLEY SMITH, *Japan: A History in Art* (1964); MITSUSADA INOUE, *Introduction to Japanese History: Before the Meiji Restoration* (1962); WILLIAM G. BEASLEY, *The Modern History of Japan,* 2nd ed. (1974); and JOHN W. HALL, *Japan from Prehistory to Modern Times* (1970). The JAPAN NATIONAL COMMISSION FOR UNESCO, *Japan: Its Land, People and Culture,* 2nd ed. (1972), is a comprehensive reference work.

Monographs: For further reading, selected monographic works on Japan (in English) are listed below, grouped according to their historical approach. *(Ancient):* GERARD S. GROOT, *The Prehistory of Japan* (1951); J.E. KIDDER, *Japan Before Buddhism,* rev. ed. (1966); ROBERT K. REISCHAUER, *Early Japanese History,* 2 vol. (1937); IVAN I. MORRIS, *The World of the Shining Prince: Court Life in Ancient Japan* (1964). *(Medieval and early modern):* G. CAMERON HURST III, *Insei: Abdicated Sovereigns in the Politics of Late Heian Japan, 1086–1185* (1976); JOHN W. HALL, *Government and Local Power in Japan, 500–1700: A Study Based on Bizen Province* (1966); KAN'ICHI ASAKAWA, *The Documents of Iriki,* rev. ed. (1955); MINORU SHINODA, *The Founding of the Kamakura Shogunate, 1180–1185* (1960); JEFFREY P. MASS, *Warrior Government in Early Medieval Japan* (1974); JOHN W. HALL and JEFFREY P. MASS (eds.), *Medieval Japan: Essays in Institutional History* (1974); JOHN W. HALL and TOYODA TAKESHI (eds.), *Japan in the Muromachi Age* (1977), the outcome of a binational conference held in 1973; DELMER M. BROWN, *Money Economy in Medieval Japan* (1951); YI-T'UNG WANG, *Official Relations Between China and Japan, 1368–1549* (1953); HELEN CRAIG MCCULLOUGH, *The Taiheiki: A Chronicle of Medieval Japan* (1959); CHARLES R. BOXER, *The Christian Century in Japan, 1549–1650,* 2nd ed. (1967), *Fidalgos in the Far East, 1550–1770,* 2nd ed. (1968), and *Jan Compagnie in Japan, 1600–1850,* 2nd ed. (1950, reprinted 1968); GEORGE ELISON, *Deus Destroyed: The Image of Christianity in Early Modern Japan* (1973); H. PAUL VARLEY, *The Onin War* (1967); CONRAD TOTMAN, *Politics in the Tokugawa Bakufu, 1600–1843* (1967); JOHN W. HALL and MARIUS B. JANSEN (eds.), *Studies in the Institutional History of Early Modern Japan* (1968): SUSAN B. HANLEY and KOZO YAMAMURA, *Economic and Demographic Change in Pre-industrial Japan, 1600–1868* (1978); HAROLD BOLITHO, *Treasures Among Men: The Fudai Daimyo in Tokugawa Japan* (1974); RONALD P. DORE, *Education in Tokugawa Japan* (1965); CHARLES D. SHELDON, *The Rise of the Merchant Class in Tokugawa Japan, 1600–1868* (1958); THOMAS C. SMITH, *The Agrarian Origins of Modern Japan* (1959); JOHN W. HALL, *Tanuma Okitsugu, 1719–1788: Forerunner of Modern Japan* (1955); DONALD KEENE, *The Japanese Discovery of Europe* (1952); JOHN A. HARRISON, *Japan's Northern Frontier* (1953); GEORGE A. LENSEN, *The Russian Push Toward Japan* (1959, reprinted 1971); WILLIAM G. BEASLEY (ed. and trans.), *Select Documents on Japanese Foreign Policy, 1853–1868* (1955); GRACE FOX, *Britain and Japan, 1858–1883* (1969); MARIUS B. JANSEN, *Sakamoto Ryōma and the Meiji Restoration* (1961, reprinted 1971); ALBERT CRAIG, *Chōshu in the Meiji Restoration* (1961). *(Modern):* CHITOSHI YANAGA, *Japan Since Perry* (1949); E. HERBERT NORMAN,' *Japan's Emergence as a Modern State* (1940); WILLIAM W. LOCKWOOD, *The Economic Development of Japan: Growth and Structural Change, 1868–1938* (1954); THOMAS C. SMITH, *Political Change and Industrial Development in Japan: Government Enterprise, 1868–1880* (1955); WILLIAM CHAMBLISS, *Chiaraijima Village: Land Tenure, Taxation and Local Trade, 1818–1884* (1965); JOHANNES HIRSCHMEIER, *The Origins of Entrepreneurship in Meiji Japan* (1964); MARIUS B. JANSEN (ed.), *Changing Japanese Attitudes Toward Modernization* (1965); KENNETH B. PYLE, *The New Generation in Meiji Japan: Problems of Cultural Identity, 1885–1895* (1969); ROBERT A. WILSON, *Genesis of the Meiji Government in Japan, 1868–1871* (1957); JOSEPH PITTAU, *Political Thought in Early Meiji Japan, 1868–1889* (1967); GEORGE AKITA, *Foundations of Constitutional Government in Modern Japan, 1868–1900* (1967); ROGER F. HACKETT, *Yamagata Aritomo in the Rise of Modern Japan 1838–1922* (1971); ROBERT E. WARD (ed.), *Political Development in Modern Japan*

(1968); F. HILARY CONROY, *The Japanese Seizure of Korea, 1868–1910* (1960); JOHN A. WHITE, *The Diplomacy of the Russo-Japanese War* (1964); SHUMPEI OKAMOTO, *The Japanese Oligarchy and the Russo-Japanese War* (1970); JAMES NAKAMURA, *Agricultural Production and the Economic Development of Japan, 1873–1922* (1966); TETSUO NAJITA, *Hara Kei in the Politics of Compromise, 1905–1915* (1967); PETER DUUS, *Party Rivalry and Political Change in Taisho Japan* (1968); ROBERT SCALAPINO, *Democracy and the Party Movement in Prewar Japan* (1962); DELMER M. BROWN, *Nationalism in Japan* (1955, reprinted 1971); AKIRA IRIYE, *After Imperialism: The Search for a New Order in the Far East, 1921–1931* (1965); JAMES B. CROWLEY, *Japan's Quest for Autonomy* (1966); FRANCIS C. JONES, *Japan's New Order in East Asia* (1954); JOHN M. MAKI, *Japanese Militarism* (1945); RICHARD STORRY, *The Double Patriots* (1957); SADAKO OGATA, *Defiance in Manchuria: The Making of Japanese Foreign Policy, 1931–1932* (1964); ROBERT J.C. BUTOW, *Tōjō and the Coming of the War* (1961) and *Japan's Decision to Surrender* (1953); HERBERT FEIS, *The Road to Pearl Harbor* (1950); THOMAS R.H. HAVENS, *Valley of Darkness: The Japanese People and World War Two* (1978). *(Postwar years):* BARON E.J. LEWE VAN ADUARD, *Japan: From Surrender to Peace* (1953); SCAP, GOVERNMENT SECTION, *Political Reorientation of Japan, September 1945 to September 1948*, 2 vol. (1949); GEORGE C. ALLEN, *Japan's Economic Recovery* (1958) and *Japan's Economic Expansion* (1965); JEROME B. COHEN, *Japan's Postwar Economy* (1958); SHIGERU YOSHIDA, *The Yoshida Memoirs* (1962); E.O. REISCHAUER, *The Japanese* (1977); HUGH PATRICK and HENRY ROSOFSKY (eds.), *Asia's New Giant: How the Japanese Economy Works* (1976).

(Ma.Ka./M.B.J.)

Japan, Sea of

The Sea of Japan is a marginal sea of the western Pacific Ocean, bounded by Japan and the Soviet island of Sakhalin to the east and by the Soviet Union and Korea on the Asian mainland to the west. Its area is 389,100 square miles (1,007,800 square kilometres). It has a mean depth of 4,429 feet (1,350 metres) and a maximum depth of 12,276 feet (3,742 metres).

Physiography and submarine morphology. The sea is almost elliptical, having its major axis from southwest to northeast; to the north it is approximately bounded by latitude 51°45′ N, while to the south it is bounded by a line drawn from the Japanese island of Kyūshū westward through the Gotb-rettb (Gotb Islands) of Japan to the Korean island of Cheju (also known as Quelpart Island) and then northward to the Korean peninsula.

The sea itself lies in a deep basin, separated from the East China Sea to the south by the Tsushima-kaikyō (Tsushima Strait) and Korea Strait and from the Sea of Okhotsk to the north by Sbya-kaikyb (Sbya Strait, or La Perouse Strait) and Tatar Strait, all of which have sill depths of less than about 700 feet. To the east it is also connected with the Inland Sea of Japan by Kanmon-kaikyb.

Component basins

Underwater, the sea is separated into the Japan Basin (about 9,800 to 11,500 feet deep) to the north, the Yamato Basin (8,200 feet deep) to the southeast, and the Tsushima Basin (6,600 feet deep) to the southwest. While a narrow continental shelf with an average width of about 19 miles fringes Siberia and the Korean peninsula, on the Japanese side of the sea there are wider continental shelves with depths at the edge of between about 430 and 1,300 feet, as well as groups of banks, troughs, and basins lying offshore. The banks lying off the coasts of Japan are divided into groups, which include Okujiri Ridge, Sado Ridge, Hakusan Banks, Wakasa Ridge, and Oki Ridge.

Geology and bottom deposits. Yamato Ridge, which has an average depth of about 1,475 feet (450 metres), consists of granite, rhyolite, andesite, and basalt, with boulders of volcanic rock scattered on the seabed. The top of the Korea Plateau is about 3,000 feet below the surface and has a depression in its central part. Geophysical investigation has revealed that, while Yamato Ridge is of continental origin, the Japan Basin and the Yamato Basin are of oceanic origin.

Bottom deposits in the Sea of Japan indicate that earth-born sediments, such as mud, sand, gravel, and fragments of rock, exist down to depths of 650 to 1,000 feet; hemipelagic sediments (*i.e.*, half of oceanic origin), mainly consisting of blue mud rich in organic matter, are found down to depths of about 1,000 to 2,600 feet; and deeper pelagic sediments, consisting of red mud, are found down to a depth of nearly 13,000 feet.

A number of submarine canyons are found on the continental slope, at depths of more than 6,500 feet on the west side of the basin, while those near the islands of Japan lie at depths of only about 2,600 feet.

Argument over the formation of the Sea of Japan has not yet ended, although there is agreement that the four straits that connect the sea either to the Pacific Ocean or to marginal seas were formed in very recent geological periods. The oldest of these straits are the Tsugaru-kaikyb and Tsushima-kaikyō, whose formation interrupted the migration of elephants into the Japanese islands at the end of the Tertiary Period (about 2,500,000 years ago); the most recent is Sōya-kaikyō, which was formed at the end of the Wisconsin Ice Age (60,000 to 11,000 BP) and which closed the route once used by the mammoths whose fossils have been found in Hokkaidb.

Climate. The Sea of Japan contributes greatly to the mild climate of Japan because of the effect exerted by its relatively warm waters; evaporation is especially noticeable in winter, when an estimated 5,000,000,000 tons of water vapour rise as steam fog near the Polar Front (*i.e.*, the frontier between the cold, dry polar air mass and the warm, moist tropical air mass). From December to March the prevailing northwest monsoon wind carries cold and dry continental polar air masses over the warmer waters of the sea, resulting in persistent precipitation in the form of snow along the mountainous coasts of Japan. In summer the southerly tropical monsoon blows from an area of higher atmospheric pressure over the North Pacific onto the Asian mainland, causing dense fog when its warm and moist winds blow over the cold currents that prevail over the northern part of the sea at that season. The winter monsoon brings rough seas and causes coastal erosion as a result of the heavy surf that breaks along the western coasts of Japan. In summer and fall typhoons occasionally occur.

The northern part of the sea, especially off the Siberian coast as well as in Tatar Strait, freezes in winter; as a result of convection, melted ice feeds the cold currents in that part of the sea in spring and summer.

Prevailing currents

Hydrography. The waters of the sea generally circulate in a counterclockwise pattern. A striking contrast occurs between the cooler and relatively fresher water in the western part and the warmer and relatively more saline water in the eastern part. A branch of the Kuroshio Current, the Tsushima Current, together with its northern branch, the East Korea Warm Current, flows north, bringing warmer and more saline water, before flowing into the Pacific through the Tsugaru-kaikyō as the Tsugaru Current, as well as into the Sea of Okhotsk through the Sōya-kaikyō as the Sbya Current. Along the coast of the Asian mainland, on the other hand, three cold currents—the Liman, North Korea, and Central (or Mid-) Japan Sea cold currents—bring cooler, relatively fresh, and turbid water southward.

Economic resources and resource exploitation. Fisheries and mineral deposits form the main economic resources. Fisheries may be divided into pelagic (oceanic) and demersal (sea-bottom) categories. Pelagic fishes include mackerel, horse mackerel, sardines, anchovies, herring, fishes of the salmon and trout family, sea bream, and squid; the demersal category includes cod, Alaskan pollack (bluefish), and Atka mackerel. Seals and whales are also to be found, as well as such crustaceans as prawns and crabs. The fishing grounds are for the most part on the continental shelves and their adjacent waters, as well as in the Polar Front zone and on the submarine banks.

Herring, sardines, and bluefin tuna have traditionally been caught, but since 1946 the fisheries have been becoming depleted. In the late 1970s squid fishing was being carried on in the central part of the sea, salmon fishing in the shoal areas of the north and southwest, and crustaceans were being caught in the deeper parts.

Mineral resources on or in the sea bottom include magnetite sands as well as natural gas and petroleum deposits off Japan and Sakhalin Island.

Commercial and historical importance. As trade increases between Asian countries, the Sea of Japan is being increasingly used as a commercial waterway. While the waters of the sea historically have served to protect Japan from foreign invasions, the southern straits have been the scene of some historical naval battles. On two occasions —in 1274 and 1281—the fleets of Kublai Khan, founder of the Yüan (Mongol) dynasty in China, attempted to cross the straits to conquer Japan, but the conqueror's forces were either destroyed by typhoons or beaten by the Japanese. In 1905, at the battle of Tsushima Strait, the Japanese navy almost completely destroyed the Baltic fleet of the Russian tsar.

Prospects for the future. It is anticipated that increasing international cooperation will result in more effective utilization of the sea's resources. Apart from the development of mineral resources, a system for forecasting oceanic, climatic, and fishery conditions may be developed. The sea may also provide an additional source of energy by harnessing waves created by the winter monsoon. (M.U.)

Japanese Language

Japanese is the language of more than 104,640,000 persons on the islands of Japan, including the Ryukyus. In addition, there are more than 607,000 Koreans and about 51,000 Chinese in Japan who speak the language. Outside Japan a considerable number of people of Japanese parentage can speak the Japanese language with some degree of proficiency; these include 491,000 persons in Brazil, 465,000 in the United States, 45,000 in Peru, and 30.000 in Canada. In Korea and Taiwan the older generations speak Japanese as a second language.

For centuries there have been many dialects in Japan differing from each other to such an extent that some of them are mutually unintelligible. Since the Meiji Restoration, which ended with the promulgation of the Japanese constitution in 1889, the rapid development of elementary education has eliminated illiteracy in the country, and a common written language has been established, based on the dialect of the residential sections of Tokyo. At present, people of the various parts of Japan can speak the common language, although with their own accents, in addition to their own dialects. Since World War II, a common spoken language based on the same dialect of Tokyo has been exerting more and more influence upon the speech of the younger generation all over Japan via radio and especially via television; as a result, the local dialects are disappearing more rapidly than before.

The genetic relationship of Japanese to other languages has not been linguistically established. It is, however, probably related to Korean and possibly to the Altaic languages, which include the Manchu-Tungus, Mongolian, and Turkic families; all have similarities in their phonological and grammatical structures. Some lexical (vocabulary) and other resemblances, however, have been pointed out between Japanese and other East Asian languages and language families—*e.g.*, Austronesian (Malayo-Polynesian), Austroasiatic, Tibeto-Burmese, and Ainu.

Two major dialect groups

Dialects. Japanese can be divided into two major dialect groups: those of the mainland and those of the Ryukyu Islands. The mainland dialects are divided by some scholars into three groups—Eastern, Western, and Kyushu. In other systems, however, they are classified into the Eastern division and the Western division, which is then split into the Kansai (including the Chūgoku and Shikoku dialects) and the Kyūshū dialects. The Tokyo dialect belongs to the Eastern group, Westernized during the last two or three centuries; except for the accent and other features, it is not very different from that of Kybto, a Kansai dialect, which was the most influential central dialect for more than 1,000 years. The peripheral dialects (*e.g.*, the dialects of remote areas—north Tōhoku and Kagoshima of south Kyūshū—as well as the Ryukyuan dialects) are very different from those of Tokyo and Kybto and are incomprehensible in those cities.

There is evidence that there may have been a language spoken by the people of the Yayoi culture in north Kyūshii about 2,000 years ago that became the Proto-Japanese tongue—*i.e.*, the ancestor of all the present-day Japanese dialects. The Yayoi culture with its rice cultivation was brought from the Asian continent to Japan. If it is assumed that the people of this culture also brought the Proto-Japanese language, then it must also be assumed that the language of the remainder (if any) of the Proto-Japanese-speaking people on the Asian continent was extinguished by another language or languages; because there is no evidence of a language on the continent that is not only similar to Japanese but also differs from it enough to reflect the 2,000 years of separation.

A more probable hypothesis is that the Japanese language became separated from Korean 5,000 or more years ago and was spoken in Japan thousands of years before the Christian Era. In that case, Proto-Japanese was probably only one of the Japanese dialects spoken contemporaneously in most of Japan, including Kansai and the area east of it. The Yayoi culture is known to have diffused rapidly eastward from north Kyushu and apparently later stimulated the development of the Tumulus culture in Kinai (*i.e.*, the "Home Provinces," including Yamato, one of the centres of which was the city of Nara). The dialects that were contemporary with and different from Proto-Japanese must have been absorbed during the subsequent 2,000 years by the dialects that branched out from Proto-Japanese.

In the 8th century, however, people noticed that the Eastern dialects were remarkably different from those around Nara (and perhaps also from the Western dialects); some of their peculiarities were recorded. These peculiarities may have been non-Proto-Japanese features, which at that time were still resisting the influence of the central dialects but which have disappeared from the present-day Eastern dialects except for the dialect of a remote island, Hachijb-jima (about 300 kilometres south of Tokyo), that preserves some of the old peculiarities.

The differentiation of social dialects has been slight. but some peculiarities of the speech of the former samurai have been reported in several cities.

Phonological characteristics. Japanese is a polysyllabic language. Simple (*i.e.*, not compound) nouns consist of one or, more often, two or three syllables, to which various particles of one or more syllables are often suffixed. Various "inflectional forms" of simple verbs and adjectives usually consist of two or more syllables and some have various endings or particles or both suffixed as well. ("Inflectional forms" in Japanese are such forms [stems] as negative, preterite, conditional, imperative, and so forth.) The structure of the syllable is rather simple; syllables are ordinarily open (*i.e.*, they consist of one consonant and one vowel that is either short or long, with or without an intervening y after the consonant). The syllable can, however, be closed with a nasal sound (indicated by linguists with the symbol /N/) or a checked sound (conventionally symbolized as /Q/), which then acquire similar phonetic characteristics or become identical to the sound that follows. For example, /N/ becomes m before p, b, m; n before *t,* d, n; *ng* before k, *g, ng;* nasalized i before y; nasalized ũ before w; and so forth. /Q/ is always followed by a consonant and is changed to an implosive *p, t,* k, *s,* etc., which forms a geminate sound (*i.e.*, a double consonant like the t's in English "cattail") with the following explosive consonant. There are five vowels, a, i, *u, e, o,* similar to those of Italian. Short vowels are pronounced very short; the short i and *u* between voiceless sounds are devoiced (pronounced without vibrating vocal cords—voiceless) or omitted in the Tokyo speech. This tendency, remarkable in Tokyo, is less prevalent in some Western dialects.

Structures of the syllable

Japanese has the following consonants: p, t, k, b, d, g, *ts* (ch), *s* (sh), *z* (*j*), m, n, *r,* h, y, w. Basically, any vowel or ya, yu, yo can follow any of the above consonants. But there are restrictions to this rule that include several unacceptable combinations of sounds, among them *ti,* tu, *di,* du, si, *zi,* wi, we, and others. Chi, *tsu,* ji, zu. *shi, ji,* i, e, etc., occur in place of them, respectively. In pronunciation, the Japanese *p,* t, and k are not as aspirated as the

Japanese Kana*

simple *kana* symbols

No	H	K	E	No.	H	K	E	No.	H	K	E	No.	H	K	E	No.	H	K	E
1.	あ	ア	a	2.	い	イ	i	3.	う	ウ	u	4.	え	エ	e	5.	お	オ	o
6.	か	カ	ka	7.	き	キ	ki	8.	く	ク	ku	9.	け	ケ	ke	10.	こ	コ	ko
11.	さ	サ	sa	12.	し	シ	shi	13.	す	ス	su	14.	せ	セ	se	15.	そ	ソ	so
16.	た	タ	ta	17.	ち	チ	chi	18.	つ	ツ	tsu†	19.	て	テ	te	20.	と	ト	to
21.	な	ナ	na	22.	に	ニ	ni	23.	ぬ	ヌ	nu	24.	ね	ネ	ne	25.	の	ノ	no
26.	は	ハ	ha	27.	ひ	ヒ	hi	28.	ふ	フ	fu	29.	へ	ヘ	he	30.	ほ	ホ	ho
31.	ま	マ	ma	32.	み	ミ	mi	33.	む	ム	mu	34.	め	メ	me	35.	も	モ	mo
36.	や	ヤ	ya					37.	ゆ	ユ	yu					38.	よ	ヨ	yo
39.	ら	ラ	ra	40.	り	リ	ri	41.	る	ル	ru	42.	れ	レ	re	43.	ろ	ロ	ro
44.	わ	ワ	wa													45.	を	ヲ	o
																46.	ん	ン	n(m)‡
47.	が	ガ	ga	48.	ぎ	ギ	gi	49.	ぐ	グ	gu	50.	げ	ゲ	ge	51.	ご	ゴ	go
52.	ざ	ザ	za	53.	じ	ジ	ji	54.	ず	ズ	zu	55.	ぜ	ゼ	ze	56.	ぞ	ゾ	zo
57.	だ	ダ	da	58.	ぢ	ヂ	ji	59.	づ	ヅ	zu	60.	で	デ	de	61.	ど	ド	do
62.	ば	バ	ba	63.	び	ビ	bi	64.	ぶ	ブ	bu	65.	べ	ベ	be	66.	ぼ	ボ	bo
67.	ぱ	パ	pa	68.	ぴ	ピ	pi	69.	ぷ	プ	pu	70.	ぺ	ペ	pe	71.	ぽ	ポ	po

digraphs representing single syllables

No.	H	K	E	No.	H	K	E	No.	H	K	E
1.	きゃ	キャ	kya	2.	きゅ	キュ	kyu	3.	きょ	キョ	kyo
4.	しゃ	シャ	sha	5.	しゅ	シュ	shu	6.	しょ	ショ	sho
7.	ちゃ	チャ	cha	8.	ちゅ	チュ	chu	9.	ちょ	チョ	cho
10.	にゃ	ニャ	nya	11.	にゅ	ニュ	nyu	12.	にょ	ニョ	nyo
13.	ひゃ	ヒャ	hya	14.	ひゅ	ヒュ	hyu	15.	ひょ	ヒョ	hyo
16.	みゃ	ミャ	mya	17.	みゅ	ミュ	myu	18.	みょ	ミョ	myo
19.	りゃ	リャ	rya	20.	り	リュ	ryu	21.	り	リョ	ryo
22.	ぎゃ	ギャ	gya	23.	ぎゅ	ギュ	gyu	24.	ぎょ	ョ	gyo
25.	じゃ	ジャ	ja	26.	じゅ	ジュ	ju	27.	じょ	ジョ	jo
28.	ぢゃ	ヂャ	ja	29.	ぢゅ	ヂュ	ju	30.	ぢょ	ヂョ	jo
3	びゃ	ビャ	bya	32.	びゅ	ビュ	byu	33.	びょ	ビョ	byo
34.	ぴゃ	ピャ	pya	35.	ぴゅ	ピュ	pyu	36.	ぴょ	ピョ	pyo

hiragana trigraphs containing long vowels §

No.	H	E	No.	H	E
2.	きゅう	kyū	3.	きょう	kyō
5.	しゅう	shū	6.	しょう	shō
8.	ち	chu	9.	ちょう	chi3
11.	にゅう	nyū	12.	にょう	nyō
14.	ひゅう	hyū	15.	ひょう	hyō
17.	みゅう	myii	18.	みょう	my6
20.	りゅう	ryū	21.	りょう	ryō
23.	ぎゅ	gyū	24.	ぎょう	gyō
26.	じゅう	jū	27.	じょう	jō
29.	ぢゅう	jū	30.	ぢょう	jō
32.	びゅう	byū	33.	びょう	byō
35.	ぴゅう	pyū	36.	ぴょう	pyō

*H = *hiragana: K = katakana; E =* equivalent. Some *kana* undergo a change in pronunciation in specific situations.

†*Tsu* is also used to indicate a doubled consonant. In such cases the *tsu kana* is written slightly below the line (in horizontal texts) or slightly to the right of the line (in vertical texts) and sometimes also in slightly smaller script. Other *kana* are also positioned in this manner when they serve special functions.

‡Romanized m before b, p, and m.

§All five vowels have long forms, which in a Japanese text can be indicated in one of several ways. Romanized long vowels are indicated by macrons, except i, which is written ii.

initial p, t, and k of English (*i.e.*, they are not pronounced with a strong accompanying puff of breath). *B, d,* and *g* are fully voiced (pronounced with vibration of the vocal cords) as in French; g between vowels is usually a nasal *ng* (as in English "sing") in many Kinki (central) and Eastern dialects. Unlike the English sounds (which are formed with the tongue touching the gum ridge behind the upper teeth), t, d, and *n* are articulated against the teeth; ch, sh, and *j* are pronounced with the front of the tongue, not with the tip of the tongue as in English; initial *z* is pronounced like dz in English "adz" in Tokyo; and *r* is a flapped sound like the American t sound in "city."

Pitch accent

The majority of dialects, including those of Tokyo and Kybto, have a word pitch accent. In Tokyo, for example, *hashi* with a high–low accent means "chopsticks," but with a low–high accent pattern it denotes "bridge"; in Kybto, on the other hand, *hashi* means "edge, end with high–high accent, "bridge" with high–low, and "chopsticks" with low–high. There are various patterns of pitch accent, and their geographical distributions are very complicated. The dialects of Tokyo and Hiroshima and those of Kybto and Osaka have patterns quite different from each other. Some dialects in Tōhoku and Kyūshū, among others, have no pitch contrast at all.

It is a common feature of all the dialects, however, that they have no word stress accent (as occurs in English—*e.g.*, háppy, fóreigner, characteristics). The sound of Japanese gives a very different impression from that of English, and it is said to be spoken with even stress and rhythm, as if a metronome were very rapidly ticking off each syllable.

Grammatical characteristics. Nouns have neither number (singular and plural) nor gender (masculine and feminine) and take no article (such as "the," "a," "an"). Case distinctions that show such grammatical features as subject and object are expressed with particles added to the ends of words; for example, *kodomo-ga* "a (the) child (children) [usually nominative]," kodomo-no "of a child," kodomo-o "a child [accusative]," *kodomo-ni* "to a child," kodomo-kara "from a child." Verbs have no person, no number, and no gender and are conjugated with the use of endings—*e.g.,* kaku "write, writes, will write," kakanai "do (does, will) not write," kake "write [imperative]," kakd "I'll write, let's write," kaite "having written, writing," kaita "wrote, has (have) written," kakeba "if . . . write (writes)." Adjectives, which have no number, gender, or case, are also "conjugated" with suffixes—*e.g.,* shiroi "is (are, am) white," shiroku-nai "is not white," skiroku-te "is white, and . . ," shirokatta "was (were) white."

There are no relative pronouns. Demonstrative pronouns have a three-way distinction rather than the two-way division of English "this" versus "that"—*e.g.,* kore "this," sore "that around you," and are "that far from both of us." Personal pronouns have a complex and intricate system: for the 1st person singular (English "I, me, my"), for example, there are watakushi, watashi, *washi,* atakushi, *atashi,* temae, boku, ore, and others; and for the 2nd person singular (English "you, your") there are *anatasa*ma, anata, anta, kimi, *omae, kisama, temē,* and others. The use of these forms depends on such factors as the social relationship of the addresser and the addressee(s) and the degree of intimacy between them, the formality of the speech, and the sex and age of the speaker. Verbs and adjectives also have conjugated forms corresponding to some of these distinctions. For example, kakimasu, kakimasen, *kakimashita* are used instead of kaku, *kakan*ai, kaita, respectively, when the utterance is addressed to those who are superior or not intimate to the addresser. On the other hand, if the person who is the performer of the action kaku ("to write") is a superior to the speaker, then oka-ki-ni naru and *okaki-ni* narimasu are used instead of kaku and kakimasu, respectively. If the action is done for a superior, then okaki *suru* and okaki *shimasu* are used instead.

Order of words in sentences

The predicate stands at the end of a sentence. The subject, the object, adverb, adverbial phrase, and other elements of the sentence precede the predicate and can be omitted when possible. For instance, Kaita. which is a sentence consisting only of a predicate, can be translated as "I (he, they, etc.) wrote (have, has written) it," according to the context or the situation. The complement precedes the copula (a linking verb such as English "is"); *e.g.,* gakusei-da "I (he, they, etc.) am (is, are) a student (students)" is composed of the complement *gakusei* "student" and the copula da "am, is, are." Sentences of the structure A-wa B-da (wa means "as for") have various structural meanings—*e.g.,* Are-wa claigaku-da "That is a university" (daigaku = "university"), Watashi-wa *dai*gakuda "I am for (from, in, etc.) the university," *Gaku*sei-wa gakusei-da "A student is nothing but a student." The expressions that modify a noun directly precede it; *e.g., chiisana hitsuji* "a small sheep," *ōkina ōkami* "a large wolf,"

ōkami-no	kutta	*hitsuji* "the sheep that a wolf ate,"
wolf-[genitive]	ate	sheep

hitsuji-o	kutta	*ōkami* "the wolf that ate the sheep."
sheep-[accusative]	ate	wolf

Vocabulary. Although the greater proportion of the basic Japanese words are native words, a large percentage of the whole vocabulary is composed of Chinese loan elements—comparable to the loanwords from Greek, Latin, and French in English. In early times Japanese apparently borrowed a number of cultural words from Korean, but from the 6th century through the 9th century the direct contact with Chinese culture had a much greater influence on Japanese; the phonology, grammar, and basic vocabulary, however, were not as strongly affected. During those centuries several Sanskrit words entered into Japanese through Chinese in connection with Buddhism—*e.g.,* danna "patron, master" (from danna-*pati*), *hachi* "bowl" (from *pātra*), kawara "tile" (from *kapāla*). Later, in the 12th and 13th centuries, Zen priests introduced several words from Middle Chinese, such as *manjū* "bean-jam bun," *yōkan* "sweet paste," udon "noodle," isu "chair," and futon "bedding." Around the end of the 16th century, Japanese borrowed several words from Portuguese, like pan "bread," kasutera "sponge cake," rasha "woollen cloth," and karuta "cards." During the 18th and 19th centuries the language acquired several words from Dutch—*e.g., buriki* "tinplate," garasu "glass pane," *rappa* "trumpet," and *zuk*ku "duck, canvas."

A large part of the Chinese loanwords in contemporary Japanese are compound terms or derivative words coined in Japan since the Meiji Era; these combine two or more Chinese morphemes (word elements) that were borrowed in ancient times. Kisha "train," for example, is a Japanese compound of this kind that consists of ki "steam" and *sha* "car"; the Chinese equivalent, ch'i ch'e, means "automobile." The newly coined Japanese compound *jidōsha* "automobile" consists of three Chinese loan elements—ji "self," *dō* "to move," and sha "car"—but Chinese has no corresponding compound such as *tzu* tung *ch'e*. Some of these new Japanese compounds have been borrowed back into Chinese through the medium of Chinese characters.

Influence of Chinese characters on word formation

The Chinese characters play a very peculiar but important role in the word formation of written Japanese, which naturally has an influence on the spoken language. Every character usually has two readings: the kun, which is an indigenous Japanese word, and the on, which is an old Chinese loan morpheme. These two readings are closely associated with each other and alternate freely in word formation. For instance, Keid *Daigaku* "Keio University" is abbreviated to Kei-dai, using the original on reading; whereas Waseda Daigaku "Waseda University" becomes *Sō-dai,* formed from the on reading for the indigenous Japanese word wase in Waseda. Wase, which means "early-ripening variety of rice," is shown by a combination of two Chinese characters meaning "early" and "rice," respectively. The on reading of the character meaning "early" is sd, hence Sd-dai.

In the vocabulary of contemporary Japanese, all the words beginning with p- are either onomatopoetic or recent loanwords from European languages—*e.g., pan* "bread (from Portuguese), pen "pen" (from English)

—because the initial p- of Old Japanese has changed into *h*- in Modern Japanese. In addition, there are very few indigenous words ending in a nasal sound. Such phonological peculiarities are now utilized by pharmaceutical and other companies, who coin for their new products names containing a p or a nasal sound so that they sound like something modern or imported. Especially since World War II, commercialism has deluged the language with English and other European words, many of which are short-lived. In general, however, almost all the foreign words have been adapted to Japanese sound patterns (*e.g.*, baiorin "violin," bisuketto "biscuit," *rejā* "leisure"), and only a few have brought new sound combinations such as ti, di (*e.g.*, *pātī* "party," birudingu "building").

Early records of Japanese

History. There is no evidence that the Japanese had their own script before they adopted the kanji (Chinese characters) early in the Christian Era. The earliest records of Japanese consist of several words found in a Chinese book of history, the Wei Chih, of the late 3rd century. A few words written with kanji are found on the swords and the mirrors of the 5th and 6th centuries, but the earliest extant Japanese documents of any length are the Kojiki (712) and the *Manyōshū* (later than 771) of the Nara period. From the 9th century on, records, mainly of the Kyoto dialect and the common written language, abound. The history of the language is usually divided into Old Japanese (to the 8th century), Late Old Japanese (9th–11th centuries), Middle Japanese (12th–16th centuries), and Modern Japanese (from the 17th century). Old Japanese was considerably different from Modern Japanese in phonology, morphology, and vocabulary, but not so much in syntax (the arrangement of words and word elements in sentences).

From Old Japanese to Modern Japanese there have been numerous sound changes, among them the shift of initial p- to h- in most of the modern dialects and the loss of three vowels, usually represented as *ï*, e, and 6. Some remnant of vowel harmony was seen in Old Japanese. In vowel harmony certain vowels are restricted by the language structure from occurring in successive syllables of a word. Thus, in one word root or stem o never co-occurred with o, and rarely with *u* and a; *e.g.*, the -ko ending in *kökö* "here," *sökö* "there" appears as -ku and -ko in *iduku* "where" and *miyako* "metropolis (*miya* "palace"). These words have become koko, soko, doko, *mi*yako in Modern Japanese.

In Old Japanese and Late Old Japanese there was a distinction between "finite" forms (which occur at the end of a sentence as the predicate) and noun-modifying forms of verbs and adjectives—*e.g.*, the finite forms *uku* "receive(s)," oku "get(s) up," shiroshi "is white" differ from the corresponding noun-modifying forms *uku*ru, okuru, *shiroki* (later shiroi). In Middle Japanese the latter began to replace the finite forms (uku, oku, *shiro*shi), which ultimately disappeared from the spoken language. In Old Japanese and Late Old Japanese there was a rule of peculiar syntactic agreement: when the predicate verb or adjective was preceded by a word suffixed with such a particle as *zo* (emphatic), ka (interrogative), and so forth, the sentence was finished with the noun-modifying form instead of the finite form. This syntactic restriction disappeared, however, as the result of the above-mentioned change in syntax. Later, on the analogy of the forms such as ukete "having received" and okite "having got up," ukuru and okuru changed into the present forms—*ukeru* "to receive," *okiru* "to get up."

The on and kun of the kanji

Writing systems. During the several centuries after the adoption of the kanji, the Japanese apparently used classical written Chinese as their formal written language. As they became accustomed to the characters, however, they tried to write Japanese with them, and in the process the on and *kun* of every kanji became established. Each kanji represents a Chinese word or morpheme, which has its own sound and meaning. The on is a Japanese imitation of the Chinese sound; *e.g.*, the 8th-century Chinese forms pat "eight" and tap "answer" became Japanese *pati* and mfu, respectively. The kun of a kanji is an indigenous Japanese word with a meaning similar to that of the Chinese; it is this reading that the Japanese are accustomed to give the kanji. Because the Chinese word pat means "eight," the kun reading of the kanji for this word is ya, which is a Japanese term for eight. The kanji for the Chinese word pa "wave" was read by the Japanese either as pa (an on) or as nami (a kun). When a kanji is used to represent a Japanese syllable by means of its on or *kun* without reference to its meaning, it becomes a kana, a phonogram. For instance, when the kanji for the Chinese word pa "wave" is used for the sound pa in such Japanese words as pana "flower" and *payashi* "fast," it becomes a kana. As early as the 6th century, there are examples of kanji used as kana. Although the Kojiki and the *Manyōshū* of the 8th century are written exclusively in kanji, the language represented is Japanese, not Chinese. The Kojiki, however, uses more kun readings, mixing them with kanji used according to the Chinese syntax, while the *Manyōshū* uses many more kana. This is why kanji used as kana are called *manyōgana* (kana often becomes *-gana* in compounds).

The growing popularity of hiragana

As it was a toilsome task to write Japanese with the kanji (which are squarish and complicated in shape), the Japanese began to write them in such a cursive and simplified way of their own that the symbols retained little or no vestige of their original shape. The resulting syllabic characters, called hiragana, "common kana" (but known as *onnade* "letters for women" in the Heian period), began to appear in the 9th century. These simplified characters were used extensively by women, who wrote many poems, diaries, and novels during that period. There is, however, evidence that men also learned and used the script, although they wrote their diaries in kanji. Parallel with this script another system of syllabic writing developed in the 9th century; it was called katakana (kata "one side, one of a pair"). When the priests of the temples in Nara read Chinese texts, especially Buddhist scriptures, they would translate into Japanese as they went along, and would jot down beside the kanji for their own memory the Japanese particles, endings, and so forth that were lacking in Chinese. This was done with symbols made for private use, mainly symbols formed by abbreviating the strokes of the kanji. Originally every sect, and sometimes every person, had a special system, so that there were various symbols for one and the same syllable. In the 10th century, however, more common features began to appear, indicating that the symbols were becoming more common and popular in use. In this way Japanese began to be written with kanji and katakana intermixed and sometimes only with katakana. In the 15th century, however, the hiragana symbols, which were cursive and fine in shape, became the more popular script; literary works were written in hiragana, while scholarly or practical books were written in katakana. Even some literary works written with kanji and katakana in earlier days were rewritten with kanji and hiragana. The present orthography is in kanji and hiragana, and only European loanwords and onomatopoetic words are written with katakana. Only katakana, however, is used in telegrams and in notes typed or printed with machines in companies and offices.

In the beginning of the Meiji Era, the grammar of the language written with kanji and kana still was archaic, based mainly on that of Old Japanese and early Late Old Japanese. Toward the end of the 1880s, however, famous writers such as Futabatei Shimei, Yamada Bimyō, and Ozaki Kōyō began a successful movement to write in the colloquial style. The orthography, with its thousands of kanji symbols, was difficult to learn and to use. Therefore, after World War II, the government carried out a series of reforms that had been advocated for many years. The kana spelling, based on the sounds of early Late Old Japanese, was changed to conform to the contemporary pronunciation; and the kanji characters, which had been used without restriction, were limited to 1,850 symbols for official and daily use, and their shapes and strokes were greatly simplified.

There also have been a number of advocates of romanization since the Meiji Era, but such a program presents many difficulties. Old Japanese syllables had a much simpler structure than those of the contemporary Chinese, which is a monosyllabic language without word endings. Therefore, many different Chinese words or morphemes became homophones (words pronounced alike) in Japanese as early as the 8th and 9th centuries; *e.g.*, the Japanese imitated the Chinese ***k, k', x*** (and ***g,*** y) sounds with only a single ***k*** sound. Moreover, sound changes in Japanese during the succeeding 12 centuries produced a great number of homophones; for example, the sound sequences ***au*** (from Chinese ***au*** and ***ang), afu, ou,*** and ***ofu*** that were distinct from each other in Old Japanese and early Late Old Japanese have all become the same *ō* today. Accordingly, Japanese now has a great many homophonous kanji, so that the Chinese loan morphemes that clearly convey certain meanings when written with kanji would very often become incomprehensible when romanized. Moreover, kanji have more characteristic configurations than would the same words written with roman letters, and thus enable rapid reading. In addition, a high literacy rate prevails throughout the entire Japanese nation, and the people are presumed to be so accustomed to the kanji and kana that such a great change in writing as romanization would undoubtably encounter very strong resistance.

BIBLIOGRAPHY. TAMAKO NIWA and MAYAKO MATSUDA, ***Basic Japanese for College Students*** (1964); and TEC COMMITTEE ON JAPANESE, ***TEC Japanese: The Basic Course, 2*** vol. (1968–70), are both recommended for learning spoken Japanese. The latter is accompanied by magnetic tapes that record all the utterances in the books in very good standard pronunciation. ELEANOR H. JORDEN and H.I. CHAPLIN, ***Beginning Japanese,*** 2 pt. (1962–63), also has good examples of Japanese sentences. HOWARD HIBBETT and GEN ITASAKA, ***Modern Japanese: A Basic Reader, 2*** vol. (1965), is a good book for learning written Japanese.

ROY ANDREW MILLER, ***The Japanese Language*** (1967), an up-to-date and comprehensive outline of the language with a good selective bibliography, includes chapters on the historical and geographical setting, genetic relationship, written systems, dialects, phonology, loanwords, "special and notable" utterances, and grammar and syntax. THOMAS A. SEBEOK (ed.), ***Current Trends in Linguistics,*** vol. 2, ***Linguistics in East Asia and South East Asia*** (1967), includes very good accounts on various aspects of Japanese linguistics, with selective bibliographies.

The following works are in Japanese: SHINKICHI HASHIMOTO, ***Chosakushū,*** 13 vol. (1946–), includes the most fundamental and comprehensive articles on Japanese and Japanese linguistics. SANKI ICHIKAWA and SHIRO HATTORI (eds.), ***Sekai Gengo Gaisetsu,*** vol. 2 (1955), contains a good outline of Japanese. HIDEYO ARISAKA, ***Kokugo On'inshi no Kenkyū,*** rev. ed. (1957), is a collection of excellent papers treating various phonetic and phonological problems in the history of Japanese; his *Jōdai On'inkō* (1955) is a fundamental detailed study on the sounds of 8th-century Japanese. ***Nihongo no Rekishi*** (1947) contains good concise articles on the various periods of the history of Japanese by several authors. Among the dictionaries, the following are recommended: KOKUGOGAKKAI, *Kokugogaku* ***Jiten*** (1955), a dictionary of Japanese linguistics and philology; KYOSUKE KINDAICHI (ed.), ***Jikai*** (1952), a comprehensive dictionary of Japanese with accent notation; ***Jidai-betsu Kokugo Daiiitsen: Jōdaihen*** (1967) lists exhaustively the vocabulary of Old Japanese; SENKICHI KATSUMATA, ***New Japanese–English Dictionary*** (1954).

(S.Ha.)

Japanese Mythology

Japanese mythology here refers to the indigenous mythology of Japan. Mythological forms and concepts derived from Buddhist or Chinese mythology are presented only insofar as they have influenced the development of indigenous Japanese mythology. This article includes an outline of some of the most important myths in Japanese religion as well as information on their origins, relation to legend, ritual, and art, and their use in politics. The myths of the Ainu people are not included.

SOURCES AND TYPES OF MYTH

Ethnic and social origins. The earliest collection of myths in Japan—the *Koji-ki* (712) and ***Nihon-gi,*** or ***Nihon shoki*** (720)—are by no means examples of purely indigenous religious belief. Even at this early time, the influence of China can be discerned. Both of these documents were written with Chinese characters, and the ***Nihon-gi*** was actually written in Chinese. Chinese ideas also permeate Japan's earliest recorded myths; this influence may be seen, for example, in Japanese notions about Yin and Yang—the division of nature into two separate, opposing categories. Furthermore, even the material that is not obviously Chinese probably had its origin far from the Japanese archipelago.

Diverse ethnic groups

The Japanese islands were inhabited as far back as the Pleistocene Epoch. Subsequently, there took place several waves of immigration from various sources, adding diverse elements to the making of the Japanese people. In earlier times, a north–south interflow of culture along the Pacific coast of Japan had created a "circa-Pacific culture." One ethnic group active in this period, the Ainu, are still found in the northernmost part of the country. Their religions and myths contain many vestiges from the days when they lived primarily by hunting and fishing. One example is their belief that game animals live in human form in another world. They come to this world as animals to be hunted and eaten by people and if the latter pay sufficient homage to their spirits, they go back to their own world with satisfaction, carrying with them gifts from their hunters. The following year they return to reciprocate the favour. This "potlatch" idea is shared by many native tribes in Eurasia and North America.

The Neolithic era lasted fairly long in Japan, where it is specifically called the Jbmon period. Archaeological artifacts from this period include female clay figures, apparently representing the Earth Mother or the goddess of fertility; earthenware decorated with viper patterns that seem to reflect a cult of the serpent; and clay masks that may have been worn by shamans (medicine men). The Japanese islands–with this aboriginal culture then received successive waves of immigrants from the nearest continent (East Asia) and the Pacific. The northern continental area of East Asia consists of steppes and deserts, and lying south of it is an agrarian area with great rivers. There was a process of confluence between the peoples and cultures of the two areas on the continent, and under the pressure of moving ethnic groups there, some of them are believed to have emigrated to Japan via the Korean Peninsula and other parts of the China coast, probably including the Shantung Peninsula.

The Jbmon period was followed by the Yayoi period (c. 250 BC to c. AD 250), in which metal articles began to be used. Around this time agriculture became prevalent in Japan. Small centres of political power emerged in various parts of the country, and finally a fairly large-scale regime headed by a shamanic female sovereign came into being. The most important farm crop in Japan, rice, is believed to have been brought initially from South China, and the religion and myth of the ancient Japanese are notably characterized by rites in which farmers paid homage to their ancestral gods by offering them samples of rice they had grown and praying for a rich crop the following year.

In the subsequent Tumulus (Great Burial) period (*c.* AD 250 to c. AD 500), the Japanese lived in a society composed of clans ***(uji),*** each worshipping its own ancestral gods (called *uji-gami*) and headed by a chief ***(uji-no-kami)*** who both presided over religious rites and held secular power. But this role was gradually divided into a political and a religious function, and those specializing in the latter further split into a group consisting of male shrine priests (called ***megi, hafuri,*** or ***kannushi)*** and another group comprised mostly of strongly shamanic females (called ***miko).*** Some ***miko*** belonged to shrines whereas others did not. The latter included ***ichiko,*** who were active among common people as mediums, through whose mouths the spirits of others, both living and dead, were believed to speak. It was also believed that people were sometimes possessed by the spirits of such animals as foxes, raccoon dogs, and snakes, and had some peculiar mental experiences in that condition. It was held as well that some families were particularly akin to such spirits. Beliefs of this sort are also found in some tribes on the Asian continent, such as the Miao. In Japan, they are apparently survivals from old tribal times. Some ritual dances and music performed by priests and priestesses were later secularized and became the origins of recreational theatricals and dances.

Most of the surviving Japanese myths are recorded in history books written in the early 8th century. They tell of the origin of the ruling class, and were apparently aimed at strengthening its authority. Therefore, they are not pure myths but have much political colouring. Nevertheless, they incorporate many typical folklore patterns, some of which are believed to have originated somewhere far on the Eurasian continent. For example, the story of Susanowo killing an eight-headed serpent and saving princess Inada belongs to the Perseus-and-Andromeda type of tales. One Japanese scholar, Seki Keigo, believes that the story is related to three universal folklore patterns: the Dragon Slayer, the Three Stolen Princesses, and the Twins or Blood Brothers. According to him, the myths concerning Susanowo and Okuni-nushi (hero of the Izumo cycle of regional myths) have the same origin. The original story was that two heroes destroyed a monster and saved a woman; then one of them betrayed the other and killed him; the latter soon came back to life and retrieved the woman by showing some evidence. To understand Japanese myth, it is essential to analyze the process of its making in the above manner and examine its components individually. But until the end of World War II, Japanese scholars were not permitted to analyze Japanese ancient history and myth with complete freedom, and little analytical work was done in that field. After the war, the new cultural freedom led to open research into Japanese mythology and a variety of novel theories about it. None has been established as universally acceptable.

Myths and genealogies

Documentary and ritual sources. Genealogies and mythological records were kept in Japan, at least from the 6th century AD and probably long before that. Even before these materials were written down, professional reciters (katari-be) attached to the great clans preserved traditions by word of mouth. By the time of Emperor Temmu (7th century), it became necessary to know the genealogy of all important families in order to establish the position of each in the eight levels of rank and title modelled after the Chinese court system. For this reason, Temmu ordered the compilation of myths and genealogies that finally resulted in the Koji-ki and Nihon-gi. The compilers of these and other early documents had at their disposal not only oral tradition but also documentary sources such as the "Imperial Sun-Lineage" and the "Ancient Dicta of Former Ages" in the case of the Koji-ki. A greater variety of sources was available to the compiler of the Nihon-gi. While the Koji-ki is richer in genealogy and myth, the Nihon-gi adds a great deal to scholarly understanding of both the history and the myth of early Japan. Its purpose was to give the newly Sinicized court a history that could be compared with the annals of the Chinese. In addition to these two books, there is also the *Kogoshūi* (AD 807), the records of traditions transmitted by the Imbe, a hereditary priestly clan. Other valuable records of early myths are found in provincial gazetteers (*fūdoki*) prepared in the 8th century by the order of the Imperial court. These provide a glimpse of Japanese myths not bent to serve the interests of the court. Mythological legends were also contained in liturgical prayers (norito) that were recited during religious services in the Imperial court.

Myths and rituals

Myths and rituals are closely related to each other. The most important ritual function of the court was to ensure a good harvest of rice. Many ceremonies were performed for this purpose, the most important being the harvest festival, or Niiname-sai.

The ancient Japanese initially did not worship their gods in shrines, but believed that the gods descended onto beautiful conical mountains, high trees, thickets, huge rocks, and other such spots, which were considered sacred. In the spring, the people invited the gods to journey from such places to their villages; in the fall, the gods were sent back. The practice of felling a sacred tree on a mountain, dragging the log down to a village, and erecting it as a mi-hashira (holy pillar) must be connected in some way to this old belief. To people living on a seashore, gods seemed to come from somewhere beyond the sea. At many shrines, the custom is still maintained of carrying a portable shrine back and forth to a beach at festival time.

MYTHIC THEMES AND THEIR CHARACTERS

Cosmology

Imperial court myths. The purpose of the cosmologies of the Koji-ki and Nihon-gi is to trace the Imperial genealogy back to the foundation of the world. In the beginning, the world was a chaotic mass, an ill-defined egg, full of seeds. Gradually, the finer parts became heaven (Yang), the heavier parts earth (Yin). Deities were produced between the two: first, three single deities, and, then, a series of divine couples. According to the Nihon-gi, one of the first three "pure male" gods appeared in the form of a reed that connected heaven and earth. A central foundation was now laid down for the drifting cosmos, and mud and sand accumulated upon it. A stake was driven in, and an inhabitable place was created. Finally, the god Izanagi ("he who invites") and the goddess Izanami ("she who is invited") appeared. Ordered by their heavenly superiors, they stood on a floating bridge in heaven and stirred the ocean with a spear. When the spear was pulled up, the brine dripping from the tip formed the Onogoro, an island that became solid spontaneously. Izanagi and Izanami then descended to this island, met each other by circling around the celestial pillar, discovered each other's sexuality, and began to procreate. After initial failures, they produced the eight islands that now make up Japan. Izanami finally gave birth to the god of fire and died of bums. Raging with anger, Izanagi attacked his son, from whose blood such deities as the god of thunder were born. Other gods were born of Izanami on her deathbed. They presided over metal, earth, and agriculture. In grief, Izanagi pursued Izanami to Hades (Yomi) and asked her to come back to the land of the living. The goddess replied that she had already eaten food cooked on a stove in Hades and could not return. In spite of her warning, Izanagi looked at his wife and discovered that her body was infested with maggots. The angry and humiliated goddess then chased Izanagi from the underworld. When he finally reached the upper world, Izanagi blocked the entrance to the underworld with an enormous stone. The goddess then threatened Izanagi, saying that she would kill a thousand people every day. He replied that he would father one thousand and five hundred children for every thousand she killed. After this, Izanagi pronounced the formula of divorce. Whatever the original meaning of this myth, it seems to be used here to account for the growth of the population of Japan.

Izanagi then returned to this world and purified himself from the miasma of Hades of the seashore. From the lustral water falling from his left eye was born the sun goddess Amaterasu, ancestress of the imperial family. From his right eye was born the moon god and from his nose, the trickster god Susanowo. Izanagi gave the sun goddess a jewel from a necklace and told her to govern heaven. He entrusted the dominion of night to the moon god. Susanowo was told to govern the sea. According to the Koji-ki, Susanowo was dissatisfied with his share and ascended to heaven to see his older sister. Amaterasu, fearing his wild behaviour, met him and suggested they prove their faithfulness to each other by bringing forth children. They agreed to receive a seed from each other, chew it, and spit it away. If gods rather than goddesses were born, it would be a sign of the good faith of the one toward the other. When Susanowo brought forth gods, his faithfulness was recognized, and he was permitted to live in heaven.

Susanowo's pranks

Susanowo, becoming conceited over his success, began to play the role of a trickster at the court of Amaterasu. Such tricksters, found in the myths and tales of many folk, often play the part of ribald anticreators who, in effect, bring forth a countercreation. Susanowo scattered excrement over the dining room of Amaterasu, where she was celebrating the ceremony of the first fruits. His worst offense was to fling a piebald horse into Amaterasu's chamber, a horse he had "flayed with a backward flaying" (a ritual offense). At the sight of this monstrosity, "the women weaving the heavenly garments were so much alarmed that they struck their shuttles (hi) into their private parts and died" (*Koji-ki* I: xv). In another version (Nihon-gi I: **37** and **43**), Amaterasu herself was wounded. According to the *Kogoshūi,* Susanowo broke down the divisions of the heavenly rice fields by filling up the irrigation channels, opening the floodgates, and sowing seed over again. Then he erected rods in the fields, causing the rice to deteriorate. Susanowo's own fields were described as barren (*yase*) places.

Librairie Larousse

Amaterasu, attracted by the laughter of the gods, emerges from her cave. Detail of a woodblock print by Kunisada Toyokuni II (1777–1835).

Enraged at the pranks of her brother, the sun goddess hid herself in a celestial cave, and darkness filled the heavens and the earth. The gods were at a loss. Finally, they gathered in front of the cave, built a fire, and made cocks crow. They erected a sacred evergreen tree, and from its branches they hung curved beads, mirrors, and cloth offerings. A goddess named Ame-no-uzume then danced half-nude. Amaterasu, hearing the 800 myriads of gods laughing and applauding, became curious and opened the door of the cave. Seizing the opportunity, a strong-armed god dragged her out of the cave. This story, like that of the Greek Demeter and other Near Eastern tales, is closely related to winter festivals, during which the barrenness of the land is associated with the grief of the goddess of earth. (Some scholars have wondered why a heavenly deity should do such earthly things as cultivate rice fields and hide in a cave and have speculated that Amaterasu was originally an earth goddess.) But this theme—of the celestial cave—has long been viewed by many scholars as representing a solar eclipse, its prototype being the old solar eclipse myth found among the Austroasiatic peoples of Southeast Asia.

Susanowo on earth

Having angered the heavenly gods and having been banished from heaven, Susanowo descended to Izumo, where he rescued Princess Marvellous Rice Field (**Kushi-inada-hime**) from an eight-headed serpent. He then married the Princess and became the progenitor of the ruling family of Izumo. Susanowo, an evildoer in heaven, became a perfect gentleman after his descent to earth. He made trees grow to provide lumber for boats, palaces, and coffins, and fathered dtoshi-no-kami (god of the harvest) and Uga-no-mitama-no-kami (god of rice), by a daughter of the mountain god. Susanowo was intimately associated with the processes of fertility. Among his many descendants were gods related to water, rivers, cereals, trees, and thunder. The most important member of the family of Susanowo was the god **Ōkuni-nushi,** the great earth chief, who assumed control of this region before the descent to earth of the descendants of the sun goddess. Molested by other gods, he once had died but was revived by the efforts of his mother. After this, he ran away to the netherworld, married a daughter of Susanowo, and returned to the upper world. Tradition says that with the help of **Sukuna-bikona-no-mikoto,** a midget god who had drifted ashore in a tiny boat, **Ōkuni-nushi** governed the country and benefitted the people by establishing the arts of healing.

Before long, Amaterasu, the leader of the celestial gods—the gods of Izumo were known as earthly gods—asked **Ōkuni-nushi** to turn over the land of Izumo, saying that "the land of the plentiful reed-covered plains and fresh rice ears" was to be governed by the descendants of the heavenly gods. After the submission of Izumo, **Amaterasu** made her grandson Ninigi (a word said to represent rice in its maturity) descend to earth. According to the Nihon-gi, Amaterasu handed Ninigi some ears of rice from a sacred rice field and told him to raise rice on earth and to worship the celestial gods. The grandson of the sun goddess then descended to the peak of Takachiho (meaning high thousand ears) in Miyazaki, Kyushu. There he married a daughter of the god of the mountain, named Konohana-sakuya-hime (Princess Blossoms of the Trees). At the time, the father of this princess had offered Ninigi his choice between his two daughters, the other being **Iwanaga-hime** (Princess Rock Durable). It was his fateful choice of beauty over durability that caused the life-span of men and even of emperors to be short from that day forth.

The lost fishhook story

When his wife became pregnant in a single night, Ninigi wondered whether the child was his own. In order to vindicate her innocence, the princess set fire to her room, giving birth to three sons in the fire. The last born was Hikohohodemi-no-mikoto (Fireshade), a hunter. His elder brother, Hoderi-no-mikoto (Fireshine), was a fisherman. On one occasion, they exchanged the implements of their respective ways of life, namely, a bow and arrow for a fishhook. The younger brother then lost his brother's fishhook in the sea. When the elder brother demanded the return of his fishhook, the younger god was compelled to go into the sea to search for it. There he married the daughter of the sea god and returned with the fishhook, which he had recovered with her assistance, and also with two magical jewels that controlled the tides. With the supernatural power of these jewels, he reduced his elder brother to poverty and finally brought him to his knees by means of a flood.

This story of Hikohohodemi-no-mikoto going into the sea to recover a lost fishhook is related to the common folklore pattern of the Lost Fishhook. Similar tales are found widely both in Oceania and on the Asian continent, and the Japanese version particularly resembles those in Micronesia and Indonesia. An interesting fact in this connection is that the descendants of the older of the brothers, Hoderi, who had lost the contest and become a subject of Hikohohodemi, were said to be the ancestors of a Kyushu tribe called the Hayato. This tribe later supplied guards for the fences of the Imperial Palace for many generations; these guards also performed a dance in the court **Daijō-sai** festival, holding shields decorated with a spiral pattern that was said to represent the lost fishhook. Moreover, the Hayato have been considered to share many of the characteristics of southern Pacific peoples.

The pregnant daughter of the sea god came ashore in Miyazaki and gave birth to a child on the beach. The child's name was Ugaya Fukia-ezu-no-Mikoto, later to be the father of the first legendary emperor of Japan, Jimmu. While this is generally regarded as the watershed between the "age of the gods" and the historical age, Jimmu's eastern expedition was also a myth. The story probably reflects the memory of the historical fact that the Imperial clan once moved eastward from Kyūshū to the Yamato province. Similarly, both the story of Prince Yamato-takeru's subjugation of Izumo, Kyūshū, and the eastern provinces during the reign of the 12th emperor, Keikb, and the story of Empress Jingō or Jingii, empress of the 14th emperor, Chiiai, conquering the kingdom of Shiragi in Korea have the air of myth though they are based on the historical memory of an actual Korean expedition and the conquest of domestic provinces. The life of Yamato-takeru was filled with suffering and ended, finally, in tragedy. He has many of the characteristics of the hero of the Western classical epics.

Folk myths and legends. Among the many myths that have been preserved, one seems to be strangely related to the myth of Hainuwele, a deity found in New Guinea. This is the myth of a god from whose sacrificed body various kinds of grains and tubers are produced. In Japan this myth seems to be associated with a group called the "celestial Kuma people," presumably archaic agriculturalists. The sun goddess, one variant of the story reveals, sent the moon god to visit the goddess of food. He was angered when she produced food for him from her mouth and killed her on the spot. From the body of the dead goddess grew millet, panic, rice, wheat, large and small beans, and even cattle and horses. The Kuma people took these things to heaven and offered them to the sun goddess. The latter was pleased and used these cereals as seed in her dry and wet fields.

While myth tells how things came to be and what is behind all things, legends tell about marvellous events in a specific place. Unlike fairy tales, both myths and legends are believed to be true. Because legends tend to flourish in closely knit social groups, the traditional Japanese village was virtually a hothouse for the production of folktales. In the villages, there grew up stories about the ***kappa,*** a goblin with a saucer-like head that holds water. When the water is spilled, the ***kappa*** loses his strength. Known to lurk about rivers, the ***kappa*** often waylays children or horses, killing them by pulling out the intestines through the anus. Another demon in Japanese folklore is the ***tengu,*** a creature with a long, pointed nose who lives in pine trees in the mountains, where he often abducts human beings. Tengu, it seems, were originally worshipped as mountain spirits, but gradually lost status until, partly under the influence of Shugen-dō (an order of mountain ascetics), they came to be regarded as evil goblins. The ***oni*** also were originally regarded as mountain spirits and played the role of farm gods. With the lapse of time, they fell in status, and under the influence of Buddhism finally turned into a sort of fiend.

Tales concerning animals

Many animals have the power to bewitch men. The cat, snake, and badger are especially important in this respect. The fox was a creature familiar to common people as it was commonly found around their villages. Hence, it came to be regarded as a messenger of Inari, one of the most popular deities in Japan. The fox was also said to bewitch people. Sometimes it disguised itself as a beautiful woman and tempted a man, who on coming to his senses some time later would find himself in an open grave or going about on all fours eating leaves, to the astonishment of his friends. Another demon is the earthquake fish, an enormous catfish whose motions cause earthquakes on the land above it. Curiously, many of these demons also have a benign aspect. The ***kappa*** and ***tengu*** sometimes teach people the arts of swordsmanship and healing. The fox has been known to sacrifice itself or its children to help cure its benefactors. The earthquake fish brings employment to impoverished artisans when it destroys buildings in earthquakes.

POLITICAL AND SOCIAL ASPECTS OF MYTH IN JAPAN

While myth in Greece evolved into literature and, in China, into history, in Japan the main collections of myth that have been preserved have been largely political documents. They have established divine origins for the Imperial and paraimperial families, transforming power into authority. In medieval times, the imperial family declined in actual power, and Buddhism and Confucianism came to influence Japanese mythology. Buddhism added stories of saints and ***bodhisattvas*** (Buddhas-to-be) to the already rich store of Japanese legends, while Confucianism added a moralizing touch. Native elements became thoroughly mixed with legendary themes from China and even India. In the middle of the Tokugawa period (1603–1867), there was a revival of interest in what was believed to be pure Japanese mythology. With the restoration of the imperial regime in 1868, this archaizing tendency in Japanese culture was given greater scope. Shintō shrines were purified of all Buddhist accretions and were reorganized under the program of State Shintō—according to which religion and politics were seen as one. Myths were included in the public school curricula and became so sanctified that they could no longer be studied critically. The literal belief in a divine emperor. who was the descendant of the sun goddess and the symbol of the sacred mission of Japan, provided a mythological foundation for the ethnocentric and nationalistic fanaticism that resulted in Japan's expansionism during the 1930s and final defeat in 1945. After World War II, mythology was largely eliminated from school curricula, while Shintō shrines lost their special government protection. More recently, however, attempts have been made to reintroduce the ancient myths into the schools. On the academic scene, scientific research on myths and folklore has been set free from the government's interference and has been stimulated by recent progress in ethnology, archaeology, history of religions, folklore studies, and other related disciplines.

BIBLIOGRAPHY. The two basic sources for knowledge of Japanese mythology are the ***Koji-ki*** and the ***Nihon-gi,*** 8th-century compilations that show Chinese influence but retain the fundamental structure of Japanese myths. The former of these, Japan's oldest written record, was first translated by BASIL H. CHAMBERLAIN as "'Ko-ji-ki' (Records of Ancient Matters)," ***Transactions of the Asiatic Society of Japan,*** vol. *10,* suppl. (1882); and more recently by DONALD L. PHILIPPI as *Kojiki* (1968). Philippi's introduction, notes, and glossary are especially helpful. The ***Nihon-gi*** complements the earlier work in many ways and was translated by W.G. ASTON and published in 1896 as ***Nihongi, Chronicles of Japan from the Earliest Times to A.D. 697;*** originally published in the ***Transactions of the Japan Society,*** suppl. *1,* and reprinted in 1956. Another valuable source is the ***Kogoshiii,*** an account written according to the Imbe clan, a Shintō priestly family; it dates from the early 9th century and has been translated by G. KATO and H. HOSHINO under the title ***Kogoshiii: Gleanings from Ancient Stories,*** 2nd ed. rev. (1925). A good but brief discussion of the topic is E. DALE SAUNDERS' contribution, "Japanese Mythology," in S.N. KRAMER (ed.), ***Mythologies of the Ancient World*** (1961). The standard study is ***Essai sur la mythologie japonaise*** by NOBUHIRO MATSUMOTO (1928). Folklore aspects of the topic may be found in studies in RICHARD DORSON's edition of translated articles by leading Japanese folklorists, entitled ***Studies in Japanese Folklore*** (1963); in KEIGO SEKI, ***Folktales of Japan*** (1963); and in HIROKO IKEDA, ***A Type and Motif Index of Japanese Folk-Literature*** (1971). C. OUWEHAND, *Namazu-e **and Their Themes*** (1964), is a fascinating study of mythic, legendary, and folkloristic components in a collection of 19th-century pictures.

(N.M.)

Japanese Philosophy

The Japanese equivalent for "philosophy" is ***tetsugaku,*** a word coined when Western schools of thought were first introduced into Japan after the Meiji Restoration (1868), the great upheaval that overthrew the shogunate (which had governed Japan for over 250 years), restored Imperial rule, and opened the country to the West. Before the Meiji era, one of the two principal schools of Japanese thought arose from Buddhism and was highly tinged with a religious character and often metaphysical; a second arose from Confucianism and was essentially a system of moral philosophy. From the 18th century on thkre were some independent thinkers who were critical of these two major schools. Since the Meiji Restoration, the philosophies of all of these schools and thinkers have been commonly classed together under the term tetsugaku. In the 20th century the influence of Western philosophy has been decisive among intellectuals, though traditional ways of thinking deriving from Shintō, Buddhism, and Confucianism still prevail among the people.

Earlier philosophical thought

THE RISE OF PHILOSOPHY IN JAPAN

The Buddhism of Prince Shotoku. Recorded Japanese philosophical thought began at the turn of the 6th century, with the writings of ShBtoku Taishi, under whom Japan became a centralized state. He lectured on the Chinese versions of three *sūtras* (classical scriptures) of Mahāyāna Buddhism, the more liberal and metaphysical branch of Buddhism. Certain commentaries on these *sūtras* that are jointly called the Sankyd gisho ("Commentaries on the Three *Sūtras*"—on the *Vimalakīrtinirdeśa-sūtra* and the *Śrīmālādevī-siṃhanāda-sūtra* and on the *Saddharmapuṇḍarīka-sūtra*, or Lotus *Sūtra*) may have been written by Prince Shbtoku. Although they are not works of systematic philosophy, they embody a great many philosophical thoughts that are valuable, whoever the author may have been. These commentaries, which are the oldest works in existence among Japanese classical writings, have been of decisive significance for Japanese Buddhism.

Seminal role of the "Commentaries on the Three *Sūtras*"

Philosophical approach. Shōtoku's procedure was dialectical in the sense that one view is made to neutralize another contrasting view in the interest of a more fundamental affirmation. He envisaged the Absolute to be Voidness (*śūnyatā*) and called it nonduality because Voidness is neither Being nor Nonbeing but is Mother, the source from which wisdom originates.

Shdtoku inherited the idea of the central Buddha from the Lotus *Sūtra*. The Mahāyāna substituted for the historical Buddha the eternal Buddha since Buddha's existence in the early form is not his true and proper mode of being. Buddha's essential (or ultimate) body (dharmakdya) had been regarded as ineffable, being located beyond the phenomenal sphere. Prince ShBtoku, however, considered it to be in the phenomenal sphere.

Prince Shbtoku distinguished source, or origin (hon), from appearance, or manifestation (*shaku*). Only what has its source as beginning (Latin, *principium*) can manifest itself in appearance; and that which does not so manifest itself cannot give evidence of its source.

Perfect-one store and Nirvana through practice

The basis of human existence is called "perfect-one store" (*tathāgata-garbha*). In its transmigration the soul continues to exist without perishing, embraced in the womb, or store, of this perfect one, which is the basis of human existence not only after man has been delivered from defilements but also even while man is still in defilements.

In traditional Buddhist philosophy, the accomplished Nirvāṇa was considered to be complete extinction of the self; but Prince Shbtoku tried to find it in the process of religious practice. The ideal status of a human being thus lies in the realization of the unity with ultimate truth "today"—*i.e.*, in daily temporal existence.

This-worldliness. It is well known that primitive Shintō, the native Japanese religion, was closely tied with agricultural rituals in agrarian villages and that Shintd gods have been symbolized, even up to the present day, as gods of production.

After the Japanese came into contact with Chinese religions, they adopted and absorbed Confucianism, which teaches the way of conduct within a concrete human nexus. Confucianism did not conflict with the Japanese thought patterns of the time.

There were problems, however, in the case of Buddhism. Buddhism declared itself to be a teaching of "otherworldliness." Its central figures have all been monks and nuns, who have freed themselves not only from their families but from any specific human relationships. But the topographical characteristics of Japan, vastly different from those of India, required men to serve their fellows within a specific human nexus. The doctrines of early Buddhism, together with those of the ensuing traditional conservative Buddhism, were despised and rejected by the Japanese under the name of Hīnayāna ("Lesser Vehicle"), whereas Mahāyāna ("Greater Vehicle") Buddhism was particularly favoured and adopted; some schools of Mahāyāna, if not all, advocated comprehending absolute truth within a secular life. In accepting Buddhism, the Japanese selected in particular the branch that asserted their own this-worldliness.

Secular life and the lay believer

This attitude in assimilating Buddhism is clearly shown in the case of Prince ShBtoku, whose selection of the three *sūtras* out of the multitude known in India and China was based entirely upon the Japanese way of thinking. The *Shōman-gyō* (*Śrīmālādevī-siṃhanāda-sūtra*) was originally preached, in compliance with Buddha's command, by Madame Shbman (Śrīmālā, or Glorious Garland), who was a queen and a lay believer. The Yuima-kyb (*Vimalakīrtinirdeśa-sūtra*) has a dramatic composition, in which Yuima (Vimalakirti, or Spotless Fame), a lay believer, gives a sermon to priests and ascetics—reversing the usual order—that commends the seeking of truth in secular life. And according to the *Hokke-kyō* (*Saddharmapuṇḍarīka-sūtra*, or Lotus *Sūtra*), all laymen who faithfully follow any of the teachings of Buddha are expected to be saved. The Crown Prince himself all through his life remained a lay believer, thus emphasizing the necessity for realizing Buddhist ideals within concrete human situations. Shdtoku asserted: "Reality is no more than today's occurrence of cause and effect." Existent beings are already qualified for Enlightenment, and there is no necessity for laymen to seek qualifying actions from Buddhas or from bodhisattvas (who forgo Enlightenment in order to minister to men).

Moral values. Prince Shbtoku esteemed actions and good deeds practiced in the worlds of life and death; for their accumulation eventually admits a person into Buddhahood. The ultimate state of religion is not bestowed upon men by divine entities that transcend them but is realized through practical behaviour within the human nexus. "The mind" of man, he said, "is the origin of all virtues." It is likely that the Prince regarded compassion or love as the basic human value.

Esteem of activity. Prince Shbtoku put special emphasis upon altruistic deeds and considered that Buddhas and bodhisattvas should serve all living beings. Shōtoku permitted the acquisition of wealth "in accordance with the law"; and traces even of Utilitarianism can be noticed in his writings.

The ruler's broad concerns and ideals

Applied to politics, this doctrine means that the ideal ruler should be beyond all the differences of dispositions and interests of the people and yet care for them all—not for the sake of their individual interests but for their ultimate welfare in spiritual communion. Thus, the people follow their ruler in full realization that his high aims are derived from the ultimate truth.

Tolerance. Tolerance was a feature of the thought of Prince ShBtoku. He did not forbid or oppress Shintō but instead put it in its proper sphere of action and kept it alive under the leadership of Buddhism. According to the Prince, there is no innate difference between the saint and the most stupid man. Everyone is primarily and equally a child of the Buddha. "Even heretics are your teachers," he said. In fact, in his eyes, there were no heretics.

Early Confucianism. The first main period in the history of Confucianism in Japan extended from the introduction of Chinese learning at the beginning of the 5th century to the spread of the so-called Sung philosophy of Neo-Confucianism at the beginning of the 13th. In this period, a tremendous influence was exerted by Confucian ethics, political theory, and legal and educational institutions, all based on the universal principle of Tao (Japanese Michi)—the effortless way by which all is done. Confucian studies were inseparably connected with the study of the Chinese language, which in turn was indirectly fostered through the introduction of Buddhism. Moreover, the tradition-oriented philosophy of Confucianism was increasingly patronized by the government as a convenient aid in support of state policy.

Ethical, political, legal, and educational theory

Confucian learning became particularly influential in the second half of the 7th century, when it motivated a series of Imperial rescripts that applied universal laws and legal concepts to human affairs. By the end of the century the resulting Taihō code comprised a total of 17 volumes. In the educational system, Chinese Confucian Classics—on ethics, literature, mathematics, and medicine—constituted the entire curriculum.

NARA AND HEIAN PERIODS

The dominant philosophical development during the two ensuing periods of Japanese history, however, was Buddhist. During the period when the capital of the country was at Nara (710–784), philosophical ideas were maintained and discussed by the so-called Six Schools of Nara: and at the beginning of the Heian period that followed (794–1185), when the capital was at Kyōto, the Tendai and Shingon sects appeared.

Six schools of the Nara period. The six schools, introduced from China, were as follows:

The Risshū, or Ritsu, school held as its main principles the observance of strict monastic discipline and, above all, the correct transmission of the holy orders. The monks of this school adhered strictly to the discipline of conservative (Hīnayāna) Buddhism of the kind that flourishes now in South Asian countries.

The Kusha school, which also taught conservative Buddhism, was based upon the Kusha-ron (*Abhidharmakośa*) composed—about AD 320 to 400—by Vasubandhu, an Indian philosopher.

The Jōjitsu school was based upon the *Jōjitsu-ron* (*Satyasiddhi-śāstra*), written by Harivarman (about AD 250–350). A school of conservative Buddhism, it was adapted to some extent to the doctrine of the "Void."

The Sanron school, derived from the Mādhyamika school in India, stressed the doctrine of "Void." *San-ron* literally means the "Three Treatises," in reference to the treatises upon which the school is based: the *Mādhyamika-śāstra* and *Dādaśamukha-śāstra* of Nāgārjuna, and the *Śata-śāstra* of Āryadeva.

The Hossō school, a kind of Buddhist Idealism derived from the Yogācāra school of India, regarded everything as a manifestation of the fundamental Mind underlying all phenomena.

Emphasis on the interdependence of all things

The Kegon school, based on the *Kegon*, or *Avataṃsaka-sūtra*, revered the Vairocana Buddha as its principal object of worship. It emphasized the interdependent relations of various aspects of actual human existence. According to its concept of interdependent origination, all existences and phenomena are interrelated: even a flower is closely connected with all of the universe and has no separate existence in the metaphysical sense. Also, a person cannot sever himself from the past or the future; and this can be said of everything in the universe, for the Kegon philosophy holds that the universe is motivated by one great compassionate soul. Further, the relationships among individual persons are governed by the theory of interdependence under the surveillance of this one great compassionate soul.

The above-mentioned six schools might be called scholastic because their sphere of influence was limited to the monks and did not extend to the common people in general.

Two sects of the Heian period. At the turn of the 9th century, with the beginning of the Heian period, the Tendai and Shingon sects came to Japan.

Tendai philosophy. The Tendai sect was introduced into Japan by Saichō, whose honorary name was Master Dengyb (Dengyō Daishi). Tendai is based on the *Hokke-kyō*, or Lotus *Sūtra*, and accordingly teaches that all men can become Buddhas and should attempt to do so.

The most remarkable aspect of Tendai is its comprehensive character. It found a place for all Buddhist scriptures, regarding them as comprising a progressive revelation, gradually disclosed by the Buddha during his life, as he found that the intelligence of his listeners ripened.

According to man's common sense, it seems that parts depend on one another and all things depend on the whole. But the so-called complete or perfect teaching (*en-gyō*) of this sect went beyond this doctrine: it taught that the whole and the parts are identical, that the whole cosmos and all of the Buddhas are present in a grain of sand or on the point of a hair. A celebrated maxim of the Tendai states that one thought is the 3,000 spheres (*i.e.*, the whole universe), and the 3,000 spheres are but one thought. That is to say, the relations involved in the simplest thought are so numerous that they imply the existence of the whole universe; man's perceptions and thoughts are thus identical with absolute reality. This teaching leads to certain doctrines of ontology, the nature of Being. There are "three forms of existence": the void, the temporary, and the middle—*i.e.*, all things that exist depend upon their relations. If a thinker tries to isolate them and to conceive of them as having no relations, they become unthinkable and, in fact, nonexistent.

The void, the temporary, and the middle

But as temporary formative parts of the whole, they do exist; and the whole could not realize its true nature if it did not manifest itself in particulars. In this sense all things exist temporarily as phenomenal beings. Things exist or do not exist according to men's view of their relations to the whole, but the middle exists absolutely. Phenomena and the one absolute truth are, if rightly regarded, synonymous. Enlightenment, as obtained by the Buddha himself, is achieved when the significance of each of the three is properly realized.

Shingon philosophy. The Shingon philosophy was introduced by Kūkai, or Master Kōbō (Kōbō Daishi), an early-9th-century scholar. It is likely that he had some contact with Manichaeism (a dualistic movement) and with Nestorian Christianity during his stay in China.

Shingon means "True Word"—*i.e.*, a sacred spell (or mantra); and indeed this sect was interfused with magical elements preserved in the popular beliefs of India. By about 700 Indian Buddhism had become a mixed creed and may have incorporated many Iranian and Central Asiatic elements. This form of Buddhism is called esoteric Buddhism (Vajrayāna). The common people of Japan in those days wanted a religion that was impregnated with magic, and that is why the Shingon sect was so strongly welcomed in Japan.

Cosmotheism

In Shingon there have been definite secret doctrines that could be communicated orally; thus, a person not yet initiated could not claim to understand the explanations. The concept of "the essential self," which had become pivotal in Buddhist philosophy, culminated in the Shingon, the chief idea of which was cosmotheism, which is reminiscent of pantheism though somewhat different from it. In this view, the whole universe is regarded as the body of Vairocana, the supreme Buddha, and is composed of six elements: earth, water, fire, air, ether, and consciousness—in short, matter (comprising the first five elements) and mind (*i.e.*, consciousness). The five elements cannot exist without consciousness; nor can consciousness exist without the five elements. The six elements can be examined from two different aspects, viz., the essential (undifferentiated), or unconditional, aspect and the phenomenal, or conditional, aspect. The former refers to the eternal, unchanging substance; the latter, to the everchanging reality.

The essence of phenomenal things lives in absolute truth (the world of the law), and such things are so perfectly interrelated as never to obstruct (or contradict) one another. It follows that mankind and Buddhas are identical in their essence. Kiikai preached that if one follows this reasoning, the formation of figures with one's hands, the recitation of incantations, and the concentration of one's mind (the three actions of man's body, mouth, and mind) would all be identified with the corresponding actions of Buddha.

In its religious practice, the Shingon philosophy ascribed mystical meanings to specific syllables. For example, the mystical syllable ham is not only the symbol but is the living breath that is itself the living substance penetrating the cosmos. This eternal living substance is called the original principle of Three Mysteries or the Three Mysteries of the original existence or, again, the three bodies of the original existence, because the Three Mysteries are the esoteric actions of doing, speaking, and thinking.

The synthesis of sundry philosophies expounded by Master Kōbō was comprehensive. He was conciliatory and tolerant toward the native faiths of the Japanese populace, and his system incorporated not only Buddhist doctrines but also a considerable amount of pagan thought. According to him, even heathen thoughts can be manifestations of the basic principles of Mahii-Vairocana Buddha.

Kamakura and Muromachi periods

RUDDHIST PHILOSOPHY

The Kamakura period (1192–1333) was a distinctive time in which a patrician society had declined, a beneficed clergy had become corrupt, and a military class had arisen. Under great distress, the people needed a simplification of religious ministrations.

Origins and distinctives of the major sects. To deal with this situation, Zen Buddhism was introduced from China at the turn of the 13th century, and the sects of Pure Land (Jōdō-shū) and of Nichiren were founded at the same time. Thereafter, no new sects were created until the Meiji Restoration, in 1868. The major sects of Japanese Buddhism that are still extant are (1) Tendai, (2) Shingon, (3) Zen, (4) Pure Land (Jbdo, Jōdo Shin, etc.), and (5) Nichiren.

As stated earlier, the Tendai and Shingon sects originated in the Heian period; the Zen, Jbdo, and Nichiren sects arose in the Kamakura epoch.

Pure Land Buddhism. Pure Land (more exactly Pure Realm) Buddhism, as such, was founded in the 1st or 2nd century in India. It was based on the larger and smaller *Sukhāvatī-vyūha-sūtra* and on the *Amitāyurdhyāna-sūtra*, which speak of the Western Paradise of the Pure Land (Sukhāvatī, or Jbdo in Japanese), where believers are supposed to be born after death by virtue of grace for their faith and good works.

Hbnen's homage to Amida Buddha

The Pure Land sect emerged in Japan in the second half of the 12th century through the efforts of Hbnen, a priest who could find no inner certainty nor deliverance from evil in the study of Tendai nor in complex rituals nor in meditations. He therefore turned to the Amitiibha Buddha, who was said to have prepared a blissful realm for all who would call upon his name in simple faith. Hōnen therefore organized his many followers around the chanting of the words *Namu* Amida *Butsu* ("Homage to Amida Buddha"), through which alone their salvation was wrought.

Nichiren sect. A prophet and nationalist of the mid-13th century, Nichiren was born the son of a fisherman. He studied all of the schools of thought widely until he decided for himself what was the true way to deliverance. He first entered the Shingon school and then studied in the Tendai school on Mt. Hiei. There, he concluded that only one scripture is needed, the Lotus *Sūtra* (*Saddharmapuṇḍarīka-sūtra,* or "The Lotus of the Good Law"), and that the deliverance of the country from the sufferings of the time could best be achieved by a vigorous campaign for a return to the Lotus *Sūtra* and the Śākyamuni Buddha. Nichiren claimed that a person could be saved only by the spiritual power of the Lotus *Sūtra*, the gospel of which he alone was entitled to spread.

Nichiren, a born religious demagogue, wandered in various districts of the country, literally banging the drum of his beliefs in all quarters. Because of his rudeness to all other sects and to the government, he soon fell into trouble with the authorities, and his life was thus a long chain of persecutions. It was among the followers of this sect, especially those among the common people in later days, that religious fervour reached its most extreme level.

Zen Buddhism. As a specific form of Buddhism, Zen first appeared in China, where it was a peculiar Chinese version of the kind of Buddhism that had been taken from India by the sage Bodhidharma in or about the year 527. Specifically, it was a variety of the Mahāyāna school that was known as Dhyāna Buddhism, called Ch'an in Chinese and Zen in Japanese. Though the nearest English equivalent of Dhyāna is "contemplation," the term Zen has acquired in Japan a static and even dreamy connotation quite foreign to that of Dhyāna.

Dhyāna, Ch'an, or Zen means immediate insight into the nature of reality, or life. All Zen sects revere the historic sage of the Śākyas (Buddha), believe that instruction from a Zen master may awaken in a disciple the Buddha mind that everyone possesses originally, and believe in expressing religious realization in daily work.

Eisai and Dōgen: Rinzai and Sōtō sects

In 1191 Eisai (or Yōsai) carried Ch'an Buddhism to Japan in a form known as the Rinzai sect (Chinese Lin-chi). Another form, called the Sōtō sect (Chinese Ts'ao-t'ung), was introduced by Dōgen, a disciple of Eisai. Dōgen's magnum opus *Shōbō genzō* ("Essentials of Truth") was a treatise on the philosophy of orthodox Zen Buddhism as expressed in the doctrines of the Sōtō sect. Zen still flourishes today in Japan, where it occurs in its most vital form and where it has had a far-reaching effect upon the national culture.

General teachings of the sects. Japanese Buddhism, as divided into these sects, has been the spiritual basis for Japanese culture for centuries past. Thus, the teachings of the three main sects require more careful study.

Deliverance by *Amitābha's* grace in Jddo *Shinshū*. Hōnen's minor, but most influential disciple, Shinran, a religious leader of the early and mid-13th century, founded, as an outgrowth of Jbdo, the Shin ("True") school—*i.e.*, the True Pure Land sect—which is the one most widely professed by the people in the present day. Shinran admitted the deep-rootedness of sin in human existence and carried the idea of Buddha's grace to an extreme conclusion. For this, and because he married a nun and spent his life as a married priest, he has often been compared to Martin Luther.

Shinran taught that men should realize what calamity is involved in the mere fact of their being alive. All living beings are sinful; indeed, no man can live without committing sins. Sin is thus deeply rooted in human existence: men are all *karma*-bound—*i.e.*, subject to latent sinful powers. Hbnen, Shinran's master, had seen "man with blind eyes, capable of doing nothing." The habitual sins of men, however, were frowned upon by Shinran, who allowed no condoning of them.

He stressed exclusively the salvation of common men by the grace of Amitiibha Buddha (Japanese Amida). One might say, "Men cannot perceive the grace of Buddha." Shinran replies, "It is men's grave sins that prevent them from noticing it." Although men are not aware of it, they are already embraced by the grace of Amitābha. Seeing his age as an especially corrupted one, Shinran held that man cannot be saved from the mundane world without relying upon the original vow of Amitābha. All men—whether they are honest or criminal—are, without any distinction, admitted to Amitābha's Pure Realm. Faith in Amitābha's grace is the one and only condition of admission. All men are equally sinful; and Amitābha is a being of compassionate love of the most sincere kind. There is no conception of punishment by Amitiibha. The Shin sect held that evil men are also rightfully eligible for salvation by Amitiibha Buddha. Earlier, one of Hōnen's sayings had been, "Even a bad man will be received in Buddha's Land, but how much more a good man." Shinran, however, turned this to the reverse: "Even a good man will be received in Amida's land, then how much more a bad man." Elaborating this saying, he argued that, whereas a good man may be able to save himself by his own merit, it is not to be expected that a bad man can save himself. He needs the grace of the Buddha, for he has no other means. Now if even a good man, who does not necessarily need grace, can be saved, how much more a bad man, who cannot be saved except by grace. The sinner has only to believe in the grace of Amitābha and the Pure Realm will be his; thus, faith became the sole requisite for salvation, and all of the other Buddhist moral philosophy was swept away. For Shinran no ceremony was necessary for salvation except genuine faith.

Grace and salvation for good and evil alike

In this connection a controversy occurred among the religions of grace in different traditions. In the Pure Land Buddhism of China and Japan, the relation between faith (shin) and works (gyd) was an issue of heated debate. Many leaders thought that both were necessary. But Shinran said firmly that pure faith alone is enough and that no one should rely upon works. The teaching that wicked persons are eligible for salvation was not intended to encourage bad actions. In fact, such a thought was forbidden as heresy by the Shin sect. On the contrary, out of pure faith good deeds emerge spontaneously. Only through meditating on the thought that men are sinful can they come to feel compassionate toward others; and the attitude of compassion can be founded only upon grace.

Way of action in Nichiren and Zen. It was natural that the Nichiren sect, which was an outgrowth of the Japanese Tendai, also laid an emphasis upon faith and grace.

For Nichiren, everyone who realizes the truth of the fundamental unity of reality is a Buddha, and everyone who lives in accordance with this enlightenment and works to propagate the lotus of truth is a messenger of the primeval Tathāgata (the title used by the historical Buddha). To such a man, all that surrounds him preaches the truth, and the place of his abode is a paradise. Nichiren asserted that the crux of Buddha's thought is revealed in the "Jyuryōbon" chapter ("Duration of Life of the Tathāgata") of the Lotus *Sūtra.* Thinkers of the Nichiren sect have stated that, whereas the Tendai sect, from its rise in China onward, took the standpoint of "action according to principles," Nichiren emphasized "action according to things"—*i.e.*, to the phenomenal world.

The absolute in the phenomenal: reality as dynamic

The method of thinking that seeks for the Absolute in the phenomenal plays an effective role in the Japanese assimilation of the Zen sect as well. For Master Dbgen, impermanence is itself the absolute state and is not to be spurned but esteemed. "Impermanence is the Buddhahood . . . . The Impermanence of grass, trees, and forests is verily the Buddhahood." Buddhahood is time, said Dōgen. He who wants to know Buddhahood may know it by knowing time as it is revealed to him. Inasmuch as time is something in which a person is already immersed, Buddhahood is not something to be sought in the future but is something already realized wherever the seeker is. The ever-changing influx of time is identified with ultimate Being itself. Again and again, Dōgen emphasized that true reality is not static but dynamic.

Way of meditation in Zen. The unique contribution of Zen to higher religion has been the many different methods of reaching and presenting the truth that have been used by the Zen masters.

In the Rinzai sect, practitioners concentrate on specific enigmatic or paradoxical, nonlogical questions called *koans*. These are substantially based upon *mondōs*, or brief dialogues between masters and disciples, which illustrate Zen's peculiar method of instruction—pointing to the truth, the real now, without interposing ideas and notions about it. For the Zen master, the best way for a man to express his deepest experiences is by the use of paradoxes that transcend opposites: "Where there is nothing, there is all." Or, "To die the great death is to gain the great life."

The way of practice in Sōtō Zen went even further than Rinzai Zen. Master Dōgen, like Eisai, made meditation the essential practice, but he rejected koans. Indeed, practitioners should not endeavour to concentrate on anything. Said Master Dōgen, "In meditation, if mind is distracted, don't try to suppress it. Let it be as it is." Sōtō Zen emphasizes silent sitting and meditating on whatever illumination or insight is received while waiting in silence.

Zen Buddhism taught intuitive knowledge of the Absolute. A well-known motto of Zen, "Direct pointing to the mind of man," emphasizes that men originally possess the Buddha mind and need only the actual experience of it—*i.e.*, the master points to the Buddha nature, or to the reality itself. Enlightenment takes place in a "Timeless moment"—*i.e.*, outside of time, in eternity; and it is an act of the Absolute itself, not of man's own doing. "Seeing into one's own nature," another motto of Zen, means that seeing the Buddha mind within one is the same thing as becoming the Buddha—that one then is the Buddha.

The Buddha within man and his daily life

Dōgen identified the practice of meditation (zazen) with enlightenment; the innate Buddha nature, he held, is the a priori basis of the practice that itself embodies enlightenment in the process of a man's endeavour. The Zen disciple does not seek for some Supreme Being above him but, instead, spontaneously finds in himself the Buddha nature as the foundation of his own existence. "Let the light be reflected so that it falls back and irradiates the self," said Dōgen. "Then mind and body will of themselves disappear and the original countenance will become manifest."

Zen dispelled all speculation about the Absolute. The Buddha dwells hidden in all of the inconspicuous things of daily life; and to take those events, just as they happen, is all that enlightenment is. Zen is spiritual freedom, the liberation of man's true nature from the burden of the fixed ideas and feelings about Reality that men accumulate through fear—through the fear, for example, that life will run away from them. The truth to which these events point, however, is radically simple and self-evident.

The changing aspects of the phenomenal world are a self-activity in which the self is mightily set forth, a procreation accomplished not under the compulsion of laws or of blind impulse but in the creative power and freedom of sublime wonder. Dōgen said,

> Being is time, and time is Being. Everything in the world is time at each moment. To practice religious disciplines and to attain enlightenment and to enter into Nirvana are nothing but to ascertain that these events are Being, time, and that all time is all Being. . . . Birth and death is the life of Buddha.

Concepts of time and change. The Japanese views of time and change are reflected in their phenomenalism and in their philosophy of history.

Phenomenalism. One of the main features of Japanese thought has been its attitude in accepting the actuality of the phenomenal world as the Absolute. The Japanese have stressed the intuitively apprehended, sensible concrete rather than the universal or abstract and have emphasized the fluid, incipient character of events. They have thus rejected the concept of an Absolute existing over and above the phenomenal world and have sometimes regarded that world itself as the Absolute. This can be seen as early as in the ancient Shintb belief that spirits reside in all kinds of things. These early Japanese personified all sorts of spirits (other than those of humans), regarding them all as ancestral gods and, in their basic reality, divine. Indeed, this way of thinking has persisted in Shinto down to the present day.

Rejection of the Absolute

This way of thinking also conditioned the reception and assimilation of Buddhist philosophy. Because of this acculturation, the Tendai sect in Japan is not the same as its counterpart in China. Thus, the Tendai scholars in medieval Japan, using the same nomenclature as that employed in continental Buddhism, arrived at a distinctively original system of thought, called Honkaku-hbmon, which asserts that the changing aspects of the phenomenal world are the Buddha.

In the Kamakura period (13th and 14th centuries), amid a prevailing anti-intellectual fideism, Buddhists pointed to the realm of phenomenal existence as the sole source of Enlightenment. This radical shift in outlook, with its aesthetic reverence for nature, exerted a significant influence on art and literature and on the "Code of Warriors" (Bushido).

Philosophy of history. Phenomenal existence, however, has a historical as well as a metaphysical dimension. It was therefore natural that, during the turbulent plague-ridden Kamakura era, when life seemed meaningless, the Japanese would be driven to seek meaning in history. The Chinese conception of history had been influential in Japan, as had Chinese forms of Buddhism. But Japanese scholars modified the notion of history introduced from abroad. Probably the most noteworthy of the many classical historical works produced by these scholars is the *Gukan-shō*, by the poet Jien, chief abbot of the Enryaku-ji of Mt. Hiei around the turn of the 13th century, which demonstrates that the concept of reality held by Japanese Buddhists is seen most clearly in their ideas of history. Jien complained that the age in which he lived was a degenerate one—a feeling that was most keenly experienced in the frequent warring around his time.

Most of the Japanese Buddhist sects have taught, as Jien had, that doctrines should always be made "a propos of

Concept of the three times

the time." Later Mahāyāna Buddhism employed the concept of the three times, the three periods that followed the demise of Lord Buddha. The first 1,000 years was called the "Period of the Perfect Law," when the religion of the Buddha was genuinely and perfectly practiced. The second 1,000 years was called the "Period of the Copied Law," when the religion of the Buddha was practiced only by imitating the practices of the sages and monks of the past. The last period, the "Period of the Latter Law" (*Mappō*), was seen as a time of open degeneration.

These ideas took deep root in Japan. Hence, time was also among the five specific criteria by which Nichiren evaluated proposed doctrines, viz., (1) the teaching of the *sūtra*, (2) the spiritual endowments of the learner ("the propensity"), (3) time, (4) the country in which the doctrine is practiced, and (5) the temporal order of circumstances affecting its practice. In the Tendai sect, Saichō had also regarded the time and the country as important criteria; but Nichiren established them as basic principles, presented in a clear and distinct form.

OTHER MOVEMENTS

Shintō philosophy. In the 15th century, Shintō, which had formerly been nothing but the popular religion of the Japanese, was given a philosophical foundation. The *Shintō* go *busho* (the "Shintō Five Books"; the Shintō "Pentateuch") were probably composed in the 14th century. In the following century Ichijō Kaneyoshi, a syncretic theorist, tried to bring to light the philosophical foundations of the Shintō religion with the help of Buddhist and Confucian thought: and Urabe Kanetomo, an esoteric theoretician, presented theories based on a group of books pretending to transmit direct revelations from the gods (see below *Shintō* revival). The methods of the philosophy of history thus came to be utilized to vindicate Japanese Imperial claims to absolute power and respect by proving the emperor's natural position in the Japanese state organization (kokutai) and his right to this position by reason of righteous succession—a cardinal tenet based upon Shintō tradition.

Confucian philosophy. The second period in the development of Confucianism extended from the diffusion of the Sung philosophy at the end of the 12th century to the rise of different philosophical schools at the beginning of the 17th century. In this period, Confucianist learning was the sole ingredient in intellectual education. The Sung philosophy was imported into and spread through Japan by Zen Buddhist monks, both Chinese and Japanese. Furthermore, its dominant figure, Chu Hsi, was not only venerated as a philosopher but also as a scholar in the philosophy of history, a field which was soon infused with the spirit of the Sung ideology.

Edo period

NEO-CONFUCIANISM

In the Edo (or Tokugawa) period (1603–1867), Confucian thought—or, more accurately, Neo-Confucianism—gradually replaced Buddhism and became the predominant system of thought that influenced the better educated, upper class Japanese, who viewed it, however, as a moral philosophy rather than a religion. As a religion, Confucianism had always been weak; indeed, images of Confucius are found today only in a few of the now-defunct schools founded by feudal lords during the 19th century. But as a philosophy of life, the sayings of Confucius and of his disciple Mencius are still taught to children, and most educated people are able to recite some of them.

The first two periods of Confucian history have already been reviewed (see above Early Confucianism and Confucian *philosophy*). The third period, extending from the beginning of the 17th century to the mid-19th century, saw the rise of schools of Confucian studies (kangaku). In the Edo period, the rigid Tokugawa shogunate, which governed the country, as well as the local feudal lords, found it to their advantage to encourage Confucian studies since they emphasized law and order. Confucian influence thus became stronger with the lapse of time, finally helping to create the mind set of modern Japan.

Three schools. There were three major schools of Confucianism in the Edo period.

The Shushigakuha ("Chu Hsi's school") taught a Rationalistic philosophy. Fujiwara Seika, living at the turn of the 17th century, is regarded as the forerunner of this school. His disciple, Hayashi Razan, was the founder of the governmental school at Edo (present Tokyo), of which his descendants became the hereditary heads. The Shushi school was officially adopted and supported as orthodoxy by the government of the Tokugawa shogunate. Though this school was the most influential of the three, its metaphysical tenets are not discussed here since they reflect those of the parent Chinese school with slight modifications (see CHINESE PHILOSOPHY; CHU HSI).

The Ōyōmeigakuha ("Wang Yang-ming's school") was named after Wang Yang-ming, an early 16th-century Chinese philosopher of the Ming dynasty, who taught a doctrine based on a peculiar Idealism. Among the representative scholars of this school were Nakae Tōju (died 1648); his disciple Kumazawa Banzan, a great reforming administrator; and, in the 19th century, Ōshio Heihachirō, who led a rebellion against the government. Moreover, several leaders of the Meiji Restoration, which overthrew the shogunate and put the Emperor in power, were greatly influenced by its thought (see WANG YANG-MING).

The Kogaku-ha (the School of Ancient Learning) taught a somewhat fundamentalist doctrine, emphasizing the original teachings of Confucius. The name Kogaku (literally "the study of antiquity") refers to a particular branch of Confucian philosophy in Japan, which, opposing the systems of Chu Hsi and Wang Yang-ming, goes back to the original texts of Confucianism, disregarding some or all of the works of later authors. The forerunner of this school was a 17th-century scholar, Yamago Sokō, who wrote extensively on the essential qualities of a knight. Sokō is considered to be the father of Bushido, or the "Way of the Samurai," which stressed frugality, loyalty, filial piety, and acts of great bravery and sacrifice. Ogyū Sorai and his senior contemporary Itō Jinsai, a 17th-century commoner, were the chief representatives of this school.

Original energy as the supreme source

Jinsai and his son. Jinsai ascribed the production of all beings to an original kinetic force called the "original energy" (genki). He wrote: "In the world there is nothing but one Original Energy. . . . For it cannot be said that the world has either beginning or end or [origin by] creation." He thus denied the creation theory of the Chinese Confucianists. "That the light and pure substances collected and ascended to form Heaven and the heavier and turbid ones congealed to form Earth," he wrote, are uncanonical vagaries that originated since the Confucianists of the Han dynasty and are all false. Jinsai aimed constantly at the concrete; and, as much as possible, he reduced to a single concept such notions as matter (ch'i) and principle (li), which the Sung philosophers had always distinguished. In his view, the Supreme Ultimate is nothing but the primordial energy considered as the supreme source of all things. The vital motion is the highest quality of the universe. He said:

> The universe is one great living being. According to the Way of the Universe, there is life but no death; there is collection but no dispersal. Death is the final stage of life; and dispersal is the completion of collection because the Way of the Universe is bent on the giving of life. Although the bodies of the ancestors pass away, their spirit is transmitted to their children who bequeath it to their own progeny. The production of living beings continues infinitely. One can rightly say that they do not die. . . . How could this be unless in the Way of the Universe there were only life and no death?

Jinsai stubbornly refused to admit the existence of evil. "In the Universe . . . there is movement and no quiescence; there is good and no evil. Goodness belongs to the category of life; evil to that of death." Human nature is fundamentally good, he held, although, through lack of culture, this original good disposition eventually may be thwarted by evil passions and bad surroundings. The will, however, is essentially bent toward the good.

Jinsai deplored the situation of the intellectuals of his day, who had divided into different contending camps. He wrote:

> Present-day Confucianists desire more than anything else to attack the Buddhists and to do away with them. The Buddhists by all means desire to attract the Confucianists and conciliate with them. They are like tigers come to grips, like oxen with horns locked.

The human virtues and love

According to Jinsai, the many human virtues can be reduced to four cardinal virtues, which can be further reduced to two important concepts: humaneness (*jen*) and righteousness (i). These, in turn, are but one, *jen*; and *jen*, which is characterized by love (ai both in Chinese and Japanese), is the sum of all virtues. "It is virtue itself, it is the Virtue . . . . But to extol it in one word, it is called Love [*ai*]." Jinsai's thought thus displayed a greater human appeal than earlier scholars. He also admitted that the providence of Heaven underlies the moral order: "The ruling of Heaven is in all things," he wrote, "punishing evil and rewarding good." He believed that the world of reality consists of change and action and that action in itself is good:

> Between heaven and earth there is only one reason: motion without stillness, good without evil. Stillness is the end of motion, while evil is the change of good; and good is a kind of life, while evil is a kind of death. It is not that these two opposites are generated together, but they are all one with life.

It6 Tōgai, Jinsai's son, succeeded his father, and hundreds of students studied under his direction. During his lifetime, in fact, which extended well into the 18th century, the School of Ancient Learning reached the zenith of its influence.

Sorai and his disciple. Ogyū Sorai, a contemporary of It6 Tōgai, was a versatile scholar who stressed the analysis of one's own experience and of that derived from history. Sorai held that the Way (or Truth) was invented by the ancient sage-kings of China and is not natural. By dint of their virtuous lives, however, the ancient kings captured the benevolence of Heaven and skillfully adapted its way to human nature. In his time, quiet sitting and the fostering of reverential love, as practiced by most earlier Chinese Confucianists (around the Sung period), was the customary method of mental training. Sorai ridiculed those Confucianists and said: "As I look at them, even gambling seems to be superior to quiet sitting and fostering reverential love in one's heart." Sorai's approach was rather sociological. He found the pattern of all good in the great examples of antiquity. History, not philosophy; practice, not speculation; sound administration, not discussion—these, he felt, will save the world. The utilitarian view of history in Sorai's system and the program of social reform that he set forth, however, threatened to engulf his philosophy.

Dazai Shundai's naturalism

Sorai's disciple, Dazai Shundai, who flourished in the first half of the 18th century, maintained a ritualistic view of morality. Calling man's natural feelings the real and legitimate feelings, which he defined as "likes and dislikes, sufferings and rejoicing, and anxiety and pleasure," he maintained that

> There is not a single human being devoid of these feelings. . . . Love of one's parents, wife and children is also the same among the noble and the common. Since these feelings are originated in the innate truthfulness of man which is free from any strain of sin or falsity, they are called the real feelings.

Shundai's standpoint was that of a pure naturalism. Believing that human desires should not only be tolerated but also fully expressed, he declared: "I would rather be a master of acrobatic feats. than to be a moralist."

SHINTO REVIVAL

A revival of Shintō, which began in the 17th century, aimed at eradicating Buddhist and Confucian influence. It was supported by philosophical studies of the ancient records that had been made in the 17th and 18th centuries by such philologists as Keichii and Kamo Mabuchi. The greatest leader of "Pure Shintō" was Moto-ori Norinaga, an 18th-century theoretician, who held that the original life of the Japanese represented the primitive purity of mankind. By rejecting all Buddhist and Confucian teachings in metaphysics and ethics as signs of degeneracy, he tried to restore the pure and ideal inheritance of humanity from the divine ages. "The pure mind is the natural mind," he said.

> The Confucian scholars who are most highly esteemed as men of wisdom, and the Buddhist priests who are revered as saints, admire the beauty of stars and flowers, but they pretend never to have taken notice of a beautiful woman. What a deception of mind.

In the first half of the 19th century, a physician, Hirata Atsutane, founded nationalistic Shintō. Being acquainted with Western science and with some elements of Christianity, especially the then-forbidden Catholicism, Atsutane maintained the idea of the Creator who is called the Heavenly-Central Lord.

NEW THINKERS IN BUDDHISM

In the Edo period many thinkers enunciated doctrines critical of former teachings. There arose, for example, the theory that, if a man pursues his own secular vocation with his whole heart and soul, he is practicing nothing other than the ascetic practice of Buddhism.

Suzuki Shōsan's sanctification of vocation

Takuan, an early-17th-century Zen priest, taught that "the law of the Buddha, well observed, is identical with the Law of mundane existence. The Law of mundane existence, well observed, is identical with the Law of Buddha." This idea was especially stressed by Suzuki Shōsan, a contemporary of Takuan, who claimed to be the first man to apply Buddhism to matters of mundane existence. In his book Banmin *tokuyō* ("The Significance of Everyman's Activities"), he discussed problems of vocational ethics. He saw an absolute significance in the pursuit of any vocation, whether that of a warrior, a farmer, a craftsman, a merchant, a doctor, an actor, a hunter, or a priest. To pursue one's own vocation, he reasoned, is to obey the Absolute One, because the essence of Buddhism consists in reliance upon the guidance of the original self or upon "the true Buddha of one's own," and every vocation is a function of this "one Buddha." Thus, he preached to farmers, "Farming is nothing but the doings of a Buddha." To merchants he taught, "Renounce desires and pursue profits single-heartedly. But you should never enjoy the fruit of your labours. You should, instead, work for the good of all others." Since the afflictions of this world, it is said, were predestined in former lives, a man should torture himself by working hard at his vocation in order to redeem the sins committed in his former life.

Even in Zen Buddhism, a critical attitude was expressed toward the founder. Dōgen, for instance, in the 13th century, denounced the theory of "perceiving one's own nature intuitively" as set forth in the *sūtra* of the Sixth Patriarch; *i.e.*, the T'an Ching (or *The* Platform Scripture) of Hui-neng. But Tenkei, his spiritual descendant (early 18th century), rejected Dōgen's opinion as "absurd sheer nonsense." According to the traditional attitude, one's own enlightenment should be conveyed face to face, from master to disciple, and it should be approved by a single master." But Tenkei presented a different interpretation. Since in this phrase "master" means "one's self," and "a single master" likewise means "one's self," the whole phrase, according to Tenkei, means "the attainment of one's own, or true self, by oneself." Thus, a seeker need not practice under the guidance of any teacher, for even by looking at peach blossoms, he can make his self clear to him. In the 18th century the venerable Jiun, the pioneer of Sanskrit scholarship in Japan, likewise stressed the idea of self-enlightenment. but with a rationalistic slant. In his view the essence'of Buddhism lies in the observance of natural law—*i.e.*, of the ten virtues. "Whether a Buddha appears or whether a Buddha does not appear, this world exists, and human beings exist, and these ten virtues will always be manifest along them [*i.e.*, so long as they exist]."

OTHER MOVEMENTS

Shingaku school. The Shingaku (Mental Learning) school was founded by Ishida Baigan, the philosopher of city people, in the first half of the 18th century. Its scholars taught that Mind alone is the basis of religion and attempted to unite the moral teachings of all of the religions and ethical systems then prevailing in Japan, setting forth a humanitarian ethics and encouraging the cultivation of the mind. Teshima Toan, a popularizer of Shingaku, and Nakazawa Dōni, an advocator of naturalism, of the middle and late 18th centuries, respectively, were leaders of this school.

Independent philosophers. There also appeared in the Edo period thinkers who did not belong to any specific tradition. These included Torninaga Nakarnoto, who advocated a relativist philosophy, or "metaphilosophy," in which the philosophical systems of different peoples are studied in the context of their respective historical and cultural settings. He was also the first to trace the development of Buddhist philosophy through history with a critical spirit.

Relativism, Materialism, and the way of nature and labour

Although Japan prior to the Meiji Restoration had no Materialists in the strict sense of the word, there were a few thinkers whose critical spirit and consciousness of ego led them to a skeptical attitude about religious authorities, finally resulting in the repudiation of spiritualism. Several scholars of the 17th and 18th centuries had prepared the way for Materialism, among them the aforementioned Sorai, Shundai, and Nakamoto, and a Rationalist, Miura Baien. And several scholars, such as Andō Shōeki in the mid-18th century, approached a Materialism; but their thoughts were limited to a small circle, and they left little influence.

Andō Shbeki advocated the way of nature and labour. He sympathized with the peasants for their miserable condition and protested against exploitation and oppression by feudal lords. Shbeki appreciated the exaltation of agriculture. He declared, "Direct cultivation and happy eating, direct weaving and happy clothing—there is no Way but this. Talking of thousands of ways is false." Such an outspoken avowal, however, was forbidden by the oppressive government of the Tokugawa shogunate.

Miura's dialectics and Ninomiya's return of virtues

Miura Baien, a younger contemporary of Shbeki, expressed a Rationalist theory of dialectics of his own. The way to understand nature (or the universe), he held, is through dialectics (jdri), the secret of which is to see a synthesis (*gōitsu*) in an antithesis (han), to abandon all one-sided preoccupations, and to make signs real (*chō*-hyd). The Yin and Yang, the all-pervading cosmic principles of ancient Chinese dualism, are antithetical to each other and thus constitute a battle. Though antithetic, however, they can be brought to synthesis. He said:

> The way to see things thoroughly is logic (jdri) and the essence of logic is nothing else but the dialectic of antithesis and synthesis, setting aside all attachments of mind and following the correct signs.

The three elements, then, that comprise the full structure of Miura's *jdri* are (1) the dialectic of things, (2) the prerequisite eliminations of bias and preoccupation, and (3) the empirical test. He also set forth a theory of the origin of price similar to that of his contemporary in England Adam Smith and established a form of Gresham's law on the circulation of bad money. Miura's logic was strongly pervaded by a Positivist and Rationalist spirit that had no respect for authority or tradition, a spirit no doubt influenced by the so-called Dutch learning (*rangaku*), a type of science imported from the West.

The so-called Peasant Sage, Ninomiya Sontoku (earlier 19th century), emphasized man's dependence on both nature and fellowship. From Ninomiya's teaching was derived the *Hōtoku* (literally "to return virtues") doctrines, addressed especially to the agricultural population, which emphasized energy and work. The purport of his teaching was, as he wrote, that "We owe our life and its preservation and enjoyment first to the benefits granted by Heaven and Creation, then to those we receive from our sovereign, our country, our parents, and other sources innumerable." Men have laws and social obligations that compel them to return, in some degree, the benefits received from parents, sovereign, and country; but there are no laws obliging them to render their gratitude for the greater benefits bestowed on them by Heaven. Men are thus prone to forget that requital for the heavenly benefits is their first duty, and so they neglect it. Man must bear in mind the will of Heaven and try to cultivate the Heaven-sent virtues that are in him, and he must work earnestly to promote the progress and development of all creation. Thus Ninomiya viewed human life as a process of cooperation and mutual helpfulness. Because this combination of moral ideas and economic measures lay at the heart of his view, his philosophy bore fruit in practical effects among the peasants.

Period of modern philosophy

The number of original thinkers revealing traces of modern thought in Japan before the introduction of Western civilization was very small; *i.e.*, the few traces of modern thought that appeared in Japan before the Meiji Restoration did so only sporadically, did not develop, and vanished in their incipient stages.

Since the Meiji Restoration in 1868, however, Western philosophy has been liberally and abundantly introduced into Japan. At first, British and American philosophies predominated; but in the course of the 20th century the influence of German philosophy became increasingly strong. Leading Japanese philosophers were especially influenced by recent German Idealism, Phenomenology, and Existentialism. Today, philosophy in the Western sense of the term is taught as a compulsory subject in most Japanese colleges and universities; and Eastern thought, under the name of Indian or Chinese philosophy, is relegated to specialized courses. In order to distinguish Western philosophy from Buddhist and Chinese thought, a new term *tetsugaku* was coined and has come rapidly into use.

MEIJI PERIOD

In 1862 Nishi Amane prepared a lecture on Western philosophy at a governmental institute. Another late-19th-century scholar, Tsuda Masamichi, also developed an early interest in philosophy.

Kato's evolutionism and Fukuzawa's illuminism

The type of thought that first came to the fore was evolutionary philosophy. Hiroyuki Kato, living prior to World War I, was the first man in Japan to construct a system of philosophy of his own. His system was founded on a theory of evolution that denies a divine Creator. Evolution comprises cosmic, inorganic, and organic development, he held; and organic evolution, in turn, comprises that of body, mind, and society. Since all organisms have a constant tendency to preserve and develop their lives, there is a severe struggle for existence throughout the world. Struggle, in fact, is necessarily indispensable, for without it human society might well have remained in its primordial condition. The sole source of right, said Kato, is power. Because morality and law have always been developed through the struggle for existence, true morality is not practicable in a great incomplete society, but only in one that is organized as a state. One generation later, Asajirō Oka, a zoologist, also advocated a distinctive evolutionism of his own.

Fukuzawa Yukichi, one of Kato's contemporaries, put forward a theory of history based on François Guizot and Henry Thomas Buckle, French and British historians, respectively. Fukuzawa stressed that civilization arose neither from upper nor lower strata but was formed by intellectuals of the middle class. This viewpoint was typical of the so-called illuminist group, which advocated a democratic government. "Heaven creates no man above man," he stated, and "no man under man."

In the field of social thought Nakae Tokusuke (pen name Chō-min), a late-19th-century scholar, was an influential political thinker of a clearly Materialistic bent. Nakae, who became famous for his translation of Jean-Jacques Rousseau's Du contrat social, forcefully expressed his atheism and antispiritualism.

Under the leadership of Kōtoku Shūsui, anarchism was also introduced into Japan. On his return from the United States, where he had encountered many American anarchists, Kbtoku advocated commonerism (*heimin-shu*gi), which esteemed the views and rights of the common people against those of the nobles and bourgeoisie. The fact that his group was executed by the state is indicative of the suppression of free thought by its power.

A forerunner of individualism was Takayama Rinjirō (pseudonym Chogyū), who flourished near the end of the 19th century and who, following the theory of T.H. Green, an Hegelian moralist, advocated an ethical theory in which the realization of self is the ideal of man. Rinjirō saw in the person of Nichiren, a 13th-century Buddhist prophet, the ideal embodiment of Friedrich Nietzsche's superman (*Übermensch*).

Enryo's panlogism and Tetsujirō's phenomenalism

At the turn of the 20th century, Inoue Enryo attempted to reform Buddhism and combine it with Western ideas. To Enryo the two Mahāyāna philosophical systems, Kegon and Tendai, represent the ultimate truth. But their emphasis on pure reason (ritai) was close to Hegel's panlogism or panrationalism. As the fundamental structure of the universe, this pure reason, for Enryo as for Hegel, transcends Materialism and Idealism. Enryo argued against Christianity because of its supposedly anti-scientific tenets.

Inoue Tetsujirō, a contemporary of Yiijirb, after studying in Germany, returned to Japan in 1890 and became a full professor of philosophy at the Tokyo Imperial University. Thereafter, the influence of German philosophy became very strong and has remained so until now. Inoue called his philosophical standpoint "the theory of phenomena as reality" (*genshō* soku *jitsuzai ron*). He believed that Immanuel Kant and Eduard von Hartmann, author of a transcendental Realism, were wrong in their attempts to separate the phenomenal world from reality. Inoue's theory is typical of the traditional way of philosophical thinking of the Japanese (see above *Phenomenalism*).

The contributions made by foreign teachers were also significant. The most outstanding of these teachers was the stimulating philosopher Raphael von Koeber, a Russian of German extraction, who visited Japan in 1893 at the suggestion of his teacher Hartmann. Subsequently, the influence of the philosophies of Arthur Schopenhauer, a voluntarist, and of Paul Deussen, a philosopher of biblical and Indian religions, endured for about four decades.

The stronghold of Pragmatism in Japan was Waseda University, a leading private institution. Tanaka Ōdō (original personal name Kiichi), having studied in the United States for nine years, was greatly influenced by the ideas of the Pragmatist John Dewey. Advocating instrumentalism, the view that ideas are instruments for action, Tanaka attempted to make Pragmatism a philosophy of real usefulness in Japan. Calling his philosophy "idealistic experimentalism," he identified the ideal with the real and stressed the impossibility of explaining the overall structure of reality.

The sociopolitical views of Miyake Setsurei (original personal name Yūjirō), an influential critic who flourished prior to World War II, were affected by his esteem for Thomas Carlyle and by Herbert Spencer's organic theory of the state.

TAISHO AND SHOWA PERIODS

Influences of Neo-Kantianism, Phenomenology, Existentialism, and Hegelianism

Imported German *and* Christian philosophy. In the Taishō period (1912–26), German philosophy exerted a great influence. In particular, Neo-Kantianism was influential among the intellectuals. The Marburg school of Neo-Kantianism was known through the works of Paul Natorp, one of its two leading figures, whereas the southwestern (Baden) branch was known mostly through the many Japanese pupils of Heinrich Rickert, an authority in the logic and epistemology of science. Among Rickert's best pupils was Sōda Kiichirō, who wrote, in both Japanese and German, on the philosophy of economics, being the first to attempt to do so in a systematic way. The *Präludien* of Wilhelm Windelband, a leading historian of philosophy, was also enthusiastically read in Japan at the time.

From 1920 onward, Edmund Husserl's Phenomenology, Heidegger's Existentialism, and Hegel's dialectic became increasingly prevalent. These trends represented counter-reactions to the fact that, until then, Anglo-American thought—from John Stuart Mill to John Dewey—had been excessively preoccupied with sociopolitical problems, almost to the exclusion of critical epistemology and metaphysics. The main Kantian philosopher of Japan was Kuwaki Genyoku, who was active prior to World War II. His way of setting forth his thought was not systematic nor logical, however, in the Western sense of the word, and in this respect he inherited a traditional way of thinking of the Japanese.

Hatano Seiichi, a Christian philosopher of the mid-20th century, analyzed the nature of time in his book Time *and* Eternity (1943). According to Hatano, time manifests itself to man in three stages: (1) as natural time, encompassing nature; (2) as cultural time; and (3) as eternity, which is characterized as a community of love (agape') involving selflessness and generosity toward others.

The work of Nishida, Tanabe, and Watsuji

Indigenous creative philosophies. Nishida Kitarō (died 1945) has usually been regarded as the greatest Japanese philosopher since the introduction of Western philosophy. In his maiden work, Zen-no-kenkyfi (1911; Eng. trans., *A Study of* Good, 1960), which was much influenced by the Pragmatist William James, the fundamental concept was "pure experience," by which he meant the direct awareness of things as they are. Thought and thinking are based on this experience. Manifestations of it are multifarious, extending from the perception of colours, or the primitive experience of the infant, to the direct experience of artistic and religious minds. In Nishida's view, all oppositions—like the subject and object of truth and falsehood—must be transcended in a unifying awareness, which alone opens the way to the knowledge of reality.

Tanabe Hajime, another mid-20th century scholar, developed a so-called logic of species, according to which mankind corresponds to a genus, a nation to a species, and an individual (a man) to a particular instance in the logical sense. Holding that, of these three, the species is the most real, Tanube urged that individuals should devote themselves to their nation. In this way a nationalism in which most individual freedoms were subjugated to the cause of the state was justified; and as such this concept was criticized later. Thus, the influence of Hegel, who in like manner especially esteemed the nation of Prussia, was conspicuous.

Tanabe's contemporary Watsuji Tetsurb, a prolific writer of great importance, occupies a unique position in the history of contemporary Japanese philosophy by virtue of his systematic treatment of human existence, which culminated in the formation of a philosophy that was termed an "anthropological ethics." His three-volume Ethics (1937–49) is the most systematic philosophical work that Japan has ever produced. Another work of his, *Ningengaku-toshite-no-rinrigaku* ("Ethics as Anthropology") is an introduction to his *magnum* opus. Key terms in Watsuji's system are ethics (rinri), man (ningen), and existence (*sonzai*). He interpreted the term ningen as signifying "to be in the world." Ethics consists essentially in the relationships involved in contacts between man and man, man and his family, and man and society. Man is not only an individual or a mere social entity but an existent essentially related to the social world in which he lives. Watsuji further explained that society is formed and perpetuated by individuals who deny their own individuality in order to blend into the group consciousness of humanity. In conceiving of society as a great void composed of self-denying individuals, Watsuji was reflecting the concept of the Void in Buddhist Mahāyāna philosophy. Starting from this fundamental principle, he theorized about various aspects of human existence. In his work *Fūdo* (1935; Eng. trans., A *Climate*, 1961), he discussed the influence of climate on human behaviour and culture, claiming that a monsoon climate leads to mercy and benevolence, a desert climate to combativeness, and a temperate climate to rationality and science.

Current types of philosophy. Japanese philosophy was thrown into a chaotic state by World War II, and it is therefore quite difficult to classify its postwar varieties. Three types of philosophy, however, seem to be predominant: German Idealism, also including Phenomenology and Existentialism; Marxism, following the philosophies of Marx, Lenin, and Mao Tse-tung; and Logical Positivism, including the philosophy of science and the philosophy of language. This last tradition, accepted by the younger scholars, is strongly influenced by American thought. With regard to Existentialism, the German philosophers Karl Jaspers and Martin Heidegger and the French philosopher Jean-Paul Sartre are influential. The philosophy of Søren Kierkegaard, which was almost completely neglected before World War II, was introduced into Japan by Watsuji Tetsurō, and under the incentive of his authority it has exerted a great influence.

In nearly all of the universities of Japan, works of Immanuel Kant and G.W.F. Hegel are assigned to philosophy students as almost compulsory reading. Many Japanese intellectuals, however, seek to formulate Marxian justifications. Moreover, there are some thinkers who are still endeavouring to raise traditional Japanese thought to a level of contemporary significance in modern civilization. The problem of philosophy East and West is of great concern to some intellectuals, who are worried about the confrontation between their own tradition and the incoming Western civilization together with leftist agitation.

At present it is difficult to identify those whom history will view as the leading philosophers of the decades since the mid-20th century.

BIBLIOGRAPHY. HAJIME NAKAMURA, *A History of the Development of Japanese Thought from 592 to 1868,* 2 vol. (1967), a history of Japanese philosophical thought up to the Meiji Restoration, explaining it by way of the problem approach; and *Ways of Thinking of Eastern Peoples: India, China, Tibet, Japan,* ed. by PHILIP P. WIENER (1964), a discussion of some major features of Japanese ways of thinking in comparison with those of other Asian peoples and of Japanese philosophy in relation to traditional ways of thinking; MASAHARU ANESAKI, *History of Japanese Religion with Special Reference to the Social and Moral Life of the Nation* (1930, reissued 1963), the most authoritative history of Japanese religion—also discusses the religious background of Japanese philosophy; WILLIAM THEODORE DE BARY (ed.), *Sources of the Japanese Tradirion* (1958), a collection of translations of philosophical and religious passages of classic Japanese works; JEAN HERBERT, *Shinto: At the Fountain-head of Japan* (1967), the most detailed work in English on Shinto; GENCHI KATO, *A Study of Shinto, the Religion of the Japanese Nation* (1926, reprinted 1971), a study of the features and problems of early Shintdism; JUNJIRO TAKAKUSU, *The Essentials of Buddhist Philosophy,* ed. by WING-TSIT CHAN and CHARLES A. MOORE (1947, reprinted 1973), a rather technical work discussing philosophical thoughts of various schools of Japanese Buddhism; JOSEPH JOHN SPAE, *It6 Jinsai: A Philosopher, Educator, and Sinologist of the Tokugawa Period* (1948, reprinted 1967), contains a good description of Japanese Confucianism; KYOSON TSUCHIDA, *Contemporary Thought of Japan and China* (1927), the first attempt in English to describe the philosophical thought of Japanese philosophers since the introduction of Western civilization; GINO K. PIOVESANA, *Recent Japanese Philosophical Thought, 1862–1962: A Survey,* rev. ed. (1968), a work in English that describes recent Japanese philosophy. See also issues of *Philosophical Studies of Japan,* a journal published by the Japanese Commission for UNESCO.

(H.Na.)

Japanese Religion

Japanese religion is assumed here to consist of the beliefs and practices of the folk religious traditions indigenous to Japan and of those great traditions, such as Taoism, Confucianism, Buddhism, and Christianity, that have migrated to Japan from the Asian mainland and other parts of the world. It also includes Japanese religious groups that are syntheses of these diverse traditions and certain new religions that have arisen in modern times.

NATURE AND TYPES OF RELIGION IN JAPAN

Folk traditions and alien traditions

Japan is a living museum of religions. There exist in it numerous survivals of archaic animism and shamanism (see below *Shintd and shamanism),* indigenous Shinto cults, Confucian ethics and religious Taoism (both of Chinese origin), Christianity, and many syncretistic new religious movements. The essence of Japanese religion may be found in the interaction of two kinds of belief systems: a folk or "little tradition" that is based on blood relations or close community ties and a "great tradition" originally introduced from outside Japan that is adopted by individual or group choice. Japanese belief patterns are complex, multilayered, and syncretistic. These patterns are based both on an indigenous religion centring in the worship of ancestors and on the various kinds of religion brought from outside Japan by missionaries or believers who belong to one of the great traditions or to more advanced folk traditions. Little tradition here refers to the indigenous or folk religion, including Shinto cults (see also SHINTO), which was shaped by ancient geographic and cultural circumstances; great tradition refers to several highly developed religious and philosophical importations such as Confucianism, religious Taoism, Buddhism, and Christianity (see also CONFUCIANISM; TAOISM; and BUDDHISM).

These two kinds of belief systems eventually became intertwined, and Japanese religion thus developed as an integration of many separate elements. At least until the end of the Tokugawa era (1867–68), each family had both a miniature Shintd shrine and a Buddhist altar in the inner part of the house, and ancestral mortuary tablets were placed on or beside the Buddhist altar. Each village, town, or city also had its own local Shintō shrine or shrines, Buddhist temples, several shaman-like magicians, diviners, fortune-tellers or healers, and, since modern times, Christian churches and churches of the other new religions. Recently many new denominations have arisen within Shintō, Buddhism, and Christianity. Even some new religions have come into being. The following table gives statistics for the distribution of religion in Japan in the late 1970s.

Distribution of Religious Affiliations in Japan

religious affiliations	number of members
Shintō	98,352,612
Shrine Shintō	63,564,075
Sect Shinto	6,005,792
New Shintō sects	2,007,295
Buddhism	87,860,100
Tendai groups	5,473,205
Shingon groups	11,854,621
Jōdo groups	21,084,723
Zen groups	9,579,450
Nichiren groups	30,579,306
Nara groups	4,697,963
Others	206,792
Christianity	973,340
Roman Catholic	356,255
Protestant	467,098
Syncretistic sects	14,213,523
Total	201,399,575

These statistics illustrate the curious fact that the total number of individual religious affiliations is much greater than the estimated total Japanese population in the late 1970s of 116,133,000. The discrepancy is explained by the Japanese system of "dual adherent" or "plural belonging," in which a person may participate simultaneously in several religious traditions. As one author has pointed out, in Japan, Confucianism and Shintō have borrowed Buddhist metaphysics and psychology; Buddhism and Shinto have borrowed many aspects of Confucian theory and ethics; and Confucianism and Buddhism, with their many and varied manifestations, have adapted themselves rather thoroughly to Japanese indigenous religion instead of maintaining their own particularities. Although there are exceptions, the characteristic feature of Japanese religion is the coexistence of different religions or heterogeneous religious elements in one family or even in a single person.

HISTORY OF RELIGION IN JAPAN

The major forms of religion during the protohistorical and prehistorical periods (prior to the 6th century AD) were naturism (worship of natural forces), animism (belief that natural objects and living beings possess souls or spirits), and shamanism (beliefs and practices surrounding spirit mediums). After the 1st century AD, however, Chinese civilization began to spread into Japan, mainly into Kyūshū and western Honshū, through the Korean peninsula.

Early historical developments (before the 16th century AD). Up until the 7th century AD the cultural contacts between Japan, Korea, and North and South China resulted in an influx into Japan of a remarkable number of immigrants from these areas. These immigrants brought with them their own beliefs, rituals, and customs. They probably introduced Chinese Confucian ethics and political ideology, as well as the philosophy and metaphysics of the Chinese Taoist thinkers Lao-tzu (traditionally 6th century BC) and Chuang-tzu (4th century BC), among the Japanese ruling classes. They also introduced among the common people some beliefs, rituals, and techniques of Chinese religious Taoism (a complex of rituals and techniques for achieving immortality), Chinese Yin-Yang magic (based on two contrary but complementary principles that pervade all things), and Korean shamanism.

Early Chinese influences

The introduction of Buddhism (a complex variety of religious schools that follow traditions derived from an Indian founder Gautama Buddha, 6th century BC) from China and Korea in the mid-6th century AD played a decisive role in the history of Japanese religion. Prince Shōtoku (AD 573–621), said to be the real founder of Japanese Buddhism, wrote commentaries on three Buddhist *sūtras* (scriptures) and understood the fundamentals of Buddhism. It is doubtful, however, that the doctrines of early Japanese Buddhism were fully understood by the majority of the Japanese people. During the Asuka and Nara periods (AD 593–784) the Buddhist sects or schools became theoretical, philosophical, and devoted to religious discipline. A variety of schools developed, including Sanron (Indian Miidhyamika, a sect that believes the "Absolute" to be inexpressible), Hossō (Chinese Fa-hsiang, a sect that believes in "consciousness only"), Kegon (Chinese Hua-yen, a sect based on the *Avataṃsaka-sūtra*, which holds that all things in the universe are interrelated and that every living being is a potential Buddha), and Ritsu (Chinese Lü or Indian Vinaya, a sect concerned with the ceremony of ordination and rules of monastic discipline) based in large part on Chinese antecedents (see also BUDDHISM and BUDDHISM, HISTORY OF). In general, Buddhism was accepted as an exotic magico-religious cult for the protection of the state, the benefit of public security, and the welfare of the nation. Buddhist rites gradually took the place of the Shintō cults.

At the beginning of the Heian period (AD 794–1185) two new Buddhist sects were introduced and came to dominate the entire period: Tendai (Chinese T'ien-t'ai, an important and highly metaphysical sect based mainly on a sacred text known as the *Lofus Sūtra)*, introduced by Saichō (Dengyō Daishi, AD 767–822) and Shingon (Chinese Chen-yen or Mantrayiina, a highly magical and mystical syncretistic sect) by Kiikai (Kōbō Daishi, AD 774–835). (Such Buddhist leaders are known both by priestly names and posthumous honorary titles.) The Heian period of Japanese religious history was characterized by the following tendencies: (1) the esoteric and magical Mantrayāna Buddhism flourished; (2) belief in the doctrines of the *Lotus Sūtra* (*Saddharmapuṇḍarīka-sūtra*) and of Amida (Sanskrit, Ambitābha and Amitāyus) Buddha spread among all classes; (3) the activities of many *hijiri* (holy men) established popular or folk Buddhism outside the orthodox ecclesiastical system, thus paving the way for a later Buddhism; (4) belief in *goryō*, which originally consisted of the malevolent or angry spirits of the dead, appeared; (5) a consciousness of the Buddhist age of the Latter Law (*Mappō*) emerged; and (6) there was commingling of various religious elements such as primitive shamanism, Shintb, Yin-Yang magic (originally Chinese), and Mantrayiina Buddhism.

Heian and Kamakura Buddhism

At the end of the Heian and the beginning of the Kamakura period (1192–1333), three new Buddhist movements appeared: the Pure Land sects (Jōdo, Shin, Ji, and others all believing in the "Pure Land" or heavenly paradise, of Amida, the compassionate Buddha who saves); Zen (Chinese Ch'an or Indian Dhyāna), both of Rinzai (Chinese Lin-chi) and Sōtō (Chinese Ts'ao-T'ung) varieties; and the Nichiren sect. The Jbdo sect was founded by Genkū (Hōnen Shōnin, 1133–1212), who was the pioneer of Kamakura Buddhism and as a reformer of Japanese Buddhism is sometimes compared with Martin Luther. The Shin sect was founded by Shinran (1173–1262), a pious disciple of Genkū and a great promoter of his teacher's ideas. The Ji sect was founded by Chishin (Ippen Shōnin, 1239–89). The Nichiren sect was founded by Nichiren (Risshō Daishi, 1222–82), who was a revivalist of the Tendai sect. He was a messianic and prophetic politico-religious leader who fought against the Kamakura shogunate. During the transition from the Heian to the Kamakura regimes, the Pure Land sects and the Nichiren sect grew up as ways of bringing salvation to the common people, who were plagued by social disorder and national crisis. Zen Buddhism was introduced from China by Eisai (1141–1215) and some other naturalized Chinese Zen masters of the Rinzai sect and by Dbgen (Shōyō Daishi, 1200–53) of the Sōtō sect. Together with the Pure Land sects and the Nichiren sect, Zen Buddhism played an important role in establishing the spiritual foundation of the warrior class, the bearers of the Kamakura feudal system.

In the Muromachi period (1338–1573) the Pure Land sects and the Nichiren sect became widespread among the common people, while Zen Buddhists became the promoters of fine art and literature. In 1549 St. Francis Xavier arrived at Kyūshū to open the first Christian mission in Japan. He and other Jesuits made numerous ardent converts among the peasants as well as among the feudal lords (daimyo) in a very short time.

During the Tokugawa era and afterward. In a sudden change in Japanese religious policy, Christianity was prohibited on penalty of death from 1616 to 1873, the greater part of the period of the Tokugawa shogunate (1603–1867). At the same time, Buddhist temples and priests were also completely controlled by the Tokugawa regime and priests were required to serve as government registrars as well as performers of funeral ceremonies. The only innovation during the Tokugawa era was the introduction of the Ōbaku (Chinese Huang-Po) Zen Buddhist sect from China by lngen (1592–1673) in 1655.

An important measure of the Tokugawa shogunate was the adoption of Neo-Confucianism, which became the officially authorized basis of national learning and morality. Learning and art as well as indigenous or nationalistic thought gradually began to flourish during this 280-year period of peace and national isolation. Orthodox Neo-Confucianism followed the lines of the Chinese thinkers Ch'eng Hao (1032–85), his brother Ch'eng I (1033–1107), and Chu Hsi (1130–1200), but an idealistic theory of knowledge, conceived by Wang Yang-ming (1472–1529) was also introduced. Several Confucianistic Shinto theories and movements (amalgamations of aspects of Chinese Confucianism and indigenous Japanese traditions) also arose, among which the Mito school played an active role in restoring national unity under Imperial rule. A more radical school promoting the same tendency was the Kokugaku (National Learning) school, which insisted on purging away all alien elements and restoring an "original" Shintō based on the myths in the *Koji-ki* (a classic relating myths, legends, and history of the Imperial court). The Mito school and the Kokugaku school were to become driving forces in the Meiji Restoration in 1868.

Rise of modern religious movements

At the same time, there had appeared several popular ethicoreligious movements, such as the Shingaku (Mental Learning) movement of Ishida Baigan (1685–1744) and the Hbtoku (Recompense or Indebtedness) movement of Ninomiya Sontoku (1787–1856). The Shingaku movement had a great influence within the merchant class, and the Hbtoku movement influenced the peasants.

In the transition from the Tokugawa regime to the Meiji era (1867–68) there appeared several new religious movements that later became independent religious bodies in Sect Shintō—a group of 13 sects, each based on the religious experiences of their individual founders. They may be classified into three categories: (1) denominations of the pure Shintō movement; (2) religious associations of popular mountain asceticism known as Shugen-db; (3) popular new religions founded by charismatic or shamanic leaders from among the common people, including Kurozumi-kyō founded by Kurozumi Munetada (1780–1850), Tenri-kyō (Heavenly Truth) founded by Nakayama Miki (1789–1887), and Konkō-kyō (Golden Light) founded by Kawate Bunjirō (1814–83). These religious movements gradually gained many ardent converts among the masses and became the pioneers of the new religious movements of present-day Japan. From the Manchurian Incident (1931) to World War II several new religious movements appeared but were usually suppressed by the militaristic government. After World War II, in the throes of the political, economic, and cultural ruin of Japan, more than 600 religious movements arose. Some of them were revivals of formerly suppressed sects, while others were new denominations developed from Buddhism, Shintō, Christianity, and the new religions; and some were newly arisen sects. The most significant phenomena of present-day new religions are (1) lay Buddhist movements especially of the Nichiren line, such as Reiyū-kai, Risshō-Kōsei-kai and Sōka-gakkai; (2) the importance of charismatic, shamanic, or prophetic founders or foundresses; (3) syncretistic doctrines; and (4) messianic tendencies.

CHARACTERISTICS AND FEATURES OF JAPANESE RELIGION

In Japan there are several kinds of rites and magico-religious customs that are thought to have originated in Japanese antiquity. The actual content of Japanese religion is not, however, limited to these ancient, indigenous elements. Japan has never been isolated from the Asian continent, from which it borrowed various magical and religious ideas and techniques. These borrowed ethical, magical, and religious elements were then blended, reinterpreted, and resystematized into a unified Japanese religion consonant with Japan's own traditions, which arose out of special sociocultural and historico-geographic circumstances.

Types of primitive Shintō

Shintō and shamanism. Primitive Shintō in ancient Japan should be classified into two categories, each distinguished by its particular notions and practices concerning kami. The term kami (or *-gami*, in combining forms), an ancient Japanese word roughly equivalent to deity or spirit in the sense of the sacred, refers to both mythological and human figures as well as natural objects. It is used in both the singular and plural. The first category of primitive Shintō may be defined as the *uji-gami* type (tutelary or guardian shrine system). It was based on the family or on the Japanese clan system, and within it each family had its own shrine as a central symbol of its solidarity, which was dedicated to an ancestral spirit or socially related kami (deity or deities) who had been worshipped and enshrined successively by their descendants. This type of religion is centred in the family group to the exclusion of other families, so that the main function is to integrate all members of the family into the patriarchal hierarchy of the family system. The heavy emphasis on ancestor worship and filial piety (*kō*) in almost all Japanese religious groups has been connected with the ancient family religious system.

The second category of primitive Shintō may be called the *hito-gami* type (man–god system), based on the close relationship of an individual kami with a charismatic figure, such as a shaman or a medicine man. The more highly integrated state systems, such as the village states or federated village states that appeared in the 3rd century AD, for example, were supposedly ruled by such charismatic or shamanic leaders. This kind of belief is characterized by the strong individuality of both the personal kami and its transmitter, who lived on in the memory of the believers. Those beliefs eventually exerted great influence on the common man. Sincere reverence and pious obedience, regardless of the believers' social origins, were the only prerequisites for gaining the favour of this type of kami.

Due to the rigidity of the ancient Japanese social structure and value system—characterized by the primacy of political values and family-centred religion—there was an increased emphasis upon on (blessings or favours handed down, not only by invisible beings but also by social and political superiors) and *ho-on* (obligation of the recipient to return something for their blessings). The Japanese social structure and value system emphasized human relationships that were strictly controlled and regulated by a patriarch according to the status of each member in the family. The patriarch, however, in turn, was responsible to the higher authorities of the Japanese nation as well as to his ancestors, and even the emperor himself has been and is held responsible by his ancestors for his behaviour. The Japanese way of thinking as it emerged from such a value system left no room for the development of the concept of an "almighty God," such as is found in Judaism, Christianity, and Islām. On the contrary, Japanese kami, as well as men, are not only regarded as independent personalities but also as lowly figures dependent on their superiors in either the divine or the social hierarchy and in need of salvation and help. In this context, the superiors, including human beings and ancestors, were believed to be semi-*kami* (demigods) or even low-ranking kami or Buddhas. Contrary to Chinese Confucian ethics, in Japan filial piety and loyalty (chzi) were respected equally and inseparably. Further, the Japanese people generally accepted the Chinese Confucian philosopher Mencius' (c. 371–289 BC) doctrine that human nature is fundamentally good. Moreover, belief in the spirits of the dead and the idea of an intimate connection between men and kami (involving the deification of human beings) have always been quite widespread and important in Japan.

Transformation of Buddhism

Buddhism **in** Japan. The same Japanese tendency to transform borrowed traditions may be seen in the case of Buddhism. The development of Japanese Buddhism has made it unique among the many forms of Buddhism in the world. Traditional Buddhist teachings concerning the *saṅgha* (monastic order) and *vinaya* (the disciplinary system) were transmitted by missionary efforts from India through China and Korea. But when these teachings were first introduced into Japan they conflicted with the existing social patterns, and inevitable modifications took place in the course of the popularization of the Buddhist religion. Under Prince Shōtoku, Buddhism developed into a religion of the aristocracy with strong lay leadership. At the same time, Shōtoku's emphasis on a text known as the Lotus Siitra promoted a social consciousness that encompassed all classes. The Lotus *Sūtra*'s promise of salvation for all mankind was in sharp contrast to the pre-Buddhistic and shamanistic folk religion. Moreover, the new Buddhist social concern was expressed in the building of temples that served as centres of philanthropic and cultural activities.

In the Asuka and Nara periods (AD 593–784), however, Japanese Buddhism was transformed into the spiritual foundation of the great family or clan system. Ancestor worship became one of Buddhism's most significant functions. Buddhist magic, commingled with Shintō and Chinese Yin-Yang magic, also flourished.

Heian Buddhism, developed in a period of intense commingling of various religious elements, shows a remarkable contrast to the Buddhist sects of Nara. Politically, the transfer of the capital from Nara to Kyōto (784–794) brought about a new mood; one purpose for the transfer was to separate church and state. Religiously, two sects of Buddhism, Tendai and Shingon, were predominant throughout the period. The Lotus *Sūtra*, the fundamental sacred book of Tendai Buddhism, gradually revolutionalized the Japanese idea of human nature; at the same time it came to serve as the foremost text for the atonement for sin and for gaining blessing and for prophecy in an age said to be the final period of the Buddha's law. Many popular practitioners of the doctrines of the Lotus Siitra appeared in the Heian period. Together with the Lotus Siitra, beliefs and practices focussing on Amida (the compassionate Buddha) arose within the Tendai sect in terms of meditation on the mystery of existence and man's ultimate destiny; belief in Amida leads to enlightenment. On the other hand, prayer (reciting praise to Amida) and contemplation on Amida (*Nembutsu*) were regarded as a means of clearing the mind and enabling it to concentrate on the presence of Amida. This also helped spirits to be saved after death by invoking Amida's merciful hand. These beliefs and practices gradually influenced the Japanese people and fostered the establishment of independent Pure Land sects in the latter part of the 13th century.

Shingon Buddhism

The central focus of the Shingon sect is *Dainichi-kyō* or Great Suit Szitra (a Japanese version of the Sanskrit *Mahāvairocanābhisaṃbodhi-sūtra*), which teaches that the phenomenal world is a manifestation of the ultimate reality known as the Great Sun Buddha, who represents two aspects of reality: the *vajra-dhātu* (diamond element) and the garbha-dhiitu (womb element). Together with the Chinese theory of yin and *yang*, the doctrine inspired a metaphysics and cosmology of balanced polarities in the Japanese way of thinking, and in turn contributed to the theoretical systematization of Shintb and the formation of the theory of the manifestation of the prime noumenon or primary reality (honji-suijaku—a theory based on Buddhist notions of incarnation according to which Japanese kami were considered to be manifestations of gods of the Buddhist pantheon, and those Buddhist gods themselves were seen as kami). Shingon theology also supported the theory of "becoming a Buddha in one's own body" (*sokushin-jōbutsu*), following the cosmological and metaphysical theories of the Great Sun Siitra and the *Avataṃsaka-sūtra.* Another focus of the Shingon is the harmonious uniformity of theory and ritual as expressed through activities of the body, mouth, and mind. Theory is expressed concretely in symbolic rituals, while rituals are performed in conformity with metaphysical doctrines.

These theories and beliefs, together with those of the Lotus Siitra and the Amida worship of Tendai Buddhism, were widely accepted not only by all the branches of both sects but also by other Buddhists and laymen. Even in Nichiren and Tendai Buddhism as well as in Shin and Zen Buddhism, the Shingon theory of "becoming a Buddha in one's own body" was accepted, and at the same time Shingon Buddhism accepted belief in Amida for the purpose of securing an individual afterlife.

The initial commingling of primitive shamanism with Yin-Yang magic and Mantrayāna Buddhism appeared in the latter part of the Nara period and developed rapidly in the Heian period. Three major streams emerged in the religious world as a result of this historical cohesion: (1) *Nembutsu*—the Buddhist Pure Land school mixed with Japanese animistic and shamanistic elements; (2) *Shugen-dd*—Buddhist Mantrayāna mixed with Shintb animism and shamanism as well as with Yin-Yang magic and Chinese astrology; (3) Onmyd-dd—religious Taoism mixed with Yin-Yang magic and astrology as well as with Shintb animism and shamanism.

Heian Buddhism in general gradually declined into occult mysticism and ritualism in response to the magico-religious needs of the people and became secularized under the influence of the Imperial family and the nobility. In response to this situation there appeared an informal movement of individuals known as hijiri (holy men) who, as lay Buddhists (ubasoku or upiisaka), resisted secularized and formalized Buddhist institutions. There were two currents in the hijiri movement, one emphasizing the Amidist tradition (following Amida) and the other the Lotus Szitra. These far-sighted hijiri intuitively perceived the religious meaning of the crucial social changes taking place in the transition from the Fujiwara regime to the Kamakura feudal regime and were torn with anxiety. Because they believed that this world was gradually approaching the fateful age of the Latter Law (Mappd), they endeavoured to find a means of self-enlightenment to cope with a hopeless and depraved age, and at the same time they sought salvation for the common people in their everyday life.

Jōdo

Genkii, the founder of the Jōdo sect and a pioneer of the new Kamakura Buddhism, merits the title of Buddhist reformer. His reformation constituted the real Japanization of Buddhism. Genkii insisted that even stupid and obstinate men could be saved; he severely criticized and rejected the old ecclesiastical institutions and the formality and aristocratic bent of traditional Buddhism. He taught that the way to the Pure Land is necessarily through simple faith in Amida's grace, which is called the "easy way" (igyd-dd) of salvation. (This term is in direct contrast with the "difficult way" [nangyd-dd] of moral perfection.) The "easy way" is also called salvation by "other power"—*i.e.,* Amida's power. Genkii's notions exerted strong influence on the reform movements in Buddhism, as well as in Shintō theories during the Kamakura period. For example, Shinran advanced Genkii'~theory several steps. According to Shinran, human nature is originally so sinful and hopeless, and the situation of the later times is so absolutely confused that no one can attain spiritual enlightenment and peace by "self power." Instead, it is necessary to throw oneself on the mercy of Amida's mysterious power. Pushing Genkfi's idea to its utmost limit, Shinran denied not only the formal temple-priest system of his times but also married and reared a family, just as any ordinary citizen or farmer. He also criticized ritualism, magic and divination, and the worship of the old pantheon. His direct descendant, Rennyo (1415–99), promoted the rationalistic tendencies of Shinran and made important advances in the ethicoreligious regulation of everyday life. He stressed the obligations of this-worldly asceticism and of bestowing blessings on the people, as well as giving thanks for Amida's blessings.

Nichiren, who predicted the Mongolian invasions in 1274 and 1281, admonished the Kamakura shogunate for its religious policy at these times of national crisis. Nichiren requested the unification of Buddhism as well as national thought on the basis of the teachings of the Lotus Siitra. He severely attacked the Jōdo sect of Genkū as well as other sects such as Shingon, Zen, and Ritsu. His teaching, focussed on a new interpretation of the Lotus Siitra, was intended to give the people peace and salvation, much like the *Nembutsu* practice of Genkii and Shinran. His apocalyptic, prophetic attitude and his violent attacks on the other Buddhist sects met stern resistance and persecution. Nichiren, however, left a vigorous sect behind him. The most extreme Nichiren group, known as the Fuju-fuse ("No give or take") branch, was severely restricted along with Christianity by the Tokugawa shogunate. The other subsects of Nichiren Buddhism, however, spread among the merchants and craftsmen in Kybto and other areas in the Middle Ages and continued to gain many ardent believers among the masses even in the Tokugawa period.

Developments since the Tokugawa regime

Curiously enough, since the end of the Tokugawa regime, several lay movements of the Nichiren line—for example, Busshb-gonen-kai, Butsuryfi-kyd, Rieyfi-kai, Risshb-Kb-sei-kai, and Sōka-gakkai—developed among the masses and are the most active and prosperous new religious movements in present-day Japan. All of this is in sharp contrast to the Shin sect, which has gone completely back to the level of a monastic religious body and has not given rise to many new religious movements.

Zen Buddhism did not spring up like the systems of Genkii, Shinran, and Nichiren. It was introduced to Japan c. AD 1200, when men's minds were weary of the old religious forms and new sects, national and Japanese, were coming into being; and in this troubled period, the contemplative and mystical doctrines of Zen, along with those of Jddo, Shin, and Nichiren, found truth in immediate experience and spiritual peace and enlightenment through meditation. Zen was accepted as the rule of life for the military class. Bushidd, the Way of Warriors, was formed from a combination of Zen Buddhism and Confucian ethics of loyalty.

The newer Buddhist sects, including imported Zen sects, were either completely Japanese or were transformed into a Japanized Buddhism. Generally speaking, these sects have the following common characteristics: (1) religious theory and form are rationalized and simplified; (2) there is a concentration of religious piety—of faith and devotion; and (3) there is an emphasis on religious practices related to the daily life of the believers and a rejection of excessive speculation and complicated magico-religious ritualism.

Christianity and Neo-Confucianism. In spite of the prohibition of Christianity from 1616 throughout the Tokugawa regime and many persecutions and martyrdoms, the few surviving Japanese Christians went into hiding in several isolated islands and peninsulas in northern Kyūshū, where they maintained their ancestors' faith until 1873. Even today there are about 40,000 of the so-called hidden Christians in these areas. Their faith and practices have been, however, so completely transformed by mixture with Japanese folk religious elements that some scholars say that their faith is not really Christianity but a unique Kirishitanism (*Kirishitan* is the Japanese name for the Christians dating from the 16th century). In this era, when Christianity was prohibited, Buddhism sank into a torpid condition, from which it was awakened only by active measures taken against it in 1868.

The Japanese Neo-Confucian school was founded by Fujiwara Seika (1561–1619) and his pupil Hayashi Razan (1583–1657). The head of the Hayashi family became the hereditary governmental university president. The Wang Yang-ming school of Neo-Confucianism had been taught first by Nakae Tōju (1608–48) and Kumazawa Banzan (1619–91), but this school was in disfavour with the authorities, and its teachers got into trouble for advocating its principles. Confucianism in Japan, however, generally insisted on ancestor worship and filial piety as well as on loyalty, based on the idea of Heaven and righteousness, and it constituted an important stimulus for new Shinto theories. Confucian Shintō was taught by Yamazaki Ansai (1618–82), Deguchi Nobuyoshi (1615–90), and Kikkawa or Yoshikawa Koretaru (1616–94), respectively, and had a strong influence on some feudal lords and the warrior class, as well as on Shintō priests of the period. Orthodox Confucianism and Confucian Shinto, in addition to influencing the warriors' discipline and their service to the existing regime, worked in subtle and unexpected ways toward undermining feudalism and restoring national unity under Imperial rule. The Mito school, founded by Tokugawa Mitsukuni, lord of the Mito fief, with its idea of loyalty to the mikado (emperor), was a representative of this tendency.

In the 18th century Shintō entered a new path and prepared for another revival. All the earlier Shintō theories had depended heavily upon either Buddhism or Confucianism and religious Taoism in interpreting Shintō ideas. In contrast, there then appeared a new Shintōistic school that purged away all alien elements and sought to restore the original religion of Japan—*i.e.*, pure Shintō. The great pioneers of this pure Shintō or Koku-gaku movement were Kamo Mabuchi (1697–1769), Motoori Norinaga (1730–1801), and Hirata Atsutane (1776–1843). They contended that the Japanese people and their Shinto, when purged of all foreign accretions and influences, represented the pure, and therefore the best, inheritance of humanity from the divine ages of Japanese mythology. Hirata pursued the contentions of Motoori to their logical conclusion, and through him the Shintō revival took definite form in vigorous propaganda. Hirata's disciples tried to carry such chauvinistic ideas into practice in the antiforeign and anti-Tokugawa movement of the 19th century. Pure Shintō and the Mito school became the dynamic forces behind the Meiji Restoration of 1868.

Koku-gaku movement

In the Tokugawa era, there appeared two remarkable popular ethicoreligious movements: the Shingaku movement and Hōtoku movement. Shingaku, the result of eclectic borrowing, aimed at uniting the moral teachings of all the religious and ethical systems prevailing in Japan and tempering them with the general principles of humanitarian ethics and mental culture. Shingaku met with great success and developed religioethical principles for merchants. On the other hand, the Hōtoku movement penetrated into the peasant class. Ninomiya, founder of the Hōtoku movement, emphasized the indebtedness of mankind to nature and fellowship among men. He taught that man's nature consists in devotion to the order of nature, which manifests itself in the moral order of human life, especially in the obligations between a lord and his subjects, parents and children, and benefactors and recipients in general.

Modern forms. After the Meiji Restoration of 1868 there were four major religious currents. First, there was the reintroduction of Catholicism and the introduction of Protestantism at the beginning of the Meiji era. Catholicism was accepted by upper class people, while Protestantism was embraced by the lower class samurai (warriors) who had been living in obscurity since the fall of the Tokugawa feudal system. Joseph Niijima was a most remarkable pioneer of Japanese Christian mission works as well as education works. His personality seems to have been formed by the Christian faith mixed with the samurai ethics based on Confucianism. Second, there was strict enforcement by the Meiji government of the separation of Shinto and Buddhism. This policy encouraged by the pure Shintoists resulted in Shintō becoming a kind of established church. Shintō served in the status of both a national cult and a community cult as a focus of the unity of the nation and community. Shrine Shinto was regarded legally not as a religion but as a national moral order. The third major current, Buddhism, after suffering from vigorous suppression, has gradually recovered its role as a family religion and has also promoted positivistic studies of Buddhism under the influence of European Indology and Buddhology. Buddhist priests have become practitioners of the funeral and memorial services based on ancestor worship. Generally Buddhism, together with Shrine Shintō, has maintained the status quo under the strong control of the government. The fourth major current is constituted by Sect Shintō and several new religious movements after the Meiji Restoration (already described above).

An eminent British scholar, Sir Charles Eliot, writing on the characteristics of Japanese religion, has said that the Japanese people are religious, "if religion is taken in the wide sense of devotion to something beyond, or readiness to sacrifice to it earthly welfare and life." In Japan there has survived a very old tradition that identifies religion with government, law, family duties, and festivities. They are to be identified in the strictest sense: religion does not regulate them, they are religion. This was true of the history of Japanese religion, but recently some serious changes have occurred in the religious world view due to rapid social change. The nature of religion and its relationship to rapid social change are important problems for the future of Japan.

JAPANESE RELIGIOUS SYMBOLISM

In Japan religious symbolism is formed from a complex of indigenous and borrowed elements, expressed both in popular forms and in abstract doctrine and iconography. Shintō possesses its own traditional symbolism. In Shintō rites a sacred tree, sacred pole, or mirror is used as a symbol, on which the kami descends or which the kami is believed to "possess." At some shrines chosen persons have played the roles of substitutes of kami or have been possessed by kami. Shinto originally did not offer an explicit moral code; its prayers and sacrifices aimed at obtaining temporal prosperity and indicated no desire for moral or spiritual blessings. Also Shintō does not advocate meditation practices but requires purification rites of all participants in a festival (see also SHINTO).

Maṇḍala and *haniwa*

Visual arts. Buddhism brought from the Asian continent several kinds of meditation techniques and esoteric rituals, as well as symbolic representations such as mandalas (Buddhist ritual meditative drawings) and images of Buddhas, *bodhisattva*s (Buddhas-to-be), and other guardian deities. In pre-Buddhistic ages, various kinds of lay images called haniwa were dedicated on grave mounds to the spirit of the noble dead, supposedly as substitutes for human and animal immolations. The introduction of Buddhist arts as symbols for worship, however, had a revolutionary effect on Japanese arts. From the Asuka period to the Nara period, almost all Buddhist arts were imported or produced by immigrant Korean artists or their descendants. But in the Heian period Buddhist arts became more symbolic and esoteric under the flourishing Mantrayāna Buddhism. At the same time, the cessation of official relations between Japan and China (in AD 894) contributed to the Japanese transformation of imported art styles. The most remarkable result was the new depiction of themes from Pure Land Buddhism in a uniquely Japanese style of picture (yamato-e). In the Kamakura period there appeared a realistic school of Buddhist arts, in contrast to the earlier esoteric and symbolic expressions of the Heian period arts. Under the Ashikaga shogunate in the 14th and 15th centuries several Zen monks engaged not only in religious activities but also in diplomacy and in creative arts such as literature, painting, architecture, gardening, Nō play, tea ceremony, and flower arrangement. Their art and literature were strongly influenced by those of the Chinese Ming dynasty and flourished around Kybto. In the Tokugawa period Buddhist art became formalized, and rather inferior and secular arts and literature were on the rise and appealed to the common people—the so-called *chōnin-geijutsu* (arts and literature for merchants).

Performing arts. Almost all Japanese performing arts originated from ancient magico-religious or shamanic séances; later they became differentiated into Shintd festivals and Buddhist rituals and finally became the theatrical arts such as Nb play, Kabuki, or Bunraku puppet show. Several varieties of popular dramas and dancing arose in the course of the popularization of Mantrayāna Buddhism and the Pure Land sects as a result of their mixture with both the pre-Buddhist shamanic séance and the Ying-Yang arts. Shintb music and dancing called *kagura* also originated from the ancient shaman's séance; they played an important role in ShintB rituals. Foreign music and drama, including Korean, Chinese, and Indo-Chinese music instruments, masks, and dramatic techniques, were introduced in the Nara period; these imported arts were frequently performed at the Imperial court as well as at Buddhist temples during both annual festivals and occasional services. On the common level, puppet dramas and dancing, Nernbutsu-dancing in procession or in circle, masked dances, and narrations with gestures were performed by such figures as lower class Pure Land priests or mountain ascetics (*yamabushi* or *shugen-ja*) and by the blind magico-religious public entertainers.

Manshichi Sakamoto

"Descent of Amida across the Mountains," painting by an unknown artist of the Kamakura period (1192–1333). In the Konkai-Kōmyō-ji, Kamakura, Japan.

Festivals. In Japan there are various annual festivals and special festivals at Shintō shrines, Buddhist temples, and in each family. The festivals represent both elite and popular ways of expressing and celebrating religious symbolism. In ShintB shrines the harvest festival named *niiname* is most important, and in the Imperial court the festival is performed by the emperor himself, who must observe rigid abstinences and purifications. The coronation ceremony named *daijō-sai* is simply the niiname festival performed on a large scale. In Buddhist temples the memorial and celebration services for Buddha Śākyamuni's birth and death and his attainment of Buddhahood and also the memorial days of the founder or founders of each sect are solemnly observed. The most important and universal annual festivals are the New Year festival *Shōgatsu gyōji*, and the Bon festival; the former is performed in a Shintd manner, the latter in a Buddhist manner. But both are said to have originated from agricultural rites mixed with ancestor worship and the belief in spirits of the dead.

Bon and New Year festivals

POLITICAL AND SOCIAL ROLES OF RELIGION IN MODERN JAPAN

There have been three major developments in Japanese religion since the Meiji Restoration (1868). First, as a result of the enforcement of a policy of the separation of Shintb and Buddhism, State Shintō became more and more bureaucratic and chauvinistic; it lost its religious elements and became formalized. Because the Meiji government recognized State Shintb as the foundation of national polity and morality (but not as a religion, in order to explain away article 28 on freedom of religion in the Meiji constitution, which stated that so long as the religious beliefs of Japanese subjects were neither "prejudicial to peace and order" nor "antagonistic to their duties as subjects," freedom of religious belief was to be guaranteed), State ShintB easily became an ideological rationale for nationalism and militarism until the end of World War II. Buddhism, having lost the leadership of the Japanese nation's religious life, became lethargic.

Second, the consequent religious vacuum after the Meiji Restoration was filled by various new religious movements founded by leaders from the lower classes. These movements spread rapidly among the common people who had suffered many social, political, and economic changes. Also at the same time a newly introduced Protestantism and reintroduced Catholicism advanced among the intellectuals and played important roles in the modernization of Japan. From the Manchurian Incident in 1931 until the end of World War II in 1945, all Japanese religions—old and new sects of Shintb, Buddhist sects, syncretistic ones, and Christianity—were severely controlled by the militaristic government and were forced to cooperate in the national war effort in support of the so-called Greater East Asia Co-prosperity Sphere. Some new religious sects such as Omoto and Hitonomichi were even persecuted and completely exterminated by the government under the pretext of offenses against the law of maintenance of the public peace and *lèse-majesté.*

Third, the period after World War II was marked by the appearance and flourishing of several new religious movements, including revivals of the suppressed sects. For example, the Omoto sect revived as Omoto-Aizen-en (Great Foundation of Love of Goodness Garden) and the Hitonomichi as the PL Kybdan (Perfect Liberty Order). Among the new religious movements in contemporary Japan, however, the most conspicuous are the Nichiren sects such as Reiyū-kai (Spiritual Friend's Association) founded by Kotani Kimi (1900–71); Risshō-Kōsei-kai (Integrative Becoming, of Nichirenism), a denomination derived from the Reiyii-kai, founded by Naganuma Myōkō (1889–1957) and Niwano Nikkyb (1906–); Sōka-gakkai (Value Creation Society) founded by Makiguchi Tsunesaburb (1871–1944) and Toda Jōsei (1900–58) and now led by Ikeda Daisaku (born 1928); and other minor Nichiren sects.

Developments since World War II

From the standpoint of the political and social roles of religion, Sbka-gakkai is the most important, not only because it is the fastest growing but because it is as well the most vigorous, dogmatic, belligerent, and self-confident group in Japan. Although Sdka-gakkai denies that it is an independent religious body, insisting that it is merely a lay association within a Nichiren sect (Nichiren-shō-shū), the fact is that its rapid growth, its penetration into labour unions, its recent successes in political activities (through its related Kbmei-t6 Party), and its aggressive methods of gaining converts have drawn Japanese national attention to it as one of the most important sociopolitical religious groups in present-day Japan. Statistics for 1968 show that the total membership of Nichiren-shb-shii is 15,729,636. This represents a remarkable growth, for in 1955 there were only 348,160 believers.

THE PLACE OF JAPANESE RELIGION IN WORLD RELIGIONS

Japanese religion is a typical example of syncretistic religion, which as an entity has become thoroughly Japanized or particularized, in spite of the fact that it is composed of such diverse elements as archaic animistic-shamanistic elements; the Shintō cult; Confucian ethics and metaphysics; Taoistic divination; rituals and astrology; the theories, practices, and rituals of Mahāyāna Buddhism in the premodern periods; and Christian denominations. Moreover, Japanese religion is characterized on the whole as a this-worldly religion. Even Jōdo Buddhist sects, which originally insisted on the denial of this impure world, have been transformed into religious traditions insisting on a this-worldly asceticism quite like the Protestant ethic in the rise of capitalism. This tendency should be recognized as one of the underlying causes of the rapid modernization of Japan since the Meiji Restoration. Non-Japanese often wonder at the fact that a single Japanese can belong simultaneously to two or more religious organizations. Such "plural belonging" illustrates the Japanese tendency to treat foreign culture as a means easily adopted for utilitarian purposes. Recently Japanese Buddhist sects such as Shin and Zen and several new religions such as Tenri-kyb and Sōka-gakkai, however, have promoted missions in foreign countries. They pay attention to foreigners who are interested in the relationship between socioeconomic changes or social crisis and religion. They are searching for new principles of salvation whereby man's humanity may be preserved in increasingly mechanized society.

Almost all Japanese religious organizations, old and new, participated in the International Congress for the Cooperation of World Religions held in Kybto in the fall of 1970 with great enthusiasm in order to discover a new approach to world peace and prosperity and establish greater cooperation among all world religions.

BIBLIOGRAPHY. On the general features of Japanese religions, the most detailed and excellent study is MASAHARU ANESAKI, *History of Japanese Religion* (1930, reprinted 1963). See also his *Religious Life of the Japanese People,* rev. by HIDEO KISHIMOTO (1961). J.M. KITAGAWA, *Religion in Japanese History* (1966), is a good guidebook to understand the Japanese religions, which mainly emphasized the characteristics of ethico-religious movements in connection with their historical contexts. R.N. BELLAH, *Tokugawa Religion: The Values of Pre-industrial Japan* (1957), is a unique study in the sociological analysis of the roles played by the ethico-religious movements in pre-industrial Japan in connection with the Meiji Restoration and the following modernization; also HIDEO KISHIMOTO, *Japanese Religion in the Meiji Era,* trans. into English by J.F. HOWES (1956). On the subject of folk religion, ICHIRO HORI, *Folk Religion in Japan: Continuity and Change* (1968), will be helpful. A very brief but relevant sketch of the history of Japanese institutionalized religions is W.K. BUNCE (ed.), *Religions in Japan: Buddhism, Shintd, Christianity* (1962).

On Shintb, W.G. ASTON, *Shintd: The Ways of the Gods* (1905), is a classic work. D.C. HOLTOM, *The National Faith of Japan* (1938), is also excellent. More recent works include: JEAN HERBERT, *Les Dieux nationaux du Japon* (1965); EDMOND ROCHEDIEU, *Le Shintoïsme et les nouvelles religions du Japon* (1968); and F.H. ROSS, *Shintd: The Way of Japan* (1965).

On Japanese Buddhism, there are many books written about Zen Buddhism or Shin sect from the philosophical and theoretical viewpoints. An excellent general history of Japanese Buddhism is SIR CHARLES ELIOT, *Japanese Buddhism* (1935); as is A.K. REISCHAUER, *Studies in Japanese Buddhism* (1917).

On Japanese Confucianism, see R.C. ARMSTRONG, *Light from the East: Studies in Japanese Confucianism* (1914).

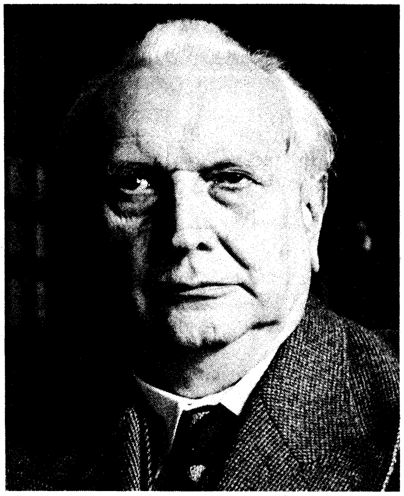
On the contemporary new religious movements, C.B. OFFNER, and H. VAN STRAELEN, *Modern Japanese Religions* (1963); H.N. MCFARLAND, *The Rush Hour of the Gods: A Study of New Religious Movements in Japan* (1967); and H.B. EARHART, *The New Religions of Japart: A Bibliography of Western-Language Materials* (1970), are useful. On the new Buddhist denominations, see W. KOHLER, *Die Lotus-Lehre und die modernen Religionen in Japan* (1963); and KIYOAKI MURATA, *Japan's New Buddhism* (1969).

(I.H.)

Jaspers, Karl

Karl Theodor Jaspers was a 20th-century German philosopher whose intellectual development led him from medicine to psychiatry and psychology and then to philosophy. He played an influential role in establishing the methodological foundations of clinical psychiatry, but his most notable work was in laying the groundwork for the modern philosophy of existence. Together with Martin Heidegger, he was one of the two most important representatives of the German-speaking world in the Existentialist movement. In his later work, as a reaction to the disruptions of Nazi rule in Germany and World War II, he searched for a new unity of thinking that he called world philosophy.

Horst Tappe—EB Inc.



Jaspers, 1968.

Early life and education. Jaspers was born February 23, 1883, in Oldenburg (near the North Sea coast). He was the oldest of the three children of Karl Wilhelm Jaspers and Henriette Tantzen. His ancestors on both sides were peasants, merchants, and pastors who had lived in northern Germany for generations. His father, a lawyer, was a high constable of the district and eventually a director of a bank. He was thus able to provide adequately for his family.

Jaspers was delicate and sickly in his childhood. As a consequence of his numerous childhood diseases, he developed bronchiectasis (a chronic dilation of the bronchial tubes) during his adolescent years, and this condition led to cardiac decompensation (the inability of the heart to maintain adequate circulation). These ailments were a severe handicap throughout his adult life.

Studies in law and medicine

During his early years of schooling, Jaspers never excelled in any subject, yet was known for his independent spirit. His dislike for harsh discipline brought him into constant conflict with the school authorities, and, in fact, he was threatened with expulsion unless he complied with the demands of the authorities. He entered the University of Heidelberg in 1901, enrolling in the faculty of law; in the following year he moved to Munich, where he continued his studies of law, but without much enthusiasm. He spent the next six years studying medicine at the universities of Berlin, Gottingen, and Heidelberg. After he completed his state examination to practice medicine in 1908, he wrote his dissertation *Heimweh und Verbrechen* ("Nostalgia and Crime"). In February 1909 he was registered as a doctor. He had already become acquainted with his future wife, Gertrud Mayer, during his student years, and he married her in 1910.

Research in clinical psychiatry. In 1909 Jaspers became a volunteer research assistant at the University of

Heidelberg psychiatric clinic, a position he held until 1915. The clinic was headed by the renowned neuropathologist Franz Nissl, who had assembled under him an excellent team of assistants. Because of his desire to learn psychiatry in his own way without being regimented into any particular pattern of thought by his teachers, Jaspers elected to work in his own time, at his own pace, and with patients in whom he was particularly interested. This was granted to him only because he agreed to work without a salary.

When Jaspers started his research work, clinical psychiatry was considered to be empirically based but lacking any underlying systematic framework of knowledge. It dealt with different aspects of the human organism as they might affect the behaviour of human beings suffering from mental illness. These aspects ranged from anatomical, physiological, and genetic to neurological, psychological, and sociological influences. A study of these aspects opened the way to an understanding and explanation of human behaviour. Diagnosis was of paramount importance; therapy was largely neglected. Aware of this situation, Jaspers realized the conditions that were required in order to establish psychopathology as a science: a language had to be found that, on the basis of previously conducted research, was capable of describing the symptoms of disease well enough to facilitate positive recognition in other cases; and various methods appropriate to the different spheres of psychiatry had to be worked out.

Jaspers tried to bring the methods of Phenomenology—the direct investigation and description of phenomena as consciously experienced, without theories about their causal explanation—into the field of clinical psychiatry. These efforts soon bore fruit, and his reputation as a researcher in the forefront of new developments in psychiatry was established. In 1911, when he was only 28 years old, he was requested by Ferdinand Springer, a well-known publisher, to write a textbook on psychopathology; he completed the *Allgemeine Psychopathologie* (Eng. trans., *General Psychopathology*, 1965) two years later. The work was distinguished by its critical approach to the various methods available for the study of psychiatry and by its attempt to synthesize these methods into a cohesive whole.

Transition to philosophy. In 1913 Jaspers, by virtue of his status in the field of psychology, entered the philosophical faculty—which included a department of psychology--of the University of Heidelberg. His academic advance in the university was rapid. In 1916 he was appointed assistant professor in psychology; in 1920 assistant professor in philosophy; in 1921 professor in philosophy; and in 1922 he took over the second chair in that field. The transition from medicine to philosophy was due in part to the fact that, while the medical faculty was fully staffed, the philosophical faculty needed an empirical psychologist. But the transition also corresponded to Jaspers' intellectual development.

In 1919 Jaspers published some of his lectures, entitled *Psychologie der Weltanschauungen* ("Psychology of World Views"). He did not intend to present a philosophical work but rather one aimed at demarcating the limits of a psychological understanding of man. Nevertheless, this work touched on the border of philosophy. In it were foreshadowed all of the basic themes that were fully developed later in Jaspers' major philosophical works. By investigating the legitimate boundaries of philosophical knowledge, Jaspers tried to clarify the relationship of philosophy to science. Science appeared to him as knowledge of facts that are obtained by means of scholarly methodological principles and that are apodictically certain and universally valid. Following Max Weber, a sociologist and historian, he asserted that scientific principles also applied to both the social and humanistic sciences. In contrast to science, Jaspers considered philosophy to be a subjective interpretation of Being, which—although prophetically inspired—attempted to postulate norms of value and principles of life as universally valid. As Jaspers' understanding of philosophy deepened, he gradually discarded his belief in the role of a prophetic vision in philosophy. He bent all his energies toward the development of a philosophy that would be independent of science but that would not become a substitute for religious beliefs. Though the resulting system presupposed science, it passed beyond the boundaries of science in an effort to illuminate the totality of man's existence. For Jaspers man's existence meant not mere being-in-the-world but rather man's freedom of being. The idea of being oneself signified for Jaspers the potentiality to realize one's freedom of being in the world. Thus, the task of philosophy was to appeal to the freedom of the individual as the subject who thinks and exists and to focus on man's existence as the centre of all reality.

Development of his basic philosophical themes

The elaboration of these germinal ideas occupied Jasper's thought from 1920 to 1930. During this decade his brother-in-law, Ernst Mayer, himself a philosopher of repute, worked with him. During these years he also enjoyed the friendship of Martin Heidegger. Somewhat later, this friendship broke up because of Heidegger's entry into the National Socialist Party.

In the early years of the 1930s the fruits of his intellectual labour became evident: in 1931 *Die geistige Situation der Zeit* (Eng. trans., *Man in the Modern Age,* 1933) was published; in 1932 the three volumes of *Philosophie (Philosophy,* 1969) appeared—perhaps the most systematic presentation of Existential philosophy in the German language. A book on Max Weber also appeared in 1932.

Conflict with the Nazi authorities. When Hitler came into power in 1933, Jaspers was taken by surprise, as he had not taken National Socialism seriously. He thought that this movement would destroy itself from within, thus leading to a reorganization and liberation by the other political forces active at the time. These expectations, however, did not materialize. Because his wife was Jewish, Jaspers qualified as an enemy of the state. From 1933 he was excluded from the higher councils of the university, but was allowed to teach and publish. In 1935 the first part of his future work on logic, entitled *Vernunft und Existenz* (Eng. trans., *Reason and Existence,* 1955), appeared; in 1936 a book on Nietzsche; in 1937 an essay on Descartes; in 1938 a further work preliminary to his logic, entitled *Existenzphilosophie* (Eng. trans., *Philosophy of Existence,* 1971). Unlike many other famous intellectuals of that time, he was not prepared to make any concessions to the doctrines of National Socialism. Consequently, a series of decrees were promulgated against him, including removal from his professorship and a total ban on any further publication. These measures effectively barred him from carrying on his work in Germany.

Banned from teaching and publishing

Friends tried to assist him to emigrate to another country. Permission was finally granted to him in 1942 to go to Switzerland, but a condition was imposed by the Nazis that required his wife to remain behind in Germany. He refused to accept this condition and decided to stay with his wife, notwithstanding the dangers. It became necessary for his friends to hide his wife. Both of them had decided, in case of an arrest, to commit suicide. In 1945 he was told by a reliable source that his deportation was scheduled to take place on April 14. On March 30, however, Heidelberg was occupied by the Americans.

Disillusioned by the events of these years, Jaspers withdrew more and more into himself. He revised the *General Psychopathology* in an effort to make it represent the high point of a free but responsible search for knowledge of man, as distinct from science, which had betrayed man. He also completed his work on logic, *Von der Wahrheit* ("Of Truth"), the first part of which was intended to throw the light of reason on the irrational teachings of the times. These works appeared in print in 1946 and 1947.

Postwar development of thought. After the capitulation of Germany, Jaspers saw himself confronted with the tasks of rebuilding the university and helping to bring about a moral and political rebirth of the people. He dedicated all of his energies in the postwar years toward the accomplishment of these two tasks. He also represented the interests of the university to the military powers. He gathered his thoughts on how the universities could best be rebuilt in his work *Die Idee der Universität* (1946; Eng. trans., *The Idea of the University*, 1959). He

Collective guilt of the German people

called for a complete de-Nazification of the teaching staff, but this proved to be impossible because the number of professors who had never compromised with the Nazis was too small. It was only gradually that the autonomous university of the pre-Nazi years could once again assert itself in Germany. Jaspers felt that an acknowledgment of national guilt was a necessary condition for the moral and political rebirth of Germany. In one of his best political works, *Die Schuldfrage* (1946; Eng. trans., *The Question of German Guilt,* 1947), he stated that whoever had participated actively in the preparation or execution of war crimes and crimes against humanity was morally guilty. Those, however, who passively tolerated these happenings because they did not want to become victims of Nazism were only politically responsible. In this respect, all survivors of this era bore the same responsibility and shared a collective guilt. He felt that the fact that no one could escape this collective guilt and responsibility might enable the German people to transform their society from its state of collapse into a more highly developed and morally responsible democracy. The fact that these ideas attracted hardly any attention was a further disappointment to Jaspers. In the spring of 1948 he accepted a professorship in philosophy in Basel, Switzerland. In spite of the apparent neglect of Jaspers' ideas of a moral regeneration of the German people, his departure for Basel was regarded as a betrayal by many of the German people. Jaspers himself hoped to find there a peace of mind that might enable him to work through and revise once again his whole approach to the entire field of philosophy.

World philosophy and world union

This revision was guided mainly by the conviction that modern technology in the sphere of communication and warfare had made it imperative for mankind to strive for world unity. This new development in his thinking was defined by him as world philosophy, and its primary task was the creation of a mode of thinking that could contribute to the possibility of a free world order. The transition from existence philosophy to world philosophy was based on his belief that a different kind of logic would make it possible for free communication to exist among all mankind. His thought was expressed in *Der philosophische Glaube* (1948; Eng. trans., *The Perennial Scope of Philosophy,* 1949) and *Der philosophische Glaube angesichts der Offenbarung* (1962; Eng. trans., *Philosophical Faith and Revelation,* 1967). Since all thought in its essence rests on beliefs, he reasoned, the task confronting man is to free philosophical thinking from all attachments to the transient objects of this world. To replace previous objectifications of all metaphysical and religious systems, Jaspers introduced the concept of the cipher. This was a philosophical abstraction that could represent all systems, provided that they entered into communication with one another by means of the cipher. In other words, the concept of the cipher enabled a common ground to be shared by all of the various systems of thought, thus leading to a far greater tolerance than had ever before been possible. A world history of philosophy, entitled *Die grossen Philosophen* (1957; Eng. trans., *The Great Philosophers,* 2 vols., 1962, 1966), had as its aim to investigate to what extent all past thought could become communicable.

Jaspers also undertook to write a universal history of the world, called *Vom Ursprung und Ziel der Geschichte* (1949; Eng. trans., *The Origin and Goal of History,* 1953). At the centre of history is the axial period (from 800 to 200 BC), during which time all the fundamental creations that underlie man's current civilization came into being. Following from the insights that came to him in preparing this work, he was led to realize the possibility of a political unity of the world in a 1958 work called *Die Atombombe und die Zukunft des Menschen* (Eng. trans., *The Future of Mankind,* 1961). The aim of this political world union would not be absolute sovereignty but rather world confederation, in which the various entities could live and communicate in freedom and peace.

Under the influence of these ideas, Jaspers closely observed, during the latter years of his life, both world politics and the politics of Germany. When the efforts toward democracy in Germany appeared to him to turn more and more into a national oligarchy of parties, he wrote a bitter attack on these tendencies in *Wohin treibt die Bundesrepublik?* (1966; Eng. trans., *The Future of Germany,* 1967). This book caused much annoyance among West German politicians of all shades. Jaspers, in turn, reacted to their unfair reception by returning his German passport in 1967 and taking out Swiss citizenship.

In 1968 his physical condition gradually deteriorated, and he died on February 26, 1969. At the time of his death, he had published 30 books. In addition, he had left 30,000 handwritten pages, as well as a large and important correspondence.

BIBLIOGRAPHY. *Karl Jaspers,* autobiographical and photographic documentary presented by HANS SANER (1970).

Autobiographical writings: "Philosophical Memoir," in *Philosophy and the World,* trans. by E.B. ASHTON, pp. 193–314 (1963); "On My Philosophy," trans. by FELIX KAUFMANN in WALTER KAUFMANN (ed.), *Existentialism from Dostoevsky to Sartre,* pp. 131–158 (1956); *Schicksal und Wille: Autobiographische Schriften,* ed. by HANS SANER (1967), with a bibliography of all autobiographical writings.

Writings: Philosophie, 3 vol. (1932; Eng. trans. by E.B. ASHTON, 3 vol., 1969–71); *Existenzphilosophie* (1938; Eng. trans. by RICHARD F. GRABAU, *Philosophy of Existence,* 1971). A complete bibliography with all translations is in preparation: GISELA GEFKEN and KARL KUNERT, *Bibliographie der Werke und Schriften von Karl Jaspers.*

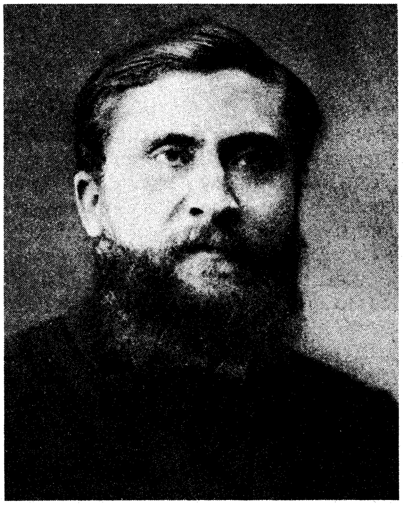
Studies: The Philosophy of Karl Jaspers, ed. by PAUL A. SCHILPP (1957), containing 24 critical contributions on different aspects of Jaspers' work—the most important publication in English; CHARLES F. WALLRAFF, *Karl Jaspers: An Introduction to His Philosophy* (1970), including a bibliography of Jaspers' writings; A. LICHTIGFELD, *Aspects of Jaspers' Philosophy,* 2nd ed. (1971); OSWALD O. SCHRAG, *Existence, Existenz and Transcendence: An Introduction to the Philosophy of Karl Jaspers* (1971); SEBASTIAN SAMAY, *Reason Revisited: The Philosophy of Karl Jaspers* (1971). References to other English-language studies are given in the monograph by Saner.

(Ha.S.)

Jaures, Jean

A principal leader of the French Socialist movement prior to 1914, Auguste-Marie-Joseph-Jean Jaurès was a man of enormous literary, philosophical, and historical erudition, as well as of prodigious eloquence. His capacity for self-sacrifice enabled him to put aside his own political beliefs in order to achieve the unification of Socialist factions into a single Socialist party, the Section Française de l'Internationale Ouvriere (SFIO).

H. Roger-Viollet



Jaures

Jaures was born in Castres (Tarn), September 3, 1859, into a lower middle class family that had been impoverished by business failures. He excelled in secondary school and was given a scholarship to attend the École

Normale Supérieure in Paris. After passing his qualifying examination, Jaurbs taught at the lycée of Albi from 1881 to 1883, and from 1883 to 1885 he was a lecturer at the University of Toulouse.

Early political activity

A convinced republican and a brilliant orator, Jaurts was more attracted to politics than to teaching and in 1885 was elected deputy from the Tarn. As he did not yet belong to any party, he took his seat in the centre of the Chamber. His election prompted the parents of the girl he loved, Louise Bois, to consent to their marriage. Madame Jaurès received as her dowry a handsome rural estate of 91 acres. Because his own political creed disclaimed the ownership of private property, Jaurbs was often reproached for his possession of this estate.

As Madame Jaurbs did not share her husband's beliefs, the Jaurbs household was not a happy one. They had two children, Madeleine, born in 1889, and Louis, born in 1898 and killed in action toward the end of World War I. His daughter's first communion brought violent attacks against Jaurès, who was strongly opposed to the influence of the Roman Catholic Church. He felt obliged to explain that, in spite of his own beliefs, he did not have the right to refuse his children a religious upbringing. Jaurts himself had left the church without abandoning his spiritual values.

His wife's indifference, coupled with his own carelessness, explains his untidy personal appearance, which provided his enemies with much material for ridicule. Short and obese, he was described as having the appearance of "a teacher who does not exercise or a fat merchant who overeats." Yet no one ever accused him of vulgarity.

Commitment to Socialism

Defeated in the 1889 elections, Jaurès returned to teaching at the University of Toulouse, and in 1891 he received his doctorate of philosophy. In 1892 he supported the striking miners of Carmaux, and that constituency elected him deputy to the Chamber in 1893. By this time he had become a Socialist, though without accepting all of Marx's ideas. Rather, of the five schools of French Socialism, he chose the least revolutionary, the Independent Socialists, led by Alexandre Millerand. During the campaign on behalf of Capt. Alfred Dreyfus, who had been convicted of treason and given a life sentence at hard labour on the basis of what later turned out to be forged evidence, Jaurks joined those demanding a revision of the trial. His position was not approved by the Marxist Socialists, who did not believe that a Socialist should defend a man who was both an officer and a member of the middle class. His book *Les Preuves,* asking for Dreyfus' retrial and rehabilitation, caused his defeat in the elections of 1898. Temporarily retired from national politics, Jaurès began to compile his monumental *Histoire socialiste de la Révolution française* (1901–07). This work, written "under the triple inspiration of Marx, Plutarch and Michelet," gave new impetus to studies on France's revolutionary period.

In spite of their dispute over the Dreyfus affair, the different Socialist factions became reconciled and held their first joint congress in 1899. But, after Millerand agreed to join the leftist government dedicated to securing the republic headed by René Waldeck-Rousseau, the Socialists divided into two groups: those who refused to cooperate with the government and advocated class war founded the Socialist Party of France (Parti Socialiste de France), and those who preached reconciliation with the state, headed by Jaurès, formed the French Socialist Party (Parti Socialiste Français). At this time Jaurès wrote many articles supporting Waldeck-Rousseau's policy of reform. After his re-election in 1902, he continued to support the leftist bloc within the Chamber of Deputies.

Formation of SFIO

In 1904 Jaurbs was cofounder of the newspaper *l'Humanité,* in which he continued to espouse the principles of democratic Socialism. That same year, the congress of the Second International, held in Amsterdam, condemned Socialist participation in bourgeois governments, thus rejecting Jaurbs's position. He acquiesced in the decision, and in 1905 the two French Socialist parties joined together to form the Section Française de l'Internationale Ouvribre (SFIO). This party remained in opposition to the government, with the result that the reform policies originally advocated by Waldeck-Rousseau were not carried out. Jaurbs's authority continued to grow within the party, though, and, on the eve of World War I, the majority of the SFIO were won over to his reformist ideas.

He fought the supremacy of the German Social Democratic Party in the Second International and, in order to deprive it of its revolutionary reputation, confronted it at the Congress of Stuttgart in 1907 with his formula "insurrection rather than war." This statement, though, did not completely summarize the whole of his political thought; he strove for the adoption of a system that would ensure "peace through arbitration" and recommended a prudent policy of "limitation of conflicts." He therefore opposed colonial expansion, such as the French invasion of Morocco, because it provided a source of international conflicts.

Hostile to the Franco-Russian alliance and suspicious of the Franco-British alliance because it seemed to be directed solely against Gemany, Jaurts became the champion of Franco-German rapprochement; as Germany was France's traditional enemy, his position earned him the hatred of French nationalists. His passion for reconciliation ultimately led to his tragic death: Jaurès was assassinated on July 31, 1914, at the outbreak of World War I, by a fanatic convinced that Jaurès was playing into Germany's hands. Up to the last moment, however, he was actively exhorting the European governments to avert a world war and to settle peacefully the conflict that followed the archduke Ferdinand's assassination at Sarajevo in June 1914. On the very day of his own assassination, Jaurts was considering an appeal to Pres. Woodrow Wilson of the United States for help in solving this crisis.

Aside from his gifts as a political organizer, Jaurès was well-known for his personal generosity, intelligence, and tenacity of purpose. An excellent scholar and polemicist, he wrote throughout his entire career. Apart from *La Guerre franco-allemande 1870–1871* (1908), *L'Armée nouvelle* (1910), which set forth an effective plan for organizing an armed nation and contained a famous study of the concept of the fatherland, and his two doctoral theses, the remainder of Jaurbs's works are collections of articles and speeches.

BIBLIOGRAPHY. Works devoted to Jaurls are very numerous, but the majority of them are polemical. The first Jaurès biography free of partisan preoccupations was HARVEY GOLDBERG, *The Life of Jean Jaurès* (1962). The *Société d'Études Jaurèsiennes* has published a bulletin, beginning in 1960, which keeps up with the state of research in the field, and has undertaken the publication of the *Oeuvres* of Jaurls (1972–). See also PIERRE RIMBERT, *Jean Jaurès: l'évolution vers le socialisme* (no. 14–17 of the OURS series; 1971), which explores new material; and JEAN RABAUT, *Jaurès* (1971).

(C.H.)

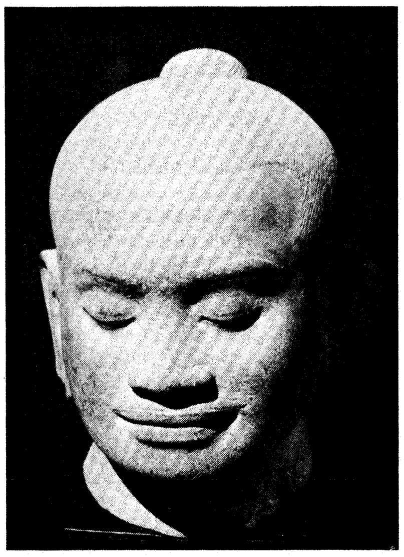
Jayavarman VII

Jayavarman VII, one of the most forceful and productive kings of the ancient Khmer (Cambodian) Empire of Angkor, expanded the empire to its greatest extent and was noted for the magnitude and frenzy of his building activities. He ruled from 1181 to *c.* 1215 (possibly as late as 1219).

Early life

Jayavarman was born in *c.* 1120 (possibly as late as 1125). A member of the royal family, he married a very religious and strong-minded princess named Jayarajadevi, who exerted an important influence on him, both before he gained the throne and during the early years of his reign. Following her death he married her older sister, a very pious and learned woman whom he had previously installed as the head professor in an important Buddhist monastery.

Though practically nothing is known of Jayavarman's childhood and youth, it is clear that during his late 30s and early 40s he settled in the neighbouring kingdom of Champa, in what is now the central region of Vietnam. At the time of the death of his father, King Dharanindravarman II (ruled 1150–60), Jayavarman was engaged in a military campaign in Champa; and after the accession of his brother (or possibly his cousin), Yasovarman II



Jayavarman VII, sandstone sculpture found at Preah Khan, Kampong Svay, Cambodia, 12th century. In the Musée National, Phnom Penh.

BY courtesy of the Musee National de Phnom Penh, Cambodia; photograph, Guy Nafilyan

(ruled 1160–66), he chose to remain there, returning to Cambodia only when he received word that a palace rebellion was in progress. Although Jayavarman arrived at Angkor too late to prevent the murder of Yasovarman and the accession of the rebel Tribhuvanadityavarman (ruled 1166–77), he decided to remain in his homeland and to await an opportunity to assert his own claim to the throne.

Some 12 years later, when Jayavarman was in his late 50s, that opportunity came as a result of a Cham invasion, which brought about the demise of Tribhuvanadityavarman, the sacking of Angkor, and its subjection to foreign rule. In this situation Jayavarman organized a struggle for independence and in less than five years' time succeeded in driving out the invaders and establishing his hegemony over all his Cambodian rivals. Finally in 1181, at the age of 61, he was crowned king of a reconstituted Khmer Empire and began a brilliant reign of more than 30 years, during which he brought the empire to its zenith, both in terms of territorial expansion and of royal architecture and construction.

Jayavarman's building program

During his reign Jayavarman continued his military activities, bringing Champa, southern Laos, and portions of the Malay Peninsula and Burma under his control. But increasingly he devoted his energies and organizational capacities to the kind of religious and religio-political construction projects that had been carried-onby his royal predecessors. He built a large number of awesome new temples, including the Bayon, a distinctively Mahāyāna Buddhist central pyramid temple designed to serve as the primary locus of the royal cult and also as his own personal mausoleum; personal funerary temples of the Mahāyāna type, which were dedicated to his mother and father; and a series of provincial temples, which housed reduced replicas of the Royal Buddha—*i.e.,* Jayavarman represented with the attributes of the Buddha, the original of which had been set up in the Bayon. He rebuilt the city of Angkor, now known as Angkor Thom, and rebuilt and extended the system of highways, which radiated outward from the Bayon and the royal palace and reached far into the provinces. In addition, he constructed more than 100 rest houses along these roads and built more than 100 hospitals, which he dispersed throughout his kingdom and placed under the protection of Baiṣajyaguru Vaiḍūryaprabhā, the Great Buddha of Healing.

Jayavarman seems to have been obsessed with the need for rapid and extensive construction. For example, the less than careful workmanship evident in the temples attributed lo Jayavarman's reign vividly points to the great haste with which they were built. Some scholars have suggested that the almost frantic sense of urgency associated with Jayavarman's works derived from the fact that, having begun his reign at a relatively advanced age, he felt that his time was short and had to be utilized to the fullest. Others have suggested that Jayavarman's concern to carry through such a vast program of largely Buddhist-oriented construction was greatly encouraged by Jayarajadevi and her sister, both of whom dedicated a tremendous amount of energy toward gaining support for Buddhism and specifically for building Buddhist temples. And finally, if scholars are correct in their surmise that Jayavarman suffered from the dread disease of leprosy, his concern to mitigate his sin and suffering through the accumulation of great merit may have given a still further impetus to his piety and zealousness. Whatever his true motivations, Jayavarman succeeded during his lifetime in creating a legacy that few monarchs in history (Khmer or otherwise) have been able to equal; he was over 90 when he died.

Modern reputation

Despite the importance of Jayavarman VII in the history of the Angkor kingdom, no memory of him was preserved in the later Cambodian chronicles. As a result of 20th-century archaeological studies, however, his name has become well-known in modern Cambodia. In recent Cambodian political rhetoric, particularly during the 1950s and 1960s (the period dominated by Prince Sihanouk), Jayavarman VII became a kind of paradigmatic national hero, who was credited not only with establishing the full greatness of the Cambodian nation but also with bringing into being a welfare state that was motivated by Buddhism and dedicated to serving both the spiritual and the physical needs of the Cambodian people. Scholars, however, have sought to maintain a more balanced view of Jayavarman, recognizing the obvious immensity of his accomplishments but also taking account of the fact that the overweening demands which he placed on the material and human resources of his kingdom were a major factor in its subsequent loss of creativity and its eventual demise.

BIBLIOGRAPHY. The classic treatment, which utilizes all of the available data concerning Jayavarman VII, is found in the chapter devoted to him in GEORGE COEDES, *Pour Mieux comprendre Angkor,* 2nd rev. ed. (1947; abridged Eng. trans. *Angkor: An Introduction,* 1963). In addition, the role of Jayavarman and his work within the broader context of Khmer history are discussed in LAWRENCE PALMER BRIGGS, "The Ancient Khmer Empire," *Trans. Am. Phil. Soc.*, New Series, 41:11–295 (1951); and in CHRISTOPHER PYM, *The Ancient Civilization of Angkor* (1968).

(F.E.R.)

Jazz

Although nobody has ever satisfactorily defined it in technical terms, jazz diverges widely, even violently, from all previous canons of musical composition and performance and is immediately distinguishable. Certainly, it is the most enigmatic of musical forms, never respecting any of the received truths about itself. It is generally accepted, for instance, that all jazz is improvised, which is untrue; that jazz is synonymous with syncopation, which is even more untrue; that jazz owes its idiosyncratic nature to the ingenious subtleties of its rhythmic pulse, whereas for at least the first half of the 20th century the precise opposite was true, its rhythmic conception being extremely rigid and formalized. As jazz is neither purely composed music nor purely extemporized music and as it cannot be accurately notated, a logical positivist would have no difficulty demonstrating that, like the beach of the German philosopher Hegel, which was neither sea nor land, jazz does not exist at all. As jazz is essentially the musical experience of a passing moment, which cannot be repeated in quite the same way, in a sense the most important figure in all its history was Thomas A. Edison, who invented the phonograph.

A loose definition that has served well enough for most of jazz history is that it is a music where the performer plays melodic variations on a given harmonic base against a regular rhythmic pulse; in the 1960s even this definition became unacceptable, because the avant-garde movement dispensed with prearranged harmonic signals, indeed, with any kind of form. Difficulties of definition are aggravated by the fact that the terminology of jazz retains its validity only within the jazz context, so that to describe, say, the leading figures of the harmonic revolution of the 1940s as "Modernists" is to beg the question of Modernism as a movement in the music world at large that has been flourishing throughout the present century. An example is the chord of the minor seventh (for instance, C–E♭–G–B♭), which the standard-bearers of jazz Modernism held aloft like a banner but which in a "classical" music context had long become venerable through the French composer Claude Debussy and the school of musical Impressionists. Indeed, much earlier still, in 1858, the Russian author Ivan Turgenev could make one of his heroes sit at the piano "entranced over minor sevenths." In the same way, other terms, such as traditional, progressive, and classical, have very specialized meanings within the jazz world.

The executant as the composer

Because of the oddly eccentric nature of its whole being, jazz has enjoyed the questionable benefits of definitions that are either tautological—"Jazz music is any music played by jazz musicians"—or confusing to the layman—"Jazz is a matter of lip-technique." What can be said with confidence is that, whereas in more conventional musical areas the artist is fundamentally an executant expressing the findings of the creative mind of the composer, in jazz the performer is usually his own composer. Within the strict meaning of the term, there can be no such thing as a jazz theme, although of course some themes will lend themselves to the jazz idiom more readily than others. The customer unable to acquire the recording of Brahms's *Fourth Symphony* conducted by Herbert von Karajan would probably settle for someone else's recording of that work, but the buyer thwarted in his attempt to buy the jazz musician Duke Ellington's version of "Caravan" might well accept as a substitute anything else played by Ellington. That is why, in jazz, there is at least one truism that has always applied: the performer playing a theme always tries to make it sound not like itself but like himself.

ORIGINS

The birth pangs of jazz are perhaps of more concern to the musicologist, the social historian, and even the anthropologist than the musician. The multiracial origins of jazz are clear enough. Had it not been for the traffic in slaves from West Africa to the United States, jazz would never have evolved, either in the United States or Africa, for jazz is the expression in music of the African native who is isolated both socially and geographically from his natural environment.

The jazz scale

Among the West African tribes supplying victims for the slave trade, music, and especially vocal and percussive music, had developed in a way quite unknown to the academic Western ear. Scales and harmony were purely intuitive, and music was deployed less as an abstract aesthetic gesture than as a specific language conveying subtle shades of meaning and emotion. By varying the pitch of a note or changing the inflection of the voice while uttering a musical sound, the performer could convey far more concrete messages than his sophisticated Western counterpart. If his art was cruder, his function was more practicable. Although most of the prehistory of jazz is speculative, the flexibility of this musical language, unfettered either by conventional ideas of correctness or by precedent, resulted in the unwitting development of a scale utterly original so far as the West was concerned. This scale, in which the third (E in the key of C), or mediant, and the seventh (B in the key of C), or leading tone, were flattened and thereby turned into what are sometimes referred to as "blue notes," became the basis of the language that eventually emerged as jazz. The transfer of these West African tribal traditions to the slave fields, railways, and rivers of the southern U.S. states was of advantage to oppressed and oppressor alike, the slave obviously taking solace in the cultural memory of his own collective past, the slaveowner encouraging work songs in the same spirit as an infantry general might approve of military bands—for the stimulus they gave to work rate. Many of the early examples of primitive vocal jazz relate closely to the labours being performed, and the content of the lyrics is a reminder that not only the cotton plantations but also the levees and railroads of the Deep South were created and maintained by slave labour. Joe "King" Oliver's "Lift 'Em Up Joe" and many of the songs of Huddie Ledbetter (Leadbelly) are examples of the work-song convention surviving as art long after its original functional need had declined.

There was another major influence on the evolution of Afro-American vocal style. If the slave could derive no comfort from the laissez-faire philosophy of his owners, at least he could draw spiritual consolation from their religion. Until the beginning of the 19th century, the established church, by performing a series of comically dishonest intellectual cartwheels, had managed to reconcile Christianity with the possession of slaves, the argument being that, although one Christian must not enslave another, it was acceptable for him to enslave a savage. This meant that the work of the missionary was actually rendered sacrilegious, and throughout the 18th century the absurdity of such an attitude became more and more exposed, until about 1790 the Methodist movement began to address itself to the redemption of the souls of men who until now had not been thought to have possessed any. This missionary campaign flourished for almost 100 years and had one of the most astounding outcomes of any evangelizing crusade in history, one that the Christian Church could never have anticipated. By attempting to convert the slaves to Christianity, the missionaries achieved the Africanization of their own hymnbooks.

Development of the spiritual

By adapting his own ritual music to the liturgy of the Christian Church, by contributing, as a member of a congregation, to the creation of new tunes, or by making his own variations on the existing ones, the slave and his emancipated descendants developed the spiritual to the point where, in the form of hymns, ring shouts, revival chants, camp songs, and funeral songs, it gradually merged into a semi-secular tradition. Significantly, ragtime, that coarse yet disarming bridge between the old songs of slavery and emergent jazz, figured unmistakably in the accompaniment of hymns such as "Good Lord'll Help Me on my Way" several years before ragtime music was published officially under its own name. This vast influence of Africanized church music on the development of jazz underlines one more fallacy about the music, which is that it was always linked irrevocably to the lowlife. Its connections with the brothels of Louisiana and the saloons of Chicago tell only half the story, for jazz has been concerned with sanctity as well as with sin, has been a sacred music as well as a profane one. Its links with Christianity and particularly with the act of worship and the rituals of birth, marriage, and death have proved so durable that they remain unbroken to this day, not only in the person of gospel singers, such as the late Mahalia Jackson and Sister Rosetta Tharpe, but also in more secular figures who insist on the church as their primary source of musical inspiration.

Improvised instruments of the early days

Not all of the early Africanized church music was vocal, although, naturally, any instrumental playing was bound to be crude in conception, execution, and instrumentation. The slave generations obviously had access to neither conventional musical education nor to legitimate orchestral instruments, and, even after the liberation that followed the War between the States, such things as education and conventional musical instruments remained, if not impossible, at least difficult to acquire. The New Orleans pioneer drummer Warren "Baby" Dodds is said to have made his first pair of drumsticks from the legs of a kitchen chair, and the semilegendary cornet player Charles "Buddy" Bolden was seen at times to thrust half a coconut shell, a bathroom plunger, and an old derby hat into the bell of his instrument to do service as mutes. The early jazz artists, it would seem, improvised not only the music they played but also the instruments with which they played it.

A key figure, because he personifies the antipathetic traditions of religious piety and worldly musical wit and also the tensions created by the unorthodoxies of musical education of so many of the pioneer jazz figures, is the pianist-composer-singer Thomas "Fats" Waller. Waller was born into a religious family the head of which was a minister of the Abyssinian Baptist Church and who regarded jazz, to quote his son, "as music from the devil's workshop." Waller, a brilliant natural musician (see below), taught himself the piano and soon developed an outstanding gift for composition. His commercial as well as artistic success was enormous, but it seems reasonable to conclude, both from Waller's own occasional remarks and from the evidence of his love for the organ and the presence in his repertory of Bach, Liszt, and Rimsky-Korsakov, that Waller was a formal musician manqué.

The generation of men such as Bolden created the earliest jazz traditions. Liberated but not freed, cast on to the open labour market for the first time, and endowed with a musical tradition self-formed but speedily secularized, the men of Bolden's time had two problems to face, no less daunting simply because they were mere subconscious anxieties. How were they to arrive at some code of artistic behaviour when making communal music and for whom were they to play the music? Clearly, if jazz were to shed its religious connotations, it must become a functional music, and, equally clearly, if formal music offered neither any precedents for procedure nor even a polite interest, then jazz would have to create its own.

The latter third of the 19th century was a crucial point in the prehistory of jazz—a time when jazz was interacting with church music, with the white commercial world of dances, soirées, drawing-room ballads, and concerts, with opera, with the theatre, with vague occasional wisps from the European tradition, and a time when the Southern Negro was learning how to live with uneasy freedom; during this period the traditions of jazz were slowly forged. It is a period and a process impossible to document with any accuracy, for obvious reasons, and about which even the most basic terms have remained in contention. The institution of ragtime piano playing, for instance, is regarded by many as no more than a subdivision of emerging jazz piano styles, but, by many others, it is regarded as a distinct, separate form, marked out from jazz because its repertory consisted largely of formal, composed works rather than slight themes for improvisation. There also exist several abstruse theories about regional variations in jazz styles. Nobody has ever decided to the satisfaction of anyone else where the "folk" music of, say, the itinerant guitarists of the Southwest ends and jazz proper begins. But, indisputably, by the beginning of the 20th century, a few jazz forms had begun to combine into a recognizable tradition, and the hub of this development was located at the Louisiana seaport of New Orleans, where the dominant form among jazz musicians was the blues.

The blues. Much of the confusion over the identity of the blues has been caused by the word itself, which has had to perform a dual purpose. It has been used for at least 200 years by writers as a synonym for a depressed mood and, more strictly, as a specific musical term, so that misconceptions have been perpetuated that all blues music must by definition be concerned with depressed subjects and even that only depressed musicians play it. Another belief about the blues, that it is the acid test of a performer's improvising ability, is more accurate for reasons intimately connected with its harmonic structure.

The blues as a test of a jazz performer

In jazz, a blues sequence extends over 12 bars, containing three harmonic crisis points, and only these give the improviser hints as to where his melodic variations should go and what kind of musical mood they should imply. The paradox is that, sparse though these are, they have a vast emotional potential for the gifted player to discover, so that, ever since the beginnings of instrumental jazz, a player's ability with a blues sequence has been a guide to his true talent. Perhaps the most extraordinary thing about the blues has been its durability; it has retained its attraction for soloists of all eras and schools, changing with new approaches and movements but remaining intrinsically itself. The Negro composer William Christopher ("W.C.") Handy was an important figure in the development of the blues. He was fundamentally a kind of folklorist-orchestrator who took his themes from the blues performers he heard around him, wrote them down, and harmonized them. Some blues was published before Handy's, but, as a result of his disseminating work, jazz players by the late 1920s were freely improvising and spontaneously inventing melodies on blues chord sequences. His most important compositions were the "Memphis Blues" (1911) and "St. Louis Blues" (1914).

Exploration and improvisation. A close textual analysis of the harmonies employed in the blues since the beginnings of recorded jazz reveals the process of ceaseless embellishment of the original sequence, and this process may stand as a symbol for the whole of jazz history. In effect, the story of jazz has been a saga of harmonic exploration. Most of the jazz pioneers were men without musical schooling or formal knowledge who had to evolve a musical language and harmonic vocabulary by trial and error. They were artists intent on hammering out some formality of procedure, some kind of convention that would reconcile the individual freedom of each player to express himself with the adherence to the demands of the ensemble without which all performances would disintegrate. This is the crowning paradox of the art of jazz: it is at once the art of the individual musician and an almost exclusively communal exercise.

Although jazz criticism has divided the music into three equal categories—traditional, mainstream, and modern—harmonic exploration and the growth of melodic vocabulary have comprised a constant, continuing process in jazz history. New labels have been attached to new styles not when some dramatic advance has been made but when the cumulative weight of several years' findings has caused people to realize that the music has steadily been changing its personality. The three compartments are useful as an approximation, not an arbitrary judgment. As the harmonic vocabulary of the jazz musician has been extended and as the degree of technical sophistication has increased accordingly, so have new conventions been forged, although it is not true that, as one style established itself, others have been superseded. Uniquely and because of the very fast pace of its evolution (covering an advance from primitivism to neoclassicism in little more than half a century), several schools of jazz have existed concurrently.

EARLY JAZZ STYLES

The New Orleans style. The rigid convention of collective improvisation, based on a specific instrumentation, flourished in Louisiana in the early years of the 20th century, and two powerful factors dictating the formulation of the convention were probably the social and the functional. In the social life of the Negro, community music played a more prominent part than perhaps can now be easily comprehended. There was almost no social activity in New Orleans then that did not imply a musical corollary. There was live music for births, weddings, christenings, funerals, picnics, parades, marches, and all kinds of celebrations. A great deal of this music was naturally played outdoors, which may explain the unchallenged dominance of the trumpet over all other instruments. All the early leaders of New Orleans jazz were trumpeters, with the line of succession passing from Bolden, through historically documented musicians such as Freddie Keppard and Joe "King" Oliver, and down to Louis Armstrong. New Orleans also produced brilliant players on other instruments, particularly a school of musicians who effected a limpid, highly attractive clarinet style, but leadership was almost exclusively the preserve of trumpeters like Bolden and Oliver.

Roles of the instruments in New Orleans jazz

In the classic New Orleans style, the trumpet's duty is to state and embellish the melodic line of the theme. The trombone stresses the harmonic root notes, providing also a solidity of resonance on which the other performers may build. Above the trumpet soars the clarinet voice, weaving further variations on the same harmonies. Thus, the three voices, linked yet independent, are able to compile between them the simple triads (chords consisting of a root and the third and fifth tone above it) that were the basis of all jazz harmony at this period of its development. And, as all three were playing together throughout the performance, the band, though small, was able to maintain a surprising degree of volume.

There was a vital reason why the ensemble convention was adhered to so faithfully. In New Orleans or, indeed, anywhere else at that time, there were few musicians capable of playing an extended solo, even had the rules of their game permitted them to do so. Apart from a handful of virtuosos, the idea that a jazz performance should consist of a succession of solos would have been unthinkable for the simple reason that there was no such thing as a succession of soloists. But the musicians were developing at an astonishing rate, and, in retrospect, it can be seen that the classic New Orleans style, rigid as it had to be, was doomed by its very nature: there was no question that, in time, a player or group of players would emerge for whom the constriction of the ensemble was intolerable. When this player arrived, then the whole New Orleans conception of tightly integrated ensemble improvisation would become obsolete. Another reason why the frailty of the New Orleans tradition is more apparent now than in the heyday of that style is the fact that jazz as a musical lubricant to oil the social machine was restricted largely to the New Orleans lowlife. Although the myth-making process has drawn a picture of jazz limited strictly to the brothels and sporting houses of Storyville, the town's bordello district, there were, of course, many instances of the music splashing over into the life of the city at large. Nonetheless, jazz, linked to the Negro performer and the social events of Negro life in the city, retained a connotation of sin and dissipation for many years after the New Orleans pioneers were forgotten. The saxophonist Sidney Bechet, one of the most gifted of all the New Orleans musicians, insisted in his autobiography that the word jazz in its original form of jass was local slang for sexual intercourse, and the evidence in favour of Bechet's assertion seems overwhelming.

Supposed origin of the word jazz

These brothels were thus a link in the jazz musician's economic chain, for many employed hands or, at the very least, a house pianist whose job was to thump out ragtime rhythms against a background of red plush and gilt. The collapse of the Storyville economy was naturally disastrous for the working musician. In 1917 the United States secretary of the Navy decreed that, in view of the repeated fighting and violence involving seamen on leave in the city, the New Orleans red-light district must be closed down. The sense of outrage and the disarming worldliness of the city are reflected in the official statement by the then mayor, Martin Behrman:

> Preterpermitting the pros and cons of legislative recognition of prostitution as a necessary evil in a seaport the size of New Orleans, our city government has believed that the situation could be administered more easily and satisfactorily by confining it within a prescribed area. Our experience has taught us that the reasons for this are unanswerable, but the Navy Department of the Federal Government has decided otherwise.

The theory that the closing of Storyville brought the heyday of New Orleans jazz to an abrupt end is one of those critical platitudes excused by the fact that it is largely true. Jazz did not, however, immediately stop in New Orleans, nor was the migration north of the musicians sudden or absolute, nor had jazz until then been unknown in the North. As early as 1917, the year of the Storyville edict, the Original Dixieland Jass Band, a group of white Southerners with a comically inflated sense of their own importance as musical innovators, had introduced jazz to the patrons of Reisenweber's restaurant in New York and recorded two compositions.

One potent evangelizing factor was the riverboat, which would ply up and down the Mississippi, often with a jazz band aboard. More than one white middle-class jazz pioneer has testified that the first jazz he ever heard came floating across the water from one of these boats as they approached the levee of some Southern port of call. The accessibility of Europe was also a factor at a surprisingly early point in the music's history. In 1919 the Southern Syncopated Orchestra, with Sidney Bechet as its star performer, played in London, there attracting the notice of a Swiss conductor, Ernest Ansermet, who was the first distinguished figure of formal music to react favourably to jazz and to discern in it uniquely vital qualities.

The spread of jazz to Europe

The main force pushing the New Orleans musician north was his need to find employment, and perhaps the most significant sequence of events after the closing of Storyville was that involving Joe "King" Oliver. Early in 1918 Oliver, acknowledged trumpet champion of New Orleans, migrated north to Chicago. By 1920 he had become a popular bandleader there, and two years later, wanting to increase the size of his band, he sent to New Orleans for the most brilliant of his disciples and, indeed, of all the jazz musicians who came out of the city, Louis Daniel Armstrong. From this point on, jazz evolved from a local musical dialect into an international language, proliferating in geographical range and in stylistic variation to a degree that astonished those of the New Orleans founding fathers who lived long enough to watch the process for themselves.

The Louis Armstrong style. Armstrong may partly stand as a representative symbol of the history of the music itself. A trumpeter of freakish gifts, he performed at least three feats—two aesthetic, one sociological—for which he will be remembered as the most influential jazz musician of all time. He took the classic style of his native city and split it at the seams through his limitless ability as an imaginative soloist. The moment his style began to mature, the convention of ensemble playing was outmoded, and he established the primacy of the improvising soloist. Then, having liberated the player from an exclusively team performance, he unwittingly codified the vocabulary of the soloist in a series of famous recordings between 1925 and 1928. These recordings, featuring his groups the Hot Five and the Hot Seven, emphasize the enormous gulf between Armstrong and even the best of his contemporaries. They also show how he summarized all that had gone before and enriched the jazz tradition with a whole range of melodic effects relevant ever after.

The Hot Five and Hot Seven recordings

Several of the performances of the Hot Five and Hot Seven are traditional blues, successively asserting the power of the idiom in a way not to be equalled for at least a generation, until the rise of Charlie Parker. Still more important was Armstrong's success as an evangelist for the jazz cause. By the late 1920s, quickly graduating from the strictly specialist environment of his formative years into being a Broadway and nightclub star, he was the first jazz personality to become a national and then an international figure. Many with no particular interest in jazz were utterly beguiled by his genius. His sense of comedy, his career in Hollywood, his spectacular physical appearance and mannerisms, and, above all, his extraordinary vocal style, all contributed to this process. By using his voice in the same way that other musicians used their instruments, Armstrong, usually dispensing with words and substituting for them an odd, wholly idiosyncratic language of sounds, popularized scat singing. Had he not drifted into the world of entertainment at large, jazz might never have won the widespread currency it did. Although from around 1930 to the end of his life he did not develop as a musician, he never really found himself outflanked by others, his vitality enabling him to retain a hold on his art almost to the end of his life in 1971.

In his failure to develop, Armstrong was typical of most jazz musicians, who have generally become reconciled to the style formulated at a comparatively early time of life. Although the music itself seems to evolve, individuals hardly ever do so. The recorded evidence to support this theory is there, but it is usually overlooked in the hunt for fresh effects and styles. Although Armstrong was to improve vastly in the years following his decision to join Oliver in Chicago, his classic style was already becoming permanent before he had ever left his home town.

Chicago style. Armstrong's arrival in Chicago, added to the burgeoning of the speakeasies of the Prohibition era (the illegal drinking establishment became the 1920s equivalent, so far as the jazz musician was concerned, of the Louisiana brothels of the previous decade), made that city the new centre of jazz. By the mid-1920s it was becoming clear that young, white, imitative musicians were quickly learning the lessons taught by men such as Armstrong. One of the better derivative groups of white musicians was the New Orleans Rhythm

Flowering of white talent

Kings, but the real flowering of white talent began a little later, farther north, in time causing the rise of a new style of playing jazz, which was a compromise between the ensemble tradition of New Orleans and Armstrong's solo power. In Chicago style, the triumvirate of trumpet, trombone, and clarinet was retained, though the saxophone was becoming much more common by the end of the decade. Although performances usually began and ended with a rousing ensemble variation of the theme, the central part of the performance usually consisted of a string of solos.

Because the term Chicago style is neither strictly geographically accurate nor musicologically precise, its umbrella usually covers stylists of widely contrasting or even antipathetic talents. The leading player of the style was Leon Bismarck "Bix" Beiderbecke, a self-taught cornetist and pianist whose pure tone and introspective curiosity about harmonic theory mark him as the exact opposite of Armstrong, whose intuitive style was rooted in a radically different racial background (Bix was middle class, the son of German immigrants).

Beiderbecke's significance

Beiderbecke, whose recordings of "I'm Coming Virginia," "Way Down Yonder in New Orleans," and "Singin' the Blues" have become classics of the genre, was also responsible for one of the most remarkable and apocalyptic of all jazz recordings, the piano solo of his own composition, "In a Mist." In this fragment are strangely fused the honky-tonk beginnings of Bix's early musical life and the subtler overtones of those modern classical composers whose recitals he attended and whose comparatively complex harmony caused him to ask the question: "Why should the jazzman be limited to the simplest triads when his conservatory-trained fellow musician has recourse to so many more variegated effects?" Although this has remained the most vital of all the questions posed by jazz, Beiderbecke did not live long enough to see it answered. He died a young man, technically of pneumonia, actually of frustration, self-neglect, disenchantment—ailments that can destroy any creative artist whose sensibilities far outstrip his technical equipment. Beiderbecke, the first patron saint of jazz, has remained the archetypal figure of the playing fool, the instinctive creator who operated against the frenetic background of illicit stills and gang warfare.

Racial segregation was long the rule in the recording studios, so posterity has tended to have a false view of the mingling of musicians in the 1920s. These players were by no means unaware of each other, as a glance at the recording catalogs of the period might suggest. Armstrong and Beiderbecke, for instance, admired each other's playing. The traffic in ideas was generally one-way, from black to white, but the white musician was occasionally something more than a plagiarist. One of these was the Texan trombonist Jack Teagarden, a musician of remarkable fluency and assurance even at the start of his career in the middle 1920s. After being associated largely with the Chicago-style groups of his first years in the North, Teagarden often worked with Armstrong, to whose All-Stars he belonged during 1947–51. Like Armstrong, he used his voice as a useful second instrument, and, like Armstrong, he retained his creative ability throughout his life. In retrospect, the recordings he made with the young white musicians of the late 1920s and early 1930s show an amazing poise and maturity. The other white Chicagoan who affected the course of jazz history was the clarinetist Benny Goodman. Born in Chicago, he was a child prodigy, a master of Chicago idiom while still in his teens, and destined to lead jazz into new areas.

Strictly speaking, Goodman's clarinet style was hardly original, consisting as it did of a brilliantly executed synthesis of all that had gone before, with particular reference to the limpid fluency of the New Orleans player Jimmie Noone; but Goodman brought to the playing of jazz a technical expertise, an academic intelligence, a speed of thought not heard before. In time his classically based style became ossified, and Goodman himself came to turn more and more frequently to playing Brahms and Mozart; but he remained the most prolific, technically accomplished, and melodically resourceful of the white musicians generally designated as the Chicagoans, though he was later dismissed by some as a peripheral rather than a central figure in jazz history.

LATER DEVELOPMENTS

Hawkins' development of saxophone technique

The emergence of the virtuoso. Throughout the 1920s, techniques were steadily becoming more efficient, thinking more sophisticated, harmonic exploration gradually more daring, so that many styles that were to flower during the soloist's golden age of the 1930s were already being shaped. One of the most gifted of all the musicians in jazz history, Coleman Hawkins, a tenor saxophonist, virtually single-handedly raised the status of that instrument. His work was characterized by a deep, passionate tone, which may be taken as the working definition of "hot" music, by sequences of ingeniously related arpeggios and, above all, by an indefinable sense of form that enabled him to transmute successions of fragmentary phrases into a corporate whole. This sense of form, unteachable but unmistakable when encountered. had already distinguished the truly gifted improviser from the merely talented, and Hawkins, particularly in slower ballads, possessed this gift to a remarkable degree. His outstanding recordings are scattered over five decades. Especially noteworthy are his version of "One Hour," made in 1929, one of the very first recordings by a racially mixed group and one of the very first on which the tenor saxophone in jazz approaches maturity; his famous recording of "Body and Soul" in 1939, a performance so finely constructed that it has long since come to acquire the standing of formal composition; and some sides he made with a cosmopolitan band in Paris in 1937, of which "Out of Nowhere" was a remarkable exposition of the art of romantic extemporization. From 1934 to 1939 Hawkins travelled in Europe, inspiring European musicians to attain fluency in what was at that time an essentially alien art.

The development of jazz piano. There have been deviations of style and approach on all instruments throughout jazz history, giving rise to rival "schools" the differing precepts of which were sometimes the centre of bitter debate, but the case of the piano in jazz is unique in the way that two antipathetic styles emerged and, at least for a time, developed concurrently but independently. These two approaches, embodying contrasting philosophies, were "stride" and "trumpet-style" piano playing. The stride style, whose roots in the early jazz past are obvious enough, is based on the premise that any instrument that makes it possible for the performer to play two or more notes simultaneously and therefore to create a harmonic as well as melodic effect ought to deploy those possibilities as far as is practicable. At their best, the stride pianists have produced work teeming with an orchestral richness, yet they have rarely been the kind of players to produce a single melodic line that might conceivably be transposed to a different instrument. The best stride exponents, pounding out ten-note chords, stressing the rhythmic pulse with great muscularity in the left hand, have been men whose music is full of harmonic possibilities and enchanting rhythmic variation.

Probably the most important stride pianist was James P. Johnson, one of the most resourceful figures in jazz history and a rare personality among the pioneers in that his musical accomplishments enabled him to compose and orchestrate at a period when such abilities were rare. Apart from his purely jazz activities, Johnson wrote some early film music, a tone poem, and a ballet; he also enjoyed great success as a songwriter, producing the "One Hour" that Hawkins (and also Louis Armstrong) recorded so finely and "Runnin' Wild."

Johnson presided over what came to be known as the Harlem group of pianists, whose stride playing in the 1920s made them virtually one-man jazz bands. Johnson's great friend and rival was Willie "The Lion" Smith, and their two unofficial pupils were Thomas "Fats" Waller and Edward Kennedy "Duke" Ellington. In both cases, the stride style underwent the most astonishing changes. Waller is perhaps the only example in all jazz of the pupil outstripping his masters. Apart from the irresistible strength and wit of his piano playing and his gifts as a composer ("Honeysuckle Rose," "Ain't Misbehavin'," "Blue Turning Grey Over You," and literally hundreds of others), Waller, like Armstrong, was a natural buffoon and stage extrovert, capitalizing on these gifts to become a popular as well as a jazz figure. Ellington's subsequent career (reviewed below) was even more remarkable, and both he and Waller acknowledged their stylistic debt to Johnson and Smith. Ironically, although the stride school avoided the dogmatic melodic statement in favour of the implied harmonic one, at least three of its greatest practitioners—Johnson, Waller, and Ellington—were also successful songwriters.

Events eventually overtook the stride school, and its eclipse may be dated from the moment in the late 1920s when Earl Hines, yet another outstanding jazz pianist, who augmented his income by working as a songwriter, saw that, in addition to stating the harmony in the left hand, the pianist might well emulate the single-note instruments, such as the trumpet, in creating with the right hand linear improvisation producing melodic single-note variations on the melody. Probably the two great influences on Hines that helped crystallize his style were the pianist Teddy Weatherford, to whom Hines acknowledged a debt of inspiration, and Louis Armstrong, with whom Hines worked during the Hot Five and Hot Seven period. Armstrong's apparently limitless powers of invention must have shown Hines how the same kind of melodic aphorisms could be produced by a pianist.

Hines's right-hand improvisations

Hines, a pianist of amazing technical command and tireless energy, had a profound influence on the development of jazz piano. His findings eventually led to the eclipse of the stride style, especially when in the early 1940s the new Modernism preferred the pianist with the ability to create right-hand, single-note lines. In the 1930s the most accomplished follower of Hines was Teddy Wilson, after which time the work of pioneer Modernists such as Bud Powell and, later, the Canadian Oscar Peterson showed at least to some degree a corroboration of Hines's methods.

Art Tatum, one of the greatest and most controversial of jazz musicians and pianists, had a style too personal to be categorized. For him, stride and trumpet-style were not so much approaches to piano playing as incidents in the course of a single casual chorus, just two of many effects that Tatum had at his fingertips. Almost totally blind, Tatum developed a technique as unorthodox as it was infallible, so that he became the despair of his fellow musicians. Because of the rococo flourish of his style, Tatum was hardly ever able to integrate his piano playing into the texture of a group and was therefore almost always heard either at the head of a trio or as a solo pianist. Although he began about the time Hines was formulating the principles of trumpet-style piano and lived on to know the violent disputes of the postwar Modernist movement and after, Tatum remained a constant. Toward the end of his life, his playing became, if anything, more florid than ever, but it hardly evolved at all, so complete and subjective was his music. Because of the technical problems presented by Tatum's all-embracing style, he has had virtually no imitators, with the notable exception of Peterson, whose work grew increasingly closer to Tatum's in spirit as the years went by.

Tatum's unorthodox technique

The era of the big bands and swing. In the first years of its history, jazz was confined almost exclusively to small-group collective improvisation, and such a conception as orchestral effect was a refinement that, for obvious reasons in a world populated by musicians illiterate at least in the conventional sense, had to come later. The first musician to make a serious attempt to organize a group of players who might keep, at least in part, to a plan was Ferdinand "Jelly Roll" Morton. Although his career has been more heavily documented and annotated than that of any other jazz musician, his recorded legacy was not large enough to decide whether he was a pioneer or a charlatan. Certainly, his announcement that he "invented jazz in 1902" did nothing to help.

The first indisputable figure in the evolution of orchestral jazz was Fletcher Henderson, an unremarkable pianist whose contribution to jazz lies in his pioneering of methods later universally adopted. Henderson, far better educated than most Negro musicians of his day, formed his own orchestra in 1923, and it became the first to gain wide fame by playing jazz. For several years after, his band employed the best Negro jazz talents, among them, at various times, Louis Armstrong (this was the last time in his life Armstrong was ever hired by a handleader) and Coleman Hawkins. Despite an embarrassment of riches in the solo department, the Henderson band is unique for the way in which its leader experimented with orchestral effects. According to many purists, orchestral jazz is a contradiction in terms, for, if it is true that jazz is improvised music, then it follows that a jazz orchestra cannot exist. Henderson was among the first to see that it is not necessarily improvisation that lends jazz its fierce vitality but the preservation of its spirit. If musicians could play written parts with the same sense of self-discovery as a solo, then the effects need not be anticlimactic, and their solo talents could shape an ensemble into the contours of a jazz performance.

Henderson's contribution to orchestral effects

Henderson also contributed the concept of sections of instruments, comprising three or four voices, playing responses to the solo voice in harmony. The soloist, used to the accompanying figures provided for him, could often space his own improvisation, shape it, and adjust its dynamics to this scored accompaniment. Henderson's harmonic vocabulary and orchestral technique, profound in the jazz context of their period, were naive in the broader musical sense, and, by an irony of fortune, his most revolutionary success happened in the end to somebody else (see below).

Many other bands of merit more or less followed Henderson's example, notably Chick Webb's band from 1926, Jimmie Lunceford's from 1927, and Bennie Moten, whose move from Kansas City eastward in 1926 coincided with the increase in the size of his group. Although none of these orchestras won the national and, in a few cases, the international fame that attended the touring orchestras of the 1930s, they were at least as good, if not much better, than their more exalted contemporaries. One of the main barriers to their being heard on a truly national scale was their colour.

As bands such as Henderson's, Webb's, and Lunceford's moved into the 1930s, the growing complexity of their harmony and technique combined with a more general process of advancing musical curiosity to produce a golden age of solo virtuosity. Some of the virtuosos remained in an orchestral setting, others were in the proliferating small groups. One of the greatest of these soloists was the tenor saxophonist Lester Young, who was associated for many years with one of the finest of all jazz orchestras, the Kansas City-based Count Basie band, which had evolved from Bennie Moten's group after the leader's death. Basie, a stride pianist, amended his style from a flow of harmony to a great plain of silence dotted with exquisite melodic epigrams neatly executed and gathered together a large orchestra that preserved the buoyancy of an improvised act by literally improvising. Many of his most effective orchestrations, such as "One O'Clock Jump," were, in effect, not orchestrations at all but "head arrangements"—that is to say, fragments of improvised music put together by the musicians in a process of trial and error, conned by rote, then gradually becoming a predictable performance. The Basie band, stressing four beats to a bar and retaining, long after others had abandoned it, the four-man rhythm section of piano, double bass, drums, and guitar, set unsurpassed standards of vigour and precision, and in this setting Young, a remarkable soloist, thrived.

Young's contribution to the solo art was twofold. First, he displayed and popularized a peculiar sensibility that paved the way for the Modernists of the following generation. Young is often credited with being the first Modernist, but, strictly speaking, he was a man of his era who did nothing to corrupt the harmonic innocence of his day. His tone, revolutionary in the 1930s in the way it distilled the hot breath of Hawkins' romanticism into something far more sinuous and oblique, set the styie for saxophonists for a generation to come. In this way he contributed

Lester Young's contribution

an alternative approach to the instrument: ever after, an apprentice could follow either Young or Hawkins, but not both. Jazz saxophone playing was thus immeasurably enriched. His other achievement was to demonstrate a new way of building an improvised solo, by using little-used chords, such as the minor sixth, by showing how silence could help a solo, by breaking down an arpeggio so that unusual intervals could express conventional harmonies, and by perfecting an ingenious series of false fingerings so that the player could produce different densities of sound on the same note. While the Modernists of a later age added to the jazz language by their increased harmonic scope, Young created a new vocabulary for the soloist within the conventional apparatus.

Many imponderables contributed to the astonishing upsurge of popularity of the big bands in the 1930s. The end of Prohibition in 1933 was a large factor; it altered the social habits of a nation and deprived jazz musicians of their biggest potential source of employment. By 1934 jazz had to find a new audience, and by a happy coincidence it reached a stage of sophistication that made the capture of a new audience possible. Benny Goodman formed his first organized orchestra in 1934, and, after an unsuccessful beginning touring the nation's dance halls, stumbled on the truth that others were soon quick to discover: that a new generation of young people was ready to patronize jazz-oriented dance bands. Jazz, though admittedly in a somewhat bowdlerized form, took to the ballrooms of the day.

Goodman and the swing age

In 1935 at the Palomar Ballroom in Hollywood, Goodman found all the factors of the swing age: massed audiences, fans clustered around the bandstand, enthusiasm for individual musicians, pressures changing bandleaders into brand names. Soon, Goodman was the "King of Swing," a Hollywood film star, a millionaire, an international celebrity. Naturally, others followed, among them a rival clarinetist, Artie Shaw, also Tommy and Jimmy Dorsey. Goodman ensured his success by recruiting the semi-retired Fletcher Henderson as chief arranger and by incorporating within his large group a smaller one (at first, a trio, later, a quartet), comprising the first racially mixed group ever to tour nationally in the United States — Goodman and drummer Gene Krupa being white, pianist Teddy Wilson and vibraphonist Lionel Hampton being Negroes.

The swing age culminated in Goodman's 1938 concert at Carnegie Hall. Then, with orchestras hit by the wartime draft and running costs rising all the time, it petered out, losing its jazz connotations. By 1941, with the hysterical career of Glenn Miller, it ran into the quicksands of commercialism. Perhaps the jazz purist would commend Goodman's career as a bandleader for discovering a hitherto obscure Midwestern guitarist, Charlie Christian, the first guitar virtuoso to use electrically amplified equipment, although others had pioneered this method. He was one of the few stars of the swing era psychologically and musically equipped to make the leap into the Modernist era of the 1940s; his early death ended a brilliant career.

Until this time, jazz had been more or less diatonic (adhering to the natural scale), restricting itself, for the most part, to cycles of resolving dominant seventh chords. The Modernist movement of the 1940s thrust it forward into unexplored realms of chromaticism (use of harmonies built on notes not in the key of the piece), which split players and devotees into for-or-against armies and caused the advanced spirits to forgo the allegiance of mass audiences, which was never recovered.

Charlie Parker's influence

Modernism. The origins of the Modernism of the 1940s, or "bebop," as it was then onomatopoetically known, are confused. One of its leaders, John Birks "Dizzy" Gillespie, said it was a device to shake off white plagiarists, but its most gifted practitioner, Charlie Parker (known to many as Bird), explained it in strictly technical terms. By evolving a system of substituted chords superimposed on the original ones and by playing in double the time of the tempo being asserted by the rhythm section, Parker, an alto saxophonist of extraordinary gifts, changed the face of jazz. His early death was symbolic of the tragic involvement of many jazz musicians of his generation with drug addiction. Although technically he made jazz more complex, emotionally he cleansed it, and his famous blues recordings looked back to Armstrong's achievements with the Hot Five, in spirit, if not in method. Parker could not have realized it, but he was making, on behalf of jazz, the last great appropriation of musical territory. To use an American pioneer analogy, after jazz had moved harmonically westward for a half a century, Parker took it to the sea. This caused heart-searching among those who inherited Parker's findings. Among his disciples was the trumpeter Miles Davis, who in 1956 made the first coherent attempt to escape from the cage of "discord to resolution," a method so persistently explored that the men of Davis' generation felt its usefulness to be ended. In an album called *Kind'* of *Blue,* Davis substituted modal (scales not based on the major and minor) patterns for the more conventional harmonic ones, the first step in a process that was to characterize jazz thereafter. The theories of men such as Davis seemed to be negatively proved by the tenor saxophonists John Coltrane and Sonny Rollins, for, in the playing of both these men, the number of harmonic changes being crammed into each theme was so big that the music was in danger of being choked to death. One possible solution came from the alto saxophonist Ornette Coleman, who abandoned all rules of discord and resolution, time signatures, and keys, in his formulation of "Free Form," which in effect was no form at all. That jazz was not yet reduced to such extremes was suggested by a few of Ornette Coleman's contemporaries, who continued to raise the technical and inventive standards of jazz. These included the pianist Oscar Peterson, the drummer Buddy Rich, the tenor saxophonist Stan Getz, and the guitarist Wes Montgomery, all of whom attained high standards of performance.

Much avant-garde jazz of the 1960s was committed art, in the sense that its creators insisted vigorously on their music as an artistic expression of racial protest, but all that their polemic suggested was that, in accusing society of anarchic tendencies, they had induced those tendencies in their own music and that art committed to a good cause is not necessarily good art.

The role of Duke Ellington

Throughout its history jazz has been abstract, in the sense that, subconsciously informed though it may be by race memory, it has been guided less by concrete or living factors than by the mathematical precision of the march from discord to resolution, the soloist being like a man working out an algebraic equation. For this reason Duke Ellington is unique, in that he has moved toward that ideal where authentic jazz performances may reflect the nuances of personality of some outside object or person. Ellington has thus attempted to make his music measure up to a constant dual standard, for not only has he meant it to be fine jazz but also intelligible program music.

Nothing comparable has been attempted in jazz by anyone else, although the work of "progressive" jazz musicians such as Thelonius Monk and the Modern Jazz Quartet in the 1950s and 1960s hinted at extramusical connotations through the relentless Europeanization of its theme titles; and the Dave Brubeck Quartet made occasional gestures of a similar nature. Ellington has attempted to make his jazz mirror the people and places that led him to the extreme sophistication of his old age. After his graduation from the Harlem group of pianists and his recruitment of a quintet, the Washingtonians, he moved steadily forward to orchestral mastery, increasing the size of his group as his palette became broader. In 1939 he recruited an orchestral assistant, Billy Strayhorn, but by this time he was already far along the road to a technique enabling him to embrace a far wider range of sound than jazz musicians normally aspire to.

By using his orchestra as an instrument on which to perform his feats of orchestrating brilliance and by assimilating the talents of each individual member of that organization as a saxophonist might acquaint himself with the workings of every key on his instrument, Ellington welded together a group inimitable by any standards. The rich impressionism of his early works, in the late 1920s, slowly evolved into the thumbnail sketches of the

1930s. In 1943 came *Black, Brown and Beige,* a musical history of the Negro in the United States, after which Ellington scarcely stopped pouring out fresh works—some whole, some mere fragments, all of them informed by the romanticism of his view of the world.

Among his most remarkable achievements are many popular songs, such as "Sophisticated Lady" and "Mood Indigo," whose instrumental structure has not prevented them from becoming world famous. Of his more ambitious projects, he has composed *Perfume Suite,* the sound track for the film *Anatomy of a Murder,* jazz versions of Tchaikovsky's *The Nutcracker* and Grieg's *Peer Gynt, Liberian Suite, The Far East Suite,* and, possibly his masterpiece, *Such Sweet Thunder* (1957), a series of 12 Shakespearean vignettes that demonstrate Ellington's working method. By settling on the stylistic idiosyncrasies of each soloist, Ellington attempts, in this work, to express his own idea of a dramatic character; thus, the elegiac fervour of his greatest soloist, the saxophonist Johnny Hodges, became the voice of Cleopatra, the stratospheric range of trumpeter William "Cat" Anderson the dementia of Hamlet, the cold academic hauteur of clarinetist Jimmy Hamilton's tone the patrician disdain of Caesar.

Blues singers

Vocal jazz. Although the essence of jazz is instrumental, one or two artists have shown that the human voice can express the spirit of the music. Apart from instrumentalists such as Armstrong and Teagarden, whose singing was an adjunct to instrumental expression, there have been two main groups of jazz singers, one rooted firmly in the folk-oriented communal past of the Southern U.S. Negro, the other committed to the world of commercial music and the fringes of show business. The male blues singers of the South, often men who earned their livelihood at some menial task, sang of life close to the bone, in terms whose graphic candour reflected the earthiness of the life they had known. Outstanding were Huddie Ledbetter, Big Bill Broonzy, Blind Lemon Jefferson, Sleepy John Estes, and Peetie Wheatstraw. Their female counterparts, much more closely linked to the professional world of vaudeville and cabaret, include Gertrude "Ma" Rainey, her young protégé Bessie Smith, Mamie Smith, and Bertha "Chippie" Hill. By far the greatest was Bessie Smith, whose vast recorded output has preserved a style outstanding for its integrity, honesty to life, and immense technical skill. Less traditional male singers have included Count Basie's robust blues shouter James Rushing and Joe Turner, whose 1938 recordings with pianist Pete Johnson coincided with Johnson's short-lived but freakishly popular boogie-woogie piano style, a variation of blues playing whose hypnotic reiterated rhythmic patterns won a world following and whose other notable exponents included Albert Ammons and Meade "Lux" Lewis. The most important female singers of this later period include Mildred Bailey, Ella Fitzgerald, and Sarah Vaughan, but by far the finest singer of modern times and, with the possible exception of Bessie Smith, the most gifted vocalist of either sex that jazz has so far produced is Billie Holiday. Whereas Bessie Smith was incomparable as an interpreter of the poetry of the blues, Billie Holiday achieved the same intensity with the far less substantial repertory of Tin Pan Alley. Many of her small-group recordings from 1935 to 1942, prized classics of the art of impromptu composition, reflect in musical terms the close personal relationship between the singer and saxophonist Lester Young. An interesting departure of the 1950s was tried by the Dave Lambert-Jon Hendricks–Annie Ross Trio, which sang lyrics composed to transcribed instrumental jazz solos.

JAZZ AS A SOCIAL FORCE

The question of the extent to which jazz reflects the society that has nurtured it has seldom been asked. The 1960s saw a belated acknowledgment of its powerful ability to evoke the urban ambience. Duke Ellington's music for *Anatomy of a Murder* (1959) was the first movie score ever composed by a jazz writer. Since then, the jazz language has overrun all spheres of popular musical expression. One of the most intriguing things about its career has been the way in which advancing technique has run concurrently with social improvement, so that, while the crude emergent music of early New Orleans was brothel music and its more polished descendant in the 1920s the music of the ginmill, the suave orchestral felicities of the swing age became the wholly respectable background music for the innocent romances of its college audiences. Finally, with the advance to chromaticism, jazz invaded the concert hall and installed itself at festivals all over the world.

For the most part, jazz has remained the province of the urban U.S. Negro, who has contributed almost every viable new idea that has helped jazz along its path to literacy and self-confidence. Although there have been astonishing advances by European musicians in the past 30 years, there remains only one non-U.S. figure in the entire history of jazz who had something original and also valid to contribute. This was the Basque guitarist Django Reinhardt, whose best work was achieved in the 1930s, before he succumbed to the blandishments of amplification.

The source of the musician's repertory has also changed radically over the years. While the blues has retained its primacy throughout jazz history, the players of the swing era, harmonically cultured, turned to the urbanities of the best musical-comedy writers, so that names such as George Gershwin and Jerome Kern occur repeatedly on recordings of all kinds. Indeed, Gershwin, with "I Got Rhythm," created a harmonic pattern whose popularity with improvisers stands second only to the blues. In recent years, however, there has been a tendency among the younger, more revolutionary players to write their own material, which has made it more difficult than ever for the onlooker to distinguish genuine inventiveness from mere charlatanism.

The influence of jazz on other music

Jazz, in watered-down form, has strongly influenced various other kinds of music, many of them commercially more viable, such as "pop" songs, rock 'n' roll, skiffle, rhythm and blues, Broadway musicals. Some critics include these in the history of jazz; others consider that anything other than original Dixieland, which had a revival in the 1950s and 1960s, cannot be considered as pure jazz. Another form, Afro-Cuban music, which has some affinities with jazz, in fact had different ethnic roots, but some jazz-musicians, such as Gillespie, incorporated its rhythms in their music making.

By and large, attempts to wed jazz with more formal methods of music making and composition have not been successful, although attempts to create a "third stream," involving musician-composers such as William Russo and Gunther Schuller, have not been without virtues to commend them. The oddest work in this direction remains Igor Stravinsky's *Ebony Concerto,* composed in 1945 for Woody Herman's orchestra. Although not in the specialist sense a jazz composition, it underlined the inherent contradictions in a marriage between the two forms of music. There appears to be a distinct possibility that jazz, faced at last with the challenge that its room to expand is finite, may eventually be merged into the world of music at large, to become an orchestral effect in a more general context. On the other hand, the tradition as represented by such artists as Armstrong, Ellington, Hawkins, Young, Tatum, Parker, and Holiday seems almost too fiery ever to be extinguished altogether.

BIBLIOGRAPHY. HUGUES PANASSIE, *Le Jazz hot* (1934; Eng. trans., *Hot Jazz,* 1936); WILDER HOBSON, *American Jazz Music* (1939); and FREDERIC RAMSEY, JR., and C.E. SMITH (eds.), *Jazzmen* (1939), although tainted by early naïveté, are indispensable pioneer attempts to formulate a critical method. ANDRE HODEIR, *Hommes et problèmes du jazz* (1954; Eng. trans., *Jazz: Its Evolution and Essence,* 1956), is an erudite attempt at close textual analysis. Among biographies, ALAN LOMAX, *Mister Jelly Roll* (1950), is outstanding for its graphic style and priceless sociological importance. BILLIE HOLIDAY with WILLIAM DUFTY, *Lady Sings the Blues* (1956); and WILLIE SMITH with GEORGE HOEFER, *Music on My Mind* (1964), successfully survive the ghosting process. STANLEY DANCE, *The World of Duke Ellington* (1970), presents a partial view of an extraordinary man. ARTIE SHAW, *The Trouble with Cinderella* (1952), is the only jazz autobiography with any real literary pretensions. NAT SHAPIRO and NAT HENTOFF (ed.), *Hear Me Talkin' to Ya* (1955), is a brilliant use of the

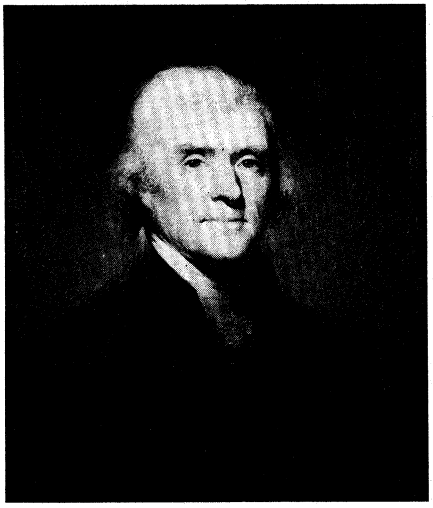
spoken word to build a picture of jazz history. The best attempts at more general critical analysis include FRANCIS NEWTON, *The Jazz Scene* (1959); BENNY GREEN, *The Reluctant Art* (1962); and HENRY PLEASANTS, *Serious Music, and All That Jazz* (1969). MARSHALL W. STEARNS, *The Story of Jazz* (1956), has the charm of modesty. WHITNEY BALLIETT, *The Sound of Surprise* (1959), is an interesting if impressionistic collection of short reviews, PAUL OLIVER, *Blues Fell This Morning* (1960), is an outstanding specialist work. LEONARD G. FEATHER, *Encyclopedia of Jazz,* rev. ed. (1960) and the *Encyclopedia of Jazz in the Sixties* (1967); and G.T. SIMON, *The Big Bands* (1967), are indispensable works of reference. PAUL OLIVER *et al., Jazz on Record* (1968), is useful but partisan. DONALD KENNINGTON, *The Literature of Jazz* (1970), is the first serious attempt to annotate all jazz literature.

(B.Gr.)

Jefferson, Thomas

Third president of the United States, principal author of the Declaration of Independence, and influential political philosopher, Thomas Jefferson was born on April 13, 1743, at Shadwell, in Albemarle County, Virginia, the son of Peter Jefferson, an early settler and leader in the county, and Jane Randolph Jefferson. Peter Jefferson was a surveyor and cartographer and was largely self-educated. Upon his death in 1757 he left his son considerable property, but the inheritance for which Thomas Jefferson expressed particular gratitude was his father's determination that he should have a sound classical education. After several years of study at local grammar and classical schools, Jefferson entered the College of William and Mary in 1760. In spite of his youth, he became a close friend of three leading residents of Williamsburg—William Small of the college faculty, George Wythe of the Virginia bar, and Francis Fauquier, lieutenant governor of the colony. These three older men gave Jefferson a taste for the pleasures of a society more urbane and sophisticated than that of rural Virginia, and Small and Wythe gave direction to his intellectual drive. Small introduced him to the natural sciences and to rational methods of inquiry; Wythe led him to see the study of law not as a narrow vocational preparation but as a means of understanding the history, culture, institutions, and morals of a people. After two years at the college, Jefferson studied law for five years under Wythe's direction and was admitted to the bar in 1767. In 1769 he entered the lower house of the colonial legislature, thus beginning a long career in politics that ended 40 years later with his retirement as president of the United States.

By courtesy of the White House Collection, Washington, D.C.



Jefferson, oil painting by Rembrandt Peale, 1800. In the White House Collection, Washington, D.C.

Author of the Declaration. When Jefferson entered the House of Burgesses. Virginia and the other colonies were already engaged in the long decade of opposition to British colonial policies that led eventually to revolution and independence. Jefferson joined with Patrick Henry and others who favoured strong resistance to George III and the British Parliament and soon became one of the leaders of this group. His political style was very different from that of Henry. He was assiduous in committee work, a skilled legal craftsman, a scholar who drew on his comprehensive knowledge of law and history to support the colonial case against Great Britain. He rarely made speeches, disliked oral dispute, whether in formal debate or informal conversation, and he recognized the necessity of consensus for effective political action; the pen was his natural means of expression, and he was a virtuoso in its use. His first major essay, "A Summary View of the Rights of British America" (1774), displayed an impressive array of learning and logic, demonstrated his capacity for intense passion and the ability to express it eloquently, and revealed an inclination to intellectual radicalism. The majority of his colleagues were not then prepared for his conclusion that the British Parliament had no authority at all to legislate for the colonies, but, as relations with Great Britain grew steadily worse, his arguments became increasingly acceptable and his language both persuasive and provocative. "The God who gave us life gave us liberty at the same time: the hand of force may destroy, but cannot disjoin them."

Delegate to the Second Continental Congress

In the spring of 1775 the Virginia legislature, sitting as a revolutionary convention in defiance of the royal governor, appointed Jefferson as a member of its delegation to the Second Continental Congress meeting in Philadelphia. There he joined with the more radical group in the Congress, and again his skills as a committeeman and stylist were recognized and used. In June of 1776, when the decision to break irrevocably with Great Britain seemed near, Jefferson was appointed to the committee assigned to draft a formal statement of the reasons for such a decision. Benjamin Franklin and John Adams were also on the committee, but they recognized the superior talent of the Virginian and gracefully bowed to it. Jefferson thus became the principal author of the Declaration of Independence. It was an official state paper, and in later life he stated that it was intended to be an expression of the American mind. That was no doubt true, but it is also true that his personal commitment to its principles was profound and intense. It was this commitment, not the mere fact of literal authorship, that rendered Jefferson uniquely symbolic of the ideals expressed in the Declaration.

Role in Virginia politics. Jefferson meant his revolutionary manifesto to be more than an eloquent justification of revolt against Great Britain. He intended to translate its principles into practice and to create in America a society in which the gap between aspiration and achievement would be narrowed. He had wanted to begin by taking part in framing the new constitution of Virginia, which was adopted in June of 1776, but his duties in Philadelphia made that impossible, and he did not enter the Virginia legislature until October. He then set in motion a plan for comprehensive reform of the laws and institutions of Virginia. Two parts of the plan show the thoroughness with which he had considered the nature of representative government and the conditions necessary to its successful operation. A third embodied his passionate commitment to intellectual freedom.

Jefferson sought and secured abolition of the laws of primogeniture and entail in Virginia in order to discourage concentration of property in the hands of a few great landowners. He believed that property was among the natural rights to which man was born and that it meant the right to a decent means of subsistence. After observing the economic conditions in France a few years later, he wrote:

> Whenever there is in any country, uncultivated lands and unemployed poor, it is clear that the laws of property have been so far extended as to violate natural right. The earth is given as a common stock for man to labour and live on. If for the encouragement of industry we allow it to be appropriated, we must take care that other employment be provided to those excluded from the appropriation. If we do not the fundamental right to labour the earth returns to the unemployed.

No society that denied this right could be just, nor was it likely to enjoy for long a republican government. Jefferson believed that the virtues required for that form of government could not flourish in conditions of extreme poverty or complete economic dependence.

The educational system proposed for Virginia was also a part of Jefferson's comprehensive plan for republican government. The lower schools would provide literacy for the entire population, which, combined with a free press, was necessary for an informed public opinion. The upper schools would develop a natural aristocracy to supply the leadership so essential to representative government, while scholarships awarded on the basis of merit would prevent identification of educational opportunity with economic privileges. Jefferson did not believe that an ignorant people could make rational and responsible decisions about public affairs, nor did he believe that men were equal in intelligence or that the operation of a government was a simple job easily mastered by the common man. He assumed that men of superior capabilities were those naturally suited for public office, and his scheme of education was intended to insure that such men, regardless of their economic circumstances, be given an opportunity to develop their talents. Jefferson's fellow Virginians were not prepared for so comprehensive a system of free public education, however, and the only part of it that he secured was the University of Virginia.

The statute of Virginia

The third and most famous reform, the statute of Virginia for religious freedom, met with bitter and persistent opposition and was not enacted until 1786, while Jefferson was in France. Although Americans had largely abandoned the gross forms of persecution common a few generations earlier, the toleration they practiced was limited and erratic. In some states, as in Virginia, a single church was established; others restricted public office to Protestants; some required belief in specific doctrines of the Christian religion, such as the divinity of Jesus, the Trinity, and immortality. The Virginia statute constituted a complete break with the traditional relationship between church and state. It prohibited support of any religion by public taxation and forbade all civil disabilities imposed on citizens because of religious belief or the lack of it. Jefferson regarded the statute as partial fulfillment of his celebrated vow: "I have sworn upon the altar of God eternal hostility against every form of tyranny over the mind of man."

After three years in the legislature, Jefferson was elected governor in 1779 and served for two years in a position characterized by much responsibility and very little power. When Virginia was invaded by British forces in the winter of 1780–81, Jefferson was unable to organize effective opposition and barely escaped capture when a detachment of troops raided Charlottesville and Monticello. His conduct during the emergency was criticized, and, although the legislature gave him a unanimous vote of confidence, he could not forget the slur cast upon his character as a public official. He refused to serve again either as governor or legislator and retired to Monticello determined to live out his life as a private citizen.

There was a reason other than wounded pride for this decision. He was worried about the health of his wife, Martha Wayles Skelton Jefferson. Since their marriage in 1772, she had borne him five children of whom only two survived, and in the fall of 1781 she was again pregnant. Jefferson's fears were justified, for she did not recover strength after the birth of the sixth child and died September 6, 1782. Jefferson's grief was incalculable.

"Notes on Virginia." After his retirement as governor and before he returned to public service in December of 1782, Jefferson wrote and revised the major portion of *Notes on Virginia*, his only book. It originated in a comprehensive but routine series of questions put to him by the secretary of the French legation in order to compile information about the new country. Jefferson's response was as revealing of himself as it was informative about the state of Virginia. In later years he learned to guard his pen carefully, especially after letters he considered to be purely private were printed in newspapers or elsewhere without his permission. The language of this book was frequently unrestrained. It was as if the *Notes*, written for the most part after his abrupt and unhappy withdrawal from Virginia politics and during the months of desperate fear for the life of his wife, provided a means for the release of otherwise restrained emotions.

Discussion of race and slavery in the *Notes*

The *Notes* include a discussion of slavery, its effects on both whites and blacks, and an attempt to delineate the racial characteristics of the latter. Although he was unalterably opposed to slavery and reiterated his reasons in this essay, he both expressed and reflected one of the principal obstacles to abolition—the belief that, because of inherent racial differences, blacks and whites could not live together in peace and harmony. Jefferson's summary of the supposed differences may now be seen as a classic example of the failure of an individual mind—and in this case one of exceptional independence and critical rigour—to transcend the cultural boundaries of its age. It is a curious blend of attempted objectivity flawed by the intrusion of unconscious prejudices and unexamined assumptions. He argued, among other points, that the blacks were inferior in physical beauty, that they might be lacking in foresight, that they were equal in memory but inferior in reason and imagination to the white race. He was aware of the influence of environment on behaviour and belief, accepted it as a general principle, and even cited it to explain the slave's alleged disposition to theft. Yet, he could not or did not apply it consistently and rigorously throughout his examination of the subject. It would appear that he clearly recognized the difficulties involved in applying the methods of scientific analysis to problems of racial 'characteristics, but they were difficulties beyond his power to resolve.

> The opinion that they are inferior in the faculties of reason and imagination, must be hazarded with great diffidence. To justify a general conclusion, requires many observations, even where the subject may be submitted to the anatomical knife, to optical glasses, to analysis by fire or by solvents. How much more then where it is a faculty, not a substance, we are examining; where it eludes the research of all the senses; where the conditions of its existence are various and variously combined; where the effects of those which are present or absent bid defiance to calculation; let me add too, as a circumstance of great tenderness, where our conclusion would degrade a whole race of men from the rank in the scale of beings which their Creator may perhaps have given them. To our reproach it must be said, that though for a century and half we have had under our eyes the races of black and of red men, they have never yet been viewed by us as subjects of natural history. I advance it, therefore, as a suspicion only, that the blacks, whether originally a distinct race, or made distinct by time and circumstances, are inferior to the whites in the endowments both of body and mind.

The *Notes* are otherwise interesting because they reveal the mind of a revolutionist in the midst of a revolution he regarded as unfinished. With some equanimity, he attributed the "very capital errors" in the Virginia constitution of 1776 to inexperience; it was with passionate outrage that he criticized proposals made twice in the Virginia legislature to follow Roman precedent and establish a temporary dictator in time of emergency:

> The very thought alone was treason against the people; was treason against mankind in general; as riveting forever the chains which bow down their necks, by giving to their oppresrors a proof which they would have trumpeted through the universe, of the imbecility of republican government, in times of pressing danger, to shield them from harm.

He urged revision of the constitution and enactment of his plans for universal education and full freedom of religion because he believed that the public virtue then prevalent among both the people and their leaders was impermanent, in part a function of the revolutionary situation, and destined to diminish. Rulers would become corrupt and abuse their power, and the people "will forget themselves, but in the sole faculty of making money, and will never think of uniting to effect a due respect for their rights." Jefferson's belief in republican government did not rest on naïve and unqualified faith in the people. Republican government would operate successfully only under certain conditions: a wide distribution of property or the availability of a substitute that provided men with a decent subsistence honestly earned; an educated and informed population; laws and institutions designed to compensate for the diminution of public virtue that Jefferson thought was sure to come when the crises of the revolutions were over.

Return to politics. In December 1782 he returned to public service and was for several months a member of the Virginia delegation to the Continental Congress. During this time Virginia ceded to the national government the area northwest of the Ohio River, which it claimed under grants made during the colonial period. In an ordinance drafted for the governance of this land, Jefferson set forth the principle that it should not be held by the original 13 states as colonial territory but should be divided into areas that, upon reaching a designated condition of population and organization, should enter the Union as states equal to the original 13. He also included a prohibition that would have forbidden slavery after 1800 in this territory and any others of which the United States might become possessed. The provision was defeated by one vote; a similar one had been incorporated in the Northwest Ordinance of 1787, but it applied only to that territory. Had Jefferson's original proposal been adopted. and had it remained in force, then slavery would have been outlawed in the whole area of the Louisiana Purchase. As he himself later commented,

> Thus we see the fate of millions of unborn hanging on the tongue of one man, and heaven was silent in that awful moment.

Visit to France

In 1784 Jefferson went to France to join Benjamin Franklin and John Adams in negotiating treaties with European powers. After a few months he succeeded Franklin as resident U.S. minister to the French government. His diplomatic duties were not onerous, and Paris offered him the intellectual and artistic society he had first glimpsed as a student at the College of William and Mary. There he could attend the theatre and opera, visit museums, keep up with science and inventions, associate freely with European scientists and intellectuals, share the politesse of French society, and indulge his passion for books. He loved France and the French, but not uncritically. His observations of economic and social conditions strengthened his aversion to absolute monarchy, and the contrast he saw between French and U.S. domestic morality led to a series of letters condemning the former and warning against the dangers of corruption should young men of his own country be sent to France for their education. (He did not want his daughters to marry abroad and so took them back to Virginia when the older was 17.) As author of the Declaration of Independence and of the Statute for Religious Freedom of Virginia, he had considerable influence with such moderate political leaders as the Marquis de Lafayette, and during the early stages of the French Revolution he was optimistic about the future of their efforts to effect gradual changes in the monarchy and its attendant laws and institutions. It was the greatest intellectual error of his life: France had almost none of the ingredients that had contributed to the success of the United States War of Independence, a fact Jefferson would surely have realized had he not allowed himself to indulge in wishful thinking. Jefferson observed only the opening stages of the Revolution, for he returned to the United States at the end of 1789.

Controversy with Hamilton. In the meantime, the Articles of Confederation had been replaced by the Constitution drafted in Philadelphia in 1787 and ratified the following year. Jefferson approved of most of that document but was critical of its lack of a bill of rights and its failure to impose limitations on the length of tenure for the presidency. Upon his return to Virginia in the fall of 1789, he was requested by George Washington to become secretary of state in the new government. With considerable reluctance, he accepted. Soon after he assumed the new office he became involved in controversy with Alexander Hamilton, who was secretary of the treasury. He opposed Hamilton's financial policies on the grounds that they exceeded the powers delegated to the central government by the Constitution, that they were contrary to the interests of the majority of the people, and that they represented a threat to republican institutions. Jefferson and Hamilton also disagreed on questions of foreign policy, with Jefferson at first leaning toward France and Hamilton toward Great Britain.

Role in the development of the two-party system

The issues between the two men were not purely personal; they extended to the country at large and led to the formation of national political parties based on policy and principle as well as personality. Thus was established the precedent and pattern of a national two-party system. Both Jefferson and Hamilton retired from the Cabinet near the end of Washington's first term, but each continued to be the symbol of the new parties, Jefferson of the Democratic-Republican, Hamilton of the Federalist. Both sides developed organizational skills among the electorate, the Congress, and the state legislatures, and both made effective use of the press. James Madison was, as usual, Jefferson's able collaborator and supplied active leadership of the party until the latter returned to the centre of national politics as vice president under John Adams in 1797. In 1798, when the United States was close to war with France, the Federalist-controlled Congress enacted the Alien and Sedition Acts. The latter, particularly as applied by Federalist judges, was used to stifle Democratic-Republican criticism of the government. Jefferson and Madison believed it to be contrary to the first amendment and therefore cons[illegible]al, a position th[illegible]d in the [illegible]inia [illegible]d Kentucky R[illegible]tions of 1798–99.

The decade ended with the defeat of the Federalists in the election of 1800. It was a critical period in the development of the new nation; politics were sharply divisive, conducted with extreme animosity, and permeated with fundamental cleavages over political principles. Jefferson regarded Hamilton as an enemy of republican government; Hamilton regarded Jefferson as a demagogic radical. Hamilton had a dream of national grandeur to which he was prepared to subordinate the immediate interests of the individual. Jefferson saw the purpose of government as the protection of the individual's right to life, liberty, and the pursuit of happiness. Jefferson's attitudes and behaviour during this period were revealing. He did not exercise an Olympian calm; his letters sometimes displayed anger and passion toward the policies of his opponents and toward some of them personally. Yet, the very bitterness of the controversy troubled him. He sensed and feared the divisive and destructive effects of unrestrained ideological conflict. Not only could the latter disrupt the social harmony that Jefferson valued so highly, but it could also conceivably rip the fabric of republican government altogether. A desire to forestall in America what had so frequently been the fate of such governments in the past seemed to influence Jefferson's conduct of the presidency during his first term.

Presidency. The Federalist candidates clearly lost the presidential election of 1800, but under the electoral system then prevailing neither of the Republican candidates, Jefferson and Aaron Burr, could claim victory. The Constitution had provided no means for electors to distinguish between their choices for president and vice president, and both candidates had received the same number of votes. The choice between them was therefore made in the House of Representatives. Partly because of the influence of Hamilton, who distrusted Burr even more than he disliked Jefferson, the latter was chosen president and inaugurated March 4, 1801.

The spirit and content of Jefferson's inaugural address were conciliatory, and so, to a considerable extent, were the policies of his first administration. There was no attempt at wholesale reversal of Federalist policies of the preceding 12 years, and in at least two instances—the Louisiana Purchase and the Embargo Act—he was said to be even more Federalist than the Federalists themselves. There was, however, an effort to nullify the Federalist attempt to fill the federal judiciary with partisan appointees holding office for life. and there was sufficient turnover in other federal offices to give some substance to the accusation that Jefferson introduced the spoils system.

But, in spite of the very bitter controversy of the preceding years, Jefferson's inauguration ushered in no drastic or radical changes. Had Jefferson been more doctrinaire or less aware of the danger of unrestrained political passion and of the delicate situation created by the first party change of administration in the new government's history, the future of U.S. politics might have been characterized by less stability than has been the case. The precedent he deliberately set must rank with the Louisiana Purchase as one of the major achievements of his presidency.

The Louisiana Purchase

The acquisition of the Louisiana Territory in 1803 was of incalculable importance, nearly doubling the size of the United States. Jefferson's original plan was to purchase merely a small area at the mouth of the Mississippi River. When Napoleon offered to sell the entire territory, Jefferson saw his chance and took it, even though, as he frankly admitted, he had no constitutional authority to do so. He believed that the purchase would contribute to the security of the United States by removing from the continent a major foreign power and that it would ensure the stability of republican government for generations to come by providing a vast reservoir of land for settlers.

Jefferson was re-elected in 1804; George Clinton replaced Burr as vice president. Jefferson's second administration was notable for his unsuccessful efforts to convict his former vice president, Burr, of treasonable acts in the southwestern territories, and for his efforts to pursue a policy of neutrality during the Napoleonic Wars and maintain the rights of neutrals on the high seas. His overwhelming desire to avoid war with either side led to charges of timidity and vacillation, and his Embargo Act (1807) was criticized as inconsistent with the principles of individual freedom and his former opposition to a strong national government. The act was securely based on the power given to the Congress to regulate commerce with foreign nations—a power of which Jefferson approved long before he became president—but the enforcement provisions of the act and its amendments can rightly be questioned as contravening the Fourth Amendment's prohibition of unreasonable search and seizure.

During Jefferson's presidency the power and prestige of the Supreme Court grew under the leadership of Chief Justice John Marshall. In the case *Marbury* v. *Madison* (1803), the court explicitly asserted the right and power of judicial review. Jefferson opposed the power of the court as the ultimate and exclusive interpreter of the Constitution and argued that such a power lodged in one department of the government whose members held office for life was irresponsible and therefore contrary to the principles of republican government.

Establishment of the University of Virginia

Jefferson might have been elected president for a third term but chose to follow Washington's example of withdrawing after two terms. On March 4, 1809, he turned the office over to his successor, James Madison, and went home to Monticello. There was one more official act he sought to accomplish, the establishment of the University of Virginia, to which he referred as "the last of my mortal cares, and the last service I can render my country." He designed the buildings and supervised their construction to the most minute detail; he gathered the faculty, planned the curriculum, and even selected the reading for some of the courses. He had never been able to persuade his fellow Virginians to support public education for elementary and secondary pupils, but the university was an appropriate conclusion to a political career remarkable for its-creativity as well as for its duration and success.

Jefferson's political career was undoubtedly impressive, but it was far from absorbing all of the energy, time, and talent of the man himself. He probably enjoyed politics more than he was willing to admit; it is also true that his often-expressed longing to retire to private life and pursue his other interests was very real. These interests were numerous and varied.

Personal and intellectual interests. He was an extraordinarily learned man, and the range of his knowledge and inquiry is scarcely credible in the modern age of specialization. He knew Latin, Greek, French, Spanish, Italian, and Anglo-Saxon and concerned himself with such questions as the difference between the ancient and modern pronunciation of Greek. At the age of 71 he tackled Plato's *Republic* in the original and found its author greatly overrated. He attempted an analysis of the New Testament in order to discover what Jesus really said as distinguished from what he was reported to have said. He enjoyed the study of mathematics and found its precision and certitude a welcome relief from the untidiness of politics and government. He was an ardent student of the natural sciences, carried on an extensive correspondence with such men as Joseph Priestley, and sometimes contributed time and money to progress in these fields. The discovery of fossil remains in various parts of the country fascinated him, and he tried to collect and classify as many as he could. He was much interested in the experiments with balloons and submarines then being made, and, while he was abroad, he sent back to his friends at home various mechanical and scientific gadgets produced in Europe, including a polygraph and phosphorus matches. His travel notes record impressions ranging from nearly ecstatic admiration of architectural monuments to sober economic analysis of the reasons for the differences in prosperity between regions producing white and red wine.

Interest in the natural sciences

He was an enthusiastic practitioner of scientific farming, conducted numerous experiments at Monticello, was always on the lookout for some new plant or seed that might contribute to the prosperity of the United States (once going so far as to smuggle a particular variety of rice across the Italian border), and kept meticulous meteorological records. His interest in architecture was intense and enduring, and his influence on the Neoclassical style in the United States was great.

The pursuit of these various interests concurrently with his political activities and the management of his estates (which included several thousand acres and at one time about 150 slaves) is remarkable. To this record of industry must be added the voluminous correspondence Jefferson maintained with extraordinary conscientiousness until very near his death. He could have accomplished so much only through rigorous self-discipline and an efficient organization of his time and activities. Yet, he was one of the most generous and approachable of men. Friends and strangers alike wrote to him for advice or came to Monticello when he was in residence. Relatives and guests filled Monticello to capacity—sometimes beds were made for as many as 50 people—and devoured his food as well as his time. For privacy he retreated several times a year to Poplar Forest, a second home built as a refuge in Bedford County.

Personal appearance and personality

Jefferson was 6 feet 2 inches in height, large boned, slim, erect, and sinewy. He had angular features, a ruddy complexion, sandy hair, and hazel-flecked gray eyes. His carriage was relaxed and somewhat awkward, and by 18th-century standards he seems to have been regarded as pleasant rather than handsome in appearance. He was sensitive and perceptive in personal relations, gracious and charming in manner (though sometimes cold upon first meeting strangers), and almost invariably even tempered. As a matter of both principle and inclination, he attempted to prevent political differences from creating personal ill will, and though he was subjected to malicious abuse during the political controversies in which he was involved, he endured it with relative equanimity and felt genuine animosity toward only a very few of his opponents and critics.

Because he was so central a figure, so widely known, so articulate, and so meticulous in preserving his letters and papers, it is possible to reconstruct a remarkably complete account of his career and his work. Yet, the man himself—the private man—remains elusive. There was a reserve of privacy that he kept inviolate. For example, no letters exchanged between him and his wife exist. Their marriage was, by contemporary accounts, an extraordinarily happy one, and it would therefore appear that Jefferson destroyed whatever letters once existed in order to keep their relationship forever private. Jefferson was, as his modern editor has suggested, ultimately a lonely man.

Ten days before his death, Jefferson replied to an invitation to join the residents of Washington, D.C., in celebrating the 50th anniversary of the proclamation of the Declaration of Independence. He could not attend because of illness, but he sent his best wishes, and, of the Declaration that was to be celebrated, he wrote:

> May it be to the world, what I believe it will be, (to some parts sooner, to others later, but finally to all,) the signal of arousing men to burst the chains under which monkish ignorance and superstition had persuaded them to bind themselves, and to assume the blessings and security of self government.

While Jefferson grew steadily weaker at Monticello, his old friend John Adams was nearing death in Massachusetts. It seems certain from the accounts of friends and relatives that both men wanted badly to live until the 50th anniversary of the day that symbolized the central endeavour and achievement of their lives. They succeeded. Jefferson died shortly before one o'clock on the afternoon of July 4, 1826; Adams died a few hours later, his last words said to have been, "Jefferson still survives." Jefferson was buried at Monticello. The epitaph that he had chosen was inscribed on his tombstone: "Here was buried Thomas Jefferson, author of the Declaration of American Independence, of the statute of Virginia for religious freedom, and father of the University of Virginia."

Deaths of Jefferson and Adams

BIBLIOGRAPHY. JULIAN P. BOYD (ed.), *The Papers of Thomas Jefferson,* 19 vol. to date (1950–74), covering the period up to March 10, 1791, is the definitive edition of Jefferson's papers. Its superlative editorial work includes extensive notes on the background, context, and significance of the documents printed, which include papers written to Jefferson as well as those written by him. Two other collections of Jefferson's writings can be used for the period not yet reached by the Boyd edition: ANDREW A. LIPSCOMB and ALBERT ELLERY BERGH, *The Writings of Thomas Jefferson,* 20 vol. (1905); PAUL LEICESTER FORD, *The Works of Thomas Jefferson,* 12 vol. (1904–05). The correspondence between Jefferson and John and Abigail Adams is conveniently reproduced in LESTER J. CAPPON, *The Adams-Jefferson Letters,* 2 vol. (1959). The letters are notable for their warmth and, especially those between the two men after 1812, for their discussions of the intellectual and political developments of the times. Much information about Jefferson's intellectual life may be found in E. MILLICENT SOWERBY (comp.), *Catalogue of the Library of Thomas Jefferson,* 5 vol. (1952), which contains a list of the books sold by Jefferson to the Congress in 1815, replacing the library burned by the British and forming the nucleus of the present Library of Congress. A convenient single-volume anthology of Jefferson's letters and papers is ADRIENNE KOCH and WILLIAM PEDEN (eds.), *The Life and Selected Writings of Jefferson*(1944).

The definitive biography for the present generation is that by DUMAS MALONE, *Jefferson and His Time* (1948–), of which five volumes had been completed by 1974, the last dealing with Jefferson's second term as president. A good popular biography is NATHAN SCHACHNER, *Thomas Jefferson: A Biography,* 2 vol. (1951). A comprehensive single-volume biography, unfortunately published without footnotes, is MERRILL D. PETERSON, *Thomas Jefferson and the New Nation* (1970). The 19th-century biography by HENRY S. RANDALL, *The Life of Thomas Jefferson,* 3 vol. (1858), is valuable because of Randall's extensive consultation with people then living who had known Jefferson personally. The influence of Jefferson in America is treated in MERRILL D. PETERSON, *The Jefferson Image in the American Mind* (1960). For more extended bibliographies, including treatments of topics in special fields, such as Jefferson's interest and achievements in architecture, education, language and literature, philosophy, science, agriculture, and civil liberties, consult the Peterson and Malone volumes.

(C.M.K.)

Jehovah's Witnesses

Since 1931 the adherents of the apocalyptic sect organized by Charles Taze Russell in the early 1870s have been known as Jehovah's Witnesses. Before that they had been called Russellites, Millennial Dawnists, and International Bible Students. Worldwide, by the 1970s, their numbers exceeded 2,130,000, of whom approximately 550,000 lived in the United States.

Animated by an urgent need to warn mankind of the impending Battle of Armageddon, the Witnesses have brought their message to some 210 countries and territories. They believe that the Watch Tower Bible and Tract Society, their legal agency and publishing arm, exemplifies the will of God and proclaims the truths of the Bible against the evil triumvirate of organized religion, the business world, and the state.

Publishing

From the earliest days of the movement the Witnesses have relied on the printed word to win converts and indoctrinate members. The output of their several printing plants is enormous. One of their doctrinal texts, *The Truth That Leads to Eternal Life,* has become the fourth largest best seller of all times, with 74,000,000 copies in print. The *Watchtower,* the society magazine, by the late 1970s reported a circulation of more than 10,000,000 copies in 79 languages; its companion periodical *Awake!* claimed almost the same circulation.

Each Jehovah's Witness is expected to spend at least 10 hours a month in home Bible studies and door-to-door preaching or distribution .of Watch Tower literature. These methods, though they may annoy some householders, have brought remarkable growth; in 1942 the sect had only 106.000 members.

The Witnesses also stand apart from civil society, refusing to vote, run for public office, serve in any armed forces, salute the flag, stand for the national anthem, or recite the pledge of allegiance.

Their religious stands have brought clashes with various governments, resulting in lawsuits, mob violence, imprisonment, torture, and death. At one time more than 6,000 Witnesses were inmates of Nazi concentration camps. Communist and Fascist states usually forbid Watch Tower activities. In the U.S. the society has taken 45 cases to the Supreme Court and has won significant victories for freedom of religion and speech. The Witnesses have been less successful in claiming exemptions as ministers from military service and in seeking to withhold blood transfusions from their children.

Members outside the U.S.

Although the largest single contingent of Witnesses is American, most Witnesses live outside the U.S. Some of the larger groups are those in Nigeria (114,000 in the late 1970s), Brazil (106,000), West Germany (102,000), Mexico (84,000). the British Isles (80,000), and the Philippines (77,000).

History. *Pastor Russell.* By the time he was 20, Russell (1852–1916) had left both Presbyterianism and Congregationalism because he could not reconcile the idea of an eternal hell with God's mercy. He had drifted into agnosticism when a chance encounter with some Adventists introduced him to the idea that the Bible could be used to predict God's plan of salvation, especially as the plan related to the end of the world. Lacking much formal education, Russell managed to master the use of Hebrew and Greek dictionaries to study the Bible. The scriptures fascinated him far more than did his haberdashery in Pittsburgh, and he left the business world to form the first Bible classes in 1872. In 1879 he began publishing a magazine, *Zion's Watch Tower and Herald of Christ's Presence.* His major exegetical work was the seven-volume *Studies in the Scriptures.*

Basing his judgment on complex biblical calculations, Russell predicted that the end of the world would occur in 1914. His movement survived the embarrassment caused by the failure of this prediction, as well as the problems caused by his separation from his wife and numerous lawsuits.

Judge Rutherford. A Missouri lawyer, Joseph Franklin Rutherford (1869–1942), who had occasionally filled in as a circuit judge succeeded Pastor Russell as head of the Watch Tower Society. In 1918 Rutherford and seven other senior officials of the sect were sentenced to 20-year terms in federal prison for sedition. They were freed after nine months, and the government eventually dropped its prosecution.

Judge Rutherford discarded some of Russell's beliefs, such as the notion that the measurements of the Great Pyramid of Egypt verified biblical predictions of the Second Advent. The judge's voluminous output of books and tracts, almost 100, supplanted Russell's writings, which were allowed to go out of print. Rutherford coined the slogan "Millions now living will never die!" which drew attention to the movement's promise that the faithful would survive the imminent battle between Jehovah and Satan. He equipped members with portable phonographs to play his "sermonettes" on the front porches and in the living rooms of prospective converts. The democratic polity devised by Pastor Russell was replaced by a theocratic system directed from the society's headquarters in Brooklyn, N.Y. Rutherford intensified the Witnesses' attacks on organized religion, especially Roman Catholicism.

Theocratic polity

Aloof and somewhat mysterious, Judge Rutherford seldom appeared in public except for annual Watch Tower conventions. It was at one of these in 1931 that he announced the new name for the movement.

The Witnesses since Rutherford. Nathan H. Knorr (1905–77) left the Reformed Church and joined the Watch Tower while in high school. He joined the sect's Brooklyn staff shortly after graduation and worked his way up the hierarchy to become president in 1942 at the age of 35. Knorr established the Watch Tower Bible School of Gilead to train missionaries and leaders, decreed that all the society's books and articles be published anonymously, and set up adult education programs to train Witnesses to deliver their own apologetical talks. Under Knorr's direction a group of Witnesses produced a new translation of the Bible. The first of six volumes of the *New World Translation of the Hebrew and Christian Greek Scriptures* appeared in 1950.

Witnesses faced active persecution in Germany and other Axis countries during World War II as well as in several Allied countries where their work was banned. In the postcolonial era, they have encountered hostility in a number of new African nations whose nationalism conflicts with the Watch Tower idea of theocracy.

When President Knorr died in 1977, the directors elected Frederick W. Franz to be the fourth leader of the Watch Tower Bible and Tract Society. Long the sect's most important theologian, Franz was the first head of the society to have attended college; he completed two years at the University of Cincinnati.

Legal Corporations and Organization. Of the three corporations that direct the activities of Jehovah's Witnesses. the most powerful is the Watch Tower Bible and Tract Society of Pennsylvania; founded by Russell in 1884, it has several hundred members, mostly Americans. The other two corporations are the Watch Tower Bible and Tract Society, Inc., of New York and the International Bible Students Association.

Most members of a local congregation are kingdom publishers, who are expected to spend five hours a week at meetings in Kingdom Hall and spend as much time as circumstances permit in doorstep preaching. Pioneer publishers hold part-time secular jobs and try to devote 100 hours a month to religious service. Special pioneers are full-time, salaried employees of the society who should spend at least 150 hours a month in this work. Each Kingdom Hall has an assigned territory and each Witness a particular neighbourhood to canvass. The sect takes great pains to keep records of the number of visits, back calls, Bible classes, and books and magazines distributed.

Kingdom publishers

The elders of the congregation select a presiding minister to serve a one-year term. He is aided by an assistant presiding minister and ministerial assistants as well as by the elders. Only men are qualified to hold teaching or administrative positions.

About 3,000 Witnesses live in various Bethel residences around the world. Of these, some 700 live and work at the Brooklyn Bethel; they write the books, set the type, operate the presses, and conduct the business of the society. Others raise food on Watch Tower farms. All residents, from the president to the newest recruit, receive the same recompense: room, board, and $20 a month.

Doctrines. *Christology.* Jehovah's Witnesses believe in a God whose personal name is Jehovah. They reject the orthodox Christian doctrine of the Trinity as pagan idolatry and hold theological views similar to those of Arius, who was condemned as a heretic by the Council of Nicaea in AD 325.

Jehovah created his Son, who was known as Michael the Archangel before his incarnation. The Son took on human form and as Jesus Christ gave his life as ransom for mankind, which had lost the right to eternal life through Adam's sin. Jesus Christ, raised as a mighty spirit creature, now rules as Jehovah's executive officer in heaven. The Holy Spirit is understood not as a separate personage in the Trinity but as the name for the power of Jehovah as he intervenes in history.

The Second Coming. Rutherford explained that Russell had been correct about 1914; the Kingdom of God did begin then, but it was an invisible event. Witnesses believe that Armageddon may break out at any moment but certainly during the lifetime of the generation alive in 1914.

Armageddon

The Last Judgment. Since the time of Christ, a select band of 144,000 persons has been gathered and will reign with Jesus in heaven. About 10,000 of this number, all of whom are Jehovah's Witnesses, are still alive. Other Witnesses will be able to survive Armageddon and the experience of death and they, along with resurrected Witnesses, will thereafter populate a paradise on earth. After the victory of Jehovah, the world will no longer see sickness, old age, crime, war, or poverty. After 1,000 years, Satan will be loosed for one more chance to seduce mankind; a few weak persons will succumb to his temptations but most will remain faithful and continue an idyllic life on earth forever. The wicked will be annihilated.

The Bible. The Witnesses claim to base all their teachings on the Bible, which they accept as literally true, and their publications are laced with scriptural references. The Watch Tower cites the Bible to condemn Christmas trees, hypnosis, hunting and fishing for sport, military service, the pope, transcendental meditation, the United Nations, birthday celebrations, gambling, abortion, Protestantism, homosexuality, evolution, tobacco, communism, and lodges. Their refusal to accept blood transfusions is based on their interpretations of Gen. 9:4, Lev. 7:26, Acts 15:29, etc.

Church and sacraments. The weekly meetings in Kingdom Halls resemble Bible study classes more than a Protestant worship service. Rather than sermons and prayers, these meetings feature lectures, reports, and discussions of Watch Tower literature. Once a year, on Nisan 14 (the eve of Passover), the society sponsors a communion service called the Lord's Evening Meal (Memorial), to which all members and sympathizers are invited but at which only the living remnant of the 144,000 actually partake of the bread and wine. In 1976 nearly 5,000,000 persons attended, and 10,187 partook of the emblems.

Lord's Evening Meal

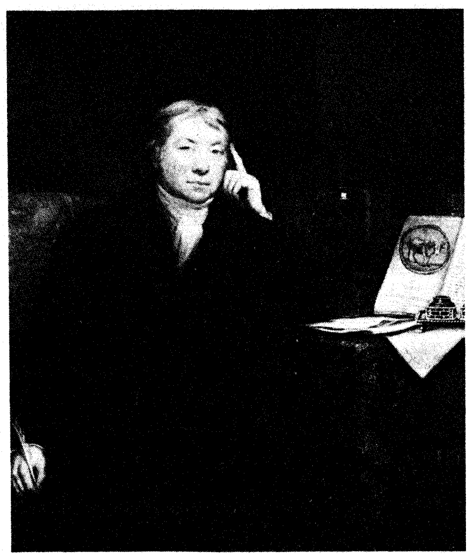
Jehovah's Witnesses rallies often feature mass Baptism. The society views Baptism by immersion as a symbol of dedication to Jehovah. Submission to Baptism not only initiates a convert into the sect but qualifies him or her to be a minister.

BIBLIOGRAPHY. JAMES A BECKFORD, *The Trumpet of Prophecy* (1975), a sociological study with additional information on the history and theology of the movement by an English scholar; GÉRARD HÉBERT, S.J., *Les Témoins de Jéhovah* (1960), theological critique by a French Canadian Jesuit; *Jehovah's Witnesses in the Divine Purpose* (1959), official history from the founding to 1959; W.C. STEVENSON, *The Inside Story of Jehovah's Witnesses* (1968), critical study by a former member of the Watch Tower Society in England; H.H. STROUP, *The Jehovah's Witnesses* (1945, reprinted 1967), objective study of particular value for the Rutherford period; WILLIAM J. WHALEN, *Armageddon Around the Corner* (1962), a study of the history, beliefs, and practices.

(W.J.WH.)

Jenner, Edward

Edward Jenner, in discovering the process of vaccination, did more than any other person to wipe out the menace of smallpox. In addition, his scientific acumen, personality,



Jenner, oil painting by J. Northcote, 1803. In the National Portrait Gallery, London.
By courtesy of the National Portrait Gallery, London

and his relations to society exemplified the best of his era.

Jenner was born on May 17, 1749, in Berkeley, Gloucestershire, at a time when the patterns of British medical practice and education were undergoing gradual change. Slowly the division between the Oxford- or Cambridge-trained physicians and the apothecaries or surgeons—who were much less educated and who acquired their medical knowledge through apprenticeship rather than through academic work—was becoming less sharp and hospital work was becoming much more important.

Jenner was a country youth, the son of a clergyman who died when the boy was only five. Brought up in a country environment, largely by an older brother who was also a clergyman, Edward acquired a love of nature that remained with him all his life. He attended grammar school and at the age of 13 was apprenticed to a nearby surgeon. In the following eight years Jenner acquired a sound knowledge of medical and surgical practice. On completing his apprenticeship at the age of 21, he went to London and became the house pupil of John Hunter, who was on the staff of St. George's Hospital and was soon to become one of the most prominent surgeons in London. Even more important, however, he was an anatomist, biologist, and experimentalist of the first rank; not only did he collect biological specimens but he also concerned himself with problems of physiology and function.

Influence of John Hunter

The firm friendship that grew between the two men lasted until the death of Hunter in 1793. From no one else could Jenner have received the stimuli that so confirmed his natural bent—a catholic interest in biological phenomena, disciplined powers of observation, sharpening of critical faculties, and a reliance on experimental investigation. From Hunter Jenner received the characteristic advice: "Why think, [*i.e.*, speculate]—why not try the experiment?"

In addition to his training and experience in biology, Jenner made progress in clinical surgery. At the end of two years in London, he returned, in 1773, to country practice in Berkeley and enjoyed substantial success. He was capable, skillful, and personally papular. In addition to practicing medicine, he joined two medical groups for the promotion of medical knowledge and wrote occasional medical papers. He played the violin in a musical club, wrote light verse, and, as a naturalist, made many observations, particularly on the nesting habits of the cuckoo and on bird migration. He also collected specimens for John Hunter; many of Hunter's letters to Jenner have been preserved, but Jenner's letters to Hunter have unfortunately been lost.

After one disappointment in love in 1778. Jenner married in 1788.

Smallpox was widespread in the 18th century, and occasional outbreaks of special intensity resulted in a very high death rate. The disease, a leading cause of death at the time, respected no social class. And disfigurement was not uncommon in patients who recovered. The only means of combatting the disease was the so-called inoculation—intentionally infecting a healthy person with the "matter" taken from a patient sick with a mild attack of smallpox. This practice, which originated in the East, depended on two distinct concepts: first, that one attack of smallpox effectively protected against any subsequent exposure; and second, that a person deliberately infected with a mild case of the disease would safely acquire such protection. It was, in present-day terminology, an "elective" infection—*i.e.*, one given to a person in good health. Unfortunately, the transmitted disease did not always remain mild, and significant mortality occurred. Furthermore, the inoculated person could disseminate the disease to others and thus act as a focus of infection.

Jenner, even as an apprentice, had been impressed by the fact that a person who had suffered an attack of cowpox—a relatively harmless disease that could be contracted from cattle--could not take the smallpox—*i.e.*, could not become infected whether by accidental or intentional exposure to the smallpox. Pondering this phenomenon Jenner concluded that cowpox not only protected against smallpox but could be transmitted from one person to another as a deliberate mechanism of protection.

Vaccination for smallpox

The story of the great breakthrough is well-known. In May 1796 Jenner found a young dairymaid, Sarah Nelmes, who had fresh cowpox lesions on her finger. On May 14, using matter from Sarah's lesions, he inoculated an eight-year-old boy, James Phipps, who promptly developed a slight fever and a low grade lesion. On July 1 Jenner inoculated the boy again, this time with smallpox matter. No disease developed; protection was complete. Jenner, in 1797, sent to the Royal Society a short paper describing his results, but the paper was refused. Then Jenner, in 1798, having added further cases, published privately a slender book entitled *An* Inquiry into the Causes and Effects of the Variolae Vaccinae, a Disease . . . Known by the Name of Cow Pox.

The reaction to the publication was not immediately favourable. Jenner went to London seeking volunteers for vaccination but, in a stay of three months, was not successful. In London inoculation became popularized through the activities of others, particularly the surgeon Henry Cline, to whom Jenner had given some of the inoculant, and doctors George Pearson and William Woodville. Difficulties arose, some of them quite unpleasant: Pearson tried to take credit away from Jenner; and Woodville, a physician in a smallpox hospital, contaminated the cowpox matter with smallpox virus.

But the procedure rapidly proved its value, and Jenner became intensely active, promoting the cause of vaccination. The procedure spread rapidly to the Continent, then to America, and soon was carried around the world.

Complications were many. Vaccination seemed simple, but the vast number of persons who practiced it did not necessarily follow the procedure that Jenner had recommended, and deliberate or unconscious innovations often impaired the effectiveness. Pure cowpox vaccine was not always easy to obtain, nor was it easy to preserve or transmit. Furthermore, the biological factors that produce immunity were not yet understood; much information had to be gathered, and a great many mistakes made before a fully effective procedure could be developed, even on an empirical basis.

Despite errors and occasional chicanery, the process of vaccination spread rapidly and the death rate from smallpox plunged. Jenner, although he received worldwide recognition and many honours, made no attempt to enrich himself through his discovery and actually devoted so much time to the cause of vaccination that his private practice and his personal affairs suffered severely. In 1802 Parliament voted him a sum of £10,000 and in 1806 a further sum of 520,000.

Jenner not only received honours but also aroused opposition and found himself subjected to attacks and calumnies. His calm, nevertheless, remained unruffled, and he continued his activities in behalf of vaccination.

His wife, ill with tuberculosis, died in 1815, and Jenner retired from public life. He died in Berkeley, January 26, 1823, one of the truly great benefactors of humanity.

BIBLIOGRAPHY. JOHN BARON, ***The Life of Edward Jenner***, 2 vol. (1827), the first biography, is the basis for all subsequent studies. W.R. LE FANU, ***A Bio-Bibliography of Edward Jenner, 1749–1823*** (1951), is the definitive study of his writings and is indispensable. For a contemporary view of smallpox and vaccination, see JAMES MOORE, ***The History and Practice of Vaccination*** (1817), as well as his ***History of the Smallpox*** (1815). There are sound and popularly written biographies: F.G.D. DREWITT, ***The Life of Edward Jenner, M.D., F.R.S.***, 2nd ed. (1933); and DOROTHY M. FISK, ***Jenner of Berkeley*** (1959). The best single short article is by E. ASHWORTH UNDERWOOD, "Edward Jenner: The Man and His Work," ***Brit. Med. J.***, 1:881–884 (1949). For the general medical background of the 18th century, see LESTER S. KING, ***The Medical World of the Eighteenth Century*** (1958, reprinted 1971). On the subject of inoculation, the definitive work is GENEVIEVE MILLER, ***The Adoption of Inoculation for Smallpox in England and France*** (1957).

(L.S.K.)

Jeremiah

Jeremiah was a major prophet of ancient Judah who lived in the latter part of the 7th and the early years of the 6th centuries BCE (before the Common Era, or BC). Closely involved in the political and religious events of a crucial era in the history of the ancient Near East, he was able to help his fellow countrymen survive disasters that included the capture of Jerusalem by the Babylonians in 586 BCE, the exile of many Judaeans to Babylonia, and Judah's loss of political independence.

Alinari

Jeremiah, detail from a fresco by Michelangelo in the Sistine Chapel, Rome, *c.* 1512.

The man and his times. Jeremiah was born (probably after 650) and grew up in the village of Anathoth, two miles northeast of Jerusalem, in a priestly family. In his childhood he must have learned some of the traditions of his people, particularly the prophecies of Hosea, whose influence can be seen in his early messages.

The era in which Jeremiah lived was one of transition for the ancient Near East. The Assyrian Empire, which had been dominant for two centuries, declined and fell. Its capital, Nineveh, was captured in 612 by the Babylonians and Medes. Egypt had a brief period of resurgence under the 26th dynasty (664–525) but did not prove strong enough to establish an empire. The new world power was the Neo-Babylonian Empire, ruled by a Chaldean dynasty whose best known king was Nebuchadrezzar. The small and comparatively insignificant state of Judah had been a vassal of Assyria and, when Assyria declined, asserted its independence for a short time. Subsequently, Judah vacillated in its allegiance between Babylonia and Egypt and ultimately became a province of the Neo-Babylonian Empire.

Beginning of prophetic career

According to the biblical Book of Jeremiah, he began his prophetic career in 627/626—the 13th year of King Josiah's reign. It is told there that he responded to Yahweh's (God's) call to prophesy by protesting "I do not know how to speak, for I am only a youth," but he received Yahweh's assurance that he would put his own words into Jeremiah's mouth and make him a "prophet to the nations." A few scholars believe that after his call Jeremiah served as an official prophet in the Temple, but most believe that this is unlikely in view of his sharp criticism of priests, prophets, and the Temple cult.

Jeremiah's early messages to the people were condemnations of them for their false worship and social injustice, with summons to repentance. He proclaimed the coming of a foe from the north, symbolized by a boiling pot facing from the north in one of his visions, that would cause great destruction. This foe has often been identified with the Scythians, nomads from southern Russia who supposedly descended into western Asia in the 7th century and attacked Palestine. Some scholars have identified the northern foe with the Medes, the Assyrians, or the Chaldeans (Babylonians); others have interpreted his message as vague eschatological (last days and Judgment) predictions, not concerning a specific people.

In 621 King Josiah instituted far-reaching reforms based upon a book discovered in the Temple of Jerusalem in the course of building repairs, which was probably Deuteronomy or some part of it. Josiah's reforms included the purification of worship from pagan practices, the centralization of all sacrificial rites in the Temple of Jerusalem, and perhaps an effort to establish social justice following principles of earlier prophets (this program constituted what has been called "the Deuteronomic reforms"). Jeremiah's attitude toward these reforms is difficult to assess. Clearly, he would have found much in them with which to agree; a passage in chapter 11 of Jeremiah, in which he is called by Yahweh to urge adherence to the ancient Covenant upon "the men of Judah and the inhabitants of Jerusalem," is frequently interpreted as indicating that the prophet travelled around Jerusalem and the villages of Judah exhorting the people to follow the reforms. If this was the case, Jeremiah later became disillusioned with the reforms because they dealt too largely with the externals of religion and not with the inner spirit and ethical conduct of the people. He may have relapsed into a period of silence for several years because of the indifferent success of the reforms and the failure of his prophecies concerning the foe from the north to materialize.

Jeremiah and the Deuteronomic reforms

Some scholars doubt that Jeremiah's career actually began as early as 627/626 BC and question the accuracy of the dates in the biblical account. This view arises from the difficulty of identifying the foe from the north, which seems likely to have been the Babylonians of a later time, as well as the difficulty of determining the prophet's attitude toward the Deuteronomic reforms and of assigning messages of Jeremiah to the reign of Josiah. In the opinion of such scholars, Jeremiah began to prophesy toward the end of the reign of Josiah or at the beginning of the reign of Jehoiakim (609–598).

Early in the reign of Jehoiakim, Jeremiah delivered his famous "Temple sermon," of which there are two versions, one in Jeremiah, chapter 7, verses 1 to 15, the other in chapter 26, verses 1 to 24. He denounced the people for their dependence on the Temple for security and called on them to effect genuine ethical reform. He predicted that God would destroy the Temple of Jerusalem, as he had earlier destroyed that of Shiloh, if they continued in their present path. Jeremiah was immediately arrested and tried on a capital charge. He was acquitted but may have been forbidden to preach again in the Temple.

The reign of Jehoiakim was an active and difficult period in Jeremiah's life. That king was very different from his father, the reforming Josiah, whom Jeremiah commended for doing justice and righteousness. Jeremiah denounced Jehoiakim harshly for his selfishness, materialism, and practice of social injustice.

The scroll dictated to Baruch

Near the time of the Battle of Carchemish. in 605, when the Babylonians decisively defeated the Egyptians and the remnant of the Assyrians, Jeremiah delivered an oracle against Egypt. Realizing that this battle made a great difference in the world situation, Jeremiah soon dictated to his scribe, Baruch, a scroll containing all of the messages he had delivered to this time. The scroll was read by Baruch in the Temple. Subsequently it was read before King Jehoiakim, who cut it into pieces and burned it. Jeremiah went into hiding and dictated another scroll, with additions.

When Jehoiakim withheld tribute from the Babylonians (about 601), Jeremiah began to warn the Judaeans that they would be destroyed at the hands of those who had previously been their friends. When the King persisted in resisting Babylonia, Nebuchadrezzar sent an army to besiege Jerusalem. King Jehoiakim died before the siege began and was succeeded by his son, Jehoiachin, who surrendered the capital to the Babylonians on March 16, 597, and was taken to Babylonia with many of his subjects.

The Babylonians placed on the throne of Judah a king favourable to them, Zedekiah (597–586 BCE), who was more inclined to follow Jeremiah's counsel than Jehoiakim had been but was weak and vacillating and whose court was torn by conflict between pro-Babylonian and pro-Egyptian parties. After paying Babylonia tribute for nearly ten years, the King made an alliance with Egypt. A second time Nebuchadrezzar sent an army to Jerusalem, which he captured in August 586.

Jeremiah's advocacy of submission to Babylon

Early in Zedekiah's reign, Jeremiah wrote a letter to the exiles in Babylonia, advising them not to expect to return immediately to their homeland, as false prophets were encouraging them to believe, but to settle peaceably in their place of exile and seek the welfare of their captors. When emissaries from surrounding states came to Judah in 594 to enlist Judah's support in rebellion against Babylonia, Jeremiah put a yoke upon his neck and went around proclaiming that Judah and the surrounding states should submit to the yoke of Babylonia, for it was Yahweh who had given them into the hand of the King of Babylonia. Even to the time of the fall of Jerusalem, Jeremiah's message remained the same: submit to the yoke of Babylonia.

When the siege of Jerusalem was temporarily lifted at the approach of an Egyptian force, Jeremiah started to leave Jerusalem to go to the land of the tribe of Benjamin. He was arrested on a charge of desertion and placed in prison. Subsequently he was placed in an abandoned cistern, where he would have died had it not been for the prompt action of an Ethiopian eunuch, Ebed-melech, who rescued the prophet with the King's permission and put him in a less confining place. King Zedekiah summoned him from prison twice for secret interviews, and both times Jeremiah advised him to surrender to Babylonia.

When Jerusalem finally fell, Jeremiah was released from prison by the Babylonians and offered safe conduct to Babylonia, but he preferred to remain with his own people. So he was entrusted to Gedaliah, a Judaean from a prominent family whom the Babylonians appointed as governor of the province of Judah. The prophet continued to oppose those who wanted to rebel against Babylonia and promised the people a bright and joyful future.

After Gedaliah was assassinated, Jeremiah was taken against his will to Egypt by some of the Jews who feared reprisal from the Babylonians. Even in Egypt he continued to rebuke his fellow exiles. Jeremiah probably died in about 570 BCE. According to a tradition that is preserved in extrabiblical sources, he was stoned to death by his exasperated fellow countrymen in Egypt.

Personality

Prophetic vocation and message. This sketch of Jeremiah's life portrays him as a courageous and persistent prophet who often had to endure physical suffering for his fidelity to the prophetic call. He also suffered inner doubts and conflicts, as his own words reveal, especially those passages that are usually called his "confessions" (Jer. 11:18–12:6; 15:10–21; 17:9–10, 14–18; 18:18–23; 20:7–12, 14–18). They reveal a strong conflict between Jeremiah's natural inclinations and his deep sense of vocation to deliver Yahweh's message to the people. Jeremiah was by nature sensitive, introspective, and perhaps shy. He was denied participation in the ordinary joys and sorrows of his fellowmen and did not marry. He thus could say, "I sat alone," with God's hand upon him. Jeremiah had periods of despondency when he expressed the wish that he had never been born or that he might run away and live alone in the desert. He reached the point of calling God "a deceitful brook, . . . waters that fail" and even accused God of deceiving and overpowering him. Yet there were times of exaltation when he could say to God: "Thy words became to me a joy and the delight of my heart"; and he could speak of Yahweh as "a dread warrior" fighting by his side.

As a prophet Jeremiah pronounced God's judgment upon the people of his time for their wickedness. He was concerned especially with false and insincere worship and failure to trust Yahweh in national affairs. He denounced social injustices but not so much as some previous prophets, such as Amos and Micah. He found the source of sin to be in the weakness and corruption of the hearts of men —in what he often called "the stubbornness of the evil heart." He considered sin to be unnatural; he emphasized that some foreign nations were more loyal to their pagan (false) deities than Judah was to Yahweh (the real God), and he often contrasted nature's obedience to law with man's disobedience to God.

Jeremiah had more to say about repentance than any other prophet. He called upon men to turn away from their wicked ways and dependence upon idols and false gods and return to their early covenantal loyalty to Yahweh. Repentance thus had a strong ethical colouring, since it meant living in obedience to Yahweh's will for the individual and the nation.

In the latter part of his career Jeremiah had to struggle against the despair of his people and give them hope for the future. He expressed his own hope vividly by an action that he undertook when the Babylonians were besieging Jerusalem and he was in prison. He bought, from a cousin, a field in Anathoth, his native town. In the presence of witnesses he weighed out the money and made the contracts and said, "Thus says the Lord of hosts, the God of Israel: Houses and fields and vineyards shall again be bought in this land." In this and other ways he expressed his hope for a bright future for Israel in its own land.

The New Covenant

Jeremiah's most important prophecy concerning the future is one regarding the New Covenant (Jer. 31:31–34). While the present literary form of the passage is probably not Jeremiah's, the thought is essentially his. He prophesied of a time when Yahweh would make a covenant with Israel, superseding the old Mosaic Covenant; Yahweh would write his law upon the hearts of men (rather than on tables of stone), and all would know God directly and receive his forgiveness. This New Covenant prophecy was very influential in New Testament times. It is quoted in the Letter to the Hebrews and lies behind words attributed to Jesus at the Last Supper: "This cup is the new covenant in my blood."

Jeremiah is sometimes called "the father of personal religion." But this is true only in part, for there was religious individualism before his time, and Jeremiah himself did not abandon the view that Israel as a nation was chosen by God. Yet he did teach that genuine religion was not dependent upon worship at the Temple of Jerusalem or residence in the land of Palestine or a physical rite, such as circumcision, but rather upon sincere personal seeking of God. His own life encouraged the ideal of individual religion, and his encounter with God found him struggling with God in a very personal way.

BIBLIOGRAPHY. J.P. HYATT, ***Jeremiah, Prophet of Courage and Hope*** (1958), a relatively brief survey of the prophet's life, personality, and message; JOHN SKINNER, ***Prophecy and Religion: Studies in the Life of Jeremiah*** (1922), an older book that is still one of the best interpretations of the prophet; S.H. BLANK, ***Jeremiah: Man and Prophet*** (1961), an excellent study, especially of Jeremiah's personality and his prayer life; J. MUILENBURG, "Jeremiah the Prophet," in the *Interpreter's Dictionary of the Bible,* vol. 2, pp. 823–835 (1962), a perceptive study of the prophet and his book; JOHN BRIGHT, ***Jeremiah: Introduction, Translation, and Notes*** (1965), a good full-scale commentary on the Book of Jeremiah.

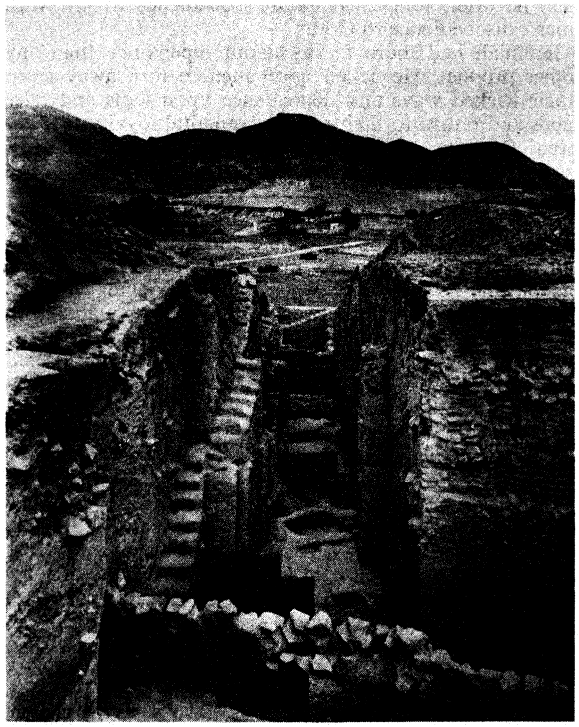
(J.P.Hy.)

Jericho

Jericho (Arabic Arīḥā) is a town in Jordan situated 825 feet below sea level on the west side of the Jordan Valley, six miles north of the Dead Sea. In 1967 it had a population of about 6,829. Jericho's great importance is that it provides archaeological evidence of one of the earliest known continuous settlements in the world and therefore of the first steps toward civilization, dating perhaps to the 10th millennium BC. It is famous in biblical history as the first town attacked by the Israelites under Joshua after they crossed the Jordan (Josh. 6).

After its destruction by the Israelites it was, according to the biblical account, abandoned until Hiel the Bethelite established himself there in the 9th century BC (I Kings 16:34). Jericho is mentioned several other times in the Old and New testaments. Herod the Great established a winter residence there, and he died there in 4 BC. The site of the Roman and New Testament Jericho is about one mile south of that of the Old Testament town.

By courtesy of the Jericho Excavation Fund



A section of the mound of Jericho. Excavations have uncovered the last wall of the Middle Bronze Age, c. 1700 BC, down to the earliest remains of the Neolithic period, c. 7000 BC.

Old Testament Jericho has been identified in the mound known as Tall as-Sulṭān (at the source of the copious spring ʿAyn as-Sulṭān), which rises 70 feet above the surrounding plain. A number of major archaeological expeditions have worked at the site, the most recent in 1952–58 under Kathleen M. Kenyon, director of the British School of Archaeology in Jerusalem; one of the main objectives has been to establish the date of the destruction of the town by the Israelites—a matter of obvious importance for the chronology of the Israelite entry into Palestine. Evidence of this destruction was thought to have been found but proved to be an erroneous identification. Most of the town of the period, including the whole circuit of the town walls, has been removed by erosion; enough survives to show that there was a town of the period, but that is all. This may have been destroyed in the second half of the 14th century BC, but the evidence is too scanty for precision. The site was then abandoned until the Iron Age. Little trace has been found of the 9th-century-BC occupation attributed to Hiel, but there was a considerable settlement in the 7th century BC, ending perhaps at the time of the second Babylonian Exile in 586 BC. The site was then finally abandoned, and the later Jerichos grew up elsewhere.

Excavations have shown, however, that Jericho had a very long history before the biblical period, and the great importance of the site is that it provides evidence of the first ,development of permanent settlements and therefore of the first steps toward civilization.

Pre-biblical Jericho

Traces have been found of visits of Mesolithic hunters, dated by carbon-14 to c. 9000 BC, and of a long period of settlement by their descendants, in which flimsy huts were the only habitations. From these huts solid houses, still with the curvilinear plan of the huts, were developed, and the settlement spread to cover an area of about ten acres. By c. 8000 BC the inhabitants had developed into an organized community capable of building a massive stone wall around the settlement, strengthened at one point at least by a massive stone tower. The size of this walled settlement justifies the use of the term town and suggests a population of some 2,000–3,000 persons. Thus, in this 1,000 years there had been a development from a hunting way of life to full settlement. The development of agriculture can be inferred from this, and grains of cultivated types of wheat and barley have been found. Jericho is thus one of the places providing evidence of the development of agriculture, and it is in fact the earliest so far known. It is highly probable that, to provide enough land for cultivation, irrigation had been invented. This first Neolithic culture of Palestine was a purely indigenous development, traceable from the Mesolithic cave dwellers of Mt. Carmel to the town of Jericho. These indigenous Neolithic occupants were succeeded c. 7000 BC by a second group, bringing a culture that was still Neolithic and still not manufacturing pottery, though it was not indigenous. This occupation probably indicates the arrival of newcomers from one of the other centres, possibly in northern Syria, in which the Neolithic way of life based on agriculture had developed. Their domestic architecture was more elaborate, with many-roomed houses, rectilinear in plan. The most remarkable achievement associated with them is a number of portrait heads, modelled in plaster over human skulls.

The first pottery users

This second Neolithic stage ended c. 6000 BC. For the next 1,000 years there is little evidence of occupation at Jericho. Only about 5000 BC did Jericho show the influences of developments that had been taking place in the north, where an ever-increasing number of villages had appeared, still Neolithic but marked by the use of pottery. The first pottery users of Jericho were, however, primitive compared with their predecessors on the site, living in simple huts sunk in the ground. They were probably mainly pastoralists. Over the next 2,000 years, occupation was sparse and possibly intermittent.

At the end of the 4th millennium BC, an urban culture once more appeared in Jericho, as in the rest of Palestine. Jericho became a walled town again, with its walls many times rebuilt.

About 2300 BC there was once more a break in urban life. The nomadic newcomers, consisting of a number of different groups, were probably the Amorites. Their successors, c. 1900 BC, were the Canaanites, sharing a culture found the whole length of the Mediterranean littoral. The Canaanites reintroduced town life, and excavations have provided evidence both of their houses and of their domestic furniture, which was found in their tombs as equipment of the dead in the afterlife. These discoveries have indicated the nature of the culture the Israelites found when they infiltrated into Palestine, and which they largely adopted.

Excavations in 1950–51 revealed something of Herodian Jericho. A magnificent facade on Wādī al-Qilt, a mile south of Old Testament Jericho, is probably part of Herod's palace; and its style, completely Italian, illustrates Herod's devotion to Rome. Traces of other fine buildings can be seen in this area, which became the centre of Roman Jericho.

Jericho of the Crusader period was on yet a third site, a mile east of the Old Testament site, and here the modern town grew up. An insignificant village in Turkish times, it became a winter resort after the British Mandate (1922). Its major expansion, however, came after its incorporation into Jordan in 1949. The establishment in the neighbourhood of two enormous camps of Arab refugees from Israel brought great activity to the town, which was largely rebuilt; the area of the oasis was expanded by irrigation.

The Arab-Israeli War of June 1967 resulted in the dispersal of much of the population.

BIBLIOGRAPHY. JOHN and J.B.E. GARSTANG, *The Story of Jericho*, 2nd ed. (1948), gives a popular account of the excavations carried out in 1930–36. The ascription of the burned and collapsed walls of the so-called City IV to the period of Joshua has now been disproved. KATHLEEN M. KENYON, *Digging up Jericho* (1957), gives a popular account of the excavations begun in 1952 up to 1956; the finds of the 1957–58 season, including those of the earliest remains, are included in KATHLEEN M. KENYON, *Archaeology in the Holy Land*, 3rd ed. (1970). The same author's "Excavations at Jericho," *Palestine Exploration Quarterly* (1952–56, 1960), contains annual reports of the excavations; her *Excavations at Jericho*, 2 vol. (1960–65), forms part of the definitive report of the 1952–58 excavations and deals with all the tombs, ranging from the Proto-Urban to the Roman period. JAMES L. KELSO, "Excavations at New Testament Jericho . . . ," *Annual of the American Schools of Oriental Research*, vol. 29–30 (1955); and JAMES B. PRITCHARD, "The Excavation at Herodian Jericho, 1951," *ibid.*, vol. 32–33 (1958), are the reports of the excavations at the site of Herodian and Roman Jericho on the Wādī al-Qilt (Wadi Qelt).

(K.M.K.)

Jerome, Saint

One of the greatest teachers of the early Western Church, St. Jerome, also known as Eusebius Hieronymus, is regarded as the most learned of the Latin Church Fathers. He is known particularly for his translation of the Bible into Latin, the Vulgate, that could be understood by the people of his day.

Life. Jerome was born of well-to-do Christian parents *c.* 347 at Stridon, probably near the modern Ljubljana, Yugoslavia. His education, begun at home, was continued in Rome when he was about 12. There he studied grammar, rhetoric, and philosophy. A serious scholar, enamoured of Latin literature, he frequented the catacombs (subterranean burial places composed of galleries), and near the end of his Roman education was baptized (*c.* 366), probably by Pope Liberius.

Early monastic experiences

He spent the next 20 years in travel and impermanent residences. At Treveris (now Trier), he was profoundly attracted to monasticism. Possibly as early as 369 he was back in the vicinity of Stridon. In Aquileia (Italy) he was linked with an ascetic elite—including Rufinus, a writer and scholar, who translated the 3rd-century Alexandrian theologian Origen—grouped around Bishop Valerianus. When the group disbanded (*c.* 373), for reasons that are unknown to historians, Jerome decided to go on a trip through the East. Fatigued by travel and by inner conflict, on reaching Antioch (374) he rested as guest of the priest Evagrius of Antioch, and there may have composed his earliest known work, *De septies percussa* ("Concerning Seven Beatings"). There also, in mid-Lent 375, during a near-fatal illness, he had a celebrated dream. In that dream, in which he was dragged before a tribunal of the Lord, he was accused of being a Ciceronian—a follower of the 1st-century-BC Roman philosopher—rather than a Christian, and he was severely lashed; he vowed never again to read or possess pagan literature.

Long afterward, in controversy with Rufinus, he minimized the dream's importance, but for years it kept him from reading the classics for pleasure, and at the time it caused a genuine spiritual crisis. One result of the dream was his first exegetical (critical interpretive) work, an allegorical commentary on the biblical book Obadiah, disowned 21 years later as a youthful production of fervent ignorance.

In 375 Jerome began a two-year search for inner peace as a hermit in the desert of Chalcis. The experience was not altogether successful to him personally, A novice in spiritual life, he had no expert guide; speaking only Latin, he was confronted with Syriac and Greek; lonely, he begged for letters; a gourmet with a weak stomach, he found desert food a penance; desperate for peace, he was racked by passion. Yet, he claimed that he was genuinely happy. His response to various temptations was incessant prayer and fasting; he learned Hebrew from a Jewish convert, studied Greek with seriousness, had manuscripts copied for his library and his friends, and carried on a brisk correspondence.

The crisis came when Chalcis became involved with ecclesiastical and theological controversies centring on episcopal succession and Trinitarian (on the nature of the relationship of the Father, Son, and Holy Spirit) and christological (on the nature of Christ) disputes. Suspected of harbouring heretical views (*i.e.*, Sabellianism, which stressed God's unity at the expense of the distinct persons)?Jerome insisted that the answer to ecclesiastical and theological problems lay in oneness with the Roman bishop. Pope Damasus did not respond, and Jerome quit the desert for Antioch.

Education and work in Antioch, Constantinople, and Rome

In Antioch his host, Evagrius, won Jerome over to the party of Bishop Paulinus, who was opposed by Basil, the great orthodox bishop of Caesarea and one of the three influential Cappadocian Fathers—the other two being Gregory of Nazianzus and Gregory of Nyssa. Recognizing his importance—since Jerome was by now recognized as a scholar and a monastic figure of significance—Paulinus decided to ordain him. Jerome accepted (378) on two conditions: his monastic aspirations would not be prejudiced, and priestly functions would not be forced on him. He attended the exegetical lectures of Apollinaris of Laodicea and visited the Nazarenes (Jewish Christians) of Beroea to examine their copy of a Hebrew gospel purporting to be the original Gospel of Matthew.

Jerome spent almost three years (379–382) continuing his pursuit of scriptural studies. An enthusiastic disciple of Gregory of Nazianzus, Jerome also came to know Gregory of Nyssa and the theologian Amphilochius of Iconium at the Council of Constantinople (381). Under such influences he improved his knowledge of Greek and developed an admiration for Origen's exegesis. He translated 14 of Origen's homilies (sermons) on Old Testament books into Latin. Here too he translated the church historian Eusebius' *Chronicon (Chronicles)* and continued it to the year 378.

But the most decisive influence on Jerome's later life was his return to Rome (382–385) as secretary to Pope Damasus. There he pursued his scholarly work on the Bible and propagated the ascetic life. On Damasus' urging he wrote some short exegetical tracts and translated two sermons of Origen on the Song of Solomon. More importantly, he revised the Old Latin version of the Gospels on the basis of the best Greek manuscripts at his command, and made his first, somewhat unsuccessful, revision of the Old Latin Psalter based on a few Septuagint (Greek translation of the Old Testament) manuscripts. He held classes for a monastic-minded circle of noble Roman widows and virgins (*e.g.*, Marcella, Paula, and her daughters Blesilla and Eustochium). He taught them the Hebrew text of the Psalms, orally and in letters, he answered their biblical problems, and he was their master in spirituality as well. Under these conditions, he wrote a defense of the perpetual virginity of Mary, Jesus' mother (383), and attacked the view of those who espoused the equality of virginity and marriage. But his preaching in support of the monastic life and his relationship with the ascetic coterie, his castigation of Roman clergy, lax monks, and hypocritical virgins, and his correction of the gospel text provoked such a storm of criticism and calumny, especially after Damasus' death, in December 384, that in August 385 he left "Babylon" (Rome) in bitter indignation and made his way to the Holy Land.

Monastic work in Palestine

In company with virgins led by Paula, Jerome made a religious and archaeological pilgrimage through all of Palestine and to the monastic centres of Egypt; he spent almost a month with the famed exegete Didymus in Alexandria.

The summer of 386 found him settled in Bethlehem. There, by 389, Paula finished a monastery for men under Jerome's direction, three cloisters for women (forming one convent) under her own supervision, and a hostel for pilgrims.

Later literary works. The literary legacy of Jerome's last 34 years (in Palestine) is the outgrowth of contemporary controversies, Jerome's passion for Scripture, and his involvement in monastic life. The controversies were varied. An anti-Origen movement in the east, famed by the anti-heretical bishop Epiphanius, turned Jerome not only against the views of Origen—whose 39 sermons on Luke he had translated c. 389–392—but against his friends Bishop John of Jerusalem and Rufinus. His petulance in early correspondence with St. Augustine, stemming from the African's strictures on Jerome's biblical efforts, imperilled their mutual respect. His catalog of Christian authors, *De viris illustribus* ("Concerning Illustrious Men"), was written in 392/393 to counter pagan pride in pagan culture. Against the monk Jovinian, who asserted the equality of virginity and marriage, he wrote a polemical diatribe *Adversus Jovinianum* (393) that was frequently brilliant but needlessly crude, excessively influenced by the 2nd- and 3rd-century theologian Tertullian, whose writings were at times unnecessarily harsh toward marriage. Against the priest Vigilantius, Jerome dictated in one night a defense of monasticism, clerical celibacy, and certain practices connected with the cult of martyrs (*Contra Vigilantiurn*, 406). The Pelagian problem—named for the heretical English monk Pelagius, who minimized the role of divine grace in man's salvation—was transplanted to Palestine from Rome with the personal appearance of the author of this heresy, and called forth Jerome's finest controversial work, *Dialogi contra Pelagianos* (three books, 415), in which his use of fictitious interlocutors makes his arguments uncommonly impersonal.

Biblical commentaries, revisions, and translations

Jerome's biblical production in Bethlehem includes two introductory works helpful to biblical scholars: *Liber locorum* ("Book of Places"), a useful translation and adaptation of Eusebius' work on Palestinian place-names; and *Liber interpretationis Hebraicorum nominum* ("Book of Interpretation of Hebrew Names"), an alphabetical list, with quite fanciful etymologies or origins, of Hebrew proper names in the Bible. Continuing his revision of the Old Latin version of the Septuagint based on Origen's *Hexapla* (an edition with the Hebrew text in Hebrew and Greek characters, and four different Greek versions arranged in six parallel columns), he revised Ecclesiastes, Proverbs, the Song of Solomon, Chronicles, and Job, and to his Roman revision of the Psalms added Origen's diacritical notes. Between 391 and 406 he produced a Latin translation of the Old Testament on the basis of the original text, a remarkable effort considering the slender instruments at his disposal and the opposition his project provoked. Jerome's Bethlehem commentaries suffer at times from hasty composition, excessive dependence on his predecessors, and a predilection for allegorical interpretation. The *Liber Hebraicarum quaestionum in Genesim* ("Book of Hebrew Questions on Genesis") gives an important place to geography, etymology, and rabbinic tradition (post-biblical Jewish scholarly views). He wrote short glosses (interpretations) on the Psalter and more scientific explanations of Psalms 10–16. His commentary on Ecclesiastes (*c.* 387) is a milestone in exegesis, because it is the first original Latin commentary that takes advantage of the Hebrew text. Perhaps Jerome's best commentaries are on the minor and major prophets of the Old Testament. As for the New Testament, he explained Matthew, accenting the literal sense; he interpreted four of the Apostle Paul's letters (Philemon, Galatians, Ephesians, Titus), with the stress on the allegorical interpretation; he recast the chiliastic commentary (on the view that Christ would return again and his followers would reign for 1,000 years) of Bishop Victorinus of Pettau (Austria), a Latin exegete martyred *c.* 304, on the Revelation to John. Many of his letters are tracts on exegetical problems, and he wrote homilies to monks under his direction on texts of Scripture.

Jerome's ascetical interests at Bethlehem are mirrored not only in his controversies but in his life of Malchus, the monk captured by Bedouins; a biography of Hilarion, with its miracles and journeyings; a translation of Coptic ascetical works (*e.g.*, the Rule of Pachomius); homilies to monks; and a significant segment of his correspondence.

Influence

Jerome is remembered for his extensive erudition, especially his understanding of the classics, the Bible, and Christian tradition. He was a learned scholar rather than a deep thinker, a sound traditionalist and not a speculative theologian, more competent as editor than as exegete. His career was a turbulent combination of scholarship and asceticism, and his correspondence is an exciting source for the historian, Scripture student, and theologian. His influence has been far-reaching and profound, on the early Middle Ages in particular: primarily through the Vulgate (the Latin version that he had translated), but importantly also through his work as an exegete and humanist and because he transmitted much of Greek thought to the West.

Jerome lived at the monastery in Bethlehem, with brief exceptions, until his death in 419 or 420. St. Jerome's feast day is celebrated in the Roman Catholic Church on September 30.

BIBLIOGRAPHY

Collections of Jerome's works: DOMENICO VALLARSI, 11 vol. (1734–42), reprinted in J.P. MIGNE, *Patrologia latina*, vol. 22–30; *Corpus scriptorum ecclesiasticorurn latinorum*, vol. 49 (revision of Victorinus on Apocalypse), vol. 54–56 (letters), vol. 59 (Jeremiah); *Corpus christianorzim, series latina*, vol. 72 (Questions on Genesis, Hebrew Names, Psalms, Ecclesiastes), vol. 73–73A (Isaiah), vol. 74 (Jeremiah), vol. 75 (Ezekiel), vol. 75A (Daniel), vol. 78 (homilies). English translations are: FREDERICK A WRIGHT, *Select Letters of St. Jerome* ("Loeb Classical Library," 1933); W.H. FREMANTLE, *St. Jerome: Letters and Select Works* ("A Select Library of the Nicene and Post-Nicene Fathers," vol. 6, 1912); CHARLES C. MIEROW, *Letters of St. Jerome* ("Ancient Christian Writers," vol. 33, 1963); *The Fathers of the Church*, vol. 48 (1964; homilies 1–59), 53 (1965; dogmatic and polemical works), 57 (1966; homilies 60–96).

Biographies and studies: PAUL ANTIN, *Essai sur saint Jérôme* (1951); YVON BODIN, *Saint Jérôme et l'Église* (1966); FERDINAND CAVALLERA, *Saint Jérôme: sa vie et son oeuvre*, 2 vol. (1922); HARALD HAGENDAHL, *Latin Fathers and the Classics* (1958); G.Q.A. MEERSHOEK, *Le Latin Biblique d'après saint Jérôme* (1966); PAUL MONCEAUX, *Saint Jérôme, sa jeunesse* (1932; Eng. trans., *St. Jerome: The Early Years*, 1933); FRANCIS X. MURPHY (ed.), *A Monument to Saint Jerome* (1952); JEAN STEINMANN, *Saint Jérôme* (1958; Eng. trans., 1959); CHARLES C. MIEROW, *Saint Jerome: The Sage of Bethlehem* (1959); DAVID S. WIESEN, *St. Jerome As a Satirist* (1964); WILFRIED HAGEMANN, *Wort als Begegnung mit Christus* (1970); extensive bibliographies in BERTHOLD ALTANER, *Patrologie* (1958; Eng. trans., 1960).

(W.J.Bu.)

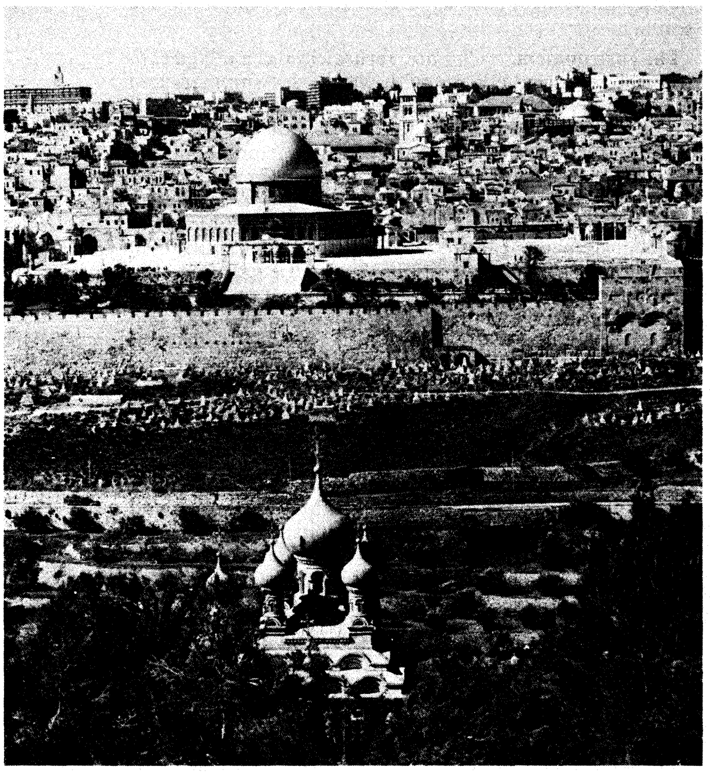
Jerusalem

Jerusalem (Hebrew Yerushalayim; Arabic Bayt al-Muqaddas, or al-Quds), claimed by Israel as its capital, was, at the beginning of the 1970s, the home of almost 300,000 people, and a symbol and inspiration for the three great monotheistic religions of Judaism, Christianity, and Islām. For Jews throughout the world Jerusalem is the focus of age-old yearnings, a living proof of ancient grandeur and a centre of national renaissance; for Christians, it is the scene of their Saviour's agony and triumph; for Muslims, the goal of the Prophet's mystic night journey and the site of one of Islām's most sacred shrines. For all three faiths it is a centre of pilgrimage—the Holy City, the earthly prototype of the heavenly Jerusalem.

From 1948 until 1967, Jerusalem was divided into Israeli and Jordanian sectors, with the Israeli sector of the city forming the capital of Israel. During the Six-Day War of June 1967, however, Israel occupied the former Jordanian sector, over which it now claims jurisdiction as an integral part of the unified city. The claim is not recognized by either Jordan or the United Nations.

An outstanding characteristic of Jerusalem, which covers an area of 42 square miles (109 square kilometres), is the variety of its people and its buildings. The Old City has Muslim, Jewish, Christian, and Armenian quarters. The old Jewish neighbourhoods and some of the new preserve much of the atmosphere created by Jews from East and West who settled there at different periods. Jews from the Orient, Europe, and the Americas; Arabs in traditional and modern dress; clergy of all three faiths and many traditions; and hosts of tourists combine in colourful, kaleidoscopic patterns. Synagogues, churches, and mosques; governmental. educational, and other public institutions; shops, hotels, offices, and dwellings in various styles—almost all built or faced with the beautiful Jerusalem stone—make up the city's unique architectural mosaic. (Ed.)

Werner Braun—Camera Press—Pix



The Old City of Jerusalem seen from the Mount of Olives. In the foreground are the spires of the Russian Orthodox Church of St. Mary Magdalene. Behind the Golden Gate (right centre) is the Dome of the Rock (centre).

HISTORY

Ancient origins. The earliest recorded masters of Jerusalem were the Egyptians. In the 14th century BC the city's ruler, Abdi-Kheba (Abdu-Heba), asked his overlord for help against the incursions of the Ḫapiru (Habiru, 'Apiru). Excavation has shown, moreover, that a settlement existed on the site south of the Temple platform, possibly in the Early Bronze Age and certainly by 1800 BG. A massive town wall still survives, just above the spring that determined the location of the ancient settlement.

The early Hebrew capital

About the year 1000 BG Jerusalem was captured by David, founder of the joint kingdom of Israel and Judah, and the city became the Jewish capital. His successor, King Solomon, extended the city and built his Temple on the threshing floor of Araunah (Ornan) the Jebusite. This foundation determined the sacred character of Jerusalem as a centre of monotheism.

On Solomon's death the northern tribes seceded. In 922 BC the Egyptian pharaoh Sheshonk I sacked the city, to be followed by the Philistines and Arabians in 850 and Joash of Israel in 786. In 701 Sennacherib of Assyria forced Hezekiah, then king of Judah, to pay tribute to him, and in 612 Assyria yielded its primacy to Babylon. Eight years later Jerusalem was despoiled, and its king was deported to Babylon. In the early 6th century BC, the city and Temple were completely destroyed, and the captivity began. It ended in 537 BC, when Cyrus II the Great of Persia had overcome Babylon. Little is known of the city's history during the next two centuries. Despite the opposition of the Samaritans (a religions community), the walls were rebuilt. and a restored Temple was dedicated.

The coming of Alexander the Great and his victory at Issus in 333 BC altered the political and cultural structure of the Middle East. Hitherto the region had been subject to one or the other of the two great "river civilizations" of the Euphrates and the Nile; now for almost a millennium, until the coming of the Arabs, it was to be within the orbit of Western power politics, a role it was to play again in the 19th and 20th centuries.

After Alexander's death, Palestine fell to the share of his marshal. Ptolemy I Soter, son of Lagus, who had occupied Egypt and bad made Alexandria his capital. In the year 198 BC Jerusalem was acquired by the northern dynasty, descended from Seleucus I Nicator, another of Alexander's marshals, which ruled from Antioch (contemporary Antakya, Turkey). The growth of Greek, or pagan, influence affronted the orthodox, whose hostility burst into armed rebellion in 167 BC after the Seleucid Antiochus IV Epiphanes had deliberately desecrated the Temple. The revolt was led by a pious countryman called Mattathias, son of Hasmoneus (Hasmon), and was carried on by his son Judas, known as the Maccabee. The Hasmoneans succeeded in expelling the Seleucids and established a state almost the size of Solomon's, which came to include "Galilee of the Gentiles," thus ensuring that all children born to Galilean parents would be brought up in a Jewish society.

The growth of Greek influence

Roman rule. Rome had for some time been expanding its authority in Asia, and in 63 BC Pompey captured Jerusalem. A clash with Jewish nationalism was averted for some time by the political skill of a remarkable family, whose most illustrious member was Herod the Great. Herod was of Edomite descent, though of Jewish faith, and was allied through his mother with the nobility of Nabataean Petra, the rich Arab state that lay to the east of the Jordan. In 40 BC Herod, who had already distinguished himself as governor of Galilee, was appointed "client" king of Judaea by the Roman Senate. He was the close friend first of Mark Antony and, after the defeat of Antony by Octavian at Actium in 31 BC, of Octavian (later Augustus) himself. He ruled for the next 36 years and largely rebuilt Jerusalem.

The kingdom of Herod

Jerusalem then profited from the offerings of Jewish communities dispersed throughout the Greco-Roman world, and thousands of pilgrims flocked to the Temple for the great feasts. Herod was succeeded by his son Archelaus, who proved utterly incompetent, and in AD 6 Judaea was absorbed into the Roman dominions as a province of the third rank under a procurator. It was under the fifth procurator, Pontius Pilate, that Jesus of Nazareth was put to death.

From AD 41 to 44 the kingdom of Herod was reconstituted for his grandson Herod Agrippa I, upon whose premature death the procurators returned. In 66 the Jews rebelled against Rome, and in 70 the city was besieged and almost wholly destroyed by the Roman forces under Titus. The Temple, Herod's greatest creation, was reduced to ashes. By 130 the city had been partially repopulated, and the Jews again revolted unsuccessfully against Rome from 132 to 135. Hadrian decided to plant a Roman city, Aelia Capitolina, on the site. The general layout of Hadrian's town has lasted into the 20th century.

Christian pilgrims early found their way to Jerusalem. It was, however, the establishment of Constantine as sole emperor of Rome in the year 324 and his conversion to Christianity that made possible the building of the famous shrines in Jerusalem including the Church of the Holy

Sepulchre and inaugurated one of the city's most splendid and prosperous epochs. In 614 this golden age was brought to an end by the Persian invasion in which the inhabitants of Jerusalem were massacred and the churches destroyed.

The Islāmic and British periods. In 638 the Muslim caliph ʿUmar I entered Jerusalem, and during the period from 688 to 691 the 10th caliph, 'Abd al-Malik ibn Marwān, built the Dome of the Rock. Both the Umayyads and their successors, the ʿAbbāsids, pursued a liberal policy toward Christians and Jews. In 969 control of Jerusalem passed to the Shi'ite Fāṭimid caliphs of Egypt, and in 1010 the caliph al-Hiikim ordered the destruction of Christian shrines. In 1071 the Seljuq Turks defeated the Byzantines, displaced the Egyptians as masters of the Holy Land, and cut the pilgrim routes, thus stimulating the movement known as the Crusades.

The Latin kingdom of Jerusalem lasted from 1099 to 1187, when it was overthrown by Saladin, whose successors ruled from Damascus. Jerusalem was again in Christian hands from 1229 to 1239 and from 1243 to 1244, when it was sacked by the Khwrirezm Turks. In 1247 the Holy City fell once more to Egypt and remained subject to the Mamlūks until 1517, when the Ottoman sultan Selim I took it and inaugurated a Turkish regime that was to last for exactly 400 years.

The Turkish era

During the next three centuries the city had almost no history. From 1832 to 1840 Palestine was governed by Muhammad 'Ali Pasha of Egypt and his son Ibrāhīm. In 1838 a British consulate (the first European representation) was opened, and in 1841 the Anglican bishopric was founded. Jewish immigration was stimulated, and archaeology began to reveal much of the city's past. In 1917 British troops entered Jerusalem, Ottoman rule came to an end, and the British administration under League mandate began.

Despite tensions between Arabs and Jews, which more than once led to open riots, improvements were made in both administration and the provision of amenities.

Post-World War II. British rule terminated in 1948; as the result of hostilities between Arabs and Zionists in that same year, the city of Jerusalem was divided. Israeli troops occupied the Old City in the Six-Day War of June 1967. Despite its annexation by Israel, negotiations continued as to the exact status of the city. (S.H.P.)

THE CONTEMPORARY CITY

The city site and traditional neighbourhoods. On the east, Jerusalem looks down on the Dead Sea and across the Jordan River to the arid mountains of Moab; on the west, it faces the coastal plain and the Mediterranean Sea. Just north of the Old City the main north–south road runs along the watershed between the Mediterranean plain and the rift valley of the Jordan and links Nābulus (ancient Shechem) with Bethlehem, Hebron, and Beersheba. The west–east road from Yafo crosses the Jordan north of the Dead Sea and runs to Amman (in Jordan; biblical Rabbah of the Ammonites) and the Arabian Desert.

The municipal boundaries, defined in 1967, enclose an area of 42 square miles that stretch from the Jerusalem airport in the north to a point about one mile from Bethlehem in the south and from the ridge of Mt. Scopus and the Mount of Olives in the east to Mt. Herzl, 'En Kerem, and the new Hadassah–Hebrew University Medical Centre in the west.

The Old City

The Old City, which is believed to have been continuously inhabited for more than 4,000 years, forms a quadrilateral about 3,000 feet (900 metres) long on each side at an altitude of from 2,370 to 2,580 feet (720 to 790 metres). It is dominated by the raised platform of the Herodian Temple Compound, the site of the First and Second Temples, known to Islām as al-Ḥaram ash-Sharif ("the Noble Holy Place"); the compound includes the Dome of the Rock and al-Aqṣā Mosque. Its Western Wall is the most sacred of Jewish shrines. The rest of the area within the walls—divided by the ancient street layout into Muslim, Christian, Jewish, and Armenian quarters—is a typical Oriental city, with its mosques and labyrinthine *suqs,* or bazaars. It is distinguished by its many churches and by the ancient synagogues and study houses of the Jewish Quarter. The first neighbourhoods outside the city walls, built from the 1860s onward, were scattered chiefly along the main roads leading into the city. The earliest, Mishkenot Sha'anannim, and Yemin Moshe have been renovated. Nahalat Shiv'a has been swallowed up by the commercial centre. Others include the Bukharan Quarter; Me'a Sheʿarim, founded by Orthodox Jews from eastern and central Europe, with its scores of small synagogues and Talmudic study houses; and Mahane Yehuda, with its fruit and vegetable market, inhabited mainly by Jews of Oriental origin. Residential quarters established between World War I and World War II include Rehavya in the centre, Talpiyyot in the south, and Qiryat Moshe and Bet ha-Kerem in the west. Some Arab districts, such as Talbieh (modern Qomemiyyut) and Katamon (Gonen), abandoned during the fighting of 1947–48, became Jewish neighbourhoods; and thousands of houses were built for new immigrants in districts to the west, newly incorporated in the city. Arab neighbourhoods outside the Old City include the American Colony, ash-Shaykh Jarrāḥ, Wādī al-Jōz, and Bayt Ḥanīnā in the north, and villages such as Silwiin and Bayt Ṣafāfā in the south.

Neighbourhoods outside the walls

The environment. *Climate.* Jerusalem has a mixed subtropical, semi-arid climate with warm, dry summers and cool, rainy winters. The average annual rainfall is about 20 inches (500 millimetres), and snow falls every two or three years. Average temperatures range from about 75° F (24° C) in August to about 50° F (10" *C)* in January. Hot spells, known as *sharavs,* or *sharqiyeh,* are fairly common in autumn and spring. Average daily humidity is about 62 percent in the daytime but may drop 30–40 percent under *sharav* conditions. Summer exposure to the sun's rays in Jerusalem is among the highest on the globe, attributable partly to the lack of clouds or humidity and partly to the angle of the sun (80") over the horizon at that season.

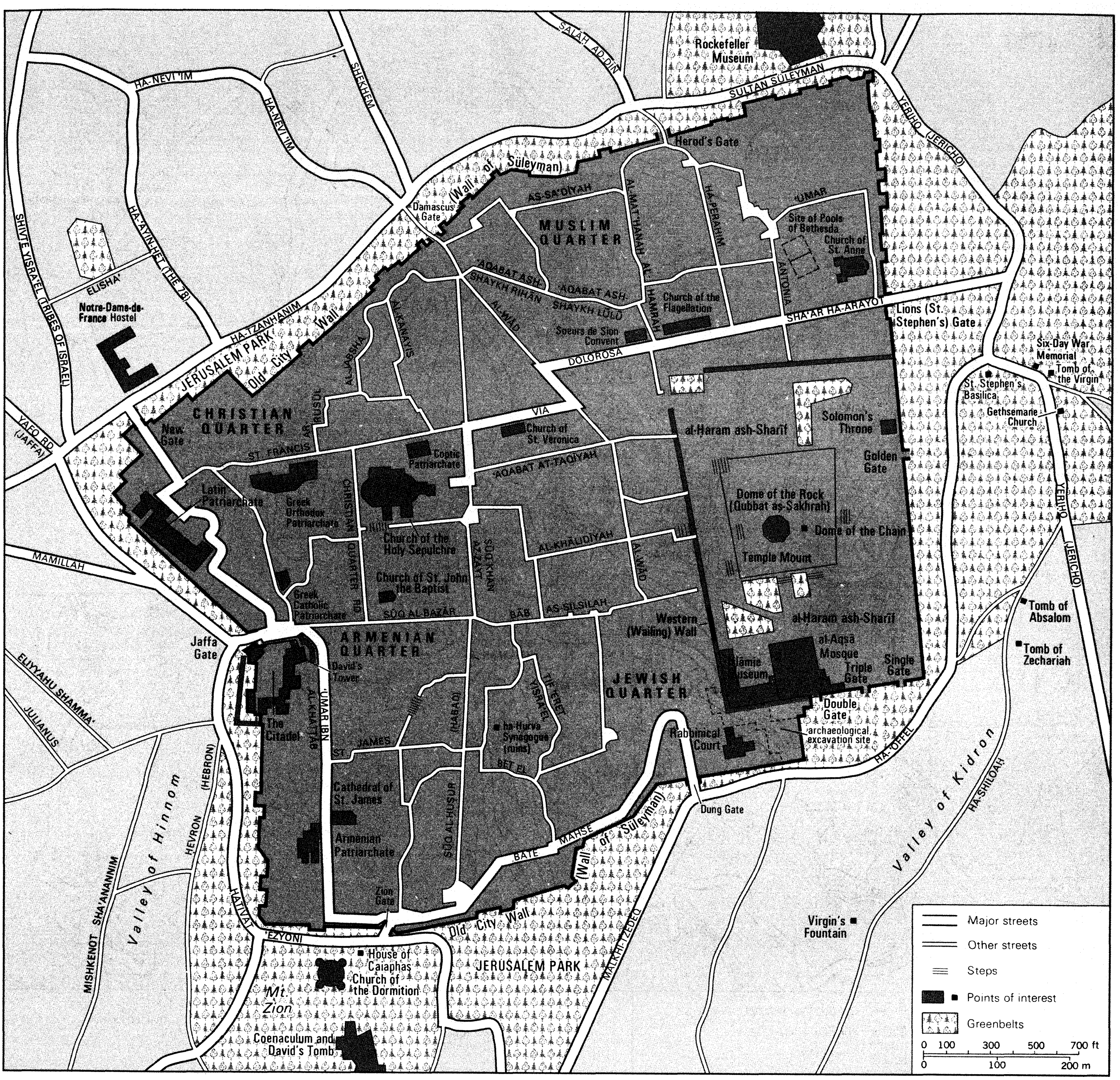
Jerusalem has no serious air pollution. Its altitude ensures the free mixing of surface air, and pollutant sources are few, for there is little heavy industry.

Vegetation and animal life. Lying on the watershed between the relatively rainy Judaean Hills and the dry Judaean Desert, Jerusalem has both Mediterranean and Irano-Turanian vegetation. The various red and brown Mediterranean soils, formed by the different types of limestone chalk covering the hills, support as many as 1,000 species of plants.

There is a great variety of birds, including 70 resident species and about 150 winter visitors. Those most commonly seen are the hooded crow, jay, swift—which nests in old walls and buildings—and bulbul. The only venomous snake is the Palestine viper; the smooth lizard and common chameleon frequent gardens.

The city plan. In 1968 a draft master plan was completed by the Jerusalem municipality for the municipal area and for the surrounding metropolitan area of about 260 square miles (675 square kilometres), which is defined as the commuting area from which the city centre can be reached within half an hour. Its aims were to establish an urban structure for the city as the capital of Israel and to ensure the preservation of historical sites and buildings. The long-term life-span of the plan was fixed at 40 years, and an operational plan was drafted to cover the period up to 1985.

The short-term plan envisages a growth of the population to 400,000 within the municipal boundaries and to 600,000 within the metropolitan area. It assumes free movement between Jerusalem and all parts of the area west of the Jordan, whatever political settlement is reached in the Middle East. There is to be a green belt around the city and a park around the Old City to be called Jerusalem Park. A central feature of the proposed road network is a wide boulevard from Mt. Herzl in the west, passing the university campus, the legislative and administrative centre, and the commercial and tourist centres, and culminating in Mt. Scopus and the Mount of Olives.



The Old City of Jerusalem.

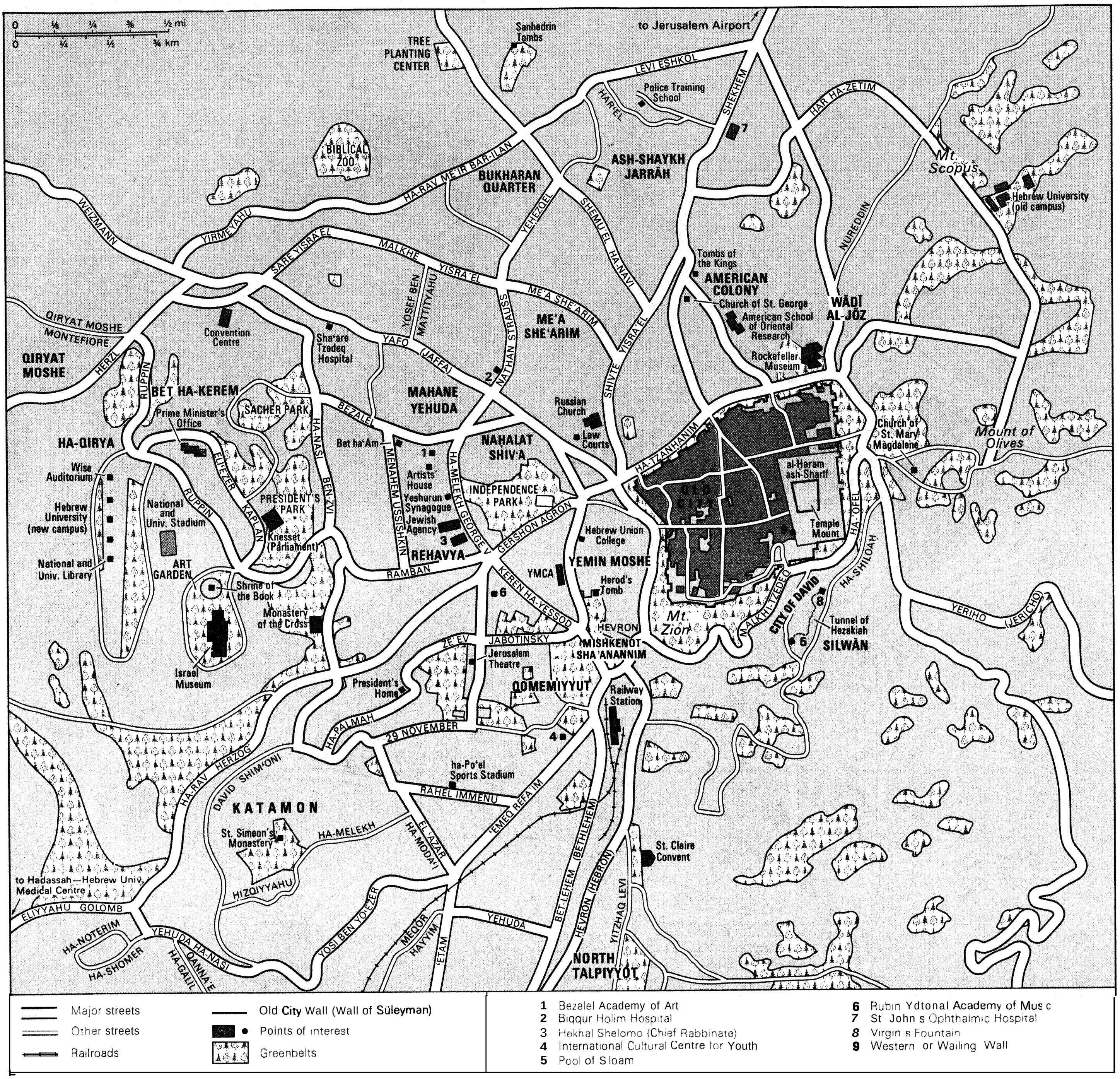
Composition of the population

Demography. In 1976 the population of Jerusalem was officially estimated at 366,300, of which about 73 percent were Jews. The Muslims were the most homogeneous of the communities, the Christians the most diversified. Among the Jews, the main divisions are between Ashkenazim and Sefardim and between religious and nonobservant; 54 percent were born in the country, 13 percent in other Asian countries, 11 percent in Africa, and 22 percent in Europe or the United States.

In 1967, 74 percent of the population lived in the former Israeli sector, 17 percent in the former Jordanian sector, and 9 percent in new extensions of the municipal area. The mean density of population was 8,721 per square mile (3,361 per square kilometre) in 1976.

Religious communities. The administration, protection, and care of holy places in Jerusalem are in the hands of the respective religious authorities. Penalties of up to seven years' imprisonment may be inflicted for desecration of these places. The rites and observances of all the faiths are publicly celebrated—the voice of the muezzin from al-Ḥaram ash-Sharif mingles with the sound of nearby church bells and of the Jewish prayer services at the Western Wall.

Jews. Jerusalem is the focus of Jewish religious reverence and aspiration. The most sacred spot is the Temple Mount, which Orthodox Jews refrain from setting foot on for fear of profaning its sanctity. In addition to the Western Wall—the most important centre of prayer and pilgrimage—other holy places are Mt. Zion, with the reputed tomb of King David, the Mount of Olives with its ancient Jewish cemetery, and the tombs of priestly and other historical figures. Ancient synagogues and study houses in the Old City are being restored; particularly worthy of mention is the group of four synagogues built in the 16th century by Jews exiled from Spain. There are scores of Jewish houses of prayer in the New City. Notable modern institutions include the Yeshurun Synagogue, where prayer is offered on state occasions, and



The modern city of Jerusalem.

the synagogue at Hekhal Shelomo, the seat of the Chief Rabbinate of Israel. Jerusalem is the world's foremost centre of rabbinic learning and contains scores of yeshivot, the Talmudic academies.

Muslims. Jerusalem is revered by Muslims as the third holiest place on earth, and the pilgrimage to Jerusalem (taqdis) completes the main pilgrimage (hajj). A Council of Waqf [religious endowment] and Muslim Affairs was created in 1967, with jurisdiction over shari'ah courts and waqf. The council assumed the responsibility for the administration of Muslim affairs that had previously rested with the Council of Waqf and Muslim Affairs in Amman.

Christians. Christians constitute the smallest but most highly diversified section of the population. The city is the seat of three resident patriarchs of the Eastern churches and many archbishops and bishops, and it has an ecclesiastical embassy for almost every sect in Christendom. The main groups, in order of size, are Orthodox, Monophysite, Latin, and Protestant. Major denominations share control over the Church of the Holy Sepulchre. Most of the church bodies in Jerusalem maintain scholarly research institutes with fine libraries.

The Greek Orthodox Church maintains a patriarchate with jurisdiction over the entire Holy Land. The Russian Orthodox churches (one governed from Moscow and one in exile) have considerable properties dating back to tsarist times. The Roman Catholic Church, first officially established in Jerusalem in 1099 during the First Crusade, controls numerous churches, hospices, monasteries and convents, schools, and hospitals. The Franciscan order appoints the father custodian of the Holy Land. Of the Monophysite churches, the Armenian is the largest, its patriarchate having been established in the 6th century. Most of the Armenians dwell in a compound around the seat of the patriarchate at the Cathedral of St. James, which constitutes the largest monastic centre in the country. The Protestant community is small

Armenian Church

but influential. The jurisdiction of the Anglican archbishop extends over the entire Middle East.

Housing. There is a great variety of housing in the city. In the Old City are antiquated buildings constructed of ancient stones; 19th-century Jewish neighbourhoods, some of which have declined into slums; modern quarters with tree-lined streets; and government-built housing projects, mainly for new immigrants. The commonest basic dwelling unit in the Old City consists of a complex of structures, often on different levels, built around an inner court that is entered through a narrow corridor. Steps are being taken to facilitate slum clearance, first in the Nahala'ot area, near Rehavya. In addition to extensive private building in many areas, large new housing projects are under construction or in the planning stage in previously undeveloped areas to the northeast and southeast.

The Wall of Süleyman

Architectural features. The outstanding characteristic of the architecture of Jerusalem is the harmonious coexistence of old and new, sacred and secular, in a variety of styles. The most conspicuous feature is the Old City Wall, erected in 1537–40 by Süleyman the Magnificent, largely on the foundations of earlier walls going back perhaps to the time of Solomon. On three sides of the Temple Compound, parts of the original supporting walls stand to this day. During the centuries when Jews were excluded from the Temple Compound, its Western Wall became Jewry's holiest shrine. Since 1967 the wall has been further exposed, and plans are being considered for the reconstruction of the present plaza and the whole area, once excavations are completed. The main buildings on the platform are the gold-capped Dome of the Rock, completed in 691, and the silver-domed al-Aqṣā Mosque, started in 710.

The Citadel (with David's Tower) beside the Jaffa Gate, which acquired its present form in the 16th century, was created over ruins from the Hasmonean and Herodian periods. Within the Old City, the religious buildings are of the walled, fortified type. Buildings of the crusader period are marked by the use of ribbed vaults. Some of them, such as the Church of the Holy Sepulchre and the church at the Tomb of the Virgin, incorporate older Byzantine remains. The best preserved are the Church of St. Anne, the Armenian Cathedral of St. James, and parts of the central bazaars. Mamlūk construction of the 13th to 15th century is easily identified by its "stalactite" ornamentation and painted brown stripes.

The rock-cut tombs east and north of the Old City exemplify architecture of the first half of the 1st millennium BC (Tomb of Pharaoh's Daughter) and the Second-Temple period (Tombs of the Kings, Tomb of Absalom, and Tomb of Zechariah). The Byzantine Monastery of the Cross, in the heart of modern Jerusalem, was rebuilt by the crusaders and restored in 1604 by the King of Georgia.

As Jerusalem spread outside the walls, the architecture came to be characterized chiefly by iron beams and red-tiled roofs. From 1930 onward there was a radical change, and flat roofs and reinforced concrete faced with naturally dressed stone predominated. There is little high-rise construction; residential buildings are seldom taller than four stories, and office buildings seldom taller than six. Tall buildings, especially hotels, rising as high as 16 stories, are under construction or being planned. Outstanding modern architectural features include the Rockefeller Museum, the Hebrew Union College, the YMCA, the Ecumenical Institute for Advanced Theological Studies, buildings on the university campuses on Mt. Scopus and Giv'at Ram, the Knesset (Parliament), the Israel Museum, and the Jerusalem Theatre.

Modern architecture

The economy. *Economic sectors.* The main source of livelihood in Jerusalem is government and public service (including academic and clerical professions). Since 1967 business activity and investment have been stimulated by the housing boom and the ever-increasing influx of pilgrims and tourists. Personal income rose steadily in the 1970s, and unemployment was marginal.

Industries include diamond cutting and polishing and the manufacture of home appliances, furniture, pharmaceuticals, chemicals, shoes, pencils, plastics, textiles, and clothing. There are also printing and publishing houses, as well as workshops for jewelry, giftware, curios, and printed fabrics.

The overall economy. In 1976 the civilian labour force numbered 108,500. According to a survey made in 1976, some 46 percent worked in public services; 19 percent in commerce, banking, finance, and insurance; 13 percent in industry; 8 percent in construction and public works; 6 percent in transportation and communications; 7 percent in personal services and entertainment; and a small percentage in public utilities and farming.

Political and governmental institutions. *National institutions.* Jerusalem is the seat of the president, the parliament, and the government of Israel. In 1947 the United Nations recommended that it be made an international city, but the proposal was opposed by both Israel and Jordan. Although a large number of countries do not recognize Jerusalem as the capital of Israel, a large proportion of the resident foreign embassies and legations in Israel are in Jerusalem. Diplomats living in the Tel Aviv area go to Jerusalem to present their credentials to the president and transact business at the Foreign Ministry. The main government offices are concentrated in ha-Qirya, the administrative centre. In addition to the Supreme Court and the Chief Rabbinate, the city also houses the head offices of many world Jewish bodies, such as the Jewish Agency and the World Zionist Organization. Yad va-Shem (official English name, Martyrs' and Heroes' Remembrance Authority), the central memorial for the Jews killed by Nazi Germany, contains extensive archives and is the major centre for research on the Nazi period.

Voting

Local government. The Municipal Council is composed of 31 members who are elected every four years by proportional representation. Permanent residents, even if not Israeli nationals, are entitled to vote. Although, for political reasons, no Arab candidates chose to stand, more than 11,000 Arabs out of 34,000 on the voter roll went to the polls in 1969. Of the administrative staff, about 20 percent are Arabs, most of whom served under Jordanian rule. Official correspondence is issued in both Hebrew and Arabic.

Services. *Public utilities.* Jerusalem has always depended on human ingenuity to bring its water from afar. The underground aqueduct built by King Hezekiah in the 8th century BC is still extant, and many reservoirs and rainwater cisterns date from ancient times. Since the 1950s the New City has enjoyed an unlimited supply from the Israeli national water grid; East Jerusalem was reconnected to the West Jerusalem system in 1967.

The construction of the city-wide sewerage system was almost complete in the mid-1970s. The six miles of ancient piping that run through the Old City, 40 feet underground, still present serious engineering problems.

Electricity is supplied by the national grid of an Israeli government corporation, as well as by a small diesel plant in East Jerusalem.

Health and safety. The Hadassah–Hebrew University Medical Centre at 'En Kerem is one of the most advanced institutions of its kind in the world. It treats patients from all over the country, as well as from the West Bank, the Gaza Strip, and Jordan. Other hospitals are Sha'are Tzedeq, which pays special attention to the requirements of orthodox Jews, Biqur Holim, St. John's Ophthalmic Hospital, Ezrat Nashim for mental patients, Alyn for crippled children, Spofford Children's Hospital, an Arab Muslim hospital named al-Maqāṣid al-Khayriyah, and an Arab Christian hospital named al-Muṭalla'.

Most families belong to one of the medical-insurance funds run by the labour federations and other nongovernmental bodies; and the municipal social-welfare department covers costs for needy nonmembers. There are municipal clinics for mothers and children. Health supervision, including dental inspection and treatment, is provided in all schools. All health services are subsidized by the government.

The police force, which is under the command of national headquarters, is composed of both Jews and Arabs. A major function of the force, in addition to keeping the peace and preventing crime, is to guarantee access to and maintain order in the holy places.

There are few fires because of the predominance of stone construction. The fire brigade is administered by the municipality.

Religious and secular education

Education. Because of the high birth rate and the strong religious convictions of many among the population, education has always involved complex problems. The language of instruction is Hebrew in Jewish schools and Arabic in Arab schools. While the great majority attend government schools, there are many private institutions maintained by Jewish, Muslim, and Christian religious organizations. State kindergartens were introduced in East Jerusalem after 1967, and 60 percent of the area's five-year-olds attended in 1971. In 1970 there were about 54,000 children in more than 280 state, municipal, and Jewish public schools and 12,000 in Arab public schools.

The Hebrew University in Jerusalem is the country's foremost institution of higher learning. It has two main campuses — at Mt. Scopus, and at Giv'at Ram in the west. The old buildings on Mt. Scopus have been renovated, and new buildings are under construction. The National Library and the University Library contain more than 2,000,000 volumes, as well as art treasures and manuscript collections on Jewish and general themes. In the mid-1970s there were about 13,000 students and more than 2,000 faculty members.

Other institutions of higher learning are the Bezalel Academy of Art, the Rubin National Academy of Music, the Hebrew Union College, several teachers' training colleges, and an Armenian seminary.

Cultural life. *Cultural institutions.* An important cultural institution is the Israel Museum, which, in addition to its general art collection, houses a comprehensive Near Eastern archaeological collection, several important Dead Sea Scrolls and other relics, a notable collection of Jewish ritual art, and Middle Eastern ethnological exhibits, a sculpture garden, and a youth wing. The Rockefeller Museum concentrates on the archaeology of the Holy Land, and there is an Islāmic Museum near al-Aqṣā Mosque.

Libraries

In addition to the National Library and the University Library, there are the Knesset Library and the State Archives (each of which receives a copy of every book printed in Israel), the Municipal Library (with branches), and libraries of various religious and public institutions.

Art exhibitions are held in the Israel Museum, the Artists' House, Hutzot ha-Yotzer (the craftsmen's centre), the International Cultural Centre for Youth, and private galleries. Concerts and theatre performances are given at Binyane ha-'Uma (the Convention Centre), Bet ha-'Am, the Khan (housed in a restored 18th-century building), and the Wise Auditorium of the Hebrew University. The Jerusalem Theatre was opened in October 1971.

The media. Newspapers published in Jerusalem include an English language daily, two Arabic dailies, and a Hebrew religious daily. The Government Press Office is located at Beit Agron, the headquarters of the Jerusalem Journalists' Association. The headquarters of the Israel Broadcasting Authority (television and radio) are also in Jerusalem. Radio broadcasts are mainly in Hebrew and Arabic, though short programs are also broadcast in a number of languages including English, French, Ladino, Romanian, Russian, and Yiddish.

Recreation. Jerusalem has about 2,000 acres (800 hectares) of parks, gardens, woodland, and newly planted forests. The largest, when completed, will be the 300-acre (120-hectare) Jerusalem Park, designed as a greenbelt to encircle the Old City walls. There are also 160 small gardens, playgrounds, and recreational areas dotted over the city. The Biblical Zoo houses specimens of all the animals that are mentioned in the Bible.

The municipality, the YMCA, and local clubs run comprehensive sports programs. The YMCA soccer field can accommodate 10,000 spectators, and the ha-Po'el (Workers' Sports Club) field 7,500. There are a number of open-air swimming pools. Community centres in the suburbs also provide sports facilities. A 100-acre sports area with a stadium for 30,000 is in the planning stage.

Archaeological excavations

Archaeology. Since 1968, extensive excavations have been carried out in the Old City on behalf of the Hebrew University Institute of Archaeology, the Israeli Department of Antiquities and Museums, and the Israel Exploration Society. The digs around the southern and western walls of the Temple Mount, going down to the Herodian pavements, have revealed the steps leading to the Temple, the priests' underground entrance to the Temple, and many religious objects. There are also impressive remains of public buildings alongside a main street. Remains found within the precincts of the First Wall in the Jewish Quarter bore the imprint of burning and destruction during the sack of the city by the Romans in AD 70. For the first time, walls were found of structures dating to the 8th and 7th centuries BC. One of these has been identified as the "Broad Wall" described by Nehemiah. A crucified body, from Roman times, with a nail still lodged in the ankle, was discovered in a Jewish tomb at Giv'at ha-Mivtar. (Ed.)

BIBLIOGRAPHY. Accounts of the city are given, by TEDDY KOLLEK and MAURICE PEARLMAN, in *Jerzisalem, Sacred City of Mankind (1968;* U.S. title, *Jerusalem: A History of Forty Centuries),*and, in the Israel Pocket Library series, *Jerusalem (1973). A* brief review, with emphasis on tourist aspects, is D. SHALEM and G SHAMIS, *Jerusalem (1970–71).*KATHLEEN M. KENYON, *Jerusalem: Excavating* 3000 *Years of History (1967)* and *Digging Up Jerusalem* (1974), describe the city from an archaeological point of view, as does *Encyclopedia of Archaeological Excavations in the Holy Land,* vol. *2* (1976), Eng. ed. MICHAEL AVI-YONAH. L. KESHISHIAN, *The Churches in Jerusalem* (1939), is still a good guide.

History: (Ancient): The Bible is the basic source, best consulted in the modern edition known as *The Jerusalem Bible* (1966), prepared in Jerusalem itself by scholars long conversant with the Holy City and its monuments. It may be supplemented by a good commentary such as JAMES HASTINGS, *Dictionary of the Bible,* 2nd ed. *(1963).* Next to the Bible, the main original source for ancient Jerusalem is JOSEPHUS, the Jewish historian who wrote under Roman patronage at the end of the 1st century AD; the Loeb edition is recommended. Of modern works, EMIL SCHURER, *Geschichte des jüdischen Volkes im Zeitalter Jesu Christi,* 2nd ed., *2* vol. *(1886–90;* Eng. trans., *A History of the Jewish People in the Time of Jesus Christ, 5* vol., *1897–98* and later eds.), is a work of such scope that it is hard to imagine its being superseded. See also the monumental *Jewish Encyclopedia, 12* vol. *(1901–05);* and the *Encyclopaedia Judaica, 16* vol. *(1972).* GEORGE ADAM SMITH, *Jerusalem: The Topography, Economics and History from the Earliest Times to A.D. 70, 2* vol. (1907–08), is a comprehensive survey by a great scholar. The works of the great Dominican scholars VINCENT and ABEL, especially *Jerusalem,* 2 vol. (1912–26), should also be read. *(Medieval)*: STEVEN RUNCIMAN, *A History of the Crusades,* 3 vol. (1951–54), provides essential background, and each volume contains an exhaustive bibliography. GUY LE STRANGE (trans.), *Palestine Under the Moslems (1890,* reprinted 1965), is the classic survey of the period. T.S.R. BOASE, *Castles and Churches of the Crusading Kingdom* (1967), is a finely illustratsd work by an outstanding scholar. *(Modern)*: The beginning of modern times was heralded by three books: EDWARD ROBINSON and ELI SMITH, *Biblical Researches in Palestine, Mount Sinai and Arabia Petraea,* 3 vol. (1841), by the founder of scientific biblical geography; JAMES FINN, *Stirring Times; or, Records from Jerusalem Consular Chronicles of 1853 to 1856,* compiled by his widow *(1878);* and CHARLES W. WILSON and C. WARREN, *The Recovery of Jerusalem* (1871), the record of the first underground survey of the ancient city. Since then, the output of books on every aspect of Jerusalem has been unceasing. For the period of the British mandate, A.M. HYAMSON, *Palestine Under the Mandate* (1950), is the standard work. STEWART PEROWNE, *The One Remains* (1954), is an eyewitness account of the succeeding period. Two general histories are JOHN GRAY, *A History of Jerusalem* (1969), an authoritative survey by a biblical scholar; and A.L. TIBAWI, *Jerusalem: Its Place in Islam and Arab History* (1969), based on original Arabic sources and on official Egyptian, Ottoman, and British records, and on UN publications and reports in *The* (London) *Times.* The latest archaeological excavations are reported in the *Israel Exploration Journal.*

(Ed./S.H.P.)

Jesus Christ

Jesus of Nazareth, the founder of Christianity, a religion that claims more than a third of the world's population in the 20th century, was born in Judaea about 6 BC and died by crucifixion about AD 30. Because of the theological motifs and presuppositions in the faith of the early church in respect to Jesus, it is difficult to write with certainty an authentic life of Jesus.

This article is divided into the following sections:

- Sources
  - Non-Christian sources
  - Christian sources
- Times and environment
  - Political conditions
  - Religious conditions
- The life and ministry of Jesus
  - The birth and family of Jesus
  - The ministry of Jesus
- The message of Jesus
  - The Kingdom of God
  - The will of God
- The sufferings and death of Jesus in Jerusalem
- The story of Jesus and faith in Jesus

SOURCES

The history of the life, work, and death of Jesus of Nazareth reveals nothing of the worldwide movement to which he gave rise. He lived and taught in a remote area (Palestine) on the periphery of the Roman Empire. His life was of short duration, and knowledge of it remained hidden from most of his contemporary world. None of the sources of his life and work can be traced to Jesus himself; he did not leave a single known written word. Also, there are no contemporary accounts written of his life and death. What can be established about the historical Jesus depends almost without exception on Christian traditions, especially on the oldest material used in the composition of the first three New Testament Gospels (Mark, Matthew, and Luke), which reflect the standpoint and outlook of the later church and its faith in Jesus.

References in Roman historical and administrative accounts

Non-Christian sources. Non-Christian sources are meagre and contribute nothing to the history of Jesus that is not already known from the Christian tradition. The mention of Jesus' execution in the Annals of the Roman historian Tacitus (XV, 44), written about AD 110, is, nevertheless, worthy of note. In his account of the persecution of Christians under the emperor Nero, which was occasioned by the burning of Rome (AD 64), the Emperor, in order to rid himself of suspicion, blamed the fire on the so-called Christians, who were already hated among the people. Tacitus writes in explanation: "The name is derived from Christ, whom the procurator Pontius Pilate had executed in the reign of Tiberius." The "temporarily suppressed pernicious superstition" to which Jesus had given rise in Judaea soon afterward had spread as far as Rome. Tacitus does not speak of Jesus but, rather, of Christ (originally the religious title "Messiah," but used very early among Christians outside Palestine as a proper name for Jesus). The passage only affords proof of the ignominious end (crucifixion) of Jesus as the founder of a religious movement and illustrates the common opinion of that movement in Rome. An enquiry of the governor of Asia Minor, Pliny the Younger, in his letter to the emperor Trajan (c. AD 111) about how he should act in regard to the Christians (*Epistle* 10, 96ff.) comes from the same period. Christians are again described as adherents of a crude superstition, who sang hymns to Christ "as to a god." Nothing is said of his earthly life, and the factual information in the letter undoubtedly stems from Christians.

Another Roman historian, Suetonius, remarked in his life of the emperor Claudius (Vita *Claudii* 25:4; after AD 100): "He [Claudius] expelled the Jews, who had on the instigation of Chrestus continually been causing disturbances, from Rome." This may refer to turmoils occasioned among the Jews of Rome by the intrusion of Christianity into their midst. But the information must have reached the author in a completely garbled form or was understood by him quite wrongly to mean that this "Chrestus" had at that time appeared in Rome as a Jewish agitator. Claudius' edict of expulsion (AD 49) is also mentioned in Acts 18:2.

References in Jewish sources

Josephus, the Jewish historian at the court of Domitian who has depicted the history of his people and the Jewish–Roman war (66–70), only incidentally remarks about the stoning in AD 62 of "James, the brother of Jesus, who was called Christ . . ." (Antiquities XX, 200). He understandably uses the proper name "Jesus" first (for as a Jew he knows that "Christ" is a translation of "Messiah"), but he adds, along with a derogatory "so-called," the second name that was familiar in Rome. A second passage in the same work, known as the "Testimony of Flavius" (XVIII, 63ff.), is without doubt a later Christian insertion, because it contains a complete summary of the Christian teaching about Jesus.

In the Talmud, a compendium of Jewish law, lore, and commentary, only a few statements of the rabbis (Jewish religious teachers) of the 1st and 2nd centuries come into consideration. Containing mostly polemics or Jewish apologetics, they reveal an acquaintance with the Christian tradition but include several divergent legendary motifs as well. The picture of Jesus offered in these writings may be summarized as follows: born as the illegitimate son of a Roman soldier called Panther, Jesus (Hebrew Yeshu) worked magic, ridiculed the wise, seduced and stirred up the people, gathered five disciples about him, and was hanged (crucified) on the eve of the Passover.

These independent accounts prove that in ancient times even the opponents of Christianity never doubted the historicity of Jesus, which was disputed for the first time and on inadequate grounds by several authors at the end of the 18th, during the 19th, and at the beginning of the 20th centuries.

Christian sources. Christian testimonies about Jesus were collected in the New Testament. Though they certainly represent only a selection from a much broader stream of tradition (Luke 1:1–4), these testimonies are a very valuable and representative selection. They are, however, of very different kinds. From many of them next to nothing can be learned about the historical Jesus.

***The* Pauline Letters.** The oldest New Testament writings, the genuine letters of Paul (written in the 50s of the 1st century), contain little information about the life of Jesus. Paul, the Apostle, who had not known Jesus personally (II Cor. 5:16), shows no interest in Jesus' biography. At the centre of Paul's thought and proclamation there stands only the theologically important significance of the death, Resurrection, exaltation, and Second Coming of Jesus Christ, contained in numerous short doctrinal and creedal formulas. These formulas the Apostle himself occasionally characterizes as being the tradition that he has received and handed on (I Cor. 11:23ff.; 15:3ff.) or they are in other ways indicated as a given tradition (Rom. 1:3ff.; Phil. 2:6–11).

The importance of the Synoptic Gospels

The Gospels. The most important sources for the life of Jesus are the Synoptic (parallel view of sources) Gospels: Mark, Matthew, and Luke. The Gospel According to John, the Fourth Gospel, assumes a special position. Though it offers some parallels to the other three, and though the independent traditions in it may in individual cases have historical kernels, the tradition in John shows that the gospel has reached an advanced theological state. Because a theological conception has been incorporated in the account to such an extent, this Gospel cannot be directly used as a historical source. It is also the latest of the Gospels, written about AD 100.

That the gospel literature was capable of developing in very different directions is also shown by the extracanonical tradition about Jesus, which is preserved in fragmentary form in quotations by the early Church Fathers and in other sources and which is marked almost without exception by legendary features and tendencies. The Coptic Gospel of Thomas (written in the 2nd century by Gnostic Christians; *i.e.*, heretical believers in esoteric, dualistic doctrines), which was found in 1945 in Naj' Ḥammādī (Egypt), is an example of such extracanonical literature. It contains 114 sayings of Jesus loosely strung together, which have some points of contact with the sayings of Jesus in the canonical Gospels. But this Gospel has no earthly, historical contours in its account of Jesus (*e.g.*, no accounts of the Passion and Easter). As a bearer of heavenly revelation in this Gospel, Jesus instructs the esoteric circle of his disciples about the foreign world of matter that they must renounce in order to participate in the imperishable, transcendent world of light from which they originate. The Gospel of Thomas, thus, is of no use as a source for the historical Jesus.

The Synoptic Gospels were originally anonymous. According to questionable 2nd-century tradition, they were written by the immediate disciples of Jesus or companions of the oldest Apostles. Most probably the Gospels were composed between AD 70 and 100. That they were written at such a relatively late time does not detract from their historical significance, however, because an older, oral tradition is collected in them and has left its traces everywhere. The character and structure of the individual traditions are incorporated into the Gospels, which definitely do not have a historical or biographical interest in facts, circumstances, and the course of events. They do not reproduce the story of Jesus as such but, instead, recount history interpreted from the viewpoint of the Christian faith. There never was a neutral reportage on Jesus in early Christianity; the Gospel accounts are therefore interlaced with theological motifs and styled accordingly. Thus, what Jesus says, does, and suffers is interpreted as the fulfillment of the Old Testament promises, and his story is slanted toward his end (the Passion and the Resurrection), his significance as the divine Saviour, and his Second Coming. In other words, the Gospel texts do not intend to describe the Jesus of the past but rather to proclaim who he is for all ages of time. These perspectives of the post-Easter church to which the writers belong and for which their reports are intended must continually be taken into consideration. To sum up, the Gospels can be characterized as follows: they are a combination, or a fusion, of history and faith, a story or report about Jesus, and a confession and testimony of faith.

A comparison of the first three canonical Gospels reveals a strange blending of agreements and differences. Mark, Matthew, and Luke contain, by and large, the same traditional material. Some parts, however, are to be found only in Matthew and Luke, and a considerable amount of material is peculiar only to Matthew or only to Luke (and a small amount to Mark, as well). According to almost all critical biblical scholars, Mark, the shortest Gospel, is viewed as the oldest—not Matthew, as was earlier assumed—and served as the main literary source for the other two. They also believe that the material common to Matthew and Luke comes from a second source (called Q, from the German *Quelle*, "source"). This second source (Q) consisted almost exclusively of sayings (logia) of Jesus and contained no Passion or Easter tradition and is therefore known among scholars as the logia, or sayings, source.

Significance of form criticism

Investigation of the Gospels by German biblical scholars such as Karl Ludwig Schmidt, Martin Dibelius, and Rudolf Bultmann—who developed what is known as form criticism, the study of the origin and development of the traditions in the Gospels—has shown that the basic stock of the tradition consisted of numerous small, self-contained units (single sayings, parables, debates, anecdotes, and miracle stories), originally without any relation to each other, and mostly without any interest in dates, places, or historical circumstances. It was the Gospel writers (or some earlier collectors) who first joined these individual pieces together editorially, forming a kind of "discourse" out of sayings and groups of sayings and, through linking individual scenes, creating the impression of a connected chain of events. They used a very modest set of tools for this; *e.g.*, short introductory and connecting phrases, stereotyped, generalizing indications of time ("next," "a few days later"), and frequently repeated, indefinite indications of place (mountain, field, road, house, lake). These editorial turns of phrase are, as a rule, easy to sever from their context and are employed very differently by the separate Gospel writers.

In methodically distinguishing and separating traditional and editorial features, form criticism of the Gospels has apparently dissolved the presuppositions for a historically sound, connected life of Jesus, which scholars have again and again attempted to write in the course of the last 200 years. But such an analysis was only a first step of research into the older material itself. Popular oral tradition, to which the Synoptic material belongs, makes use of fixed forms appropriate in each case to the contents, so as to be easily fixed in the memory. The tradition about Jesus offers many examples of this: prophetic sayings, the Beatitudes, pronouncements of woe, wisdom sayings similar to proverbs, legal sayings, church rules, dialogues, and others. In a corresponding way, many miracles of Jesus are narrated by means of motifs and other features also known from reports of other miracle workers. From this one perceives that this tradition is interested not so much in what was historically unique as in what was typical. Thus, with regard to the Gospels, it has to be considered that their tradition was formed and collected from the point of view of the faith of the post-Easter church, under the influence of its ideas and ways of thought and in close connection with its vital interests and the ways in which its life found expression. When interpreting the texts, scholars must therefore be concerned with the question of their setting in life (*Sitz im* Leben) in the church as well.

This critical survey of the sources shows that there are limits set on a portrayal of the historical Jesus. Many questions are still under debate or have to remain open.

TIMES AND ENVIRONMENT

Political conditions. Politically, the small Jewish nation in Jesus' time was rent and impotent. Always situated in an area of tension between the great empires of the ancient world (*e.g.*, Egypt, Assyria, Babylonia, Persia, and Syria) as they struggled with each other and succeeded one another, it had already lost its political independence since the time of the Babylonian Exile (586–538 BC) and had come under changing foreign domination: in the Hellenistic period, first under the Egyptian Ptolemaic dynasty (3rd century BC) and then under the Syrian Seleucid dynasty (2nd century BC), and, finally, until its ultimate overthrow (AD 70), under the Romans, who continued to rule the area. Only for a short interim was there a Jewish kingdom. The Maccabees, a priestly family, reigned after their revolt (168–165 BC) against Antiochus IV (Epiphanes), the Syrian king. Their rule, however, came to an end as a result of internal disintegration and violent struggles for the throne.

Judaea under Roman domination

Initially courted by the rival parties, the Roman general Pompey marched into Palestine, capturing Jerusalem in 63 BC, and reduced the Jewish territory to Judaea, without the coastal cities and the confederacy of towns of the Decapolis (central Transjordan). Several other smaller regions—*e.g.*, inland Galilee of the northern province around Lake Gennesaret and Peraea, east of the Dead Sea—were left to the Jews. By exploiting the threat to the Roman Empire from the Parthians and by adapting himself skillfully to the changing power situations after the murder of Julius Caesar (44 BC), the clever and adroit Herod I (reigned 37–4 BC) managed with the help of the Romans to become "king of the Jews" and to extend the Jewish state over almost all of Palestine again. His regime was decidedly progressive. He promoted Hellenization (*i.e.*, emphasizing Greek culture) by modern building projects, the founding of towns, and in other ways. But he also attempted to win the favour of the Jews, above all by rebuilding Solomon's Temple in ostentatious form and on an enormous scale. It was begun in 20 BC but was not finally completed until AD 64, a few years before its destruction in AD 70 by Titus, who became emperor of Rome nine years later.

Though the Jews demanded of the Romans the abolition of Herodian rule after his death, the Romans divided the land up among the sons of Herod the Great. The most important and largest part, Judaea, with Jerusalem, Samaria, southern Judaea, and Idumaea, was granted to Archelaus, who was deposed by AD 6. His area was integrated into the Roman administration under a governor (procurator), who controlled military, taxation, and judicial affairs. As was their custom, the Romans allowed the Jews to practice their religion and to exercise restricted powers of administration and jurisdiction. Some of the procurators, however, did not hold themselves strictly to these principles. Pontius Pilate, who is designated in an inscription found in 1961 as *praefectus Judaeae*, ruled (AD 26–36) ruthlessly and with bursts of cruelty. He was dismissed for this reason. The reigns of Herod's other two sons were of rather longer duration: Philip (4 BC–AD 34) ruled as tetrarch of the non-Jewish region northeast of Lake Gennesaret, and Herod Antipas (4 BC–AD 39) served as ruler of Galilee and the remote Peraea.

As far as Jesus' story is concerned, the conditions in Galilee, the land of his origin and his ministry, are of paramount importance. Thoroughly changed in character by the settlement of foreign colonists, although again in the process of being re-Judaized: Galilee was held in contempt by the Judaeans. Though the land's culture and civilization were in large measure Hellenistic, especially at the court of Herod Antipas, in individual towns and among the owners of large estates, the Jewish population, which spoke Aramaic, lived with its own, largely unaffected religious traditions. At the time of Jesus, Galilee was known as a seat of Jewish resistance to Rome.

According to Josephus (Antiquities XVIII, 18 ff.), Herod Antipas—whom Jesus called a "fox" (Luke 13:32)—held John the Baptist, the prophet who preached repentance, to be politically dangerous, had him put in prison, and had him executed for this reason. The Synoptic tradition, however, gives the Baptist's harsh criticism of Herod's unlawful second marriage as the reason (Mark 6:17–29).

Information about political conditions in Palestine at the time of Jesus is found mainly in non-biblical sources, especially in Josephus. Only a few details are mentioned in the Gospels. Such information is nevertheless significant as background for the story of Jesus, even if it does not contribute much to an understanding of his teaching. The attitude of the Jewish people to the foreign rule of the Romans was not uniform. There were conformists, especially among the priestly aristocracy in Jerusalem, and there were those who exhibited concealed and open resistance.

Various sectarian Jewish groups at the time of Jesus

Religious conditions. Judaism in the time of Jesus presents a disunited, fragmented picture, composed of widely different groups.

The Pharisees. In the reports of the Synoptic Gospels, the Pharisees serve almost entirely to exemplify his opponents. They are incensed by his preaching and behaviour, spy on him, press from the very beginning to have him done away with, and are, conversely, themselves attacked by him most fiercely as being self-righteous hypocrites. Debates with the Pharisees without doubt played an important role in Jesus' life. From the Gospels there has developed a crude popular view that "Pharisee" is synonymous with "self-righteous hypocrite." The New Testament sources, however, are to be used with discretion in this respect for the following reasons: (1) the later narrators were in large measure no longer conversant with the historical circumstances, especially because they were themselves outside the region of Palestine. As a rule, the Pharisees are introduced as a collective quantity in the Gospels but, in reality, were not a unified group. There are also sporadic references in the Synoptic tradition to the fact that Jesus maintained table fellowship with Pharisees (Luke 7:36; 11:37; 14:1). It is also worthy of note that they play no part in the Passion tradition (with the exception of the later legend in Matt. 27:62ff. about the Pharisees' requesting a guard at Jesus' tomb). (2) A Synoptic comparison reveals the tendency to give Jesus' opponents more concrete form, but in a schematic way. In the later texts, the Pharisees are frequently introduced as the constant foil for Jesus, whereas the older tradition speaks of Jesus' opponents in an indefinite way. (3) Matthew, especially, reflects the sharpened conflicts between Jews and Christians in the period after the destruction of Jerusalem (AD 70), when a theologically narrower brand of Pharisaism was finally asserting itself in the course of the religious reconstitution of Judaism. This later picture dominates the Talmudic tradition, but it may not be projected back into the time of Jesus.

Originating in the time of the Maccabees (or earlier, according to some scholars), the movement of the Pharisees (*i.e.*, the "separated ones") formed itself into a religious association composed chiefly of laymen from varied classes and callings. Its aim was strict adherence to the Torah (Law) in even the most remote areas of daily life, in order to realize the true Israel of God. This especially included the scrupulous observance of the individual ritual commandments for the practice of prayer and fasting, cultic purity, and the avoidance of all contact with the cultically unclean, whether that be lawless persons, sinners, corpses, animals, or unclean utensils. In the Pharisees' piety there was also to be found an eager longing for the future world of God, a doctrine of the resurrection of the dead, and a hope in the promised Davidic Messiah, who would establish his rule in Jerusalem and destroy the power of the heathen.

Jesus' relation to the Pharisees

In view of this religious situation, it is difficult to arrive at a uniform judgment on Jesus' relation to the Pharisees. Points of contact in matters of teaching definitely are present; *e.g.*, in the expectation of the resurrection of the dead, which they hold in common (Mark 12:25–27). Again, there are critical statements about formalized and hypocritical piety in Jewish Talmudic tradition, and not just in sayings of Jesus. It would therefore be unjust to judge all Pharisees to be alike. Obviously, many sayings of Jesus have parallels in Rabbinic tradition. Nevertheless, there is no question that Jesus rejected their claim to righteousness and their ideal of representing the true Israel, that he characterized their "tradition of the elders" as human tradition in contrast to the commandment of God, and that, through his attitude to tax collectors and sinners, he must have given them offense. Because of such opinions of Jesus, they probably would have influenced the people against Jesus. That certainly need not mean, however, that the Pharisees, who were politically anything but dominant, aimed at Jesus' crucifixion from the start (contrary to what is said; Mark 3:6).

The Sadducees. A party of quite another kind was that of the Sadducees, who belonged to the Jerusalem priestly caste. They carried much less authority among the people than the Pharisees. As a theologically conservative school, they differed from the latter also in their rejection of the additional "traditions" and the new doctrine of the resurrection of the dead. Because of the Sadducees' hierarchical tradition and their readiness to adapt themselves to the current political conditions, their influence in Jesus' time, before the destruction of the Temple, is not to be underestimated. Besides the Pharisees and the elders of the people, they had a decisive voice in the supreme religious and judicial authority, the Sanhedrin. A close relation probably existed between them and the Roman rulers. They did not, however, survive the catastrophic outcome of the war and the end of the Temple (AD 70).

The scribes. The scribes are frequently mentioned in the Gospels. In later Judaism, which, since the time of Ezra (5th century BC), was committed to the Mosaic Law, they formed a most respected class of the teachers. Corresponding to the normative significance of the Law for all religious, moral, social, and legal questions of Jewish life, the scribe was a combination of theologian and lawyer. Social origin and membership of a particular party played no role in this group. In Jesus' time, there were, apart from Pharisees and priests, also Sadducees and Zealots among the scribes. They were not paid as a professional class but, instead, had to find their own living. As scribes, they had to expound the Torah and give directives for daily life. Those who had undergone the long and careful training in their schools were accorded the status of scribe, wore the long robe of the scholar (Mark 12:38), were respectfully addressed as "rabbi" (Matt. 23:7), and were allowed to sit in a place of honour in the synagogue. Jesus, like the scribes, sat to teach (Matt. 5:1; Luke 4:20), engaged in debate, gave his opinion on the diverging doctrinal propositions of particular schools (Matt. 19:3ff.), and gathered disciples about him. The stereotyped way in which, particularly in Matthew, Pharisees and scribes are grouped together reflects the conditions obtaining at the time of the Gospel-writers, in which it was the Pharisees who controlled the instruction in the synagogues exclusively. But earlier, in the time of Jesus, the scribes were a more motley group. Also, it is not allowable to conclude from the fact that Jesus is frequently addressed as "rabbi" and "teacher" that he himself was a member of this profession.

Theocratic ideals and zeal for the Law among the Zealots

The Zealots. The involvement of the religiopolitical movement of the Zealots, a revolutionary group, in the historical development of Palestine was disastrous to the nation. No longer contented with the passive resistance of the Pharisees, out of whose ranks they certainly gained many adherents, the Zealots took the ideal of a theocracy and zeal for the Law extremely seriously. The first outbreak of their activities occurred in AD 6, when the Syrian legate Quirinius ordered the population in Judaea to register. This aroused indignation and was the signal for an insurrectionist movement, which confined itself initially to scattered individual acts of revolt but soon expanded, took military form, and finally instigated the Jewish–Roman war (AD 66–70). Biblical and nonbiblical sources name Judas, a Galilean scribe from Gamala, as founder of the Zealots. Like him, other fanatical messianic prophets also found significant followings. In Jesus' time, the conflict had not yet reached its zenith. The Zealots carried out sudden raids on the Roman occupation forces and conducted a guerrilla war from their hiding places in the wilderness. The Romans correspondingly held the land under strict control, reinforced their troops in Jerusalem at the times of the Jewish festivals, when great crowds of pilgrims gathered in the city, and took drastic and ruthless action if they anticipated sedition. This situation illuminates the events leading to Jesus' death. The Zealots' goals were political and, primarily, religious: the realization of a Jewish theocracy, the rule of the promised Messiah, and the destruction of the heathen regime.

The thesis that Jesus belonged to the Zealots or founded a related movement was first advanced in the 18th century and has repeatedly been supported in recent times. The most important point in its favour is Jesus' execution on the cross, a punishment that only the Roman authorities could inflict and did frequently against rebels. There were two others executed in the same manner with Jesus, and they, like Barabbas, who was granted amnesty in Jesus' place (Mark 15:15), are referred to as "robbers" (Mark 15:27), a customary term for rebels at this time. This could indicate that, at that Passover time, when many Jews were in the city, a Zealot revolt had been planned and was bloodily suppressed but also that Jesus had actually been willing to play a leading part in it.

Jesus' messianic entry into Jerusalem and the cleansing of the Temple (Mark 11) are also interpreted along these lines, the latter being understood as an attack on the dominant priestly class that sympathized with the Romans. Some also see a connection with the fact that one of the disciples was carrying a weapon when Jesus was arrested in Gethsemane (Mark 14:47). The later Christian tradition has, it is claimed, for apologetic and theological reasons, altered the true historical state of affairs until it has become unrecognizable. But isolated hints have nonetheless been preserved in it; *e.g.*, Jesus' critical sayings about that "fox" Herod (Luke 13:32) and the violent earthly rulers (Luke 22:25); similarly, the way he attracted Zealots, documented by the fact that among his disciples at least one, called Simon (Luke 6:15; Acts 1:13), was a Zealot.

There are, however, no sufficient reasons to support the hypothesis of Jesus belonging to the Zealots. The undeniable fact that he was crucified by the Romans as a political messianic pretender only proves that he was held to be a Zealot and was probably denounced as an enemy of the state, but not that he really was. The most important and decisive argument against the Zealotism assumption is found in Jesus' message of the dawning of the Kingdom of God, which belongs to the best established items in the tradition. It lacks any politico-nationalistic features and expressly says that God alone, and not any human activity, establishes his Kingdom (Mark 4:26–29) and offers his salvation to all without exception. If Jesus were directly or indirectly to be counted among the Zealots, this would mean at the same time that he must have fought to have the Law rigorously carried into effect and must have strictly avoided associating with sinners, especially with the tax collectors, who stood in the service of Rome. In the dialogue on paying tribute to Caesar (Mark 12:13–17), Jesus even expressly rejected rebellion against the Roman emperor, without thereby glorifying his regime.

Characteristics of the Qumrān community

The Essenes. Far removed from the above-mentioned religious groups were the sectarian, separatist Essenes, most probably identical with the sect of Qumrān (near the northwest bank of the Dead Sea). The sensational discovery in 1947 of many of their original writings (the Dead Sea Scrolls) and the later excavations of their settlement have extended knowledge of the Jewry of those times to an extraordinary degree and have occasioned the suggestion that both John the Baptist and Jesus came from this sect or were, at the least, heavily dependent on their teaching. Important arguments, however, speak against this assumption. This sect had arisen, like the Pharisees, in the 2nd century BC out of a conflict with the official priesthood in Jerusalem but had nevertheless preserved the priestly traditions and, at the same time, developed a strongly ritualistic practice of the Law. Characteristics of the Qumrān community are: its monastic seclusion from the outside world, including the rest of the Jews; the way it termed itself the "children of light" in contrast to the "children of darkness"; its rigid organization and discipline; and its apocalyptic expectations—centring on the intervention of God in history, along with dramatic and cataclysmic events—and other special features of its theology. Although the new texts found at the Dead Sea show numerous individual parallels to the Jesus tradition of the Gospels, there are already fundamental differences between the Qumrān sect and John the Baptist. His eschatological (last times) message of repentance addressed to the nation as a whole fits in with the sect as little as does his unique kind of Baptism—which one underwent once and for all—with the Essenes' regular ritual washings. Nor does John's temporal and geographic proximity to the strictly esoteric Qumrān sect justify asserting close relations between them. There also are diametrical differences between the views of the sect and the range of Jesus' ministry, his message of salvation, his understanding of God's will in a way free of all casuistry, and, especially, the radical character of his commandment of love and his fellowship with sinners and social outcasts.

THE LIFE AND MINISTRY OF JESUS

The birth and family of Jesus. The birth of Jesus. The life of Jesus will be dealt with in this section up to the beginning of his Passion. The course of his life and the geographical setting of his ministry can only be given in rough outline. The details are surrounded by many uncertainties. The period within which his ministry and death occurred may, however, be narrowed down with considerable accuracy on the basis of a synchronistic dating of the appearance of John the Baptist in the 15th year of Tiberius (Luke 3:1)—*i.e.,* AD 28/29—which is confirmed by nonbiblical sources. But the year and place of Jesus' birth are uncertain. Mark and John say nothing about them. The only sources for them are the widely divergent birth and childhood legends in Matthew 1 and 2, on the one hand, where Jesus' birth and early lot are set in the time of Herod I and the change of regime (4 BC), and, on the other hand, the narrative of Luke 2, which links Jesus' birth with the first registration in Judaea under the emperor Augustus (AD 6). (The use of BC [before Christ] and AD [Anno Domini, or "in the year of the Lord"] was not common until the Middle Ages.)

The tradition of Bethlehem as the place of Jesus' birth, which is found only in Matthew 2 and Luke 2, has its source in all probability in the Old Testament conception of the Messiah as a descendant of David. Early Christian-

Davidic descent of Jesus

ity took this view from the beginning. "Son of David" is found in many texts (*e.g.*, Mark 10:48) alongside other titles of Jesus. Its original political and national sense was abandoned. even though it is still recognizable in the acclamation of the people (Mark 11:10). The theological motif of Jesus' Davidic descent, however. did not necessarily involve the idea that he was born in Bethlehem, David's hometown. That is only the case in the opening stories in Matthew and Luke. They differ in that, in Matthew. Bethlehem is thought of as the parents' original place of residence, which they soon change to Nazareth because of the dangers threatening their child (*e.g.*, the flight to Egypt), whereas, in the Lucan story, Jesus' parents really live in Nazareth but stay in Bethlehem temporarily because they are obliged to register at the Davidic family's place of origin. Both traditions are to be judged as legendary variations of the theological theme of Jesus' messiahship, even though each in its own way assigns to his birth a place in history. The extent to which these texts are marked by theological motifs, above all by the thought that Jesus as Messiah fulfills the promises of the Old Testament and the hope of Israel and the world, is shown by the numerous quotations woven into the stories.

The widely differing genealogies in Matthew 1 and Luke 3 also belong in the context of the doctrine of the Davidic descent of the Messiah (Christ). They are the only New Testament evidences for genealogical reflection about Jesus' messiahship. The two texts, however, cannot be harmonized. They show that originally a unified tradition about Jesus' ancestors did not exist and that attempts to portray his messiahship genealogically were first undertaken in Jewish Christian circles with use of the Septuagint (Greek translation) text of the Old Testament. Both texts have to be eliminated as historical sources. They are nevertheless important for the development of Christology (doctrines on the nature of Christ), because they reveal the difficulty of reconciling the genealogical proof of Jesus' Davidic descent with the relatively late idea of his virgin birth.

This last tradition, too, is only recorded in two stories—in Luke 1 and Matthew 1—and was originally quite unconnected with the frequently found motif of Jesus' divine Sonship. Paul, John, and the rest of the New Testament writers are not acquainted with the idea. Also, it has left no traces in the rest of the Synoptic tradition, not even in the story of Jesus' birth (Luke 2:1–10), where Joseph and Mary appear as his natural parents. In Matthew 1, Jesus' miraculous birth is presupposed, and, in Luke, 1, it is explained more closely. This tradition is not to be traced back directly to the idea, widely held in classical antiquity, of heroes and great men who derived from the union of a deity with a human woman. In other words, Jesus is not characterized as a demigod here. What underlies this tradition is, rather, the concept of the creative power of God and his Spirit, which is known from Hellenistic Judaism. This theological, not biological, motif has been applied to Jesus and, with the greatest probability, only secondarily combined with the Greek version of the messianic promise of Isa. 7:14 (in the Septuagint, the Hebrew word *ʿalma—i.e.*, "young woman"—is translated as "virgin"), and in this way the Christian story came about. According to a very old, reliable tradition, the village of Nazareth, which lay in the Galilean hill country, had a Jewish population, and was untouched by the influence of the Hellenistic cities, was the hometown, and then certainly also the birthplace, of the "Nazarene" (Mark 1:24; 10:47; 14:67; 16:6).

The Jewish character of Jesus' parents, brothers, and sisters

The family of Jesus. Four of Jesus' brothers and several sisters are mentioned in Mark 6. (There is no basis in the text for making them into half brothers and half sisters or cousins, and to do so betrays a dogmatic motive.) All his relatives' names testify to the purely Jewish character of the family: his mother's name was Mary (Miriam), his father's, Joseph, and his brothers', James (Jacob), Joseph, Judas, and Simon (names of Old Testament patriarchs). The same is true of the name Jesus. In the Septuagint it is the customary Greek form for the common Hebrew name Joshua; *i.e.*, "Yahweh helps." It is also mentioned in Mark 6 that Jesus or his father (there are variant textual versions) was a carpenter. There are several not unimportant pieces of information preserved about the later history of the family. Of his father, who probably died early, little is mentioned. His mother, brothers, and sisters did not join his movement at first but, rather, disapproved of his behaviour (Mark 3:31–35). Mary is, however, mentioned as a member of the Christian Church after his death (Acts 1:14). The same is true of his brother James, whom Paul names among the witnesses of the Resurrection (I Cor. 15:7) and who was the leader of the Jerusalem Church after Peter (Galatians, Acts). The author of the Letter of James has taken a brother's name for himself, as did the author of the Letter of Jude in respect to another brother. According to a later nonbiblical account (in the *Ecclesiastical History* of Eusebius, a 4th-century historian of the church), grandchildren of Jude (who otherwise remains unknown), who were living in Galilee, were summoned by the emperor Domitian as "descendants of David," but then released as representing no political danger.

Jesus most likely grew up in the piety that was cultivated in the home and in the synagogue (including Bible study, obedience to the Law, prayer, and expectation of the final coming of the Messiah) and also took part in pilgrimages to Jerusalem. From these scattered reports it is possible to gain some information about Jesus' background and theological education. The latter also comes to light in his teaching and in the frequently attested honorific form of address "rabbi" (teacher), which, in the language of the time, was not yet confined to members of the trained and ordained profession of the scribes. Nothing is precisely known, however, about Jesus' youth and inner development. What is known is contained in the sole narrative in Luke 2:40–52 (the boy Jesus in the Temple) and the legendary apocryphal gospels, which, after the manner of legend, sought to illumine the obscurity of Jesus' childhood.

The Christian significance of John the Baptist

The ministry of Jesus. ***The role of John the Baptist.*** The Gospel accounts of the appearance and activity of John the Baptist and of Jesus' Baptism at his hands first establish a historically safe basis for knowledge of Jesus' life and work. Significantly, the oldest Gospel writer calls these events "the beginning of the gospel of Jesus Christ" (Mark 1:1), indicating that his would be a message about Christ, not a description of the contemporary background for Jesus' life. The Baptist is, therefore, represented exclusively from the Christian point of view. His place in the Christian history of salvation is that of a forerunner or pioneer; or he is a witness to Jesus, as in the Gospel According to John. But the tradition has nevertheless preserved unchallengeable information about John, especially in Q. Josephus characterizes him as a mere moral teacher and his Baptism as merely ritual washing. In reality, however, he made his appearance in the desert as a prophet of the imminent Last Judgment, calling all without exception to repentance in the eleventh hour, and baptized those who were ready to repent, in order to prepare them for the baptism of fire of the mightier one coming from heaven and to preserve them from his annihilating wrath (Matt. 3:7ff. and Luke 3:7ff.). His dress and diet as an ascetic nomad and, above all, the location of his ministry (the Judaean desert and the Jordan steppes) far away from the institutions and places of traditional religion and secularity illustrate the earnestness of his eschatological preaching and his attack on all conventional piety: but they also correspond to the old prophetic promise that God would encounter his people in the Last Days, as he did once before, in the desert. Historically, all these features may not be understood immediately in Christian perspective; *i.e.*, as pointing to Jesus as the Messiah. The tradition of the Gospels visibly and increasingly interpreted the history of the Baptist in retrospect, and not least for the reason that there still existed for a considerable time alongside the disciples of Jesus a rival body of disciples of the Baptist.

That Jesus was baptized by John, as all the Gospels record, indicates that in all probability Jesus initially be-

The significance of Jesus Baptism

longed to John's movement. The account of Jesus' Baptism is styled in the Gospels as an "epiphany (or manifestation) story" and deals with Jesus' installation at this time as Messiah (Mark 1:9–11). The announcement of the Kingdom of God by John and his call to repentance retained decisive significance for Jesus. His high estimate of the Baptist emerges unambiguously from the fact that he placed John above the prophets and called him the greatest among men (Matt. 11:7–11). He saw the signs of the approaching Kingdom of God in the work of the Baptist as in his own work, and he recognized the authority given John as being from heaven (Mark 11:27–33). These words carry all the more weight historically because the tendency of the context here is to proclaim Jesus as the Messiah and to place the Baptist, as the lesser, in Jesus' service. It is significant that John himself is nowhere attacked in the Synoptic texts, nor is he designated as a follower of Jesus. Wherever polemic can be recognized in the Gospels (especially in John), it is always directed against the false belief, doubtlessly held by the (later) Baptist disciples, that John was the promised Messiah. The extent to which the close connection between Jesus and John occupied the theological reflection, apologetics, and imagination of the Christian Church is shown by several passages and, above all, by the cycle of legends in the introduction to Luke (chapter 1). Regardless of the close relationship between Jesus and John the Baptist, especially in their prophetic announcement of the approaching Kingdom of God and their call to repentance (cf. Matt. 3:2; 4:17), there are also radical differences.

The *beginning* of Jesus' *ministry*. At the latest, after the Baptist's imprisonment (as the Synoptics state), possibly even earlier (according to John), Jesus began as a grown man (Luke 3:23) an independent public ministry, but in the villages of his Galilean homeland and—sporadically—in the neighbouring countryside, rather than in the wilderness, as did John. The real area of his ministry was the district on the northwest bank of the Lake of Gennesaret (or Sea of Galilee; the towns of Beth-saida, Chorazin, and Capernaum). The change of scene is significant in itself. Jesus did not call the people into the desert. He sought men in their settlements and took part in their ordinary life, and not as an ascetic, like John the Baptist (Matt. 11:18). He worked among them as a wandering preacher (Matt. 8:20) and charismatic miracle worker, without, however, baptizing like John. But the image he presents is nonetheless highly peculiar. He taught not only in the synagogues but likewise in the open air, on the shore of the lake, and on the road. There also were strange people in the group surrounding him: women, children, and many who were viewed as godless or unclean. Further, the manner of his teaching is surprising. He did not derive it from the Holy Scriptures, although he was familiar with them, esteemed them, and appealed to them here and there. Instead, he constantly presented the reality of God and the validity of his will in an immediate way and made them comprehensible to his hearers without using the established structure of sacred texts and traditions and without presupposing a conventional religious point of view. His metaphors, parables, and proverb-like utterances were not used to explain traditional teachings of biblical theology but, instead, appealed directly to the everyday experience and the understanding of his hearers, and they are therefore characterized by a unique self-evidence and a disarming simplicity.

Jesus' relationships to various people

This corresponds to the manner of his behaviour in his meetings with other people. The Gospels portray this in a large number of separate scenes. These persons vary considerably: pious and impious, rich and poor, respected and outcast, healthy and ill. In every encounter, Jesus' amazing sovereignty with which—free of all prejudices—he mastered the situation is made visible. He saw through his opponents' attempts to corner him in debate, disarmed their objections, saw the needs of the possessed and the sick who crowded around him, and associated with those who were avoided by others. Some of the scenes may only have been added or filled out in later popular tradition, but they clearly demonstrate the power with which Jesus helped people by word and deed, whether he grew passionately angry over the power of disease or over the pride and lovelessness of the "righteous" or whether he commanded the demons or blessed children and laid hands on the sick.

Origins of the disciples

The calling of Jesus' disciples. According to the unanimous witness of the Synoptic Gospels, Jesus gave rise to a movement in Galilee and found numerous followers, although not without provoking rejection as well. This movement cannot yet be called a "church." (This concept first appears in the later tradition.) To spread his message and movement, he called on his disciples, for the sake of the approaching Kingdom of God, to resolutely surrender all ties of family and work (Mark 8:34ff.; Matt. 10:37ff.; Luke 14:26ff.) and to follow him and to become "fishers of men" (Mark 1:17; Luke 5:10). Many of his words are of extreme sharpness and do not conceal how difficult the disciples' road will be (Luke 14:25–33). But, at the same time, the patent immediacy of Jesus' sovereign power comes to light in these texts. In the scenes mentioned, it is Jesus who makes the decision. He calls, appoints, and selects particular men, without regard to their origin and previous training. There are fishermen (Andrew, Peter, James, and John), a tax collector (Matthew), and Zealots (Simon and, perhaps, Judas Iscariot) among them, perhaps also a few craftsmen and peasants. Whether it was a circle of 12 disciples from the start is questionable and under debate. It is clear, however, that he commissioned and authorized his disciples to preach and to drive out demons (Mark 3:14). Some of these disciples are well noted in the Synoptic tradition (*e.g.*, Peter and Judas Iscariot). In the Gospel According to John, others come into the foreground, including some from among the followers of the Baptist. Of others, only their names are known (*e.g.*, Thaddaeus). A characteristic of these companions of Jesus is that their discipleship is not, as with the rabbis, a transitional stage that ends with their "training." None of them moves up after sufficient study to the status of "master" (Matt. 23:8). Even if accounts of the calling of disciples have, in general, been styled in the later tradition as examples of what it means to be a Christian and individual scenes have been added to the original stock of stories, the recollection of incidents that occurred during Jesus' ministry in Galilee is doubtlessly preserved in the texts.

The Galilean period. The loose and often differing order of the individual scenes only entitles scholars to speak of a rather ambiguous Galilean period of Jesus' activity: they cannot say with certainty how long it lasted. Because the Synoptic Gospels mention only one trip of Jesus to Judaea and Jerusalem, with the Passion following it, the impression is created that the period lasted no longer than one year. Editorial and theological considerations have, without question, also played a part in this presentation (*e.g.*, Jesus' activity in Galilee and his sufferings in Jerusalem). Scholars offer several good reasons, however, to support the assumption that the Synoptic outline still deserves to be preferred to the widely differing one in John. In the latter, Jesus is in Jerusalem for three celebrations of the Passover (John 2:13–23; 6:4; 11:55), as well as for one Sukkot (Feast of Tabernacles; John 7:2) and one Hanukka (Feast of Dedication; John 10:22). This involves a period of more than two full years. It is doubtful, however, that John is based on an independent tradition, because the indications of time referred to serve the Evangelist as a means of changing the scene of Jesus' ministry between Jerusalem and Galilee. (The centre here is Jerusalem.)

THE MESSAGE OF JESUS

***The* Kingdom of God.** Jesus announced the approaching Kingdom of God and therefore called people to repentance. The first two Gospels have set this at the beginning in a programmatic saying as a summary of his preaching and have thus characterized the central and dominant theme of his mission as a whole (Mark 1:15; Matt. 4:17). Thus, the Kingdom of God, or Kingdom of Heaven (a Jewish circumlocution for God preferred by

The central and dominant theme of Jesus' teaching

Matthew), does not just denote a final chapter of his "system of doctrine" (a concept that cannot be applied to Jesus, in any case). The underlying Jewish word (*malkhuta)* means God's kingship, and not primarily his domain. This meaning prevails in the New Testament texts. But Kingdom of God or Heaven is also used in a spatial sense ("Enter . . ."). The burning expectation of the Kingdom of God was widely spread in contemporary Judaism in manifold form, based on the Old Testament faith in the God of the fathers, the Creator and Lord of the world, who had chosen Israel to be his people. But with this faith there had united itself the contradictory experience that the present condition of the world was ungodly, that Satanic powers reigned in it, and that God's kingship would only manifest itself in the future. In wide circles, this expectation had the form of a national, political hope in the Davidic Messiah, though it had expanded this hope in apocalyptic speculation to a universal expectation. In each case it was directed toward the Last Days. Likewise, in Jesus' message, the expression Kingdom of God has a purely eschatological—*i.e.*, future—sense and means an event suddenly breaking into this world from the outside, through which the time of this present world is ended and overcome.

These traditional motifs of the end of the world, the Last Judgment, and the new world of God are not lacking in the sayings of Jesus preserved in the Gospel tradition. Thus, Jesus has not by any means changed the Kingdom of Heaven into a purely religious experience of the individual human soul or given the Jewish eschatological expectation the sense of an evolutionary process immanent in the world or of a goal attainable by human effort. Some of his parables have given rise to such misunderstanding (*e.g.*, the stories of the seed and harvest, the leaven, and the mustard seed). In such cases, the modern thought of an organic process has been wrongly introduced into the texts. People of classical and biblical times, however, heard in them connotations of the surprising and the miraculous. The Kingdom of God, thus, is not yet here. Hence the prayer, "Thy kingdom come!" (Matt. 6:10; Luke 11:2), and the tenses, for example, in Jesus' Beatitudes and predictions of woe (Luke 6:21–26). The poor, the hungry, and the weeping are not yet in heaven. The petitions of the Lord's Prayer presuppose the deeply distressing circumstance that God's name and will are abused, that his Kingdom is not yet come, and that men are threatened by the temptation to fall away.

Eschatological motifs

In regard to Jesus' preaching, one cannot, therefore, speak of a realized eschatology—*i.e.*, the Last Times are now here (according to the view of C. H. Dodd, a British biblical scholar)—but of an eschatology "in process of realizing itself" (according to the view of Joachim Jeremias, a German biblical scholar); for God's Kingdom is very close. It is on the threshold, already casts its light into the present world, and is seen in Jesus' own ministry through word and deed. In this, his message differs from the eschatology of his time and breaks through all of its conceptions. He neither shared nor encouraged the hope in a national messiah from the family of David, let alone proclaimed himself as such a messiah, nor did he support the efforts of the Zealots to accelerate the coming of the Kingdom of God. He also did not tolerate turning the Kingdom of God into the preserve of the pious adherents of the Law (Pharisees; Qumrān sect), and he did not participate in the fantastic attempts of the apocalyptic visionaries of his time to calculate and thus depict in detail the end of the present world and the dawn of the new "aeon," or age (Luke 12:56). Nor did he undertake a direct continuation of the Baptist's preaching.

All the ideas and images in Jesus' preaching converge with united force in the one thought, namely, that God himself as Lord is at hand and already making his appearance, in order to establish his rule. Jesus did not want to introduce a new idea of God and develop a new theory about the end of the world. It would therefore be incorrect to understand his preaching in the Jewish apocalyptic sense of immediate expectancy, coming, as it were, to a boiling point. The proximity of the Kingdom of God actually means that God himself is at hand in a liberating attack upon the world and in a saving approach to those in bondage in the world; he is coming and yet is already present in the midst of the still-existing world. In Jesus' message, God is no longer the prisoner of his own majesty in a sacral sphere into which pious tradition had exiled him. He breaks forth in sovereign power as Father, Helper, and Liberator and is already now at work, as is indicated by Jesus' proclaiming of his nearness and by Jesus' actions in entering the field of battle himself, to erect the signs of God's victory over Satan: "But if it is by the finger of God that I cast out demons, then the kingdom of God has come upon you" (Luke 11:20). For this reason, Jesus called out: the shift in the aeons is here; now is the hour of which the prophets' promises told (Matt. 11:5; Isa. 35:5). This "here and now" carries all the weight in Jesus' message: "Blessed are the eyes which see what you see! For I tell you that many prophets and kings desired to see what you see, and did not see it, and to hear what you hear, and did not hear it" (Luke 10:23–24). In answer to the Pharisees' question about when the Kingdom of God is coming, Jesus therefore said, "The Kingdom of God does not come in an observable way, nor will they say, 'Look, here it is!' or 'There!' For look, the Kingdom of God is within your reach" (Luke 17:20–21; another translation: "in the midst of you").

The dominant feature of Jesus' preaching is the Heavenly Father's turning in mercy and love to the suffering, guilty, outcast, and to those who, according to the prejudices of the "pious," have no right to receive a share in the final salvation. Numerous parables described how God behaves toward them and shows himself as Lord and King (*e.g.*, Luke 15; Matt. 18:23ff.; 20:1ff.). They all speak of God's action in images drawn from daily life, so that everyone can understand. They belong to the uncontestedly oldest stock of the Jesus tradition. But Jesus did not only teach this, he practiced and illustrated it himself by his own behaviour and thereby offended the pious, who claimed the Kingdom of Heaven for themselves.

The message of repentance

In this message of the approaching Kingdom of God, Jesus' call to repentance is grounded. He called on all not to miss the hour of salvation (Luke 14:16ff.; 13:6ff.), to sacrifice everything for the Kingdom of God (Matt. 13:44ff.), and to receive it like a child (Mark 10:15), without the presumptuous and desperate conceit that one might win it and realize it by one's own works (Mark 4:26ff.; Matt. 13:24ff.). Jesus' summons to be wise, to be on the watch (Luke 16:1ff.; 12:35ff.; Mark 13:33ff.; Matt. 24:45ff.), and to surrender the fiction of one's own righteousness (Luke 18:10ff.) belongs here, too. In Jesus' preaching, repentance does not mean a prerequisite or precondition or even a penitent contemplation of oneself but, rather, a consequence of the proximity of the Kingdom of God (Matt. 4:17) and an opening of oneself for his future, a movement not backward, but forward. Jesus in this way binds future and present insolubly together. The apocalyptic's question about how much time still has to elapse before the new world of God is here is thus rendered meaningless. He who asks this only proves that he understands neither the future nor the present properly; namely, God's future as the salvation that is already dawning and one's own present in the light of the coming Kingdom of God.

Jesus therefore rejected the demand that he produce "signs" as proof of the dawning of the time of salvation (Matt. 12:38ff.; Mark 8:11). He himself is to be viewed as the "sign," just as once Jonah, the prophet of repentance, was the only sign given to the people in Nineveh (Luke 11:29ff.). The sign is not identical with the thing signified, but it is a valid indication of it.

According to the Synoptics, Jesus never made his "messiahship" the subject of his teaching or used it as legitimation for his message. It is significant that the "I am" sayings of John, which bear the stamp of Christology throughout, are not found in the Synoptic tradition. That does not in any way affect the fact that Jesus in a decisive way included his own person as eschatological prophet and charismatic miracle worker in the event of the Kingdom of God: "And blessed is he who takes no offense at me" (Matt. 11:6).

Eschatology and ethics

The will of God. In Jesus' teaching, the nearness of God is itself viewed as a moving force. It creates, as it were, a field of force and challenges the whole person to obey the will of God unconditionally ("Let your loins be girded and your lamps burning"; Luke 12:35). As little as Jesus tolerated attempts at calculating the time when the Kingdom of God should come, so much the more did he demand that men reckon with its coming. The relation between eschatology and ethics in Jesus' teaching, however, needs to be further clarified. His commandments nowhere have the character of prophetic sayings, and their content is not given eschatological basis even where Jesus linked them with the promise of heavenly reward and, correspondingly, with the threat of damnation in the Last Judgment (*e.g.,* Matt. 24:24ff.; Luke 19:11ff.). God's will is valid in itself, always and everywhere. For this reason, it is incorrect to characterize Jesus' demands as "interim ethics"; *i.e.,* as exceptional emergency laws in the situation of the world that lies in the blaze of the cosmic catastrophes accompanying the shift of the aeons and the speedy dawn of the Kingdom of God (as did Albert Schweitzer, a great Alsatian theologian, medical missionary, and Nobel laureate). Jesus did not draw arguments for his ethical demands from the perishing order but, rather, from the existing world, the Old Testament commandments, the creation, and experiences known to everyone. Thus, he did not aim at forming a "holy remnant," which would escape the rejection awaiting others in the Last Judgment, on the basis of some kind of select monastic rule.

The certainty of God's nearness is, nevertheless, the open or concealed point of reference for Jesus' exposition of the will of God and explains his attitude to the Old Testament Law. Corresponding to the character of the Old Testament legal tradition, he refers to the will of God in single sayings and in comments in relation to individual commandments, and, it should be noted, he did not develop these into coherent "moral teaching." Rather, he took up quite different kinds of commandments as concrete examples, above all from the Decalogue and related texts, concerning one's behaviour toward one's fellow human beings (on murder and anger, adultery and divorce, oaths, retaliation, love for others; see Matt. 5: 21ff.) and also ceremonial commandments (concerning the Sabbath, prayer, fasting, and defilement) and other cultic duties. Jesus always went to the root of these commandments, and he did not content himself with the mere letter of the Law but disclosed within the Law—sometimes even against the letter of the Law (Mark 10:1ff.)—the genuine will of God. Though Jesus respected the Law, it was no longer for him the only source of knowledge of God's will and no longer the intermediate authority that exclusively mediates people's relation to God. From this basis are to be understood both Jesus' exposition of the Law and also his criticism of all formalistic casuistry, which is for him only "human tradition."

Jesus thus brings about a confrontation between the reality of God, which is no longer disguised by holy letter and tradition, and the similarly undisguised reality of humanity. People also can no longer delude themselves into believing that their pious works would represent them before God and thus keep on piling them up, as it were, like the Pharisee (Luke 18:11ff.). What God wants from humanity is not something but humanity itself, unconditionally and undividedly. The classic passages for these thoughts are the antitheses of the Sermon on the Mount (Matt. 5:21–48). They sharpen God's demands to the utmost extreme and leave no room for merely legalistic behaviour. Their leitmotiv is: "Not only, but even. . . ." Even anger, the lustful look, the "legal" divorce, retaliation that keeps within the limits prescribed, and love that excludes the enemy are against God's will.

These extreme demands are meant not so much to be paradoxically overdemanding as, rather, liberating. Firstly, they are formulated in a way that everyone can understand. They include numerous references to the natural, unperverted practices of people in their daily lives. Secondly, the demands do not describe an unattainable distant goal, which all human action must of necessity fail to meet. Rather, Jesus pointed again and again to what the heavenly Father has done, does, and will do with his children and to God's possibilities, which are unlimited, whereas a person might despair of his or her own limited possibilities and impotence (Mark 10:27). Jesus' sayings about faith (Mark 9:23ff.), prayer (Luke 11:1ff.; Matt. 6:1ff.), or worry (Matt. 6:25ff.) are examples of this. Wherever Jesus calls on people to decide for themselves for God, he bases the argument on the fact that God has already decided for humanity. The unlimited readiness to forgive that he calls for also has its motivation in the limitless mercy of God, which he demonstrates toward the guilty in unfathomable measure (Matt. 18:23ff.). Jesus draws his hearers into this relation to God and, therefore, does not engage in abstract reflections about whether his demands are capable of fulfillment. In this way, what a person loses is the characteristic of being able to attain meritorious achievements (Matt. 20:1ff.). On the other hand, Jesus certainly did not give up the thought of "reward." The reward, however, is not a material prize, although images of this kind are not lacking, but the confirmation and perfection of the relation to God (Matt. 25:14ff.). The idea that human beings could claim and charge payment from God is for Jesus completely excluded (Luke 17:10).

The nearness of God, the real God, also brings humanity, no longer graded and classified in traditional categories, into urgent and imperious proximity. How much Jesus was concerned with human beings is shown especially by his commandment of love, which he not only taught but also practiced to the point of offensiveness. In it is concentrated the "better righteousness" that he demands of his disciples (Matt. 5:20). Jesus has taken over the Old Testament dual commandment of love of God and one's neighbour (Deut. 6:5; Lev. 19:18), which is also in Judaism a summary of the whole Law. But it is characteristic of Jesus' preaching (1) that he consistently subordinated all other laws—*e.g.,* the Sabbath commandment—to this highest critical standard (*e.g.,* Mark 2:27; 3:4), and (2) that he extended and heightened love of one's neighbour to love of one's enemies (Luke 6:27ff.), and (3) that his commandment does not have the abstract ideal of a general philanthropy at its root. Rather, he directed his hearers into the situations—always eventful and concrete—where they encounter their enemy (Matt. 5:38ff.) and their fellowman in need (Luke 10:25ff.). Behaviour toward one's fellows is so important for Jesus that it is all that is spoken of in many of his utterances, without any mention of the first commandments of the Decalogue concerning behaviour toward God (*e.g.,* Matt. 5:25ff.; 7:12; 19:16ff.).

Individual and social ethics

The distinction that modern moral philosophy makes between individual and social ethics has, in respect to Jesus' teaching, only limited application. To be sure, Jesus did not draw up a program for a new order for the world and the nations, he did not demand a more just distribution of property, did not fight against the differences existing between masters, slaves, and hired workers, and did not give any directives for a better administration of justice. The world he had before his eyes was the world as it was, within the horizon of Palestinian Jewish rural conditions, and not the world as it ought to be. His sayings, parables, and illustrations show how keenly he assessed everyday life and how clearly he described it in his graphic, vigorous way—not glorifying this world as an eternally valid, divinely willed order, and also not getting morally indignant about it. Rather, he calls on people to behave in this given world in conformity to the original will of God and his dawning Kingdom; *e.g.,* to renounce the reign of mammon (Matt. 6:24; Luke 16:9ff.). Jesus did not, however, require a complete surrender of property from everyone. His followers were not to avail themselves of the legally regulated facilities for asserting one's own rights and were not to conform to the ways of customary behaviour in the world. The assertion that the world cannot be governed with the Sermon on the Mount is thus not to be denied. Jesus' sayings about retaliation and his commandment of love are not juristically practicable as they stand, because they can only serve as a guide for the one who has been wronged by someone else or who is required to divide his possessions with another person. Legislators and judges have to decide exclusively about the rights of others and must restrain evil for the sake of the general social order. But the truism about the impracticability of the Sermon on the Mount conceals the fact that Jesus' teaching contains strong impulses toward social criticism.

Jesus unmasks as hollow conventions many ostensibly valid standards, explaining the Law according to the norm of the commandment of love and applying it to concrete situations. For this reason he also resists egocentricity, not only of individuals but of entire religiously and socially privileged groups, and joins with discriminated-against people (*e.g.*, heathens, Samaritans, tax collectors, and harlots). Thus, Jesus calls on men to live a life that corresponds to the proximity of God's Kingdom, although the validity and urgency of his commandments require no apocalyptic basis. The act of their proclamation, however, stands nonetheless close to Jesus' own mission (Luke 11:32ff.). Whether, and in what way, he expressed the fact of his mission by the use of Christological titles is not thereby decided. Jesus knew that he had been sent to the "lost sheep of the house of Israel" (Matt. 15:24; 9:36). His ministry, seen as a whole, was confined to the sphere of his own people. Only a few significant words and scenes point forward to the inclusion of non-Jews, in a new, eschatological people of God (Matt. 8:11ff.). Jesus, however, did not organize a mission to the heathen (Matt. 10:5ff.) nor a worldwide "church." The only saying of this kind, spoken to Peter (Matt. 16:17ff.), has been placed in the mouth of the earthly Jesus by the later church and clearly reflects its situation, doctrine, and discipline. But Jesus certainly did call into existence a movement in Galilee and allowed at least the narrower circle of his disciples, if not all of his followers, to share in his wanderings and ministry. Later tradition first identified the latter group alone with the Apostles (authorized emissaries), the circle of whom was, however, not originally restricted to that group (*cf.* I Cor. 15:5ff.). The number 12 symbolizes the 12 tribes of Israel. If Jesus appointed these disciples himself, he thereby demonstrated his eschatological claim on the whole of Israel. According to the saying in Matt. 19:28 and Luke 22:30, which was probably not formulated until later, he conferred on them the office of ruling and judging the perfected Israel of the new aeon.

THE SUFFERINGS AND DEATH OF JESUS IN JERUSALEM

Jesus' decision to go to Jerusalem

Jesus' decision to go to Jerusalem is the turning point in his story. The events it set in motion soon came to have decisive significance for the faith of his followers. It is not coincidental that the Gospels narrate this period of his life in disproportionate breadth. Despite the many points of agreement among the Gospels, there also are considerable discrepancies within the tradition of the Passion. Thus, one cannot expect the tradition of the Passion to provide historically accurate reports, for it has been formed from the viewpoint of the church and its faith in Christ. The most important theological motifs in the narratives include the intention of presenting Jesus' sufferings and death as the fulfillment of God's will, the decision, in conformity with the words of the Old Testament Prophets and Psalms, to proclaim him as Messiah and Son of God, despite his brutal end. Nevertheless, important historical facts may be inferred from the texts.

Jesus probably went to Jerusalem with his disciples for the Passover in order to call the people of Israel gathered there to a final decision in view of the dawning Kingdom of God. He must have been aware of the heavy conflicts with the Jewish rulers that lay ahead of him. The story of the cleansing of the Temple, in particular, shows that Jesus did not avoid these conflicts. The later tradition, stylizing the story, gives as Jesus' sole motive for going to Jerusalem his desire to die there and to rise again in accordance with the will of God (Mark 8:31; 9:31; 10:32ff.). The best clue for a reconstruction of the outward course of Jesus' Passion is given by his Crucifixion. It proves that he was condemned and executed under Roman law as a political rebel. All reports agree that he died on Friday (Mark 15:42; Matt. 27:62; Luke 23:54; John 19:31). They differ, however, in that, according to the Synoptics, this was the 15th of Nisan (March/April); *i.e.*, the first day of the Passover. But, according to John, it was the previous day; *i.e.*, the one on which the Passover lambs were slaughtered and on which the festival was begun in the evening (in accordance with the Jewish division of days) with a common meal. Thus, according to John, Jesus' last meal with the disciples was not itself a Passover meal but took place earlier. Each of these datings may be theologically motivated, whether it be that the Eucharist is to be represented as the Passover meal (Synoptics) or whether Jesus himself is to be shown as the true Passover lamb, who died at the hour when the lambs were slaughtered (John). Historically, the Johannine dating is to be preferred, and the 14th Nisan (April 7) is to be regarded as the day of Jesus' death. The question of the occasion for Jesus' execution and the role that the Jews played is thereby more difficult and more important.

The date of Jesus' death

The way the Gospels present the facts of the case, Jesus was actually condemned to death by the supreme Jewish tribunal (Mark 14:55ff.). Pilate, on the other hand, was convinced of Jesus' innocence and made vain attempts to release him but finally yielded to the Jews' pressure against his better judgment (Mark 15:22ff.). The historical reliability of this account has rightly been questioned. First, the Synoptic reports differ among themselves. According to Mark and Matthew, the Jewish supreme court had already gathered in the home of the High Priest after Jesus' arrest in the night of Holy Thursday to Friday and condemned him to death as a blasphemer at that point (Mark 14:64). Thereafter, they resolved to hand Jesus over to Pilate in a new session in the early morning (Mark 15:1). Luke knows of only one session and has the interrogation take place in the morning (Luke 22:66), but he says nothing about Jesus' condemnation (Luke 22:71). John deviates even more; here, only the high priests Annas and Caiaphas are involved in the interrogation of Jesus (John 18:13ff.). Secondly, with regard to all the Gospel accounts, the question arises, what earwitness can be supposed later to have given the disciples an exact report? Thirdly, the jurisdictional competency of the Jewish Sanhedrin is disputed. In the opinion of some scholars, the Jewish authorities were permitted to pronounce sentence of death and to carry it out by stoning in the case of serious religious offenses (blasphemy). In the opinion of others, though, this required the confirmation of the Roman procurator. Also, trials of this kind were not to be conducted during the period of the festival.

The strongest argument against the Synoptic presentation is, however, that it is styled throughout in a Christian, and not in a Jewish, way; *i.e.*, on the basis of scriptural proof and the Christian confession to the messiahship and divine Sonship of Jesus. The High Priest's question, "Are you the Christ, the Son of the Blessed?" (Mark 14:61), is unthinkable from the viewpoint of Jewish premises, because Son of God was not a Jewish title for the Messiah. Thus, the account reflects the controversies of the later church with the Judaism of its day.

There also is in the Gospels a tendency to exonerate Pilate at the Jews' expense. His behaviour, however, does not match the picture that nonbiblical sources have handed down about him. But everything speaks for Jesus' having been arrested as a troublemaker, informally interrogated, and handed over to Pilate as the leader of a political revolt by the pro-Roman priestly and Sadducean members of the Sanhedrin, who were dominant in Jerusalem society in those days. The cleansing of the Temple and a prophetic, apocalyptic saying of Jesus (John 2:19; cf. Mark 14:58; Acts 6:14) about the destruction of the Temple may thereby have played a role. It can hardly be assumed that each and all of the Pharisees, who were without political influence at that time, were involved in the plot. Nor are they mentioned as a separate group in the Passion narratives alongside the priests, elders, and scribes.

The other scenes in the Passion story do not need to be listed here separately. They relate more to the theological meaning of Jesus' Passion and are, to a large measure, formed in a cultic manner, even though they refer to events that are certainly historical; *e.g.*, Judas' betrayal, Jesus' last meal with his disciples, and Peter's denial of Jesus. The traces of an eyewitness account are perhaps still recognizable at certain points (Mark 14:52; 15:21).

The last words of Jesus

The accounts differ in their presentation of Jesus' death, especially in their rendering of his last words. It is only in Mark and Matthew that Jesus dies crying out the prayer from Psalm 22: "My God, my God, why hast thou forsaken me?" The distinction between the repentant and the defiant thief is only found in Luke. Jesus' last words are given differently in Luke ("Father, into thy hands I commit my spirit!") and John ("It is finished"). Each of these accounts, as also the testimony of the Roman centurion ("Truly this man was the Son of God!"; Mark 15:39), gives expression to the significance of Jesus and his story.

THE STORY OF JESUS AND FAITH IN JESUS

Did Jesus' violent death render his mission and story meaningless? In other words, did he enter definitively into the past as a failure and thus in this sense become the "historical" Jesus? For Pilate and the Romans, as for Jesus' Jewish opponents, there was no longer any problem. The decision had been taken. But Jesus' disciples were faced with this pressing question all the more. Their hopes were bitterly disappointed (Luke 24:13ff.). According to the unanimous witness of the New Testament texts, they did not find the answer themselves but were given it soon after Jesus' death through the Easter (Resurrection) appearances of Jesus (I Cor. 15:3ff.; etc.) and the experience of his Presence in the Spirit. The faith of early Christianity, with all of its practical and theological manifestations, grew out of this. This faith was not the preserve of a few enthusiasts or the personal opinion of individual Apostles. Wherever there were early Christian witnesses and communities, they were all united in believing and acknowledging the risen Lord (I Cor. 15:11).

The forms and ideas in which this faith found expression were various. According to the oldest view, Jesus' Resurrection meant his exaltation to divine lordship and was not necessarily connected with the tradition of the finding of the empty tomb, as the Gospels variously relate it. The theory of the resurrected one's having walked on earth for 40 days and only subsequently ascending into heaven is only found in Acts (1:3). Thus, there exists an undeniable tension between the unequivocal nature of the Easter *message*, on the one hand, and the equivocal nature of the Easter *accounts* and the historical problems connected with them, on the other. But the phenomenon of the whole Gospel tradition, rightly understood, is an expression of the faith in the living Christ without which neither a single word or deed of Jesus nor his Passion would have been handed down at all. The New Testament tradition does not aim at preserving the memory of Jesus as a figure of the past and telling only who Jesus *was*, but it wants to proclaim who Jesus is.

It may seem surprising that the question of Jesus' awareness of being the Messiah has been scarcely discussed in this article, let alone been given a precise answer. Usually, decisive significance is assigned to this question. Many scholars believe that access to the historical Jesus is only to be gained through the fact that Jesus had such an awareness in association with particular titles, such as Son of God, Son of man, and Messiah. In just the same way, they believe that the rise of the early Christian faith can only be understood by the same means. In light of the fact that the Gospels portray Jesus as the Christ (Greek term for the Messiah) and that numerous other titles of a similar kind are also found in them, the importance of this question is not to be underestimated. But it must nevertheless be noted that the Gospels are interested in the fact that Jesus was and is the Messiah, but not in his "consciousness" and inner development in a modern sense. The stories of Jesus' Baptism, temptation, and Transfiguration, for example, are not reports of experiences. But the question of whether Jesus applied one or several of those titles to himself still needs to be examined carefully, for each of them implies thoughts and concepts that must be of considerable relevance for his preaching and ministry. On this question, the opinions of scholars diverge widely, but it is uncontested that Jesus related his mission and activity in a unique way to the dawn of the Kingdom of God. It is another—and doubtful—question, however, that he expressed this understanding of himself through any traditional title.

Jesus' use of messianic titles

Three observations are important for the discussion of these problems. Firstly, in the incontestably authentic texts in the Synoptics, Jesus never makes his own status a special topic of his teaching or the recognition of his rank a condition of salvation. Secondly, it is not only presumed but—by means of a comparison of the parallel texts and their modification from one Gospel to the next—often capable of proof and, in other cases, requiring to be assumed that the faith of the later church has had a major influence on the formation of the Christological texts. A third observation is also not without its importance. Wherever in such texts Jesus talks about the Messiah and the Son of David, the Son of man, the Son of God, and the Lord, there is never any indication that he is using these titles in a completely new sense. The meaning they have, however, is no longer congruent with the ideas that Jesus' contemporary hearers must have connected with the titles, to the extent that they were not completely unknown to them. Because the historical Jesus indubitably wanted to be understood, the critical question necessarily arises about texts of this kind reflecting the views of the later church and its environment.

Views about traditional messianic titles

Some of the traditional titles could not possibly have been used by Jesus with reference to himself. In those days his hearers could only have understood "Messiah" or "Son of David" in the political or national sense, which conflicted with Jesus' intentions. Also, the exclusive title "Son of God" must have been incomprehensible to the Jews of Palestine, although not to the later hearers of the Christian missionary preaching in the non-Jewish, Hellenistic world. The same applies to the expression "the (divine) Lord," which for Jews was reserved for God alone. Some scholars believe that Jesus understood himself to be the suffering servant of God, of whom Isaiah 53 speaks. But in the Gospels reference is hardly made to this important chapter. The sole passage (Mark 10:45) does not, at least in the form handed down, reproduce an authentic saying of the Lord but contains an interpretation of Jesus' death that goes back to the Greek-speaking Jewish Christian Church.

Thus, the problem is narrowed down to the question about Jesus' calling himself the "Son of man." This concept, which is frequently found in his statements about himself, is a title of sovereignty. It stems from Jewish apocalypticism and means not a normal human being but, rather, the mythical figure of the Judge of all the world, who will come on the clouds of heaven at the end of the days (Dan. 7:13ff.; etc.). An early group of Jesus' sayings (Mark 13:26; 14:62; Matt. 24:27) speaks of the Son of man in this eschatological, future sense and always in the third person, yet in some texts in such a way that Jesus does not explicitly identify himself with this Son of man (Mark 8:38; Luke 9:26; 12:8ff.). Two other groups of sayings speak of him quite differently. One speaks exclusively of his suffering, dying, and rising again in accordance with the will of God (Mark 8:31; 9:31; 10:33ff.); the other, of his authoritative work and wanderings on earth (Mark 2:10–28; Matt. 8:20; 11:19), both without a view of the Last Judgment. No Jewish hearer of Jesus could have recognized the apocalyptic Son of man in these sayings. Loosed from the ideas traditionally linked with it, the concept has here been given completely new contents in a retrospective view of Jesus' ministry and end. Thus, both of these groups of sayings are only to be understood from the point of view of the later church. Therefore, only some sayings of Jesus of the first group probably come into question as authentic ones. If Jesus spoke of the coming Son of man, those sayings prove he was speaking in the apocalyptic language and concepts of his day in order to express the promise that his disciples' loyalty to him will be recognized and confirmed in the Last Judgment. The relation of his earthly person to the figure of the coming Judge is not thereby made the subject of reflection. The Jesus tradition has gone through a process of modification, and the faith of the later church has made a major contribution to the formation of the tradition, whatever its precise extent may be.

The post-Easter message of salvation

In the post-Easter message of salvation, the eschatological here and now belongs inseparably to Jesus' message of the Kingdom of God and was being realized in him. In the face of unbelief and doubt, the Gospels do not just offer an account of the history of Jesus as it transpired, but they interpret it as God's history with the world, as the decisive, redemptive, and ultimate act and word of God for the world. All titles of sovereignty that faith has assigned to Jesus express the fact that in him the turning point of the ages, the inauguration of salvation. and the nearness and presence of God have arrived. The special character of the Gospel tradition should therefore be understood in this sense. This tradition has not replaced the historical Jesus with a mythical Christ but has made explicit the Christology that was secretly implicit in Jesus' words, works, and way, although without titles of sovereignty and supernatural traits. The question appropriate to the Gospel tradition would, therefore, not be about what has happened to Jesus of Nazareth in the course of the development but, rather, why the first Christians held fast to him. To ask in this way and to accept the answer of the Gospels are matters for faith. It goes beyond the limits of historical research. (G.Bor.)

BIBLIOGRAPHY

Times and environment: BO REICKE, *Neutestamentliche Zeitgeschichte* (1964; Eng. trans., *The New Testament Era*, 1968).

The life and ministry of Jesus: Valuable surveys of premodern material are provided by R.M. GRANT, *The Earliest Lives of Jesus* (1961), for the early patristic period; and HARVEY K. MCARTHUR, *The Quest Through the Centuries* (1966), especially for the 14th and 16th centuries. Modern historical study of the Gospels dates from the 18th century and was characterized in the 19th and early 20th centuries by the Life-of-Jesus movement. The literature is authoritatively surveyed in ALBERT SCHWEITZER, *Von Reimarus zu Wrede* (1906; 3rd ed., 1954; Eng. trans., *The Quest of the Historical Jesus*, 1910); and C.C. MCCOWN, *The Search for the Real Jesus* (1940). MAURICE GOGUEL, *La Vie de Jésus* (1932; Eng. trans., 1933), is a late biography especially valuable for the inclusion of detailed evidence frequently assumed in subsequent works. JOSEPH KLAUSNER, *Jesus of Nazareth* (1925; Eng. trans. from the first Hebrew ed., 1922), remains important for its collection of relevant rabbinic materials. S.J. CASE, *Jesus: A New Biography* (1927. reprinted 1968), in spite of its title, is a socio-historical treatment of the subject that had already moved away from the biographical format. More recent lives include VINCENT TAYLOR, *The Life and Ministry of Jesus* (1954); ETHELBERT STAUFFER. *Jesus: Gestalt und Geschichte* (1957; Eng. trans., *Jesus and His Story*, 1960); and DAVID FLUSSER, *Jesus* (1968; Eng. trans., 1969). The first half of the 20th century witnessed a marked decline in the appearance of biographies and studies of the historical Jesus. Several factors contributed to the decline, of which the most important was the rise of form criticism in the first quarter of the century as exemplified in MARTIN DIBELIUS, *Die Formgeschichte des Evangeliums*, 2nd ed. rev. (1933; Eng. trans., *From Tradition to Gospel*, 1934); and RUDOLF BULTMANN, *Die Geschichte der synoptischen Tradition*, 3rd ed. (1958; Eng. trans., *The History of the Synoptic Tradition*, 1963). RUDOLF BULTMANN, *Jesus* (1926; Eng. trans., *Jesus and the Word*, 1934), presents a so-called encounter with the message of Jesus, but without sharp differentiation of that message from the church's "earliest tradition." The period 1950–70 produced a resurgence of interest in the historical Jesus. Stimulus for the resurgence is often credited to a paper by ERNST KASEMANN, "The Problem of the Historical Jesus," in W.J. MONTAGUE (trans.), *Essays on New Testament Themes* (1964). The "new quest," as it came to be called after JAMES M. ROBINSON, *A New Quest of the Historical Jesus* (1959), arose as a question concerning continuity between Jesus and his message, on the one hand, and the early church's proclamation of Christ, on the other. Some of the alternative positions concerning the possibility of a biography of Jesus may be seen in CARL E. BRAATEN and ROY A. HARRISVILLE (eds.), *The Historical Jesus and the Kerygmatic Christ* (1964), a collection of essays by major figures. Distinctive positions are to be found in ERNST FUCHS, *Zur Frage nach dem historischen Jesus* (Eng. trans., *Studies of the Historical Jesus*, 1964); JOHN KNOX, *The Church and the Reality of Christ* (1962); HUGH ANDERSON, *Jesus and Christian Origins* (1964); and LEANDER E. KECK, *A Future for the Historical Jesus* (1971). Accurate popular presentations are in HEINZ ZAHRNT, *Es begann mit Jesus von Nazareth* (1960: Eng. trans.. *The Historical Jesus*, 1963); and in the moderately conservative JOACHIM JEREMIAS, *Das Problem des historischen Jesus* (1960; Eng. trans., *The Problem of the Historical Jesus*, 1964). The entire trend of Gospel study since the rise of form criticism is challenged by BIRGER GERHARDSSON in *Memory and Manuscript* (1961), who argues that Jesus himself taught his disciples, in the manner of a Jewish rabbi, to memorize and transmit traditions in a fixed form. J. ARTHUR BAIRD, *Audience Criticism and the Historical Jesus* (1969), explores the use of the computer in literary analysis but is somewhat ambiguous with respect to total historical methodology. The standard contemporary study is GUNTHER BORNKAMM, *Jesus von Nazareth* (1956; Eng. trans., 1960).

The message of Jesus: Many of the titles listed in the preceding section include discussions of the teachings of Jesus. NORMAN PERRIN, *Rediscovering the Teaching of Jesus* (1967), is the important technical introduction. T.W. MANSON, *The Sayings of Jesus* (1949), is a valuable commentary on the so-called Q material. Since JOHANNES WEISS, *Die Predigt Jesu vom Reiche Gottes* (1892; Eng. trans., *Jesus' Proclamation of the Kingdom of God*, 1971). it has been recognized that the eschatological Kingdom of God was the centre of Jesus' message. Whether the message implied a wholly futuristic expectation, a "realized" eschatology, or a future kingdom with present manifestations is still discussed. Convenient summaries of alternative interpretations are presented in NORMAN PERRIN, *The Kingdom of God in the Teaching of Jesus* (1963): and GOSTA LUNDSTROM, *The Kingdom of God in the Teaching of Jesus* (Eng. trans. 1963). Perrin's work also includes a survey of interpretations of the phrase Son of man and other titles related to Jesus' vocation. Representative recent studies include W.G. KUMMEL, *Verheissung und Erfüllung*, 2nd ed. (1953; Eng. trans., *Promise and Fulfillment*, 1957); RUDOLF SCHNACKENBURG, *Gottes Herrschaft und Reich* (1959; Eng. trans. *God's Rule and Kingdom*, 1963); GEORGE ELDON LADD, *Jesus and the Kingdom* (1964), written from a conservative perspective; and H.E. TODT, *Der Menschensohn in der synoptischen Überlieferung* (1959; Eng. trans.. *The Son of Man in the Synoptic Tradition*, 1965). Modern investigation of the parables derives from C.H. DODD, *The Parables of the Kingdom*, rev. ed. (1961). The definitive work is JOACHIM JEREMIAS, *Die Gleichnisse Jesu*, 6th ed. (1962; Eng. trans., *The Parables of Jesus*, rev. ed., 1963). DAN O. VIA, *The Parables* (1967), is one of several recent books in which literary critical insights are employed to advance interpretation beyond strictly historical critical study. ETA LINNEMANN, *Gleichnisse Jesu*, 3rd ed. (1964; Eng. trans.. *Jesus of the Parables*, 1967), advances views along the lines proposed by Ernst Fuchs. MARTIN DIBELIUS, *The Sermon on the Mount* (1940); HANS WINDISCH, *Der Sinn der Bergpredigt*, rev. ed. (1937: Eng. trans., *The Meaning of the Sermon on the Mount*, 1951); and W.D. DAVIES, *The Setting of the Sermon on the Mount* (1964). offer varying interpretations of the most familiar body of Jesus' teaching.

The sufferings and death of Jesus: EDUARD LOHSE, *Die Geschichte des Leidens und Sterbens Jesu Christi* (1964; Eng. trans., *History of the Suffering and Death of Jesus Christ*, 1967), provides a convenient overview of the problems. Of the massive recent literature on the trial and Crucifixion. JOSEPH BLINZLER, *Der Prozess Jesu*, 2nd rev. ed. (1955; Eng. trans., *The Trial of Jesus*, 1959); and PAUL WINTER, *On the Trial of Jesus* (1961), have been very influential. The range of scholarly opinion about the trial is well represented by eight responsible essays in *Judaism*, 20:6–74 (1971).

The story of Jesus and faith in Jesus: REGINALD H. FULLER. *The Mission and Achievement of Jesus* (1954) and *The Foundations of New Testament Christology* (1965), partly because of the author's shift in viewpoint, exhibit a range of opinion on the relation of Jesus' sense of vocation to the church's Christology. A more difficult study is that of FERDINAND HAHN, *Christologische Hoheitstitel* (1963; Eng. trans., *The Titles of Jesus in Christology*, 1969).

(M.J.Su.)

Jet Engine

A jet engine is a cigar-shaped propulsion device that draws in air from the atmosphere, compresses it, heats it by combustion with a fuel, and ejects the resulting gases with sufficient force that the reaction produces a substantial thrust in the opposite direction. While the same principle can be used with other fluids—for example, water ejection to propel a boat—the term jet engine normally refers to an air-breathing device. It is distinct from a rocket engine, which also produces thrust by ejecting gas, but all of whose fuel elements are stored within the fuel supply.

This article deals with the jet engine used in aircraft, whose introduction in the 1940s and '50s produced radical changes in military and civil aviation. Previously, aircraft propulsion had been by means of a piston engine whose reciprocating motion produced intermittent bursts of power that had to be converted by mechanical means to a rotating crankshaft and thence to a propeller. A jet engine reduces this cumbersome process to a much simpler one, resulting in a much smaller and lighter power plant. In addition, the jet engine eliminates the restriction on power output imposed by the propeller, opening the way to significant increases in aircraft speed, size, and operating altitudes.

Early developments

Like many other inventions, jet engines were envisaged long before they became a reality. The earliest proposals were based upon adaptations of piston engines and were usually heavy and complicated. The first to incorporate a turbine design was conceived as early as 1921, and the essentials of the modern turbojet (see below) were contained in a patent in 1930 by Flight Lieut. (later Sir) Frank Whittle in England. His design was first tested in 1937 and achieved its first flight in May 1941. In Germany, parallel but completely independent work followed issuance of a patent in 1935. It proceeded more rapidly, and the very first flight of a turbojet-powered aircraft, a Heinkel He 178, came in August 1939. By the end of World War II, these prototype aircraft had developed into a few operational turbojet squadrons in the German, British, and American air forces.

In the military area, jet fighter aircraft developed rapidly and were in use during the Korean War (1950–53), flying at speeds of 600 miles (1,000 kilometres) per hour. During the next decade they overcame the sound barrier and established normal operations up to more than twice the speed of sound (Mach 2). Bomber and transport jet aircraft were also able to reach and cruise at supersonic speeds—*i.e.*, speeds greater than the speed of sound.

The first civil jet transport, the British de Havilland Comet, flew in 1949 and regular transatlantic jet services were started in 1958 with the Comet 4 and the American Boeing 707. By 1974 more than 90 percent of hours flown throughout the world were flown by jets; the first supersonic airliner, the Anglo-French Concorde, flying at more than twice the speed of sound, entered regular service in January 1976.

The jet engine has developed along several different but clearly related lines, which can be classified broadly as turbojet, fan-jet, turboprop, and other types.

TYPES OF JET ENGINES

Turbojet. *Operating principles.* The turbojet is an internal combustion engine in which the fuel is burned inside the engine and the heat energy released is converted to mechanical energy. In the process, the working fluid undergoes a cycle of pressure and temperature changes, starting with compression of atmospheric air, followed by the addition of heat by combustion of the fuel at a nearly constant high pressure, the expansion of the products of combustion during which the heat energy is converted to mechanical or kinetic energy, and final exhaust at a fixed low pressure. The elements of an engine needed to achieve this cycle are shown in Figure 1, with typical changes in the temperature, pressure, and velocity of the working fluid as it passes through the engine.

The thrust developed derives from the momentum of the exhausted gas. Since momentum is the product of mass times velocity, it is significant that the gas is greater in mass and has substantially greater velocity than the air taken in. The net thrust achieved by a jet engine is the result of the difference between the mass times the velocity of the ejected combustion products and the mass times velocity of the incoming air.

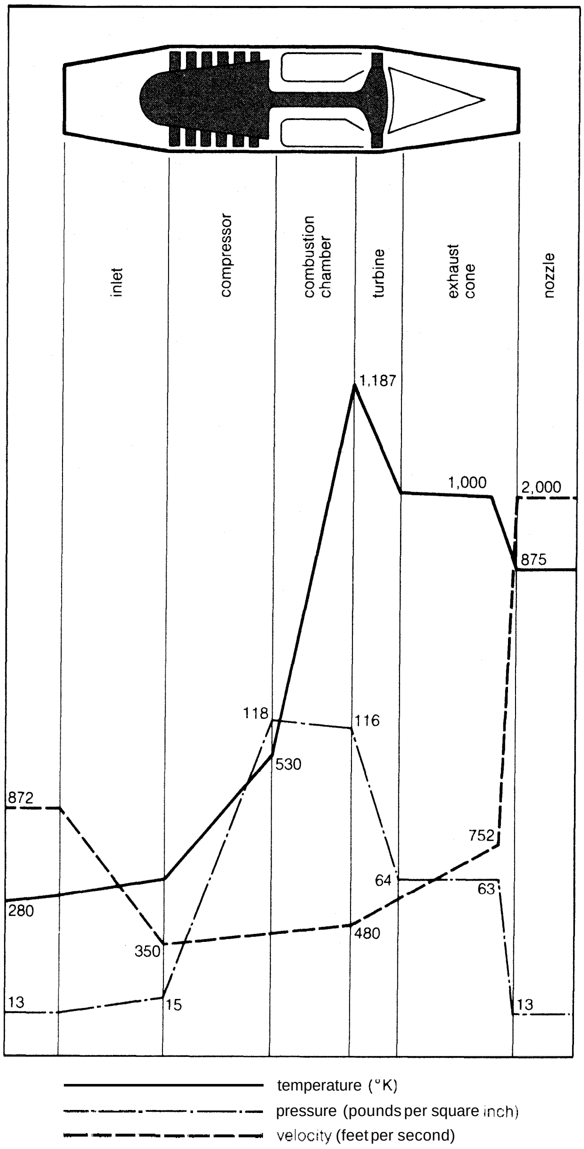


Figure 1: Cross section of a turbojet and (below) graph of typical operating conditions for its working fluid.

This is expressed in the equation

$$F=\frac{1}{g}\,(m_gV_j)-\frac{1}{g}\,(m_aV)$$

in which F is the engine net thrust in pounds; m_g is the flow rate of gas leaving the nozzle in pounds per second; m_a is the flow rate of air entering the inlet in pounds per second; V_j is the velocity of the exhaust jet in feet per second; V is the velocity of the air at inlet in feet per second; g is the gravitational constant of 32.2 feet per second per second.

Other things being equal, thrust will be increased by increases in either the mass of the ejected gas or its velocity or both; this is the basis of all changes to improve performance of the simple jet or its derived forms.

Principal components. Turbojet components and their functions, unlike those of the piston engine, are distinct and can be described and analyzed individually: (1) Inlet: The velocity of the incoming air relative to the engine depends on flight conditions, especially speed and altitude of the aircraft. It may vary from zero to much greater than the speed of sound. The air inlet is provided with control devices to adjust this velocity to values acceptable to the compressor. (2) Compressor: The early Whittle turbojets used centrifugal compressors, in which the air is forced radially outward by whirling blades, like the water in a centrifugal pump. Such an arrangement was found to suffer from low efficiency and required a compressor of awkwardly large diameter; it has since been superseded by the axial-flow compressor in which the air is forced through several rows of alternately fixed and moving blades. Each row of blades increases pressure moderately; the overall effect is a substantial increase in pressure and the configuration is well suited to engine design. Aerodynamic limitations in accelerating the engine arise when pressure ratios of the order of 7:1 are reached, but these can be overcome by such modifications as "blow-off" valves and variable settings for the stationary blades. Alternatively, and more usually, the restriction is avoided by using two independently driven compressors in series. They are referred to as low-pressure and high-pressure units and their combined pressure ratio may be 25:1 or more. (3) Combustion chamber: The combustion chamber ensures that the energy of the fuel is converted to heat energy and added to the compressed air to produce high-temperature gas. It is required to do this over a wide operating range, with maximum efficiency and minimum pressure drop, within a limited volume and without large temperature variations across the outlet. The temperature rise is generally of the order of 600"–700" C (1,100"–1,300" F) and the outlet gas temperature, depending upon engine design requirements, is in the range 927" to 1,177" C; (1,701" to 2,151° F). Because of the engine layout, an annular (ring) form of combustion chamber is most logical, but it was only extensively adopted in the 1960s because of earlier development, production, and assembly problems. (4) Turbine: The turbine, in its simplest form, consists of a rotating disk (rotor) with blades or "buckets" mounted on its rim and a ring of stator blades or nozzles upstream. Its purpose is to drive the compressor. The high-pressure gas from the combustion chamber flows through the nozzles and by impingement on the rotor blades imparts its energy to them. The energy may be transferred purely by a change of gas velocity and direction through the blades or by a change of both velocity and pressure. The former is known as an impulse effect and the latter as a reaction. In practice the rotor is designed to produce a combination of the two effects. In order to obtain greater power output, the gas may additionally pass through a second or third stage fixed to the same rotor or, when driving a separate compressor, through an independent stage. (5) Exhaust cone and propelling nozzle: The final element is the exhaust cone and propelling nozzle in which the residual gas energy is converted to kinetic energy and ejected in a high-velocity jet. This component may also contain silencing devices, a thrust-reversing mechanism, and variable-area nozzles.

Advantages of axial-flow compressor

The compressor, combustion chamber, and turbine are together referred to as the gas generator. They form the basis of all variations of turbojet engines. The gas generator pumps energy into the air by compression and combustion, but more than 60 percent of this is extracted again by the turbine driving the compressor. Only the remainder is available for external work.

Performance. The main criteria of turbojet performance are thrust F (in pounds or kilograms), fuel consumption Q_f (in pounds or kilograms per hour), and engine weight W (in pounds or kilograms). These are assessed as specific thrust—thrust divided by engine airflow rate (F/m_a pound per pound per second), specific consumption (Q_f/F pound per pound per hour), and specific weight (W/F pound per pound).

These performance characteristics depend on the maximum gas pressure and temperature developed in the engine, the separate efficiencies of the various engine components, the mechanical design and the external flight conditions of velocity, ambient temperature, and altitude. Increasing maximum temperature increases the specific thrust and increasing the pressure ratio improves specific consumption, but in combination they reach optima at different values. Design is, therefore, a compromise the direction of which depends on whether thrust or consumption is the major consideration. Improvements in pressure ratio were achieved, first by adding more stages to both compressor and turbine and later by using two independent rotor systems. Increases in temperature have relied upon the development of improved high-temperature materials, and recently, the introduction of turbine-blade cooling techniques. These have raised engine design temperatures from 750" to 1,177" C (1,400" to 2,151" F) and it is anticipated that future designs will reach higher than 1,377" C (2,511" F).

Reduction of specific weight

In parallel with these improvements specific weight has been reduced from more than 0.5 pound per pound of thrust in early engines to about 0.2 pound per pound in modern engines. In some special engines designed for vertical takeoff operations values lower than 0.1 pound per pound have been achieved.

The effect of altitude is to decrease thrust because of the lower density of the air entering the engine. This reduces the mass flow and hence the jet momentum. In the same way, forward speed increases the momentum of the air entering the engine and reduces the thrust. At the same time the greater kinetic energy of this air, converted to pressure energy, tends to increase thrust. The differential effect is first a fall and then an increase in thrust. This continues until the temperature rise, which accompanies the increase in pressure, reaches a point at which the fuel added in the combustion chamber must be reduced to avoid exceeding the maximum permissible gas temperature. At this stage thrust begins to drop again, but this only occurs at high supersonic flight speeds.

Though appreciable increases of thrust can be obtained by afterburning or reheating and by water-methanol injection, these procedures are unsuitable for continuous use because of the high additional consumption. Afterburning consists of injecting fuel into the hot gas–air mixture entering the exhaust system so that its temperature is raised still further. This accelerates the gas in the jet and so increases its momentum and the net thrust. It has the advantage that it does not affect the conditions upstream and can produce increase of 40 percent or more in static thrust. This thrust increase becomes proportionately greater with increasing forward speed. Its main disadvantage is the weight and complication of the variable-area nozzle necessary to deal with the greater volume of gas, and a small permanent loss in thrust caused by the combustion equipment.

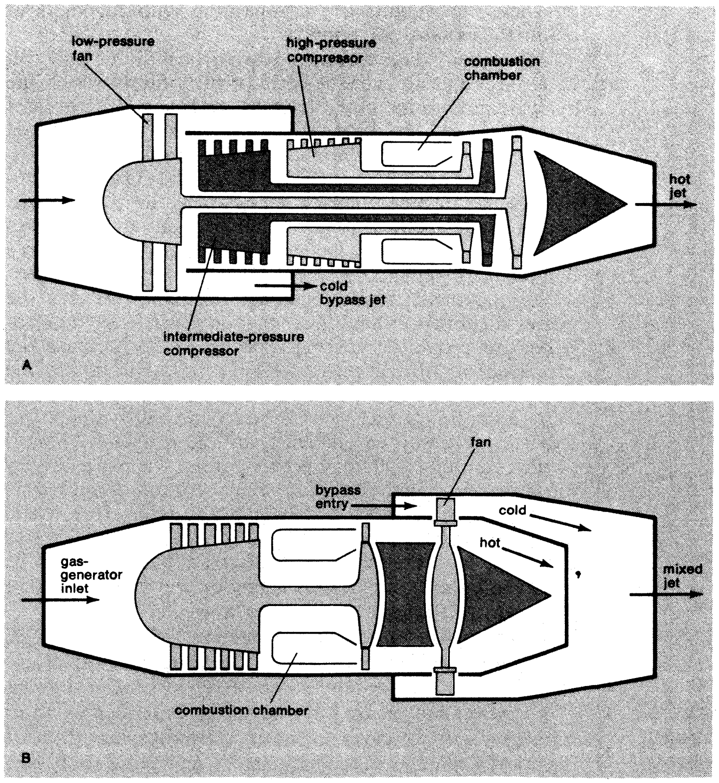


Figure 2: (A) Front fan-jet with separate jets and three independent rotors. (B) Aft fan-jet with mixed jet.

Water-methanol thrust boosting acts, in contrast, by injection at the inlet or immediately after compression. The liquid evaporates and reduces the air temperature, thus increasing density and mass flow. The thrust increase by this means is restricted to about 10 percent and is usually applied at takeoff only.

Fan-jet. In the turbojet, all of the air entering the intake passes through the gas generator, but in the fan-jet only a portion does so. The remainder passes through a fan or low-pressure compressor and is ejected directly as a "cold" jet or mixed with the gas-generator exhaust to produce a "hot" jet (Figure 2). The object of this "bypass" system is to increase thrust without increasing fuel consumption. It achieves its object by increasing the total air-mass flow and reducing its velocity within the same total energy supply.

Bypass ratio

Configuration. The fan-jet may be designed with a front or aft fan as indicated in Figure 2. The former is more usual and the size of fan or compressor depends upon the bypass ratio. For large ratios only one or two fan stages are necessary, but for smaller ratios a higher jet velocity and more compression are required. For a "hot" jet, the pressures of the two streams have to be designed to be the same at the mixing section.

The aft fan has an advantage in that it can be made mechanically independent of the gas generator and coupled only by ducting. It has problems of installation that are more difficult to overcome, and it has not been used extensively.

Performance. The main performance characteristic of the fan-jet is its low specific fuel consumption, which may be 25 percent less than comparable turbojets. Consumption at sea-level takeoff conditions for a 5:1 bypass ratio is of the order of 0.45 pound per pound-hour (pounds per hour for each pound of thrust), increasing for subsonic cruising at high altitude to 0.65 pound per pound-hour. Detracting from this, specific thrust is less than half that of modern turbojets and may be as low as 30 pounds per pound per second. This low figure is due to a large frontal area that can lead to high aerodynamic drag at forward speeds. The engine thrust also decreases rapidly with speed and is, in general, unsuitable for high supersonic flight conditions. The larger diameter and additional power delivered to the fan tend to raise the overall engine weight, but this is balanced by the extra thrust, and engine specific weights are comparable with turbojets at about 0.2 pound per pound.

Turboprop. The turboprop or turboshaft engine can be regarded as an extreme form of turbofan in which the bypass ratio has increased from the order of 5:1 to greater than 50:1, and the residual energy of the hot exhaust jet is reduced to a very minor value. The fan becomes a propeller and is no longer enclosed in a duct. It may be coupled to the turbine on the same shaft as the compressor, as in a single-spool turbojet, or independently, as in a two-spool engine. In the latter case it is described as a "free turbine" engine, and this form of drive permits more flexibility in control. In both cases the rotational speed of the power turbine is much higher than is suitable for the propeller and reduction gearing is necessary. This entails the provision of lubrication and oil cooling and adds to the design complication and weight.

There is a limit to the size of the engine, set by the power that can be absorbed efficiently by the propeller. This is of the order of 4,000 horsepower but may be increased by the use of counterrotating propellers. A more usual way is to couple two engines in tandem through a free wheel device to concentric propellers. In this way if the full power of both engines is not required, part load can be achieved by shutting down one engine and using the other nearer to full load and obtaining a greater efficiency.

Turboprops have a high efficiency at speeds below 500 miles (800 kilometres) per hour. In recent years their application has been chiefly as power units for helicopters.

Turboprop performance

The power output of a turboprop is expressed as shaft horsepower nr in cases in which a useful proportion of the energy remains in the exhaust jet, as equivalent horsepower. The latter is the sum of the shaft power and a theoretical power derived from the assumption that approximately three pounds of residual jet thrust (*F*) is equal to 1 shaft horsepower at takeoff.

The specific consumption of the actual engines is approximately 0.60 pound per hour for each equivalent horsepower at takeoff, but this is with relatively low pressure ratios. In general, designs with pressure ratios comparable to turbofans have not been developed, although these would give specific fuel consumption as low as 0.40 pound per hour for each shaft horsepower. The second characteristic of the turboprop is its high output at low forward speeds, and this may reach more than 160 shaft horsepower for each pound per second of air flowing through the gas generator. The third feature is specific weight, which may be 0.45 pound per equivalent horsepower. Except at low flight speeds this is heavier than the turbojet or turbofan, largely because of the weight of gearing required for the propeller. It is, however, only one-third of the weight of the piston engine it replaces.

When used in helicopters, all reduction gearing is regarded as part of the lifting rotor system and the turboshaft engine specific weight is reduced to the order of 0.25 pound per shaft horsepower.

Special types. ***Ramjet.*** This engine with no major moving parts is probably the lightest and simplest form of air-breathing jet engine. Its application is severely restricted by the fact that its compression ratio depends entirely on its forward speed. It develops no static thrust and very little thrust below sonic speeds. A ramjet aircraft must therefore be brought up to speed by some special launching means, such as another aircraft.

From Figure 3 it can be seen that it has a design similar to a turbojet in which the rotor system has been omitted. Because of this there is not the same limitation on maximum gas temperature, and temperatures beyond 1,930° C (3,510° F) are permissible. These boost the thrust output in the same way that afterburning augments the turbojet thrust, but they carry the same penalty of high specific consumption.

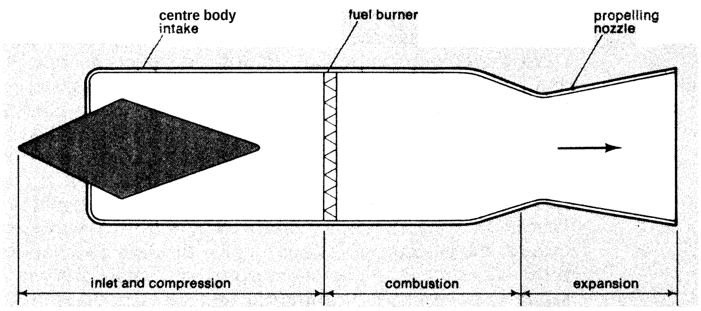


Figure 3: Arrangement of ramjet engine.

Consumption is sensitive to intake efficiency, and performance falls rapidly both above and below the designed speed unless the complications of variable geometry intakes and exhausts are accepted. The result is that the ramjet becomes competitive only at speeds approaching three times the speed of sound and above. Its main application has been in missiles where its low weight and simplicity have advantages.

Turboramjet. Because of the limitations of turbojet engines, primarily associated with speed and the maximum temperature of the rotating components, various forms of hybrid engine have been proposed. One such is the combination of turbojet and ramjet into a single power plant (Figure 4). The turbojet portion supplies static thrust and power for accelerating to supersonic speeds and is then closed off from the air flow, which bypasses it as in the fan-jet. A combustion system like a very large afterburner is operated for this bypass air and the whole engine becomes a ramjet.

This system requires air-flow shutoff mechanisms for the turbine, and a feathering arrangement for the fan blades so that they can be turned out of normal position to minimize drag. Even with these arrangements there are

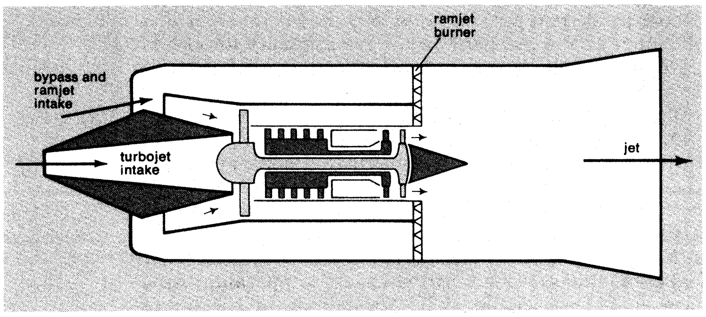


Figure 4: Internal organization of a turboramjet.

some aerodynamic losses in the ramjet operation, in addition to the extra weight. In the early 1970s aircraft speeds did not yet justify development of such an engine.

DESIGN PROBLEMS AND OTHER ASPECTS OF JET TECHNOLOGY

Jet aircraft engines operate under extreme conditions of temperature, pressure, and rotational speed in order to achieve the maximum output and minimum weight. These induce stresses and vibrations and other effects that must be taken into account in a successful design. This is done in two ways—first, by the inspection and maintenance system applied to engines to detect and remove any incipient component failures, and second, by design to limit the risks and consequences of a failure.

Past practice has been to fix the time between overhauls at a conservative figure and increase it in the light of experience. A well-established engine may now be expected to operate for more than 5,000 hours between overhauls. An alternative is to use continuous overhaul or "on-condition" maintenance. By means of built-in inspection equipment and continuous recording instrumentation it is possible to pinpoint any incipient troubles in the engine and take action to overcome them. This is facilitated in the latest designs by a form of module construction that enables major sections of the engine to be removed and replaced while the engine itself remains in its mountings on the aircraft.

The most serious sudden failure that may occur is the breaking of a turbine blade. Due to its weight and the speed of rotation, a turbine blade has great momentum when it breaks away from the rotor and could cause serious damage to anything in the plane of rotation. Design has been successful in reducing this risk by use of a containment shield that stops blades penetrating the engine casing and by ensuring that no vital fuel or oil pipes are in the plane of the blades.

Hazards from birds and icing

At the compressor end, there is a risk of foreign bodies being drawn into the intake and striking the compressor blades. The most likely object to be ingested is a bird, and all engine designs have to be strong enough to withstand the entry of a heavy bird at high speed without catastrophic effect on the compressor blading. A similar problem is created by icing conditions in which ice is deposited on the stationary parts of the intake and compressor guide vanes. In addition to reducing the air-flow area and thus reducing thrust, any ice that breaks off will enter the compressor. Protection is by anti-icing equipment, usually in the form of heat supplied to the sensitive area.

With the development of large engines, noise has become a serious nuisance and legislation to restrict it has been established in many places. There are two major noise sources in the jet engine—high-frequency noise from the compressor and low-frequency noise from the jet. The latter predominates at takeoff and the former during landings. Much of the compressor noise can be suppressed by lining the intake with sound-absorbing materials. Jet noise is largely a function of velocity and the most effective way to reduce it is by lowering the jet velocity. For pure turbojets this reduces the performance seriously but in fan-jets considerable noise reduction is possible without loss of performance.

Jet fuel. The fuel used in the earliest turbojets, domestic paraffin (kerosene), still remains the basis of all jet fuels. It is a liquid hydrocarbon similar to gasoline, but with some variations in physical properties. Gasoline is derived from crude petroleum by distillation within a limited temperature range (25°–150° C; 75°–300° F), but for kerosene this range is higher (150"–265" C; 300°–510" F). The effect of this is to endow kerosene with a higher flash point (temperature at which the fuel ignites), a higher freezing point, and higher specific gravity (weight per unit volume). The first of these ensures that kerosene will be much safer in accident conditions but presents slightly more difficult combustion problems, which are often manifested by smoke in the exhaust jet. The second produces difficulties when an aircraft is operating at low ambient temperatures for prolonged periods, and the third gives an advantage in the weight of fuel available from a fixed fuel-tank capacity. Since the calorific (heat) value of both fuels is almost the same, kerosene thus has almost 10 percent greater range or endurance.

Advantages of kerosene

Attempts have been made to develop higher calorific value fuels both chemically and by additives, but increases have usually been accompanied by uneconomical costs.

High-temperature materials. The continued improvement in turbojet performance over the years is, in large measure, due to operation at higher temperatures. This, in turn, has been made possible by the development of materials capable of withstanding stresses at the higher temperatures. The most highly stressed parts of the engine are the rotating components, and of these the turbine blades are subjected to the highest temperatures. They are required to resist oxidation, maintain strength, especially against fatigue. avoid embrittlement, and have reasonable hot-working and machining properties for manufacture. It was discovered early that nickel–chromium base alloys were better than steels and a series of alloys with approximately 80 percent nickel and 20 percent chromium were used. Improvements in these have permitted maximum gas temperatures to rise by an average of approximately 10" C (18" F) each year since the earliest days, but a limit to these alloys has now been reached. Cast, instead of forged, materials have extended the range up to over 1,025" C (about 1,880" F), but beyond that either cooling or refractory materials are necessary. Cooling has allowed the maximum gas temperature to be raised a further 200" C (360" F) without metal temperature rising. Molybdenum, tungsten, and niobium alloys could operate with higher surface temperatures, but all the associated problems have not yet been solved.

Turbine disks

Turbine disks, while not exposed to the direct impingement of hot gases, are subject to temperatures approaching 625" C (1,160" F) at their rims while their hubs may be some 450" C (800" F) lower. To overcome the resultant stresses the disks are always machined from forgings and the best materials are found to be steels with high nickel–chromium content.

Although compressors are at the air-entry end of the rotor system, blades are subject to increasing air temperature as flight speed increases and the compression ratio rises: temperatures of nearly 400" C (750" F) can be reached. Normally, aluminum alloy is used for the blades, but this begins to lose strength above 180" C (360" F), and it is necessary to change to steel in some high-pressure stages, resulting in an increase in weight. Efforts to reduce weight, particularly in the larger fan blades now being used, have led to the use of titanium and experiments with nonmetallic carbon-fibre material. Both of these are expensive but may be used increasingly in the future.

Future developments. Neither military nor civil aircraft designs appear to have reached their limits and turbojet engines can be expected to continue development along the lines followed steadily since their introduction. First, greater thrust without comparable increases in weight and size will be required together with operation at higher supersonic flight speeds and altitudes (at least for military aircraft), and second, greater economy and cheaper maintenance for airline operations.

In technical terms this will involve refinements of design rather than major changes, as performance approaches more closely to that of the ideal cycle. It is probable that pressure ratios will rise from the present 25:1 to 40:1, that bypass ratios will rise from 5:1 to 10:1, and that maximum gas temperature will rise from 1,450" to 1,800" K (1,180" to 1,525" C; 2,150" to 2,780" F) by the 1990s. There will also be significant but relatively small improvements in the efficiency of compressors, turbines, and combustion systems. Together, these will result in engines with specific fuel consumptions 15 percent less than the present best, with similar reductions in specific weight. Some of the weight reduction will arise from the introduction of new composite nonmetallic materials. It may be foreseen that engines having 50 percent more thrust than in the early 1970s will be built, and others designed, for operation at 3.5 times the speed of sound. In many cases the servicing and maintenance will be largely assessed by computer records, and a new concept of engine life will be established. Finally, assisted by the rise of higher bypass ratios and intensive research, a substantial reduction in noise levels will be achieved.

BIBLIOGRAPHY. C.G. SMITH, *Gas Turbines and Jet Propulsion*, 6th ed. (1955), an early detailed history of jet engines; N.E. MORGAN, *Turbojet Fundamentals* (1958), an elementary explanation of theory and operation; B.H. JENNINGS and W.L. ROGERS, *Gas Turbine Analysis and Practice* (1953), a standard textbook; O.E. LANCASTER (ed.), *Jet Propulsion Engines*, vol. 12, *High Speed Aerodytzamics and Jet Propulsion* (1959), an authoritative study of all aspects; A. SILVERSTEIN, *Progress in Aircraft Gas Turbine Engine Development*, a paper given at the 5th Congress of the International Council of the Aeronautical Sciences (1966), a review and discussion of recent progress.

(A.D.B.)

Jet Streams

Atmospheric jet streams were discovered toward the end of World War II by U.S. bomber pilots over Japan and by German reconnaissance aircraft over the Mediterranean. The World Meteorological Organization defines a jet stream as a strong, narrow air current that is concentrated along a nearly horizontal axis in the upper troposphere or stratosphere (10 to 50 kilometres altitude), characterized by wind motions that produce strong vertical and lateral shearing action, and featuring one or more velocity maxima. Normally a jet stream is thousands of kilometres long, hundreds of kilometres wide, and several kilometres deep. The vertical wind shear is on the order of five to ten metres per second per kilometre, and the lateral shear is on the order of 5 m/sec per 100 km. An arbitrary lower limit of 30 m/sec is assigned to the speed of the wind along the axis of a jet stream.

With abundant radiosonde data now available over the Northern Hemisphere it is possible to map the jet streams in the upper troposphere (near 10 to 12 kilometres) in their daily occurrence and variation and to forecast them reasonably well with numerical prediction techniques. Upper-air information from the Southern Hemisphere is still sparse. Constant-level balloons (the so-called GHOST balloons) and satellite information on temperature structure and characteristic cloud formations in the atmosphere are serving to close the data gap on the global jet stream distribution.

The strongest winds known in jet streams have been encountered over Japan, where speeds up to 500 kilometres per hour (close to 300 knots) occur. A persistent band of strong winds occurs during the winter season over this region, flowing from the southwest and leading tropical air from northern India into juxtaposition with polar and Arctic air from Siberia. A similar region of confluence of air masses with vastly different temperatures exists over the central and eastern United States, leading to a maximum frequency of occurrence of jet streams during winter and spring (Figure 1).

The main impact on weather and climate (*q.v.*) comes from two distinct jet stream systems: the Polar-Front Jet Stream, which is associated with the air-mass contrasts (the fronts) of middle latitudes and which gives rise to the formation of squalls, storms, and cyclones in this latitude belt; and the Subtropical Jet Stream, which lies over the subtropical high-pressure belt, and which is characterized by predominant subsidence motions and, hence, by fair weather. During summer a belt of strong easterly winds is found over Southeast Asia, India, the Arabian Sea, and tropical Africa. This tropical, easterly jet stream is tied in with the weather disturbances of the Indian and African summer monsoons and their heavy rainfalls.

Because of their strong winds, jet streams play an important role in the economy of air traffic. Head winds must be outlasted by extra fuel, which takes up useful cargo space. Clear air turbulence (CAT) is often associated with the strong vertical wind shears found in the jet stream region. It is a hazard to passenger and crew safety, and, because of the increased stresses on the air frame, it decreases the useful life of the aircraft.

From Kao and Hurley, *Journal of Geophysical Research*, vol. 67, pp. 4235-36 (1962)

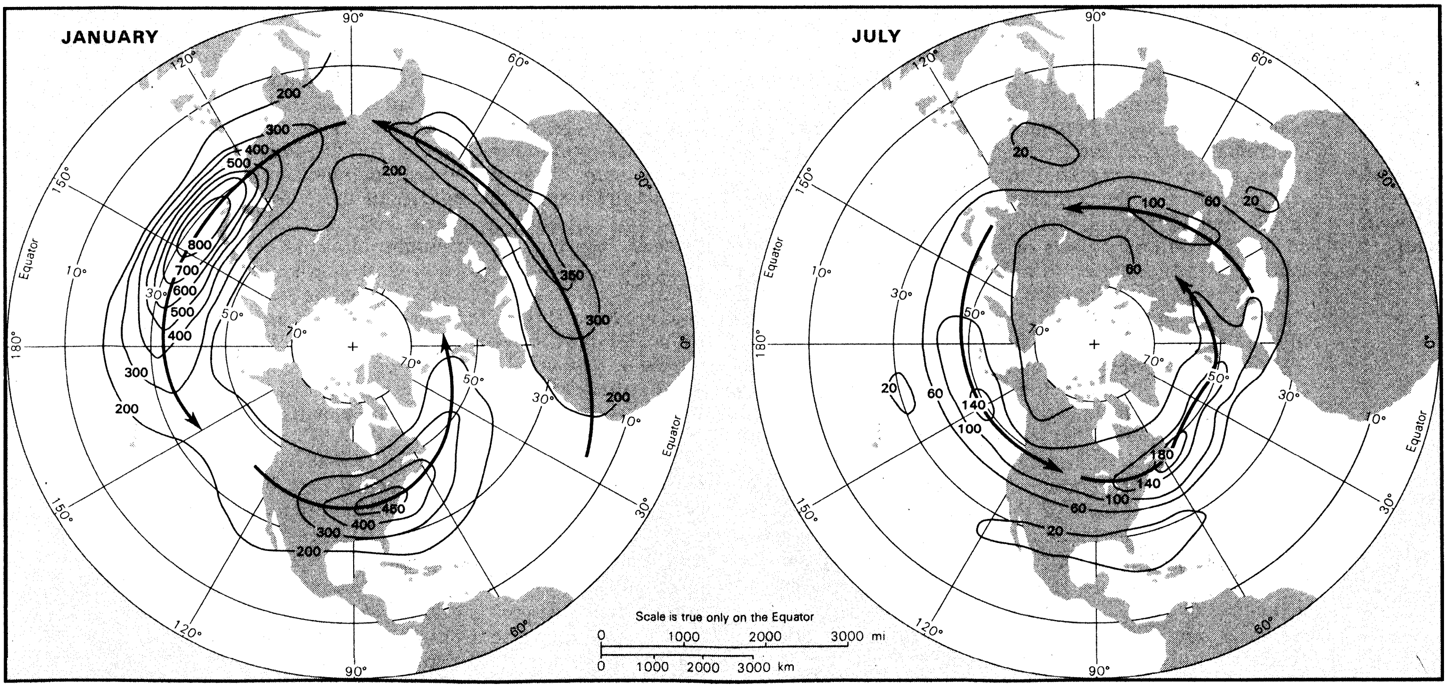


Figure 1: Mean kinetic energy in kilograms per metre per second at 300 millibars during January and July. Heavy lines and arrows indicate jet axes.

Jet streams are characteristic of a heat-driven circulation between Equator and pole that takes place in a rotating coordinate frame. Ocean currents, such as the Gulf Stream or the Kuroshio, bear a striking similarity to atmospheric jet streams, even though the wind-driven circulation in the trade-wind belt is of greater importance in their generation than the heat exchange between low and high latitudes.

This article treats the characteristics, properties, occurrence, and mode of formation of jet streams. For information on the general role of the jet stream in meteorology, see WINDS AND STORMS; CYCLONES AND ANTICYCLONES; CLIMATE; MONSOONS; and HURRICANES AND TYPHOONS. See also ATMOSPHERE and, for information on similar phenomena in water, OCEAN CURRENTS.

THE CAUSE AND FORMATION OF JET STREAMS

Relative motion of air parcels

The ultimate source of the kinetic energy that is concentrated in the jet streams is the positive radiation balance in the tropics and the negative balance in polar regions. For stable climatic conditions, heat transport from the Equator to the poles is required. On a non-rotating earth this would lead to a direct circulation between low and high latitudes, defined as a circulation that releases potential energy by the rising of warm air and the sinking of cold air and subsequently converts it to kinetic energy of air motions. In such a direct, or Hadley, circulation warm air would rise in the tropics, would move poleward near the top of the planetary layer in which strong convective mixing occurs, sink again at high latitudes, and then move equatorward near the earth's surface (Figure 2). This layer of convective mixing is called the troposphere. Its upper boundary is the tropopause. On a rotating earth the motions described above, forced by the necessity of heat transport between Equator and pole, must transport absolute angular momentum This quantity is defined as $M = \omega r^2$ in which ω is the angular velocity of a rotating point, r is the radial distance from its axis of rotation, and M is the angular momentum. An air parcel at the Equator that rotates together with the solid earth has a rate of rotation $\omega = 7.292116 \times 10^{-5}$ radians per second and a radial distance from the earth's axis equivalent to the earth's radius $R = 6{,}371$ km. If the air parcel rotates with the solid earth, there will be no wind because the parcel does not move relative to the earth.

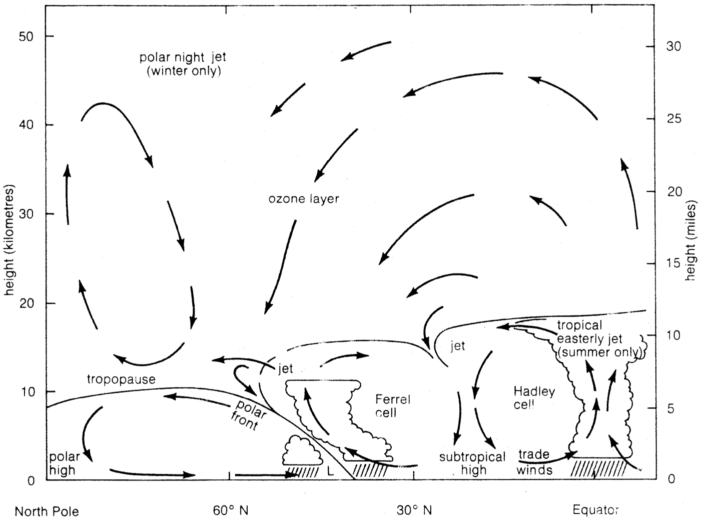


Figure 2: Positions of jet streams in the atmosphere. Arrows indicate directions of mean motions in a meridional plane.

If the air parcel under consideration is forced by heat convection to rise to the tropopause (16 to 18 km in height over the Equator), its radial distance from the earth's axis will increase slightly. In the absence of external forces such as pressure forces or frictional drag, the air parcel would conserve its absolute angular momentum as given by the above equation. For this reason, its angular velocity would have to decrease in order to compensate for the slight increase in radius. The parcel, in other words, would rotate more slowly from west to east than the solid earth underneath. The motion of the air parcel relative to the earth, therefore, would constitute a weak east wind.

During its poleward journey near the tropopause, the air parcel would be subject to shrinking radial distances from the earth's axis because of the spherical shape of the earth. Thus if M is to remain constant in the above equation, the decrease in r has to be compensated by an equivalent increase in ω. The parcel will rotate faster than the earth. and the difference will increase as the parcel moves farther poleward. Because the radial distance from the axis of rotation decreases with the cosine of latitude (φ), the expression for absolute angular momentum may be written as

$$M = (\omega_e + \omega)R^2 (\cos \varphi)^2,$$

in which ω_e is the angular velocity of the solid earth, ω is the angular velocity of the air motion relative to the earth's rotation, and R is the radius of the earth. It may be seen from this equation that a relatively short meridional displacement of an air parcel under conservation of absolute angular momentum will lead to strong westerly winds. An air parcel starting at 30° N latitude with zero westerly speed, for example, and traveling to 40° N under conservation of absolute angular momentum, will arrive at this new latitude with a west-wind component of 99 m/sec or 192 knots.

This example shows that the circulation ensuing from the poleward transport of heat and its associated transport of absolute angular momentum will generate a strong jet stream in subtropical latitudes. Because the air flowing northward near tropopause level, as it acquires strong westerly velocity components, converges with this jet stream band a high-pressure ridge will form. The northward flow of air, therefore, will have to overcome pressure forces. Absolute angular momentum will not strictly be conserved in this flow, but the tendency to such conservation, nevertheless, exists.

The Coriolis effect

The deflection of flow toward the right in the Northern Hemisphere, and toward the left in the Southern Hemisphere, results in strong west-wind belts in both hemispheres; it is caused by the earth's rotation. Such a deflection not only acts upon south winds but upon any arbitrary wind direction. The apparent force producing this deflection is called the Coriolis force, although in the

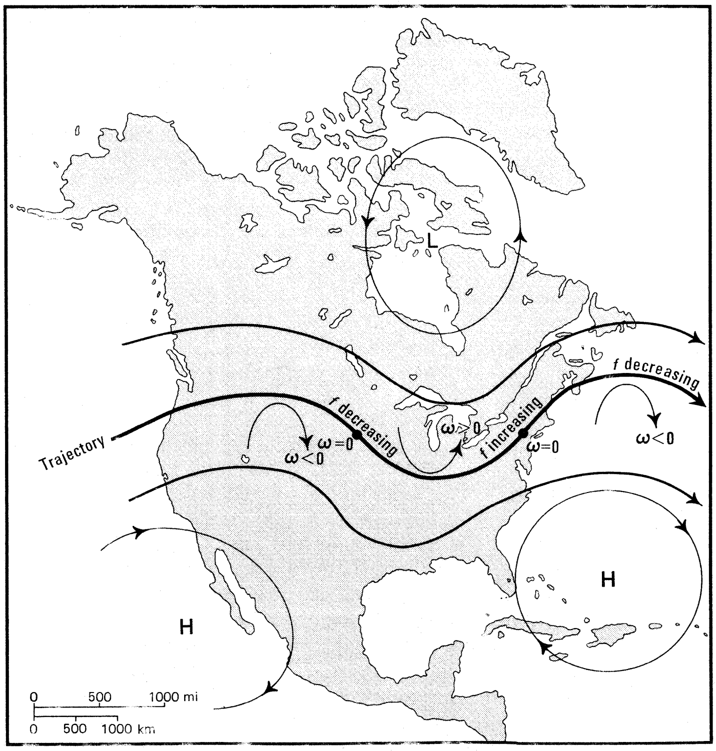


Figure 3: Changes in the Coriolis parameter, *f*, along the trajectory of an air parcel lead to changes in the rotation of the air mass (ω), thereby generating Rossby waves.

strict sense it is an effect. Its action upon a unit of mass is measured by the product of velocity times the Coriolis parameter,

$$f = 2\omega \sin \varphi,$$

in which φ is latitude. There is no deflecting effect at the Equator ($\varphi = 0$). The deflection reaches a maximum at the poles ($\varphi = 90°$, $\sin \varphi = 1$). This deflecting force prevents the air above the friction layer near the ground from flowing directly from a high-pressure region into the adjacent low-pressure region and causes it to adjust the flow direction until it parallels the isobars (lines of constant pressure). Because this deflection of flow prevents a rapid filling of cyclones and a rapid destruction of anticyclones, such pressure systems may remain active for many days as will the jet streams with which they are associated (Figure 3).

LABORATORY STUDIES OF JET STREAMS

The effects of meridional heat and absolute angular momentum transports on the formation of jet streams may be studied in so-called dishpan experiments. A flat pan (15 to 20 cm in diameter in some experiments) is filled with about 4 cm of water. Heat is applied through a heating coil at the bottom of the pan around the "equator" (*i.e.*, the outer rim of the pan) and is extracted at the pole (the centre of the pan) through a cooling cylinder, which is continuously flushed with cold water. The temperature gradient between equator and pole brings about a thermally driven circulation similar to the one described above. If the pan is rotated slowly, absolute angular momentum will be transported poleward in a manner described before, leading to the formation of a jet stream band that surrounds the polar cylinder symmetrically. Because of frictional forces acting in the vicinity of the polar cylinder, the jet stream will not touch the cylinder but will establish itself at some distance from it in high latitudes. If the rotation rate of the dishpan is increased slightly, the velocity of west winds in this jet stream will increase because of the transport of larger quantities of absolute angular momentum. At the same time, the jet stream band will move to slightly lower latitudes because the increased velocities will cause stronger frictional forces near the polar cylinder.

Thermal gradients and wave formation

A further increase of the rotation rate will require even higher transport rates of absolute angular momentum, enforced by the thermally driven circulation. Such increased transport rates, it turns out, can no longer be handled by a simple meridional circulation with rising motion at the equator, northward flow near the top of the water layer (equivalent to the tropopause), sinking near the pole, and southward flow near the bottom of the dishpan. The symmetric jet stream that surrounded the polar cylinder breaks down into regular and symmetric waves that progress slowly eastward. These waves, or eddies, support the meridional circulation described above in transporting heat and angular momentum: the southwesterly flow on the east side of a low-pressure trough brings warm water with high westerly angular momentum from the equator in rapid contact with the polar cold source. The northwesterly flow on the west side of these troughs carries cold water from the pole to the equator, thus also effecting the transport of heat toward the pole.

Rotation rate effects

Relatively slow rates of rotation generate one such symmetric wave. Higher rates of rotation cause two, three, four, or as many as six or seven symmetric waves to form, depending on the increased requirement for poleward angular momentum transport. If the rate of rotation is increased still further, the symmetric waves can no longer fulfill the requirements of such transports. They break down into strongly tilting and asymmetric troughs, whose axes extend in a direction from the southwest to the northeast. On the east sides of such tilting troughs, strong west winds are carried northward. On the west sides of the troughs, however, only weak westerlies, or even easterly winds, are carried southward. Thus, there will be vast net transport of westerly momentum poleward. The same consideration holds for the poleward transport of heat. These tilting troughs also reveal a tendency to break down into cutoff cyclones and anticyclones.

These vortices isolate themselves from the jet stream flow in middle latitudes. The cold cutoff lows become absorbed in warm subtropical regions, whereas the warm cutoff anticyclones merge with the cold water at high latitudes. This contributes significantly toward a net transport of heat and angular momentum poleward.

This stage of events, encountered in dishpan experiments at relatively high rates of rotation, closely resembles the weather patterns observed in the atmosphere in middle and high latitudes. The earth rotates too rapidly in these latitudes to permit the formation of a simple meridional circulation system. Eddy transport processes, brought about by low-pressure troughs, high-pressure ridges, cyclones, and anticyclones, play the dominant role in the transport of heat and absolute angular momentum. Only near the Equator, where the Coriolis force approaches zero, can a Hadley cell of meridional flow develop; but even there it is strongly influenced by the eddy transport processes of middle latitudes. Strong troughs, developing in the jet stream belt of middle latitudes, can sometimes reach far into the tropical and subtropical latitudes. Conversely, subtropical high-pressure cells occasionally may extend far into middle latitudes.

The necessity of eddy transport processes to maintain a climatic balance of temperature between Equator and pole, forced by the relatively high rate of rotation of the earth, exercises a strong effect upon the jet streams found in the atmosphere. It is because of these eddy processes that several jet stream systems exist in nature, rather than one simple jet stream band.

MEAN AND EDDY TRANSPORTS IN THE JET STREAM REGION

The Subtropical Jet Stream shown in Figure 2 may be considered an outgrowth of the northward transport of absolute angular momentum in a thermally driven Hadley circulation. It rests over the subtropical high-pressure belt, indicating the presence of convergence near tropopause level, if one averages flow conditions around the hemisphere, and of divergence near the earth's surface. The subsidence motion that dominates beneath the Subtropical Jet Stream gives rise to predominantly fair weather.

Jet streams and fronts

Farther to the north the Polar-Front Jet Stream occurs. It is associated with an indirect circulation cell, or Ferrel cell—that is, a circulation that consumes kinetic energy and generates potential energy. Such an indirect meridional circulation cannot maintain itself, together with a strong jet stream, against the forces of internal friction. The supply of energy maintaining the Polar-Front Jet Stream must originate in eddy processes. These are given in the form of migrating cyclones and anticyclones. The sinking of cold air behind a cold front and the rising of warm air along the warm-frontal surfaces in the cyclones of middle latitudes release the required amounts of potential energy to maintain the jet stream flow. The confluence of air masses from different latitudinal regions, which occurs near the earth's surface and which is shown in Figure 2, leads to the formation of strong fronts. This is in contrast to the flow regime found underneath the Subtropical Jet Stream, where diffluence tends to destroy and dissolve any fronts that may have been present.

The strong horizontal temperature gradients that occur in these frontal zones cause strong horizontal pressure gradients that increase in intensity up to the tropopause level. Hence, an increase of winds up to this level is required by a balance between pressure gradient force and the deflecting Coriolis force. Such strong temperature and pressure gradients are found throughout the troposphere beneath the Polar-Front Jet Stream; only in the upper troposphere do they occur beneath the Subtropical Jet Stream. There the lower troposphere is subject to frontolysis (the dissolution of fronts). Hence the jet streams are in balance with the horizontal temperature and pressure gradients in the layers underneath the maximum wind level. Because the formation of jet streams may be explained by a thermally driven circulation under conservation of absolute angular momentum, it may be argued that the jet streams generate their own frontal zones, and not vice versa.

Velocity and vorticity of flow

Polar-Front and Subtropical Jet Streams do not surround the hemispheres in bands of uniform wind speeds. Well-defined wind-speed maxima and minima are superimposed upon the jet stream belts. Thus, the air flowing in the jet stream region is subject to accelerations and decelerations, which, in turn, produce divergent and convergent flow patterns. These are shown schematically in Figure 4 in their relation to an individual jet- or wind-speed maximum. The curvature of flow that may be encountered in the jet stream region introduces a certain rotation in the flow pattern. So do horizontal wind shears. The latter may be realized by placing a hypothetical stick across one side of the jet stream shown in Figure 3. If the north side is used, the stick would rotate in a counterclockwise, or cyclonic, rotation because of the fast flow in the centre of the jet and the slower motion farther to the north. The rotation of flow is measured in terms of vorticity, defined as twice the angular velocity of the rotation. Vorticity is produced by the curvature of streamlines and by wind shears. The changes of vorticity, Q, and divergence of flow, D, are approximately related by $dQ/dt = -QD$ where dQ is the change of absolute vorticity (defined as the vorticity of the air motion plus the vorticity of the earth at the appropriate latitude, $2\omega \sin \varphi$) during the time interval dt. $D > 0$ measures divergence, $D < 0$ indicates convergence. Thus, if the absolute vorticity decreases along a given streamline ($dQ/dt < 0$), D has to be positive; divergent flow conditions must prevail because Q itself is positive everywhere. Such a divergence at jet stream level will cause a pumping out of vertical air columns in the locations indicated in Figure 4. This, in turn, will lead to falling surface pressure and to cyclogenesis. Under the regions of convergent flow at jet stream level anticyclones will build up. It therefore is apparent that jet maxima are instrumental in shaping cyclonic and anticyclonic weather conditions in temperate latitudes.

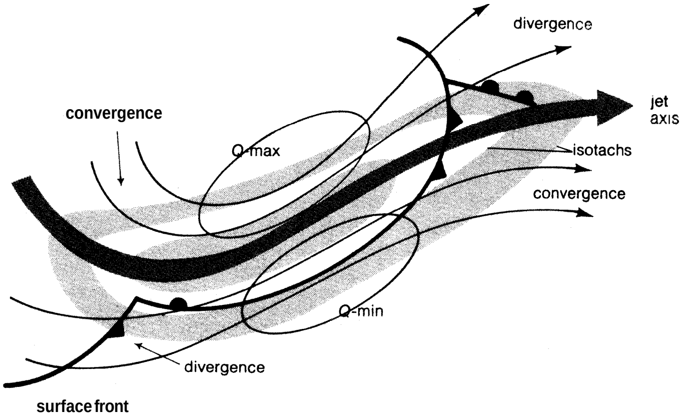


Figure 4: Vorticity and divergence distributions near a jet maximum. The streamlines (arrows) and convergence and divergence areas shown are characteristic of the jet stream in the upper troposphere (see text).

From the above equation it is clear that, if flow conditions are neither divergent nor convergent (that is, if $D = 0$), the absolute vorticity should not change with time but remain constant because $dQ/dt = 0$. This explains the reason for the formation of long planetary waves, the so-called Rossby waves, in the upper-tropospheric flow patterns: Q may be considered as the sum of the rotation of the air mass plus the earth's vorticity, given by the Coriolis parameter $2\omega \sin \varphi$. If an air mass moves southward, the Coriolis parameter decreases. If Q is to remain constant, the vorticity of the air-mass rotation must increase in proportion. This means that the air mass is acquiring more cyclonic rotation and therefore will return to higher latitudes. This reverses the process: because φ increases, the Coriolis parameter increases; consequently the rotation of the air mass will decrease and eventually become anticyclonic, thus repeating the whole process. The initial impulse, which forces the air mass to deviate from a purely westerly flow, may come from forced motion over or around large mountain ranges, such as the Rocky Mountains or the Himalayas. In the lee of these ranges are quasi-permanent low-pressure troughs in the upper flow patterns, which may be explained in the above manner and which greatly influence the climate in these regions.

Relation to topography and climate

The confluence of cold air, flowing down along the west side of these troughs, with warm tropical air moving from a southwesterly direction accounts for the strong horizontal temperature gradients prevailing there, as well as for the frequent occurrence of strong jet streams over the eastern United States and Japan (see Figure 1).

The divergence and convergence patterns near the jet stream, caused by the mass flow through the vorticity pattern, give rise to important vertical motions in the atmosphere. Upward motion leads to cloud formation and eventually to precipitation. Such motions occur beneath divergence regions at jet stream level.

Downward motions, on the other hand, lead to dissolution of clouds and fair weather. These occur beneath the convergent regions. Such downward motions may transport stratospheric air into the troposphere and toward the ground. During periods of nuclear testing, stratospheric air generally was contaminated by radioactive debris. Vertical transport through the jet stream was discovered to be an important factor in causing radioactive fallout at the ground.

OCCURRENCE AND DISTRIBUTION OF JET STREAMS

Polar Front and Subtropical Jet Streams, both occurring near the tropopause, have been described in detail above. In the Northern Hemisphere they are best developed during winter and early spring. The larger meanders, which both jet stream systems describe, fuse them into one single jet stream on monthly or seasonal mean maps or on hemispherically averaged cross sections through the atmosphere. On daily weather maps, however, two separate systems are distinctly present over wide regions of the Northern Hemisphere. The Subtropical Jet Stream during winter forms rather regularly three planetary waves with ridges over eastern China and Japan, over the United States, and over the eastern Mediterranean. In these three regions the Subtropical Jet frequently merges with the Polar-Front Jet Stream, giving rise to excessively strong jet maxima.

During summer in the Northern Hemisphere, the Subtropical Jet Stream loses its global identity and occurs only in sporadic velocity streaks.

In the Southern Hemisphere two distinct jet stream systems are present during winter, even on mean meridional cross sections through the atmosphere. This indicates that the meandering tendency of the two flow systems is not as strong there as in the Northern Hemisphere. During summer the Polar-Front Jet Stream system seems to dominate in middle latitudes, with stronger winds during this season than during winter.

Over the central United States a low-level wind maximum is frequently associated with an inversion near the 800- or 850-millibar (mb) level. This wind maximum is often referred to as the Low-Level Jet Stream. Because the flow in this wind system is usually from the south, it imports moist air from the Gulf of Mexico into the central United States. This rapid moisture transport plays an important role in the formation of severe thunderstorms and tornadoes. It is presently believed that this Low-Level Jet Stream is related to the diurnal variation of wind speed in the lower troposphere. This, in turn, is produced by frictional coupling and uncoupling of the atmospheric flow with the rough terrain during the day and night, and by the presence or absence of convective motions. The terrain slope from the Rocky Mountains into the Mississippi Valley also exercises an influence on the formation of the Low-Level Jet Stream.

The Tropical Easterly Jet Stream occurs near the tropopause over Southeast Asia, India, and Africa during the summer season. It is caused by southward flow under quasi-conservation of absolute angular momentum of air that travels out of the subtropical high-pressure cells (located over Tibet and the Sahara during summer), and moves Equatorward. Such motion will cause east winds. The Tropical Easterly Jet Stream is closely connected with the Indian and African summer monsoon.

During winter the polar stratosphere cools appreciably because of the lack of solar radiation. The ensuing strong horizontal temperature gradient in the stratosphere, directed from the Equator toward the pole, gives rise to a strong jet stream with a westerly wind direction that blows near the 50-kilometre level in high latitudes and meanders around the winter pole. It is called the Stratospheric Polar-Night Jet Stream.

Over the summer pole, where the stratosphere is exposed to continuous sunshine, temperatures are relatively warm. Thus the temperature gradient is directed from the pole toward the Equator, resulting in summer east winds in the stratosphere, which normally do not quite reach jet stream proportions.

The winter stratospheric circulation establishes itself gradually during autumn. It breaks down rather abruptly near the spring solstice in the Southern Hemisphere, and off and on during the middle and later winter in the Northern Hemisphere. The breakdown of the polar-night vortex leads to the abrupt warming of portions of the stratosphere at high latitudes in the winter hemisphere. The downward motions in midlatitudes, which predominate during this breakdown season, carry ozone to lower levels of the stratosphere. This leads to the observed spring maximum of ozone in middle latitudes.

Jet stream systems appear to be present in the upper mesosphere and in the thermosphere as well. In these regions of the high atmosphere, diurnal motions produced by tidal forces begin to dominate over the synoptic disturbances that characterize the weather in the troposphere and stratosphere. Because charged particles participate in these flow processes, at ionospheric levels, the name "electrojet" has been coined for some of these motion patterns. This term should be applied, however, only to motions in which the neutral gases of the high atmosphere participate in a significant proportion. Organized electron motions are more properly called currents.

See IONOSPHERE.

BIBLIOGRAPHY. E.R. REITER, *Jet-Stream Meteorology* (1963), a comprehensive text on the structure, dynamics, and occurrences of jet stream systems in the troposphere and stratosphere, with more than 1,800 literature references; *Jet Streams: How Do They Affect Our Weather?* (1967), a description of the discovery, structure, and dynamics of jet stream systems, more elementary than the above-cited work; *Atmospheric Transport Processes*, vol. 1 (1969), a thorough review of energy processes in the atmosphere that maintain the general circulation and jet streams, including rigorous mathematical treatment of the problem and extensive references to pertinent papers; S.-K. KAO and W.P. HURLEY, "Variations of the Kinetic Energy of Large-Scale Eddy Currents in Relation to the Jet Stream," *J. Geophys Res.*, 67:4233–3242 (1962), mathematical derivation of kinetic energy near jet streams with actual data on the summer and winter distribution of this energy over the Northern Hemisphere.

(E.R.R.)

Jewelry

Jewelry consists of objects of personal adornment prized for the craftsmanship that went into their creation and generally for the value of their components as well. In the prehistoric era and in modern primitive cultures, the materials considered rare and beautiful were shells, bones, pebbles, tusks, and wood; in more advanced civilizations, precious metals, precious and semiprecious stones, pearls, corals, enamels, vitreous pastes, and ceramics. Even in advanced civilizations where precious materials are available, in certain eras artist-craftsmen have sometimes placed less emphasis on the intrinsic value of materials than on their aesthetic function as components contributing to the effect of the whole. Thus, they might fashion a brooch out of steel rather than gold or platinum.

This article will cover the uses, forms, styles, materials, and techniques that have characterized jewelry throughout its 40,000-year history.

Materials and methods

MATERIALS

The first materials used to make objects for personal adornment were taken from the animal and vegetable world. The material taken from the animal world, in a natural or processed form, constituted the actual adornment, whereas vegetable fibres served as its support. A great variety of shells and pieces of shell were used during the prehistoric age and are still used by primitive tribes on the islands and coastal zones of nearly all the continents to make necklaces, bracelets, pendants, and headdresses. In the inland regions, the first materials used for personal adornment came from mammoths' tusks, the horns of reindeer and other animals, and, later on, amber and lignite.

Gold,
silver,
and
platinu

Metals. Of gold's properties, when it was first discovered (probably in Mesopotamia before 3000 BC), it was the metal's malleability that was a new phenomenon: only beeswax, when heated to a certain temperature, could be compared to it. Gold's molecules move and change position in accordance with the stresses to which it is submitted. so that when it is beaten it gains in surface area what it loses in thickness.

In modern jewelry, gold can take on a variety of hues when it is alloyed with other metals: water green, dead-leaf white, gray, and blue.

After gold, silver is the metal most widely used in jewelry and the most malleable. The discovery of some silver artifacts at Hisarlık in Asia Minor, which was also the location of the Homeric city of Troy, shows that this metal was known during the Copper Age. Before the classic age, however, silver made only rare appearances in jewelry. In general. silver was, and still is, used in jewelry for economic reasons or to obtain chromatic effects. It was often used in the 17th, 18th, and 19th centuries, however, as support in settings for diamonds and other transparent precious stones, in order to encourage the reflection of light.

Another rare metal, whose use in jewelry is fairly recent, is platinum. From the 19th century onward, this metal was used ever more frequently in jewelry because of its white brilliance and malleability, as well as its resistance to acids and high melting point.

Modern jewelry, like that designed by early-20th-century artists, introduced nonprecious metals such as steel.

Gems. In addition to gold, silver, and platinum, the precious materials most widely used in jewelry are gems—any precious or semiprecious stone. By definition this group also includes some animal and vegetable products with precious characteristics, such as amber, pearls, and coral. Conventionally, the following are classified as precious stones: diamonds, rubies (corundum), emeralds (beryl), and sapphires (corundum). To these, however, can be added chrysoberyl, topaz, and zircon because of their hardness and their refraction and transparency index.

Use of
diamonds,
rubies,
and
sapphires

Diamonds have the highest refraction index, and those used for jewels are very transparent. Diamonds from Indian deposits were known in ancient times; in the West the limited use of diamonds began in the late Middle Ages. Diamonds for jewelry are graded on the basis of colour from blue-white to yellow. Grading is also done on the basis of purity, which varies from perfectly clear, extremely pure stones to those with many impurities and flaws. Large demand provided an incentive for the production of false diamonds (as well as other stones) as early as 1675 in Paris.

Mogok rubies, from Burma, are the most highly prized because of their bright red colour (pigeon's blood). Those from Thailand are usually a more brownish colour, while those from Ceylon tend toward violet. Production of synthetic stones is far greater than the supply of natural rubies. The physical and optical properties of synthetic and natural rubies are so similar that it is difficult to distinguish between them.

The sapphire (blue variety of corundum) is considered one of the most valuable of precious stones. A sapphire's colouring usually indicates its origin. Those from Burma are deep blue. The Kashmir (Indian) sapphire is cornflower blue and is highly prized, being quite rare. Sapphires from Thailand are very similar in colour to those from Burma; those from Ceylon are of different shades but incline toward violet. Sapphires, like rubies, can be cut so that, in the light, a beautiful, luminous six-pointed star appears on the surface of the gem. Star sapphires and rubies are semiopalescent. Synthetic sapphires and rubies are produced by the same industries.

Use of emeralds

The green emerald is a precious stone used since very ancient times. There is documentation of its presence in Egypt during the life of Pharaoh Sesostris, in 1650 BC. At the end of the 16th century, emeralds from South America began to be brought into Europe. On the American continent, the first peoples to use emeralds were those belonging to the pre-Columbian civilizations, in particular the Incas. In 1935 in the U.S. (Chatham) and in Germany (Farbeindustrie), synthetic emerald crystals were made with characteristics similar to natural ones.

Among the beryls, mention must be made of morganite (pink beryl) found in various shades of peach-blossom pink. The main deposits are in California and Madagascar.

The two best known and most widely used varieties of chrysoberyl are alexandrite (transparent) and oriental cat's-eye (opaque). Because of its great power of absorption of certain colours, alexandrite looks green in daylight and reddish purple in artificial light. The cat's-eye is a yellowish-green colour and is characterized by a luminous line. The intensity of the light in this line varies according to the brightness of the rays of light that strike it.

One of the most important gems with pure crystals is the topaz, used a great deal in jewelry. The honey-yellow variety is the best known, but there are also pink, red, blue and the less used colourless stones. The oriental topaz (a corundum) and citrine quartz are also widely used. They are less rare than other kinds of topaz and, therefore, less expensive but create a similar effect.

Among the less important and less rare stones, the zircon is quite widely used in its three varieties: orange, blue, and colourless. The orange variety is called jacinth and was used to a great extent in classical antiquity. The blue variety is called starlite or Siam zircon, while the third type is called Ceylon or Matara diamond.

Among the semiprecious stones used in jewelry are amethyst, garnet, aquamarine, amber, jade, turquoise, opal, lapis lazuli, and malachite.

Varieties of pearls

The pearl is one of the oldest gems known to man. Its colour varies according to the waters from which it comes. Pearls from the Persian Gulf are usually cream coloured; those from Australia are white with greenish or bluish shades; golden-brown pearls come from the Gulf of Panama; those from Mexico are black or reddish brown; pink pearls are from Ceylon; and those from Japan are cream coloured or white with greenish tones. The main characteristic of the pearl is its iridescence. Baroque pearls are those with defects in their outer layer. In modern times, baroque pearls are rounded off artificially but, in the 16th and 17th centuries, their irregular form was exploited in jewelry by using them to make up parts of animals or other figures. After huge quantities of cultivated pearls invaded the world market, interest in real pearls underwent a considerable decrease.

Other materials. Some of the less highly prized materials used for jewelry are coral (corallium rubrum), glass, wood, and ceramic. The most recent material used in creative jewelry is plastic, which, when exposed to heat or cold in a mold, can take on and retain any shape desired. In addition, when unmolded and cooled, the material can be subjected to further working or can be combined with other similar or different materials.

The imagination shown by jewellers is often related to material of local origin. In Malaysia, for example, use is made of tigers' claws bound in gold, with inserts of precious stones, to create pins in the shape of a half-moon. The use of brightly coloured feathers from rare birds was also quite common, especially in pre-Columbian civilizations. Used traditionally only in Chinese jewelry are the breast feathers of a rare seabird similar to the fisher martin (Fei Ts'ui), which are mounted on a leather support and then on gold or silver.

PROCESSES AND TECHNIQUES

All materials that have been used over the centuries for the manufacture of jewelry have undergone to some extent mechanical, physical, or chemical treatment for the purpose of transforming their raw shapes into shapes which, in addition to being functional, also satisfy certain aesthetic concepts.

Metalwork. The basic components of jewelry have always consisted of metal sheet, metal cast in a mold, and wire (more or less heavy or fine). These components take on the desired shape by means of techniques carried out with the help of tools. Gold in its natural state was beaten while hot or cold and reduced to extremely thin sheets (this operation could be performed with stone hammers). The sheets were then cut into the desired sizes (with flint knives).

Metal sheet, embossing, openwork, and granulation

Examination of the most ancient pieces of jewelry shows that one of the techniques used most widely in decorating metal sheets for jewelry was embossing (relief work). Through the centuries embossing techniques have remained substantially unchanged, although in modern times mechanization has made possible mass production of decorative parts of jewelry, with great savings of time and labour but with a corresponding lack of art.

In repoussé, the relief is pressed (in a negative mold) or hammered out from the reverse side of the gold sheet and then finished off on the right side with a hammer or engraving tool. For half-modelled or completely round reliefs, the gold leaf was pressed onto wooden or bronze models. Completely round objects were made in two pieces and then welded together.

Another embossing, or relief, technique is engraving, which involves impressing designs into the metal with a sharp tool.

Decorative openwork designs can be created by piercing the gold leaf. In the Roman period this technique was called *opus interassile*.

Granulation is a decorative technique in which small or minute gold balls (with diameters ranging from $\frac{1}{60}$ to $\frac{1}{180}$ of an inch) are used to form silhouettes on smooth or embossed metal.

Casting from precious metals has always been rare. When the relief was to be visible only from one side, the metal was poured into the cast and, when hardened, touched up with a graver. When the relief was to be fully modelled, the cire perdue (lost-wax) process, involving casting from a wax mold, was used.

Metal wire and filigree

Gold and silver wire, according to its function, can be made into various sizes, shapes, sections, and weights. It can serve to join, to support pendants of varying importance, to make necklaces and bracelets, or to alternate with other decorative components.

From the 3rd millennium before Christ through the present day, chains—ranging from the simple type, consisting of a series of round or oval rings, to one of the oldest elaborations, the "loop in loop," or square, chain—have offered goldsmiths the widest field for decorative imagination.

Filigree is a form of decoration made exclusively from fine gold or silver wire welded onto the surface of an object made of the same metal or done in openwork (without a background). The decoration to be carried out is designed first on a model with a flat or curved surface identical to that on which the completed filigree is to be welded or to the unsupported shape that it must assume. It can be made from smooth wire or from a ropelike plait or from a series of small hemispheres. A more complicated type of filigree consists of metal wire made in the shape of beads called granulated filigree.

After having been prepared separately, the different parts that make up a piece of jewelry are put together. In primitive jewelry this was done mechanically, by inserting beaten pins, by bending and beating the parts to be fastened together, or by binding them with gold wire or tape. Welding is a technique belonging to a more highly developed stage of ancient goldworking, (end of the 3rd millennium BC).

Enamel work. In enamel work, powdered glass coloured with metal oxides diluted with water and adhesive is applied to certain parts of the piece of jewelry that have been cut lower or surrounded with a raised rim made of gold, silver, or copper. The object is then heated until the glass melts and adheres to the metal. As the enamel gradually cools, it crystallizes and, when smoothed, takes on greater lustre and colour. The enamel applied to jewelry can be opaque or translucent. By letting light through, transparent enamel catches reflections from the metal to which it is applied and makes visible any engraving done on the metal. Enamel is also distinguished according to the way it is applied, as in cloisonné, champlevé, basse taille, painted, and *plique à jour.*

Enamelling preceded the polychromy created by precious stones. In the beginning, in Egypt, Greece, and the Sāsānid period in Iran, unpolished enamelled parts of jewelry were often used to imitate lapis lazuli or malachite.

Niello technique

To a limited extent, jewelry was also decorated with the niello technique (from the Latin *nigellus*, adjective derived from *niger* = black). This consists of cutting grooves in gold or silver with a graver and then filling these with a powder made of red copper, silver, lead, sulfur, and borax. When heated, the powder melts and fills the grooves, adhering to the metal. After the piece has cooled, the surface is smoothed and polished, and the design shows up in black.

Gem engraving, setting, and cutting. The most ancient technique of stone engraving is intaglio-incised carving. This technique at a certain point gave rise to the idea of exploiting the different coloured layers of certain minerals to create contrasting figures (cameo): the background was cut down to the lower level, of a different colour or shade, in order to make the subject stand out chromatically. The stones that have properties suited for this purpose are sardonyx, agate, and onyx.

The cameo is usually one of the components for necklaces, bracelets, and rings or is included in medallions with a jewelled frame. The art of cameo in jewelry was most highly developed during three periods: the late Roman period, the Renaissance, and the neoclassic period in the 18th century.

Gem setting

The evolution of techniques of setting has followed that of stonecutting. The insertion of gems in jewelry can be done in various ways. The setting can have a round, square, oval, or rectangular collet (rim); in periods in which gems were mounted in their own irregular shapes, the collet followed this form. Usually, on the inside of the collet a short distance from the edge, there is a protrusion on which the stone rests. The edge is pounded down around the gem to ensure its stability. In coronet settings the form may be conical or pyramidal, solid or pierced. The edge is first shaped into a row of teeth, which are later hammered down onto the gem in order to hold it in place. Until fairly recently, nearly all gems were mounted on a metal base; and transparent stones, according to their colour, were placed on a gold or silver base to increase the amount of light reflected. As new cuts were developed for stones, setting techniques also progressed, especially for those jewels in which important stones like diamonds, emeralds, and rubies form the main theme. The tendency was to leave the stones as visible as possible (especially in rivière necklaces and bracelets made only of diamonds) by mounting them with a very small ring of white gold or platinum fitted closely against the back of the stone. Three claws, attached to this ring, hold the stone in place.

Pearls, like some coloured stones, in ancient classic times were pierced with a drill, the hole going half or all the way through according to whether the pearls were to be strung on a necklace or fastened onto a jewel.

Gem cutting

Until the 15th century, stones were only polished or the part to be left visible was rounded into a dome shape called cabochon. The cutting known as faceting gradually developed from the first attempts in the 15th century. During the 16th century the simple rose cut began to be used, after which there were no new developments until 1640 when, under the patronage of Cardinal Mazarin, the first brilliant cut was carried out (also called the Mazarin cut). Toward the end of this century, a Venetian succeeded in obtaining the triple brilliant cut, which is still used. The numerous cuts used for diamonds today are usually applied to other precious and semiprecious transparent stones as well. For emeralds, rubies, and other coloured stones the square or rectangular cut with a stepped bulb or the cabochon cut are usually used.

Historical development

During most of its history, jewelry has been worn as a sign of social rank—forbidden by sumptuary laws to all but the ruling classes—and as a talisman to avert evil and bring good luck. During the Middle Ages, for example, a ruby ring was thought to bring its owner lands and titles, to make him virtuous, to protect against seduction and to prevent effervescence in water—but only if worn on the left hand. Purely decorative jewelry has existed only since the period of the French Revolution. The possibility of tracing jewelry's historic itinerary derives primarily from the custom, beginning with the most remote civilizations, of burying the dead with their richest garments and ornaments. This source will last as long as archaeological diggings continue. The diggings have also brought to light occasional ancient "treasures"; that is, they have discovered hiding places where precious objects were secured for safekeeping during invasions and that were never recovered because of the death of those few who knew their location. Plastic and pictorial iconography—painting, sculpture, mosaic—offer abundant testimony to the jewelry worn in various eras and, most important, to the jewelry rich in precious components. Because such pieces were often disassembled and their components recast into new forms or sold, little has come down to us.

Prehistoric era

It is probable that man thought of decorating his body before he thought of making use of anything that could suggest clothing. Before precious metals were discovered, men who lived along the seashore decorated themselves with a great variety of shells, fishbones and fish teeth, and many-hued pebbles carried up onto the beach by the waves, exercising their aesthetic sense in making a selection and in the techniques for fastening and arranging these materials on their bodies. Men who lived inland used as ornaments materials from the animals they had killed for food: reindeer antlers, mammoth tusks, and all kinds of animal bones. After they had been transformed, with great patience, from their natural state into various elaborate forms, these materials, together with animal skins and bird feathers, provided sufficient decoration.

Most of the evidence that has survived from this period consists of cave paintings and carvings. Some of these represent women decorated with long necklaces, bracelets, and headdresses. This era was followed by one that saw a transition from a nomadic life to a settled social order and the subsequent birth of the most ancient civilizations. Most peoples settled along the banks of large rivers, which facilitated the development of agriculture and animal husbandry. Indirectly, this also led to the discovery of virginal alluvial deposits of minerals, first among which were gold and precious stones.

Over the years the limited jewelry forms of prehistoric times multiplied until they included ornaments for every part of the body. For the head there were crowns, diadems, tiaras, hairpins, combs, earrings, nose rings, lip rings, and earplugs. For the neck and torso there were necklaces, fibulae (the ancient safety pin), brooches, pectorals (breastplates), stomachers, belts, and watch fobs. For the arms and hands, armlets, bracelets, and rings were fashioned. For the thighs, legs, and feet craftsmen designed thigh bracelets, ankle bracelets, toe rings, and shoe buckles.

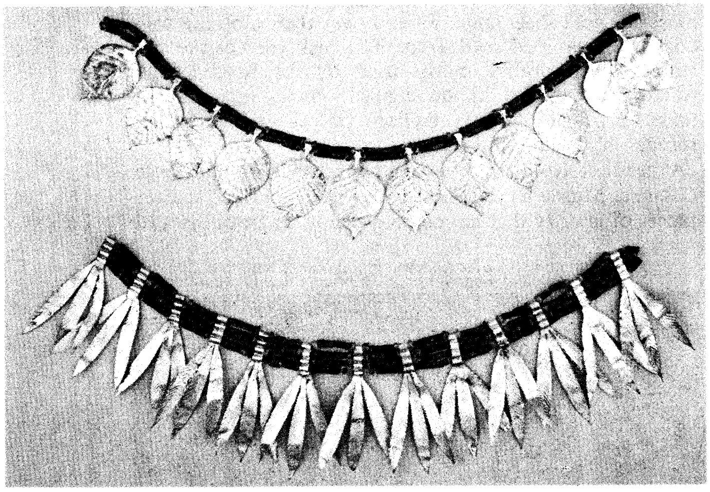
WESTERN CULTURE

Antiquity. *Sumerian.* The most ancient examples of jewelry are probably those found in Queen Pu-abi's tomb at Ur in Sumeria (now called Tall al-Muqayyar), dating from the 3rd millennium BC. In the crypt, the upper part of the queen's body was covered with a sort of robe made of gold, silver, lapis lazuli, cornelian, agate, and chalcedony beads, the lower edge decorated with a fringed border made of small gold, cornelian, and lapis-lazuli cylinders. Near her right arm were three long gold

Prehistoric African women wearing jewelry perform a ritual dance; rock painting, Tassili des Adjjer, Algeria.
J.D. Lajoux

pins with lapis-lazuli heads, three amulets in the shape of fish—two made of gold and one of lapis lazuli—and a fourth amulet of gold with the figures of two seated gazelles. On the Queen's head were three diadems, each smaller than the one next to it, fastened to a wide gold band: the first, which came down to cover the forehead, was formed of large interlocking rings, while the second and third were made of realistically designed ivy and willow leaves. Above the diadems were gold flowers, on drooping stems, the petals of which had blue and white decorations. On the back of the headdress was a Spanish-type comb, with teeth decorated with golden flowers. Huge golden earrings, in the shape of linked, tapered, semitubular circles, completed the decoration of the head. On the neck was a necklace with three rows of semiprecious stones interrupted in the middle by an openwork flower in a gold circle. Many rings were worn on the fingers. There were large quantities of other jewels—among them wrist and arm bracelets and pectorals—belonging to the handmaidens, dignitaries, and even the horses that formed part of the funeral train. As was the custom, the queen's attendants had killed themselves in the crypt after the burial ceremony.

By courtesy of the trustees of the British Museum



Sumerian gold and faience diadems from Queen Pu-abi's tomb, Ur, c. 2500 BC. In the British Museum.

As this description suggests, Sumerian jewelry forms, much more numerous than those of modem jewelry, represent almost every kind developed during the course of history. Nearly all technical processes also were known: welding, alloys, filigree, stonecutting, and even enamelling. Sources of inspiration, aside from geometry (disks, circles, cylinders, spheres), were the animal and vegetable world; and expressive forms were based on an essential realism enriched by a moderate use of colour.

The Tutankhamen treasure

Egyptian. The sensational discovery of the tomb of the pharaoh Tutankhamen (18th dynasty: 1567–1320 BC) revealed the fabulous treasures that accompanied an Egyptian sovereign, both during his life and after death, as well as the high degree of mastery attained by Egyptian goldsmiths. This treasure, now in the Egyptian Museum, represents the biggest collection of gold and jewelry in the world. The Pharaoh's sarcophagus was made entirely of gold and the mummy was covered with a huge quantity of jewels. More jewels were found in cases and boxes in the other rooms of the tomb. The diadems, necklaces, pectorals, amulets, pendants, bracelets, earrings, and rings are of superb quality and of a high degree of refinement rarely surpassed or even equalled in the history of jewelry.

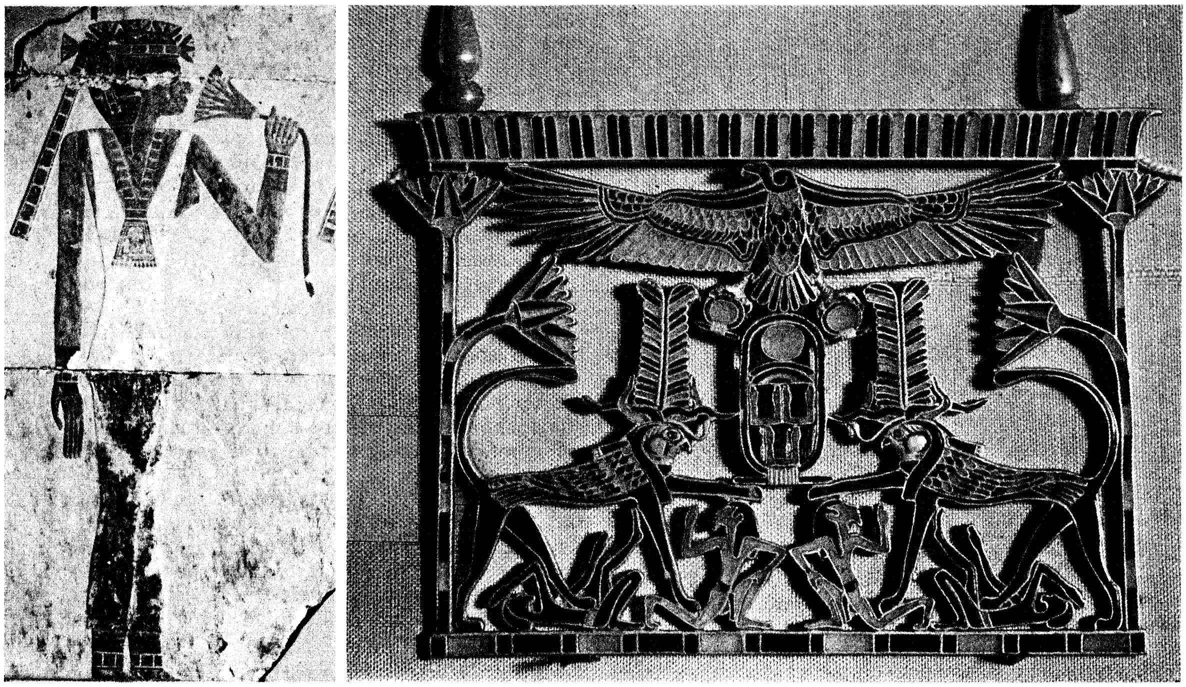
The ornaments in Tutankhamen's tomb are typical of all Egyptian jewelry. The perpetuation of iconographic and chromatic principles gave the jewelry of ancient Egypt—which long remained uncontaminated in spite of contact with other civilizations—a magnificent, solid homogeneity, infused and enriched by magical religious beliefs. Ornamentation is composed largely of symbols that have a precise name and meaning, with a form of expression that is closely linked to the symbology of hieroglyphic writing. The scarab, lotus flower, Isis knot, Horus eye, falcon, serpent, vulture, and sphinx are all motif symbols tied up with such religious cults as the cult of the pharaohs and the gods and the cult of the dead. In Egyptian jewelry the use of gold is predominant, and it is generally complemented by the use of the three colours of cornelian, turquoise, and lapis lazuli or of vitreous pastes imitating them. Although there was a set, fairly limited repertoire of decorative motifs in all Egyptian jewelry, the artist-craftsmen created a wide variety of compositions, based mainly on strict symmetry or, in the jewelry made of beads, on the rhythmic repetition of shapes and colours.

Forms of Egyptian jewelry

The concept of symmetry was utilized on the small pectoral or pendant (85 × 61 millimetres or 3.3 × 2.4 inches) that belonged to Sesostris III in the 12th dynasty (1991–1786 BC). Here the superbly rhythmic composition is framed by an architectonic design obtained by leaving open all of the nonfigurative part. The jewel is coloured with cornelian, turquoise, and lapis-lazuli inlays, while the function of the gold separating these materials is limited to creating the design. The victorious Pharaoh is represented by two lions with the plumed heads of falcons in a symmetric position in the act of trampling conquered Nubians and Libyans. Over the scene is the protective vulture of Upper Egypt with wings outspread, the only nonsymmetrical component in the ornament (Egyptian Museum, gen. cat. 52002). These memorial or dedicatory pendants, as well as other small jewels such as earrings, bracelets, and rings, consist exclusively of symbols.

Necklace beads—generally made of gold, stones, or glazed ceramic—are cylindrical, spherical, or in the shape of spindles or disks and are nearly always used in alternating colours and forms in many rows. The necklaces have two distinct main forms. One, called menat, was the exclusive attribute of divinity and was therefore worn only by the pharaohs. Tutankhamen's menat is a long necklace composed of many rows of beads in different shapes and colours, with a pendant and with a decorated fastening that hung down behind the shoulders (Cairo, Carter cat. No. 269K). The other, much more widely used throughout the whole period, was the *usekh,* which, like the vulture-shaped necklace from the tomb of Tutankhamen, also has many rows and a semicircular form (Cairo, Carter cat. No. 256MMM).

Of the many diadems made by Egyptian artist-craftsmen, one of the earliest was discovered in a tomb dating from the 4th dynasty (2613–2494 BC). It consists of a gold band supported by another band made of copper, to



Egyptian *jewelry* of the Middle Kingdom period.
(Left) A daughter of the noble Dhut-hotpe wearing a pectoral, headdress, and wrist and ankle bracelets. Painted limestone relief from al-Barshā', 12th dynasty. (Right) Gold pectoral with semiprecious stones belonging to King Sesostris III, 12th dynasty. In the Egyptian Museum, Cairo.

(Left) John Ross, (right) Hirmer Fotoarchiv, Munchen

which three decorative designs are applied. In the centre is a disk worked with embossing in the form of four lotus buds arranged radially. On the sides are two papyrus flowers linked horizontally at the base by a disk with a cornelian, while the upper line of the flowers comes together to create a kind of nest in which two long-beaked ibis crouch. The floral and animal symbology is carried out with a style that interprets and characterizes the theme.

Among the treasure discovered in the tomb of Queen Ashhotep (18th dynasty) is a typical Egyptian bracelet. It is rigid and can be opened by means of a hinge. The front part is decorated with a vulture, whose outspread wings cover the front half of the bracelet. The whole figure of the bird is inlaid with lapis lazuli, cornelian, and vitreous paste.

Foreign influence

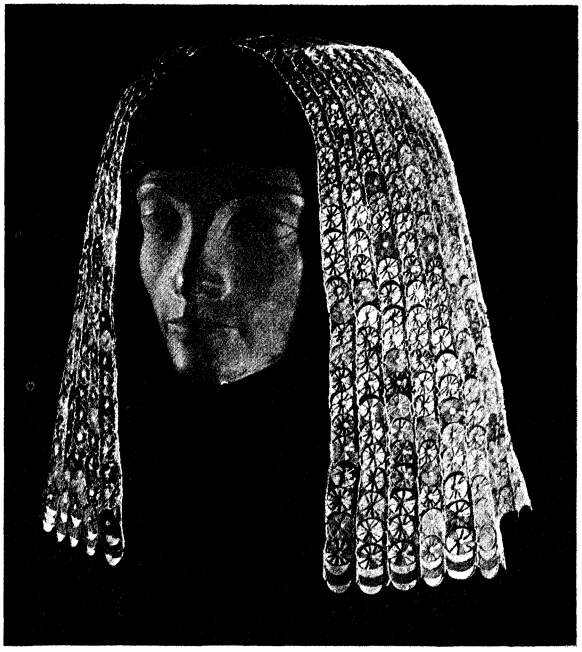
A first sign of outside influence occurs in the 18th dynasty and consists of earrings, which are imported jewels, unknown in classical Egyptian production. Another evidence of the influence of foreign styles in some of the jewelry of the 18th dynasty is a headdress (reconstructed) that covered nearly all of the hair, made of a network of rosette-shaped gold disks forming a real fabric (New York, Metropolitan Museum of Art). Foreign influence increased to an ever greater extent during the last dynasties and with the arrival of the Greeks. Like all other forms of artistic expression, in spite of three centuries of Ptolemaic dynasty (up through 30 BC), the great artistic tradition of Egyptian jewelry slowly died out, and the introduction first of Hellenism and then of the Romans led to the definitive decline of the most monumental cultural and artistic structure known throughout all history.

Mycenaean. Because the Mediterranean island of Crete lay near the coasts of Asia, Africa, and the Greek continent and because it was the seat of prosperous ancient civilizations and a necessary point of passage along important sea-trading routes, it developed a level of wealth which, beginning about 2000 BC, stimulated intense goldworking activities of high aesthetic value. From Crete this art spread out to the Cyclades, Peloponnesus, Mycenae, and other Greek island and mainland centres. Stimulated by Cretan influence, Mycenaean art flourished from the 16th to the 14th century, gradually declining at the beginning of the 1st millennium BC.

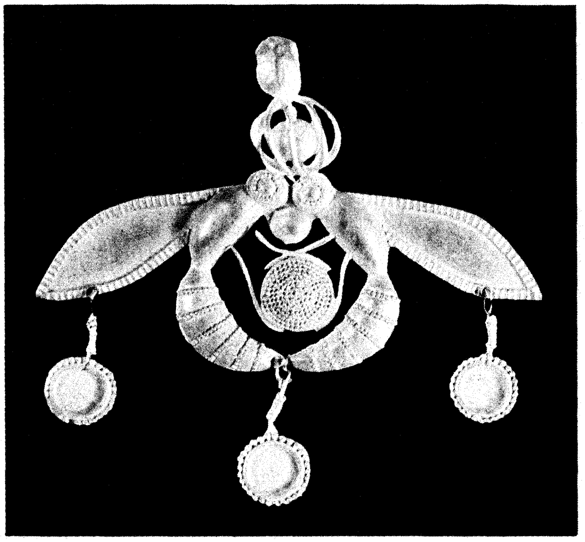
Among the techniques used in goldworking were granulation and filigree, but the most widely used was the cutting and stamping of gold sheet into pieces to form necklaces and diadems, as well as to decorate clothing. The kings from Period I of Mycenaean civilization (c. 1580–1500 BC) discovered in their burial places, wore masks of gold sheet and scattered over their clothing were dozens of stamped gold disks. The disks reveal the rich variety of decorative motifs used by the Mycenaeans: round, rectangular, ribbon shaped—including combinations of volutes, flowers, stylized polyps and butterflies, rosettes, birds, and sphinxes.

A pendant from a Mycenaean tomb at Mállia, Crete, (Candia Museum) is one of the most perfect masterpieces of jewelry that has come down to us from the 17th

By courtesy of the Metropolitan Museum of Art. New York, purchased with funds given by Henry Waiters and Edward S. Harkness, 1926



Headdress of a lady of the Egyptian court during the Nai Kingdom period. Gold, originally inlaid with cornelian and glass; from Thebes, reign of Thutmose III, 18th dynasty. In the Metropolitan Museum of Art, New York.



Mycenaean gold pendant of bees encircling the sun, showing the use of granulation; from a Mycenaean tomb at Mállia. 17th century BC. In the Candia Museum, Crete.
Dimitri

century BC. The sun's disk is covered with granulation and is held up by two bees, forming the central part of the composition. Ring bezels (tops of the rings), with relief engravings of highly animated pastoral scenes, cults, hunting, and war, are also fine. Like those of the other jewelry forms, the ornamental motifs of the necklaces are varied, including dates, pomegranates, half-moons facing each other, lotus flowers, and a hand squeezing a woman's breast. During the late Mycenaean period, earrings appeared in the shape of the head of a bull, an animal frequently represented in early gold plate.

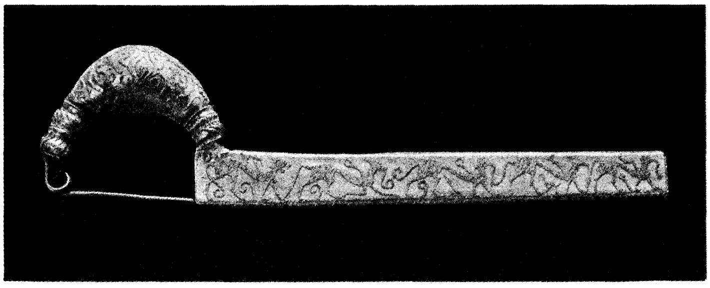
Orientalizing period

Phoenician. Phoenicia, a centre for both the production and exportation of jewelry, was not a source of great originality. It is to the trading done by this people throughout the Mediterranean, however. that we owe knowledge of the products of the most highly developed civilizations in the most remote lands—northern Africa, Sardinia, Spain, and Italy. The period in the 8th and 7th centuries BC, during which Scythian–Iranian oriental objects with their animalistic motifs were spread and consequently imitated throughout the Mediterranean countries, especially in Greece and Italy, is called the orientalizing period.

Etruscan. In Etruria, to a much greater extent than elsewhere, the stimulus provided by the jewelry imported by the Phoenicians led to emulation that soon had imposing results. Alongside imported objects and mechanically repeated oriental motifs, original forms, techniques, and styles developed that were the result of Etruscan taste. There was an entirely new concept, in which the goals of magnificence, impressive size, and a great wealth of decoration led to some of the most outstanding achievements in the history of jewelry. Technical virtuosity exploited all of the resources available to filigree and above all to granulation, carried out with gold alone without chromatic inlaying.

Fibulae began to be made in forms other than the single oriental leech shape: with a dragon bow, lozenge shaped, with a long foot. Like such ornaments as pendants and the heads of pins, fibulae were often decorated with gold dust, in which opaque granulated figures—ibexes, chimeras, sphinxes, winged lions, centaurs, horsemen, and warriors, nearly all of oriental derivation—stand out against the smooth surface of the gold (one notable example is the fibula from the Lictor's tomb in Vetulonia; Florence, Museo Archeologico).

SCALA, New York



Etruscan fibula of sheet gold decorated with animals made by the granulation technique, from the Lictor's tomb, Vetulonia, 7th century BC. In the Museo Archeologico, Florence.

The most elaborate, complicated examples of orientalizing Etruscan jewelry consist of very large brooches with fully sculptured decoration applied to a combined tubular and plate structure. The minutely designed granulated figures of sphinxes, winged lions, chimeras, winged griffons, and human heads—set in series in alternating rows—form a plastic fabric, the details of which are of astonishing technical ability, while at the same time they suggest the evocative, mysterious animalistic symbolism of near-Asian civilizations.

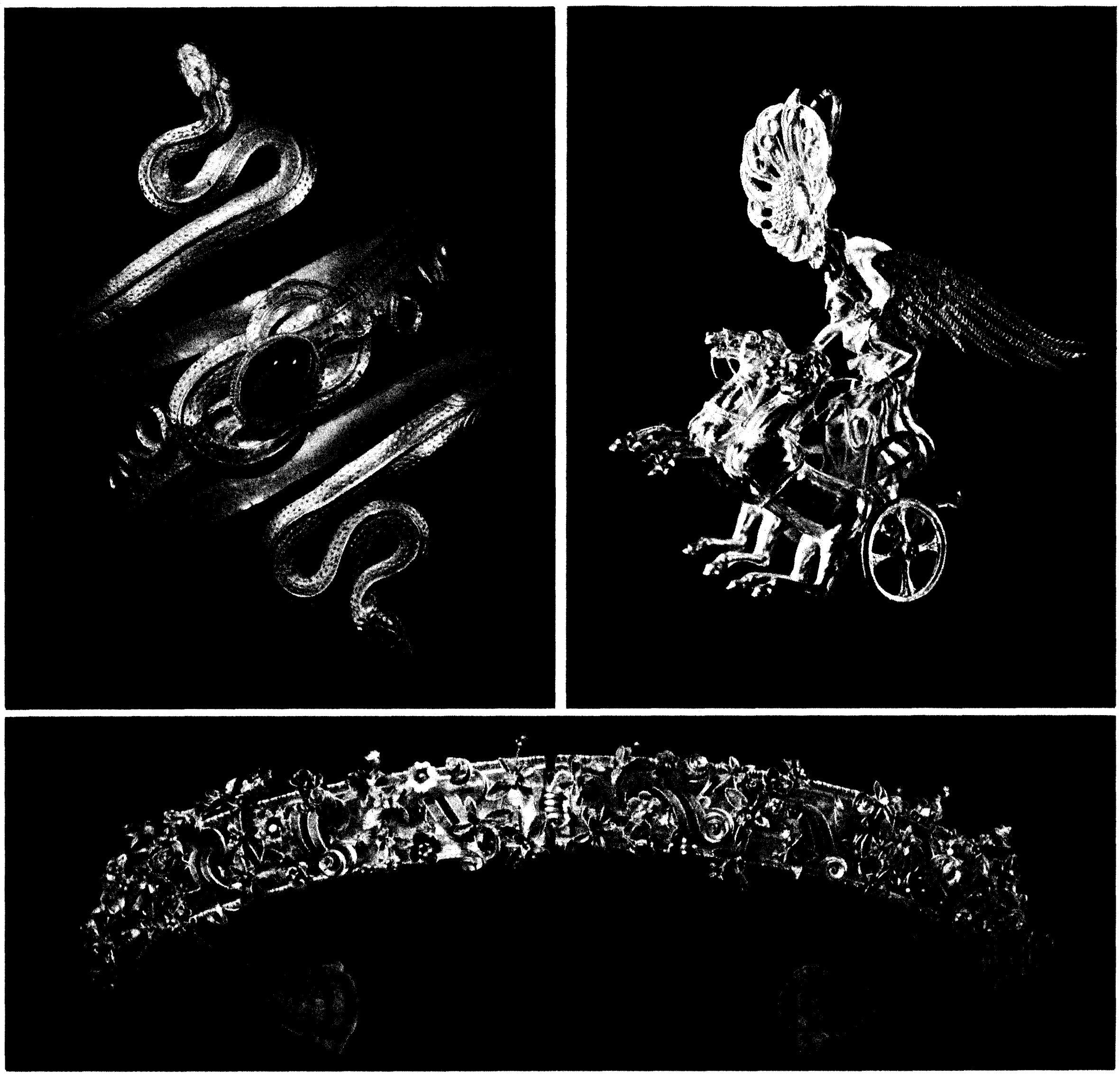
In the period that followed the orientalized one, Etruscan jewelry revealed Ionic influence (6th-5th centuries BC). The most beautiful examples are necklaces made of many flexible chains that cross each other and bear different rows of embossed pendants in the shape of harpies, mermaids, Gorgons, and Sileni, interspersed with others such as pomegranates, acorns, lotus flowers, and palms. These show the clear influence especially in the modelling of the pendant heads of the Greek severe period, an influence that spread throughout the entire Etruscan territory, from Spina on the Adriatic coast of Italy to southern Italy. Even clearer evidence of the acceptance of imported forms is provided by a new shape, the bulla, a pear-shaped sphere used to hold perfume. Its surface was decorated with embossed and engraved symbolic figures.

Greek In the 7th and 6th centuries BC the jewelry produced in Attica and the Peloponnese shows evidence of strong oriental stylistic influence the same influence which in Etruscan territory turned up in a much more magnificent form. In the 5th century BC the Ionic style became predominant, taking the place of the too showy oriental style. War scenes and animals of oriental origin disappeared for example. from the wide oval ring bezels and were replaced exclusively by the human figure: naked riders on galloping horses; seated and standing maidens, clothed or naked; deities and mythological figures. This extremely refined repertoire in reality was more closely related to sculpture and to classic ideals of beauty than to decoration. Indeed, in its long evolution, Greek jewelry has the predominant character of sculpture in miniature and represents isolated figures or religious, mythological. or heroic scenes.

Hellenistic jewelry

Greek expansion into Asia Minor to the east, southern Italy to the west, and the Balkan peninsula to the north resulted in the Hellenization of this entire area. Under the reign of Alexander the Great, a magnificent era for jewelry began. Hellenistic jewelry, much more so than painting and sculpture, underwent flourishing development in the art centres of the different regions under Greek rule. In the 3rd and 2nd centuries BC, the technical ability of Hellenistic goldsmiths reached the highest levels ever attained. A style both sumptuous and full of plastic vigour was created, in which meticulous arrangement of the decorative motifs resulted in the contrast and harmony, clarity and unity, rhythms and cadences that make some of these jewels complete works of art. The very fine technique and virtuosity in miniature is reflected in the creation of the first cameos and in disk earrings bearing pendants, often of minute proportions. A real masterpiece is an earring with a winged figure of a woman driving a two-horse chariot (Boston, Museum of Fine Arts). The precision of its tiny details, the severity of style with which it is modelled and the rhythmic dynamism of the figures make this earring a microscopic monument of sculpture.

Also worthy of high consideration are the magnificent diadems that came into wide use as a result of the Persian conquests made by Alexander the Great. One type is a rigid elliptical shape with a Hercules knot in the centre and pendants hanging down over the forehead (The Her-



Greek jewelry.
(Left) Gold spiral bracelet of two snakes whose tails are tied in a Hercules knot that is decorated with a garnet from Eretria in a bezel setting, 4th–3rd century BC. In the Schmuckmuseum Pforzheim, West Germany. (Right) Gold earring with Nike driving a chariot; from the Peloponnese, 4th century BC. In the Museum of Fine Arts, Boston. (Bottom) Gold diadem embellished with blue, green, red, and white enamelled flowers; from a tomb at Canossa, 3rd century BC. In the Museo Nazionale di Taranto, Taranto, Italy.

By courtesy of (left) the Schmuckmuseum Pforzheim. Germany, (right) the Museum of Fine Arts, Boston: photograph (bottom) Hirmer Verlag. München

cules knot was the most famous one used in ancient times, as it was considered a magic knot and, in jewels, took on the significance of an amulet. It was also used on bracelets, belts, and rings during this period.) One of the most impressive of these diadems is the one from Kerch (James Loeb Collection, Murnau am Staffelsee). Above the Hercules knot is a sculptured figure of Nike flanked by two sea serpents.

One type of necklace during the Hellenistic period was made of gold pieces, often hollow or filled with resin, in the shape of acorns, amphorae and rosettes, sometimes alternated with stones or vitreous paste. In the 3rd century BC, the bracelet in the shape of a serpent originated and remained popular through the Roman period. The serpent motif was also used for rings.

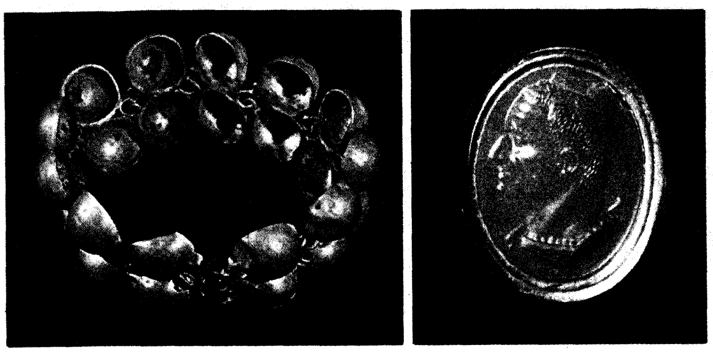
Roman. In ancient Rome, jewelry was used to an extent never seen before and not to be seen again until the Renaissance. Imperial Rome became a centre for goldsmiths' workshops. Together with the precious stones and metals that were brought to the city came lapidaries and goldsmiths from Greece and the oriental provinces. The gold ring, which under the republic had been a sign of distinction worn by ambassadors, noblemen, and senators, gradually began to appear on the fingers of persons of lower social rank until it became common even among soldiers. The great patrician families in Rome and the provinces possessed not only jewels but also magnificent gold and silver household furnishings, as shown by the objects found in Pompeii and Boscoreale (Louvre).

Stylistic characteristics of Roman jewelry

From the standpoint of style, Roman jewelry in its earlier phases derived from both Greek and Etruscan jewelry. Later it acquired distinctive features of its own, introducing new decorative themes and attaching greater importance to sheer volume (such as massive rings), in keeping with the rather pompous rhetorical spirit displayed at that point in cultural history.

The motif of a serpent coiled in a double spiral, copied from Hellenistic models, was frequently used for bracelets, rings, arm bands, and earrings. The Romans also used Greek geometric and botanical motifs, palmettos, fleeting dogs, acanthus leaves, spirals, ovoli, and bead sequences. From Etruscan gold jewelry the Romans took the strong plasticity of the bulla, which they transferred to necklace pendants sparely decorated with filigree or combined in completely smooth hemispheres in bracelets, headdresses, and earrings.

In Pompeii and Rome, jewelry began to take on Italian characteristics. New decorative motifs of a magical na-



Roman jewelry.
(Left) Gold bracelet of thirteen pairs of plain hemispheres. (Right) Gold ring with cameo from the House of Menander. Both are from Pompeii, 1st century BC–1st century AD. In the Museo Archeologico Nazionale, Naples

Arnoldo Mondadori Editore—C E A M

ture began to appear, such as the half-moon and the wheel with four spokes. In addition, as Roman jewelry freed itself of Hellenistic and Etruscan influences, greater use was made of coloured stones: topazes, emeralds, rubies, sapphires, and pearls. A strong preference was shown for engraved gems, so much so that they were considered collectors' items by wealthy people, including Caesar himself. The stones were set in bezels or supported by pins that passed through them, Another new technique also came into use called *opus interassile*, with which a flat or curved metal surface was decorated with tiny pierced motifs. Many pendants were used in the earrings: from a ring a series of pieces hung down with square bezels or bands of small bullas alternating with stones, which in turn supported pendants in different shapes.

There was an extremely varied production of gold mesh and chains, often containing inserted bezels set with stones or half pearls, while others had ivy or laurel leaves attached to them. Although pendants were not used on necklaces in the beginning, later examples have pendants in the form of embossed medallions like the one from Tortosa, Syria, with the head of Medusa (British Museum, No. 2730). Precious stones, vitreous pastes, and cameos with golden frames also served as pendants for necklaces. Toward the end of the 3rd century AD, necklaces often bore medallions or gold coins with portraits of the emperors.

Middle Ages. *Byzantine.* The fabulous phenomenon that was ancient Rome, the bearer of civilization throughout almost the entire known world, began to lose vitality and decline near the end of the 4th century AD though not without leaving indelible traces of its unique culture and civilization. The Roman Empire embraced Christianity, although in reality it was the papacy that embraced the Roman Empire, acquiring ever more ample powers. The intention of the Byzantine court (at Constantinople, the new seat of imperial power) to maintain Roman supremacy in the field of the arts was forced to give way to a style more closely related to that of the near Orient. Partly for religious reasons, this style soon developed a new spirit and its own distinctive characteristics. The wave of Iconoclasm—the controversy in the 8th and 9th centuries about the depiction of images in religious art—gave the decoration of jewelry, too, a basically ornamental nature, in which the techniques used to the greatest extent were filigree and enamelling, as well as the copious application of precious stones and pearls. Very complex decorations and arabesques were obtained with filigree, while enamelling was favoured for representations of flowers and birds. Typically Byzantine were the half-moon-shaped earrings that were in wide use up through the 12th century. There are examples with pierced decoration, with filigree basketwork, and with the figures of enamelled birds facing each other on a golden half-moon (such as the one in the British Museum). The court jewels, if credit can be given to the figures shown in the mosaics in S. Vitale at Ravenna, must have been of astonishing splendour. Although the mosaics give only a sketchy idea, on the figures of Justinian, Theodora, and their retinue, precious ornaments can be distinguished that were of ceremonial magnificence suited to their rank.

Iconoclasm

For all practical purposes, Constantinople's artistic activities came to an end when it was conquered and looted by the Crusaders in 1204.

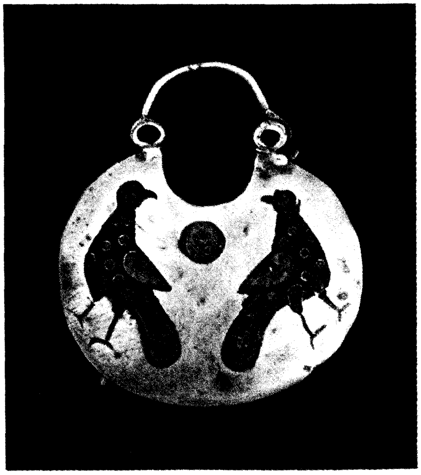
Islāmic. After the Arab conquest of Iran brought it into the Islāmic community of peoples, rings, pendants, earrings, and necklaces of gold continued to be worn, and the Iranian tradition of animal art persisted, modified to some extent in order to conform to the canons of Islām, which forbade the making of images. A 12th-century gold pendant in the form of a lion is a highly schematic rendering of this animal; it is decorated with granulation. Other techniques were filigree, encrustation with precious and semiprecious stones, and the use of niello (a mat-black substance). From the 14th century onward, manuscript illustrations give some idea of the kind of jewelry worn by Persians. In Mongol and Timurid times, jewelled coiffures for women and diademed headdresses for men seem to have been fashionable in court circles. Under the Safavid rulers, jewelry became more sumptuous and elaborate. In the 19th century, native traditions were corrupted by European influence. Traditional designs, however, have persisted in Zinjaniib and among the Kurdish mountaineers of northwest Iran. Silver decorated with twisted wire arranged in scrolls is a feature of the former. The Kurdish goldsmiths also work in silver, which they decorate with chased or repoussé designs, sometimes reminiscent of motives found on Sāsānian metalwork.

Jewelry worn by men and women in Turkey during the Ottoman period was probably influenced by the fashions current in Iran. Objects of adornment were jewelled turban aigrettes, rings, earrings, necklaces, and armlets. A technique popular in Turkey from the 16th century onward was the encrusting of jade and other hard stones with jewels attached to the surface by delicate floral scrolls in gold. Unfortunately, not many surviving pieces are earlier than the 19th century, when native tradition had been stifled by a taste for Rococo jewelry.

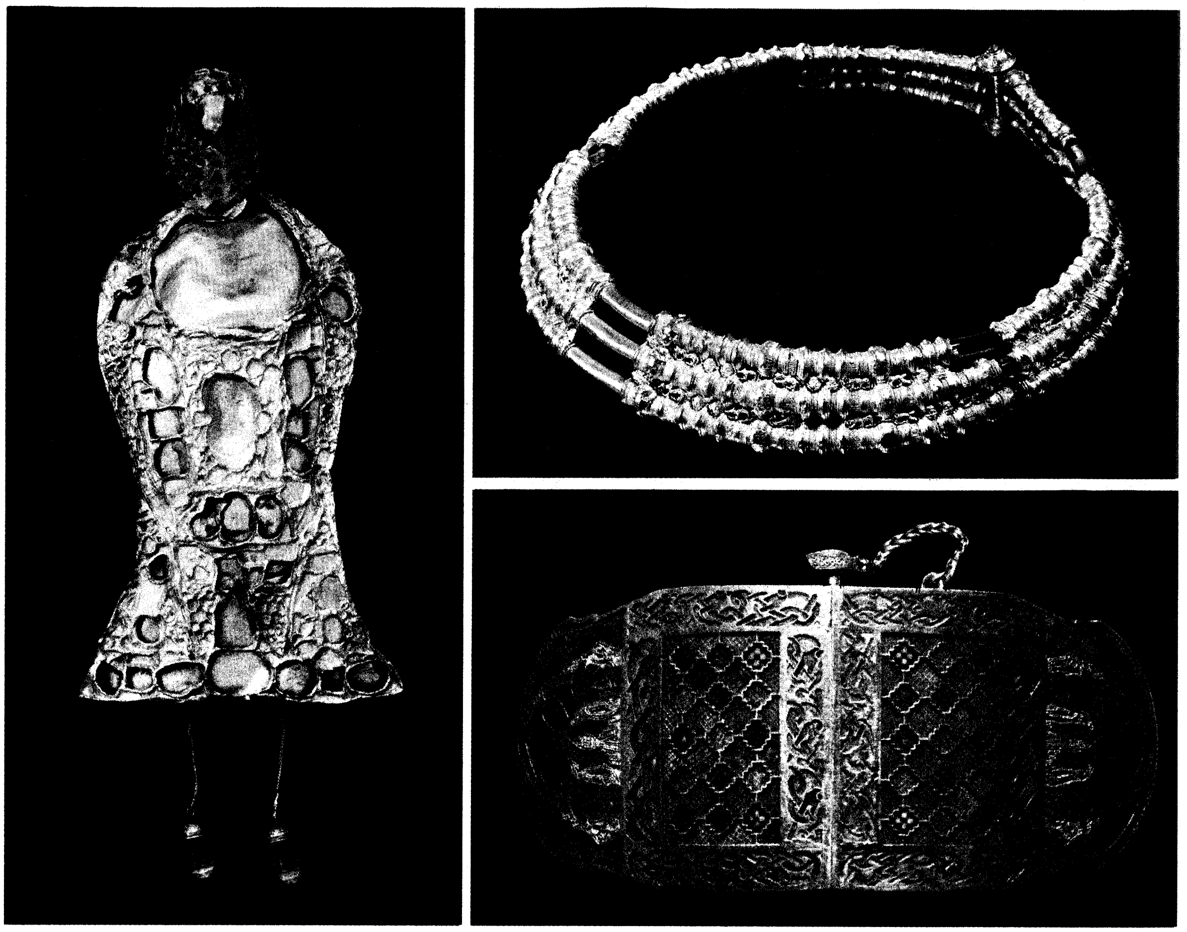
In north Africa an independent tradition has been maintained by the Berber and Arab tribes. In design the jewelry of southern Morocco shows curious analogies to Byzantine jewelry—heavy silver plaques decorated with niello or cabochons, serving as diadems or headbands. In other parts of Morocco, in Algeria and Tunisia, popular forms are headbands, breast ornaments, brooches, pendants, and a characteristic triangular-shaped shawl pin.

Teutonic. While in the Byzantine area classic forms of expression were being wiped out by the development of a skillful class of artisans who impressed their entirely ornamental taste on jewelry produced solely for decorative purposes, in the rest of romanized Europe a huge, complex movement of peoples was taking place. Bringing their tradition of polychrome decoration with them, these peoples swarmed over the old declining Greek–Roman

By courtesy of the trustees of the British Museum



Gold earring with enamelled bird, Byzantine, 12th century. In the British Museum.



Early medieval jewelry.
(Left) Fibula modelled into the shape of a bird from gold sheet and originally set with stones and cloisonné enamel: Petroasa treasure. Germanic. 4th century. In the National Museum of History, Bucharest, Romania. (Top) Gold necklace from Alleberg, Vastergotland; Viking, 6th century. In the National Museum of Antiquities. Stockholm. (Bottom) Gold shoulder clasp with garnets and cloisonne enamel; from the Sutton Hoo ship burial, Anglo-Saxon, *c.* 650. In the British Museum.

By courtesy of (left) the National Museum of History, Bucharest, Romania, (top) the National Museum of Antiquities, Stockholm. (bottom) the trustees of the British Museum

artistic civilization. Goths, Vandals, Huns, Franks, and Lombards emigrated, extending their conquests into central, northern, and southern Europe beginning in the 4th century AD, and they remained there until the 9th century. In accordance with an ancient definition, they were called barbarians; that is, not Christians, foreigners. They were also considered barbarians because they were thought to have destroyed the classical art of the Roman world.

Throughout all the provinces of the Roman empire, these Teutonic tribes produced gold ware that shared a common, well-defined style moderated according to the tastes of the particular regions in which they settled. The blend of Teutonic and Iranian, Scythian, Sarmatian, or Celtic styles produced ornaments far from the great classic tradition. Precious ornamentation, which represented the main artistic ambition of these nomadic peoples, was achieved with faience (decoration made of opaque coloured glazes), jewels, and enamels. Dominant also is braiding, which was done with strips of embossing, with bands of stones or enamel set in bezels, and also with filigree.

Teutonic fibulae

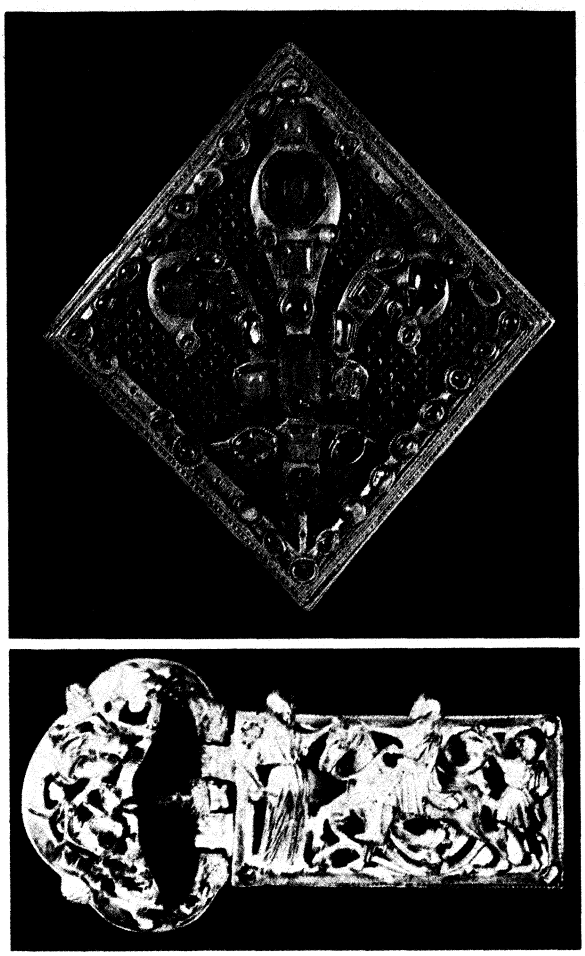
There was a highly varied production of fibulae. One of the most impressive for its size (35 centimetres or 14 inches) is the one in the shape of a bird found in Petroasa, Romania (Bucharest, National Museum), whose body is covered with sockets of different sizes and shapes in which stones and enamel were meant to be set (*c.* 4th century). The most widely used type of fibula was the so-called buckler variety, with a fan head arched bridge, and flat or molded foot, with pierced work in various shapes. Equally common were disk fibulae, either flat or with concentric embossing, while S-shaped fibulae and belt buckles were rarer.

Rigid necklaces, made up of several circles with much decoration, were typical. The most magnificent examples are those from the 6th century from Alleberg and Färjestaden, Sweden (Stockholm, Museum of National Antiquities). A ring with zoomorphic braiding (Milan, Museo Poldi Pezzoli) was found in the same region. This technique was most widely used in the Celtic and north Germanic regions of Europe, while in the British Isles, to judge from the magnificent jewels in the Sutton Hoo burial-ship treasure (London, British Museum), it was the technique of enamelling that reached extremely high levels. In northern Europe and Scandinavia the main goldworking techniques were filigree, embossing, and turning on a lathe.

As time passed, the different products of barbaric goldworking art took on a more definite stylistic identification according to the various races and locations.

Western European. The widespread adoption of Christian burial rites put an end to the custom of burying the dead with all of their jewelry. Thus, beginning with the 8th century, almost the only important gold products handed down to modern times were those preserved in abbey and cathedral treasures or by imperial and royal courts; among these gold products are very few pieces of jewelry. As the graphic and plastic arts gradually developed, however, they documented the jewelry in use at the time. According to these sources, little jewelry was worn in the Romanesque period (*c.* 950–*c.* 1150).

In the 11th century, monastic workshops for the service of the church began to decline, disappearing one after another to be replaced by secular workshops. Goldworking activities in western Europe gradually freed themselves from the centralizing patronage of the church in



Gothic jewelry.
(Top) Fleur-de-lis brooch of the French regalia in gold and precious stones; 14th century. In the Louvre, Paris. (Bottom) Rhenish belt buckle in gilded silver; from the island of Visby, Sweden, c. 1340. In the Historiska Museet, Stockholm.
By courtesy of (bottom) the Historiska Museet. Stockholm; photograph (top) Marc Garanger

order to serve the numerous courts and noble families, and in the 12th century the first goldsmiths' guilds were organized.

One of the most widely used ornaments in medieval Europe was the ring. To it was attributed ever more symbolic and religious value, as well as ever greater importance as a talisman, good omen, and sign of office; and, as always, it served as a seal.

Another widely used ornament was the brooch. Most popular was the medallion type, which might be round, star-shaped, or pentagonal, while the diamond shape was less common. A magnificent example of the diamond-shaped brooch is the fleur-de-lis brooch that was part of the French crown jewels and is now in the Louvre. Standing out in relief against a blue enamelled background is a large gold lily covered with important precious stones that create a rich chromatic effect.

Flamboyant Gothic style

A fine example that typifies the plastic decorative repertory of the flamboyant Gothic style is a silver belt buckle from Sweden (Stockholm, Historiska Museet). Modelled in high relief on the buckle plate is a gentleman on horseback approaching a lady and followed by his servant. The three-lobed buckle ring is modelled in a complex design that includes a seated person and a man kneeling in front of him (*c.* 1340).

Renaissance to modern. *15th and 16th centuries.* The "rebirth" of classicism, which combined all artistic expression in a single orderly, rational approach, found a fertile creative field in gold jewelry. During the Renaissance the jeweller's art reached truly high levels—particularly in Italy in the grand duchy of Tuscany. Eighteen centuries after the great flowering of Hellenistic jewelry, Italian Renaissance jewelry once again achieved an expressive form worthy of comparison with the figurative arts. There was, in fact, no sharp division between the two. Nearly all of the most famous names responsible for the Renaissance artistic revival—Ghiberti, Brunelleschi, Pollaiuolo, Botticelli—served apprenticeships in the goldsmiths' workshops, where gentlemen went to order medallions for their hats and where ladies went to have their jewels set.

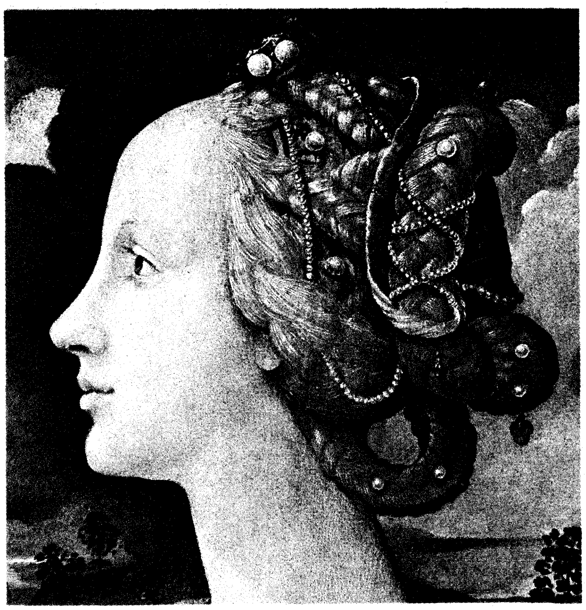
Because of their elaborate workmanship, which meant that their artistic value was far greater than the intrinsic value of their materials, many pieces of jewelry have been handed down to modern times in public and private collections. Even more extensive evidence, however, is provided by paintings from this period which show the jewelry worn by both men and women. From portraits by Botticelli and Piero di Cosimo, it can be seen, for example, that as early as the second half of the 15th century the elaborate decoration of women's hair with precious materials had become a real art, in which goldsmiths and craftsmen carefully worked out every line of the often extremely complicated ornamental design that had to harmonize with the movement of braids or unbound hair,

Increase in the use of jewelry during th Renaissance

The Renaissance saw an enormous increase in the use of jewelry throughout Europe. The courts of England, France, and Spain, the French duchy of Burgundy, and the Italian duchy of Tuscany indulged in real contests, trying to outdo each other in the display of gold, gems, and pearls, a phenomenon that for centuries had not occurred on such a large scale. The nobility and the rich middle class followed this fashion, and even the youngest scions were covered with jewels, as shown in the portrait of the Medici princess by Il Bronzino (Florence, Uffizi), as well as many others. Francis I of France surrounded himself with famous artists such as Benvenuto Cellini and Leonardo da Vinci. In Paris, artists such as Jean Duvet, Étienne Delaune, and the Fleming Abraham de Bruyn were the outstanding creators of designs for jewelry. It was Hans Holbein the Younger who was mainly responsible for the introduction of Renaissance jewelry into England, where he found fertile ground, thanks to Henry VIII's great passion for jewels—a passion surpassed only by that of Elizabeth I. Henry possessed more than one magnificent parure, or set of matching jewelry, designed for him by Holbein, as well as several hundred rings.

As Holbein's portrait of Henry VIII suggests, the custom of wearing bejewelled clothing, which had begun gradually in the 14th century, flourished in the Renais-

Lauros—Giraudon



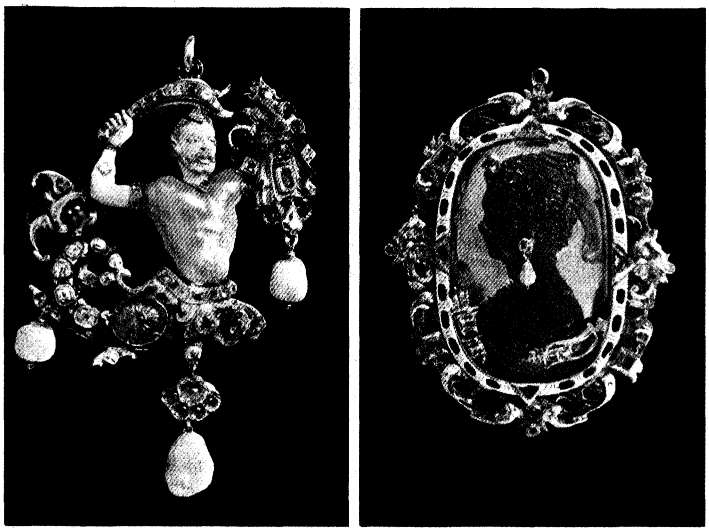
Simonetta Vespucci wearing a pearl headdress with her braids, typical of elaborate Renaissance hair ornamentation. Detail of her portrait, panel painting by Piero di Cosimo; c. 1498. In the Musée Conde, Chantilly, France.

sance. Even hat brims were decorated, with designs in pearls as well as with pendants of great value.

In Holbein's portrait there is also a magnificent example of a popular necklace of the period. It consisted of wide gold bands decorated with embossing that formed medallions, in the center of which were large stones. From the necklace hung a pendant. Only rarely were women content to limit themselves to a single necklace, usually wearing a choker-type necklace made of pearls, with or without a pendant, together with a longer second necklace made of gold, with or without the inclusion of gems. A third necklace was often hooked to the clothing, on the shoulders, and formed a double loop, being lifted up in the centre and fastened to the bodice with a jewelled pin.

The precious ornament on which the artist-jewellers lavished all of their creativity and technical ability was the pendant. At first this consisted of a decorative medallion enclosing a cameo with figures and subjects of classic derivation, such as busts of women and pagan deities. These figures were later enriched with inserts of gold, enamel, and gems, which enhanced the polychrome effect. Still later, the figures were freely modelled in brilliant polychromy with a great variety of subjects—animals, Tritons, mermaids, ships, sea monsters, and symbolic figures—in complicated openwork compositions made of several linked pieces, in which the irregular shape of a large baroque pearl was often used for the body of an animal or a centaur's torso.

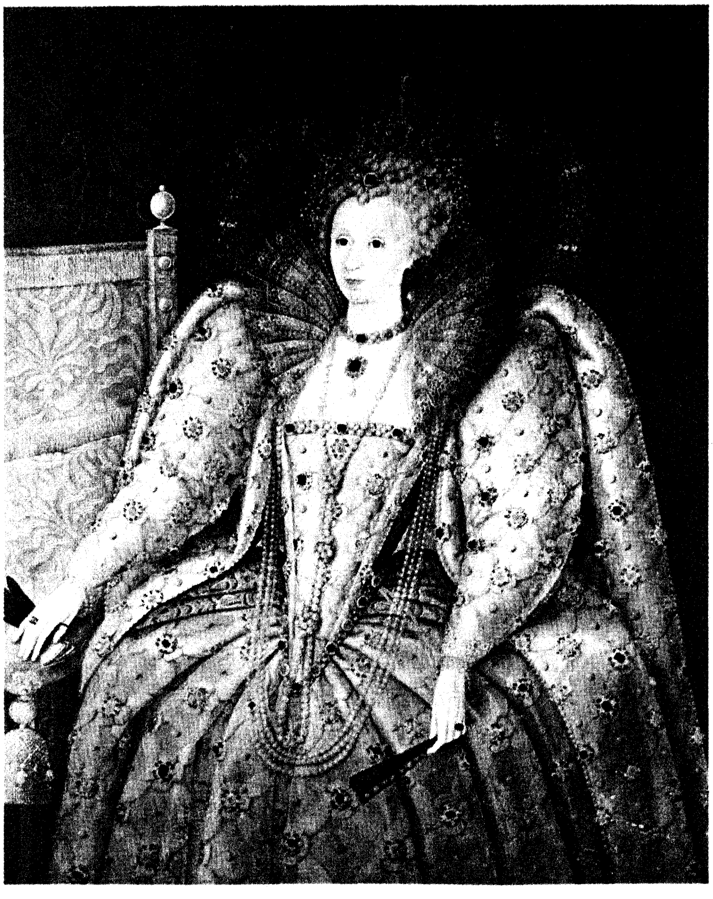
Throughout Europe the ring enjoyed wide popularity in an unlimited variety of types, including those with a bezel that could be opened and used as a container for relics, symbols, or—as romantic tradition has it—poison.



Renaissance pendants.
(Left) The Canning Jewel, a pendant of gold, enamel, rubies, diamonds, and baroque pearls; German. 16th century. In the Victoria and Albert Museum, London. (Right) Onyx cameo pendant of the goddess Diana wearing a pearl earring enclosed in a gold and enamel frame; Italian, 16th century. In the Kunsthistorisches Museum, Vienna.

By courtesy of (left) the Victoria and Albert Museum, London, (right) the Kunsthistorisches Museum, Vienna

Carlo Bevilacqua—SCALA



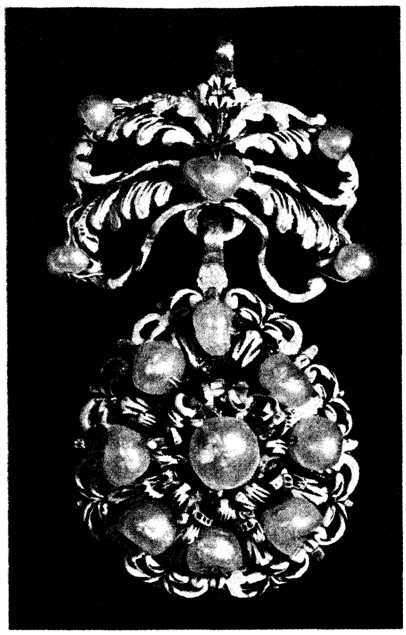
"Queen Elizabeth of England." adorned in Renaissance fashion with pearl choker and pendant and a series of longer necklaces; portrait in oil by an unknown artist, English, 16th century. In the Pitti Palace, Florence.

17th century. Toward the end of the 16th century, the Renaissance style blended gradually into the manifestations of the Baroque period, which arose at different times in different countries. This gradual change in the style of jewelry was conditioned mainly by two factors. The first was of a technical nature and concerned improvements in the cutting of precious stones, while the second consisted of a great vogue for the cultivation of flowers. Floral and vegetable decoration therefore became the most fashionable theme for jewelry designers, and its popularity spread throughout all Europe. The ornamental motifs of knots, ribbons, and Rococo scrolls also saw a considerable development. There was a corresponding decrease in the amount of figurative decoration, which finally completely disappeared. At first these ornamental forms were carried out in openwork gold jewelry, the majority of which was coloured with enamel; later diamonds and other precious stones became the real protagonists in the composition of jewelry.

Period of Louis XIV

During the 17th century, the number of pieces of jewelry worn decreased, as did the fashion for male adornment. The last monarch to make heavy use of jewels was Louis XIV, and the word heavy is used here in a literal sense, the great weight consisting mainly of gems with which the monarch covered himself for official ceremonies. He had his own personal jeweller, Gilles Légaré, who was a guest in the Louvre palace. He was not the only sovereign, however, who enjoyed showing off his jewels nor was Versailles the only court in Europe to follow the king's example. Those of London, Madrid, and Munich were not far behind. The precious ornaments worn by women started on the hat, on the side of which at least one striking aigrette (spray of gems) was fastened. Then came two or three heavy necklaces, each of which might have a pendant, then a belt that followed the pointed shape of the bodice. Other jewels were inserted along the armholes, shoulders, and wrists, and at least four rings were worn on the hands. Often the heavy fabrics used for the clothing were embroidered with gold thread. It was during this period that a spectacular form of jewelry was created in Spain, which in a more subdued form spread throughout Europe: the stomacher brooch, which covered a woman's entire bodice, from neckline to waist. With its heavily bejewelled composition of scrolls, leaves, and pendants on a gold framework that followed the curves of the body, even extending under the armpits, this jewel usually contained no less than 50 precious stones of different sizes. A famous example is the one in emeralds from the treasure of the Virgin of Pilar (Victoria and Albert Museum, London).

18th century. About 1725. Brazilian diamonds in large numbers began to be imported into Europe, and, during the course of the century, this stone became so popular that imitations began to be produced. The jewelry of this



Baroque earring of pearls, white enamel, and light gold; Spanish, mid-17th century. In the Museo Poldi Pezzoli, Milan.

Arnoldo Mondadori Editore-C.E.A.M.

Use of Brazilian diamonds

period seems to have been created to glorify and exploit the cutting of diamonds and other precious stones. The dense forms of Baroque jewelry were replaced by an entirely different conception, in which the design was to appear in gems alone, while the metal setting was concealed to the greatest extent possible. The greater lightness that resulted was increased by the large number of empty spaces in the composition as well as by its lack of symmetry in many cases. Wide choker necklaces with pendants were very popular, and the stomacher brooch remained in style but in a lighter, airier form (such as the one in Munich, Schatzkammer der Residenz). The jewelled stems of the aigrette were often made so that they could sway back and forth in order to show off the sparkle of the diamonds that covered them, The brooch in the shape of a bouquet of flowers began to become fashionable. As in the 17th century, both men and women wore jewelled buckles on their shoes.

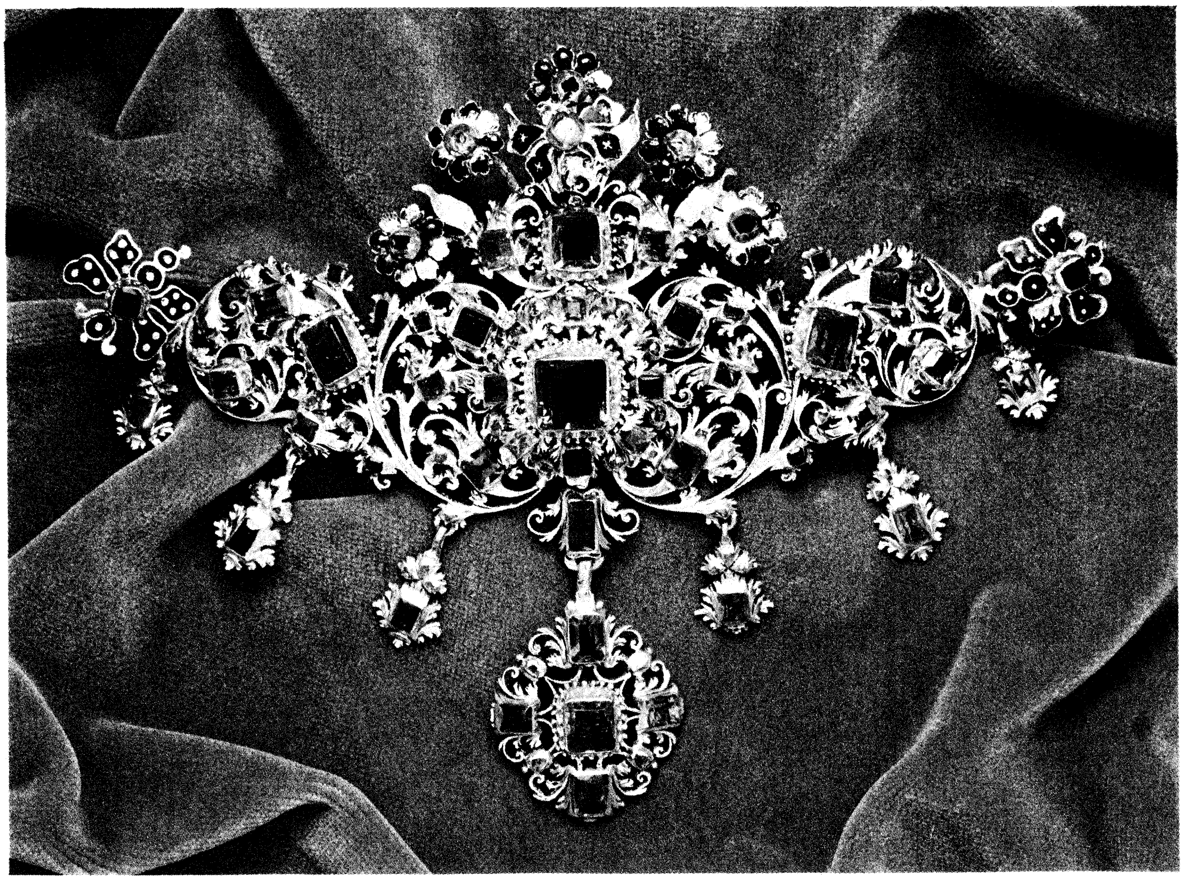
A piece of jewelry that was widely used for daytime wear during this century was the chatelaine, on which, together with the watch case, goldsmiths lavished some of their most highly refined work. The chatelaine was a pendant made of jointed, embossed gold components of different shapes and sizes, with scenes and designs in elaborate frames. It was fastened by means of a hook to the belt or waistcoat pocket, and from its protruding points hung decorative chains of various lengths, on which men fastened their watches, the keys for winding them, and other accessories. Women, on the other hand, used the chatelaine to carry keys, scissors, and other more or less useful objects.

Classical ideals and Josiah Wedgwood

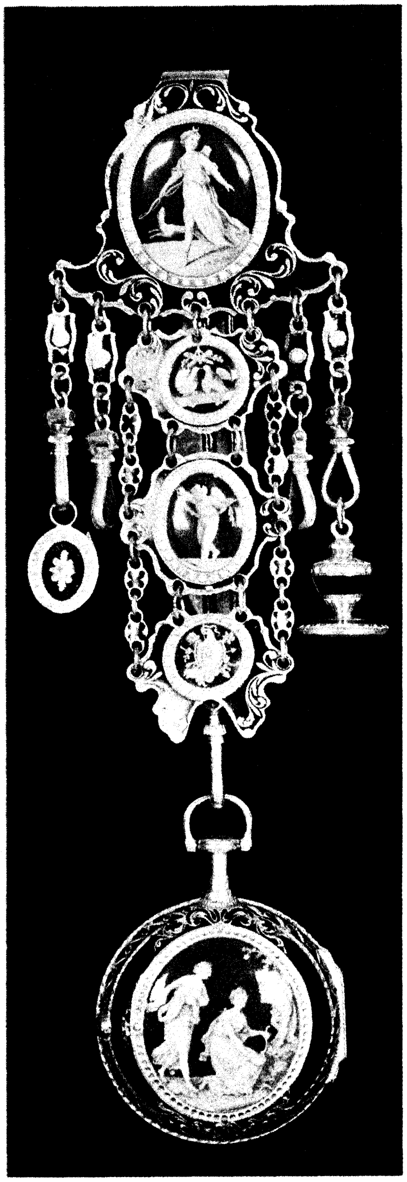
During the last 30 years of the 18th century, the great sensation caused by the archaeological discoveries in Pompeii and Herculaneum caused art forms to turn towards classical ideals of harmony and brought about a decisive change in European tastes and decorative forms. Curved lines no longer appealed in the ornamental repertoire, the new Neoclassical style being characterized by greater simplicity, together with severity of composition. Jewelry forms. too, were influenced by decorative motifs based on Greek and Roman models, and the cameo became fashionable once again.

An English pottery manufacturer, Josiah Wedgwood, made a big contribution to the popularization of the new jewelry forms. An expert technician, he produced reproductions of classic cameos, calling upon sculptors like John Flaxman to work with him on the execution of oval, round, and octagonal plaques with figures done in relief

By courtesy of the Victoria and Albert Museum, London



Stomacher brooch with emeralds and enamel flowers on gold; from the treasure of the Virgin of Pilar, Spanish, mid-17th century. in the Victoria and Albert Museum, London.



Chatelaine of gold and enamel; French, late 18th century. In a private collection, Rome.
Arnoldo Mondadori Editore-C.E.A.M.

in a white paste on a light blue, green, black, or pink background. These plaques, framed in gold, were used for all sorts of jewelry—medallions, pins, pieces of diadems, belts, bracelets, rings (a notable example is the ring with a Wedgwood medallion in the Musée des Arts Décoratifs, Paris).

*19th **century.*** The Industrial Revolution destroyed forever the ancient role of jewelry as a symbol of social rank. The social evolution created a market for a vast quantity of jewelry at prices the middle class could afford; and so jewelry, too, succumbed to the machine. Hundreds of different components for ornaments were produced by machines, an electric gold-plating technique was invented, and the production of imitation stones increased. Despite the growing dominance of the machine, however, the goldsmiths' technical ability remained at a high level.

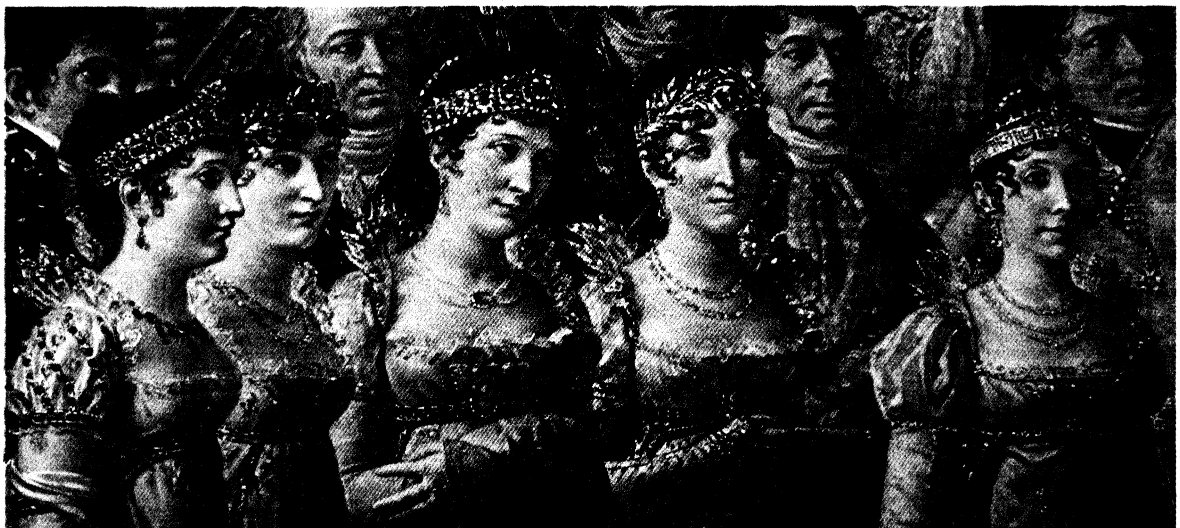
The jewelry produced in the 19th century is characterized by a stylistic eclecticism that takes its inspiration from all past styles—Gothic, Renaissance, Greek, Etruscan and Roman, Rococo, Naturalistic, Moorish, and Indian, all tinged with the Sir Walter Scott–Lord Byron Romanticism of the period. The futility of transferring forms of artistic expression from an era in which they were the result of organic aesthetic development and of adopting them for objects that reflect only a gesture of romantic admiration is evident in the painting by Jacques-Louis David (Louvre, Paris) immortalizing Napoleon's coronation ceremony (1804). The painting provides documentation on the precious ornaments worn by the ladies who were present. In their jewelry, the conventional, rhetorical Empire style appears as a strict, uninspired interpretation of classic motifs, a far cry from the exquisite Neoclassicism of the 18th century.

Artistic commercial jewelry firms

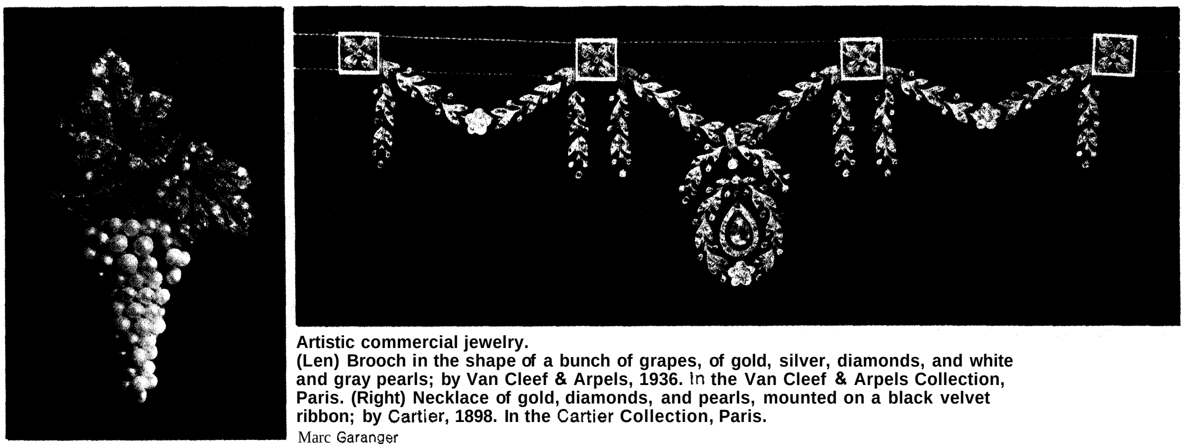
Besides mass production, the 19th century saw the establishment of large artistic commercial firms that produced high quality jewelry suited to the requirements of the prosperous new bourgeois class. While always satisfying very high standards in regard to technique and materials, these firms tended, from the aesthetic point of view, to reflect the tastes of a bourgeois clientele, which are usually quite traditional.

The oldest of the firms was the one founded by Peter Carl FabergC in St. Petersburg (now Leningrad) in 1870, which took over from the firm his father had started in 1842. FabergC attained great renown at the Universal Exposition in Paris in 1900, where for the first time he put on display all the imperial Easter eggs that he had created, together with a selection of other "luxurious objects." On that occasion he was proclaimed "Master" and was awarded the Legion of Honour. FabergC used a

J.E. Bulloz



Empire style tiaras and garland diadems, necklaces, earrings, and jewelled belts; detail from "The Coronation of Napoleon," oil painting by Jacques-Louis David, 1804. In the Louvre, Paris.



Artistic commercial jewelry.
(Len) Brooch in the shape of a bunch of grapes, of gold, silver, diamonds, and white and gray pearls; by Van Cleef & Arpels, 1936. In the Van Cleef & Arpels Collection, Paris. (Right) Necklace of gold, diamonds, and pearls, mounted on a black velvet ribbon; by Cartier, 1898. In the Cartier Collection, Paris.
Marc Garanger

greater variety of precious and semiprecious stones than almost any other jeweller in history. He had a strong preference for the Louis XVI style but also made numerous objects in the Italian Renaissance, Rococo, and medieval styles, as well as in the "old Russian" style, which is a mixture of Byzantine and Baroque. Decoration with enamelling, too, was one of the main specialties of the Fabergé firm.

In Paris in 1898 Alfred Cartier and his son Louis founded a jewelry firm of great refinement. The firm was distinguished for a production characterized by very fine settings, largely of platinum, which were designed so that only the precious stones, always selected from the very purest, were visible.

At the beginning of the 20th century, Cartier was the most famous jeweller in the world, supplying jewels to the king of Portugal, the duke of Saxe-Coburg-Gotha, the grand dukes and princes of Russia, the prince of Wales, and other notables.

In the United States in about 1850, Charles Lewis Tiffany (father of Louis Comfort Tiffany, one of the most original of the Art Nouveau artist-craftsmen) produced silverware in New York. In 1886 he introduced the Tiffany setting, a special type of fork for the setting of diamonds. Among his clients were Sarah Bernhardt and President Lincoln.

Other high-quality jewelry firms founded in the 19th century were Van Cleef & Arpels in Paris, Bulgari in Rome, Asprey & Company in London, Black, Starr & Frost in New York, and Patek Philippe in Geneva.

Art Nouveau

The development of the movement called Art Nouveau at the end of the 19th century represented a reaction against the imitation of ancient styles and the emphasis given, in the creation of jewelry, to precious stones. The material used for Art Nouveau jewelry was prized not for its intrinsic value but for its ability to render a design or to carry out chromatic effects. The new jewelry was made from any material that would best express the new symbolic or decorative ideas. Vegetable and animal components, together with the feminine figure, formed the basis for compositions made of flowing lines of rich plastic and chromatic effect and antistructural, dynamic design on a high artistic level.

In Paris, through the works he presented at the Salon du Champ de Mars in 1895, René Lalique (1860–1945) achieved a position of European renown and importance. He personified the Art Nouveau jeweller-artist, his works providing evidence of such highly personal taste that they can be compared to Renaissance jewels. They lean toward a symbolism carried out by the use of milky or watery blue-green colours; of stones such as the opal; of disquieting animals such as the snake, the owl, the octopus, and the bat; and of feminine figures, usually enigmatic, mysterious, and dreamy. Enamel, ivory, vitreous paste, and engraved glass were often used by Lalique to obtain pictorial and plastic effects in his jewels.

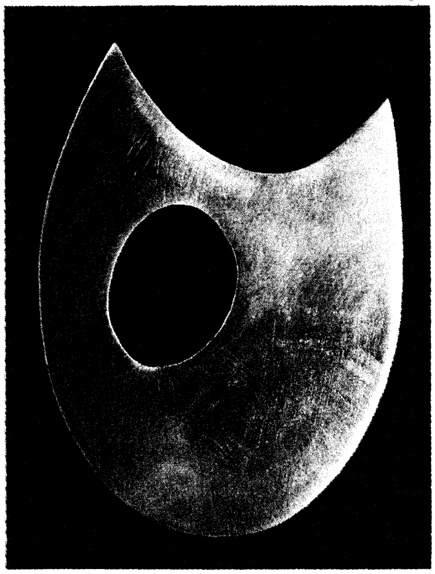
Unlike Lalique, the jewellers Georges Fouquet (1858–1929) and Henri Vever (1854–1942) expressed themselves through more synthetic geometric forms. The pendant representing a butterfly by Fouquet and the bracelet and ring for the actress Sarah Bernhardt (both in the Périnet Collection, Paris) show a very carefully thought-out stylization.

The Czechoslovak graphic designer Alphonse Mucha (1860–1939), who worked in Paris, created a number of jewelry designs, transferring his brilliant talent as an illustrator to precious stones and metals.

In the United States, the floral style in jewelry found one of its most highly personal interpreters in Louis Comfort Tiffany (1848–1933), one of the greatest of all American decorators. In the creation of jewelry he expressed himself at first by transferring to Art Nouveau forms the colourful Oriental and Byzantine style that so fascinated him. Later he adopted Lalique's French symbolism, on which he set his own stylistic mark.

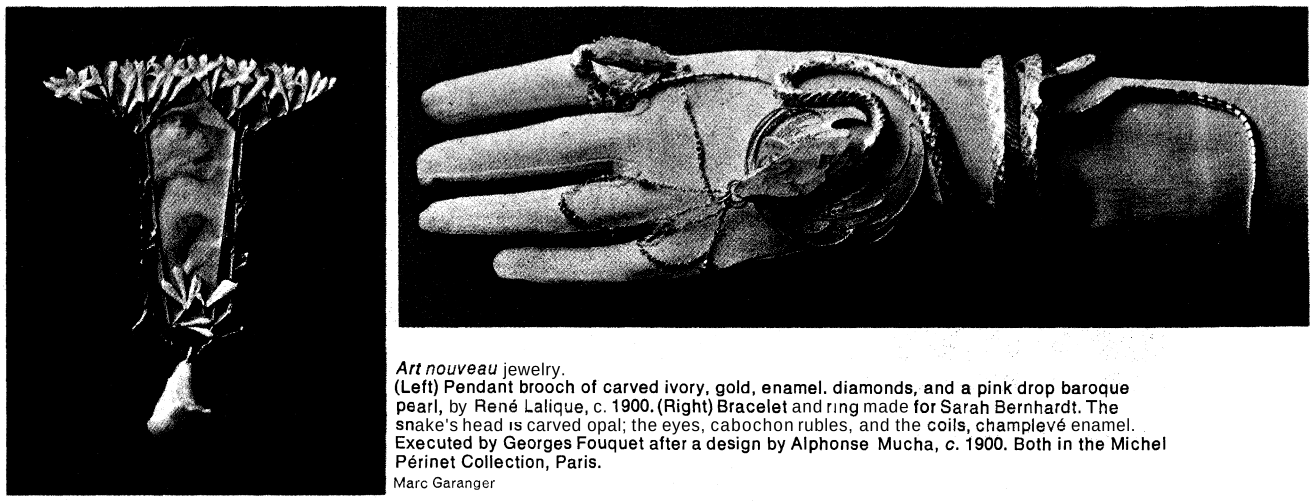
Modern. The Art Nouveau movement came to an end at the beginning of World War I. The years that followed the war's end seethed with new excitement. In this new phase, the stylistic trends—particularly the nonfigurative—that began to emerge in the most advanced jewelry creations were closely linked to those of painting and sculpture. Cubism, Futurism, the abstractionism of Piet Mondrian and other artists of the de Stijl group, Paul Klee's paintings, and above all the Bauhaus school (which aimed at integrating artistic disciplines with one another and with industrial techniques) provided a basis for the new forms used in avant-garde jewelry.

Marc Garanaer



"Cutout," made in gold by François Hugo after a design by Jean Arp, 1966–67. In the Galerie le Point Cardinal, Paris.

Compositions were based mainly on the interplay of geometric forms. Like Art Nouveau jewelry, creations of the Art Deco movement (named for the art displayed at the 1925 Paris exposition) used materials suitable for



Art nouveau jewelry.
(Left) Pendant brooch of carved ivory, gold, enamel, diamonds, and a pink drop baroque pearl, by René Lalique, c. 1900. (Right) Bracelet and ring made for Sarah Bernhardt. The snake's head is carved opal; the eyes, cabochon rubies, and the coils, champlevé enamel. Executed by Georges Fouquet after a design by Alphonse Mucha, *c.* 1900. Both in the Michel Périnet Collection, Paris.
Marc Garanger

expressing the new stylistic language. Preference was given to the smooth, polished, satined surfaces of precious metals or even of steel. Diamonds and other precious stones were used sparingly, functioning largely as chromatic accents. In the same piece of jewelry, coral could be combined with diamonds, regardless of the great difference in intrinsic value, because their sole purpose was to satisfy the aesthetic requirements of the nonfigurative styles.

During this period there were outstanding artist-jewellers such as Raymond Templier, Jean Fouquet, and René Robert in France, H.G. Murphy in England, and Wiwen Nilsson in Sweden.

Later, artists of great international renown devoted some of their creative efforts to the art of jewelry. Some—such as Georges Braque, Jean Cocteau, Max Ernst, Jean Arp, Man Ray, Salvador Dali, Yves Tanguy, and Jean Dubuffet—designed jewelry, while others—including Pablo Picasso, Alexander Calder, Alberto Giacometti, Gio Pomodoro—made jewels.

One of the most recent developments in modern mass-produced jewelry is the use of plastic. This material, as well as providing colour, can have mineral fragments or dust embedded in it or can be used in combination with more or less valuable metals, producing pieces of jewelry whose composition may call for considerable effort and which may be of much interest.

OTHER CULTURES

East Asian. *Chinese.* Much of Chinese jewelry, both of recent and early date, displays the familiar manipulative skill of the Chinese craftsmen; yet the work of the goldsmith or lapidary applied to personal ornament does not represent so distinct a branch of craft as it does in the west and is accorded no special attention by the native connoisseurs and writers on the arts. Most of the jewelry is designed to adorn the costume rather than the person, and much of it has a fulsome and insubstantial quality that is not immediately pleasing to Western eyes. Necklaces, bracelets, and earrings are comparatively rare, headdresses and elaborate hairpins being the more common forms attached to the person. In the traditional costume of recent times, ornate hooks and buckles were used to attach girdles, and women wore strings of beads, often multiple and variously spaced, with decorative plaques and other larger ornaments interspersed. The beads might be attached to the neck, head, or waist, and their purpose was to dignify the whole figure, rather than to display the fine quality of a curiously wrought gem. In any case, the splendour of the stuff of the costume, with richly woven or embroidered ornament, provided the distinctions of rank and wealth, and jewelry was often dispensed with altogether. The long sleeves and high collar of the garment left little of the person exposed for ornament set against the skin, in the manner favoured in the West.

Jewelry to adorn the costume rather than the body

In the time of the Shang dynasty, in the last centuries of the 2nd millennium BC, bone and ivory hairpins with ends carved as birds or abstract figures were a popular adornment. The many finely wrought, small jade plaques of the period, depicting animals in profile, are in many cases clearly intended for sewing to the costume. The earliest evidence of gold ornaments belongs to the time about 400 BC, though these are harness mounts, or weapon parts, rather than jewelry in the normal sense. The latter is better represented by the belt hooks (said to have been adopted from the nomads of inner Asia) that were probably worn by both men and women. They were mostly made of bronze, with fine cast ornaments usually of abstracted dragon and bird heads. These belt hooks were inlaid with gold or silver foil, polished fragments of turquoise, or more rarely with jade or glass; sometimes they were gilded.

Toward the end of the Han dynasty, probably not before the later 2nd century AD, the art of granulation was communicated to China from the Hellenized region of the Black Sea coast. Granulation can be traced in China until about the 10th century AD, its discontinuation in the East curiously coinciding with the loss of the technique in the West. Granulation was combined with filigree; and hairpins, combs, ear-rings, and costume plaques survive in some quantity, particularly from the richly furnished tombs of the T'ang dynasty (AD 618–907). There are plaques with birds and flowers delineated by soldered wire, inlaid with turquoise, on a ground of fine granulation that appears like a dust of gold.

The employment of the repoussé technique in gold and silver, particularly on the heads of combs, can be attributed to the T'ang period but became more common in the Sung dynasty (AD 960–1279). Meanwhile, hairpins of filigree, with heads shaped as butterflies or flowers, sometimes with pearls or small jade additions, continued the age-old fashion. A scented hairpin takes the place of the scarf or ring of European romance. They were called *pu yao* ("shaking while walking") and were loosely made so as to sway when the wearer moved. Gilded bronze and silver were the principal materials. There are accounts of elaborate headdresses, some no doubt of the kind representing a complete phoenix such as are to be seen on clay tomb statuettes of the T'ang period, but no surviving examples of these can be attributed with certainty to the Sung period. Jade ornaments were still attached to the costume.

Jewelry survives in greater quantity from the Ming dynasty (AD 1368–1644) and gives an impression of greater taste for elaborate figural and floral designs in high repoussé relief and for the effect of semiprecious stones. The latter were prized for their colour rather than their luminosity or rarity. They are never elaborately faceted, being merely ground flat and beveled at the edge for the most part and are set nearly always en *cabochon*, with barely a preliminary polishing, sometimes even retaining the irregularities of the pebble. The stones are invariably semiprecious or even commonplace: amethyst, agate, chalcedony, pink and other quartzes, and, of course, jade. Until modern times, this last has been the most admired

Ming dynasty jewelry

of the stones, especially the white variety, which was used for spacers and linking pieces in the silk and beaded hangings of elaborate costumes. The plaques of silver repoussé with flowers and scenes of people were probably used only by men as belt ornaments. Apart from the signet ring, the use of which may not go back beyond Ming times, the male could affect jewelry only in his accoutrement. An unusually splendid piece of personal jewelry is represented by the openwork gold plaque set with rough stones that is said to have come from the tomb of Emperor Hsiian-te (1425–35).

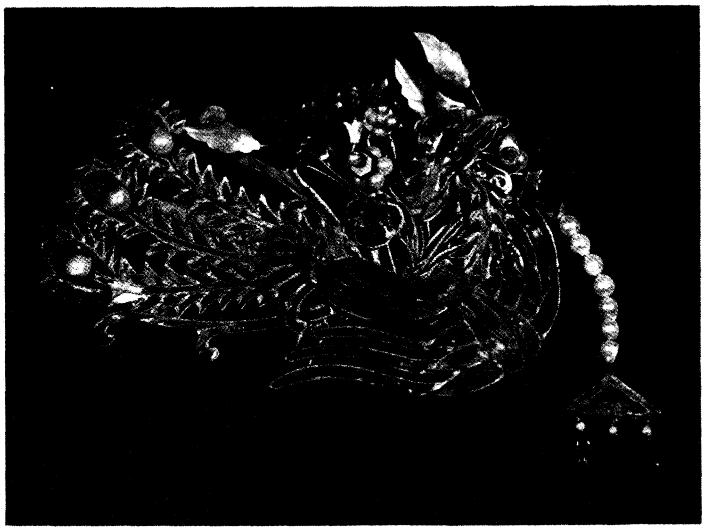
Magatama, inrō, and netsuke

Japanese. In the later prehistoric and early historical period of Japan (c. 2nd century BC–6th century AD), comma-shaped objects—not usually more than an inch in length, called *magatama*—were made of green jade and eventually even imitated in glass. They were sometimes pierced to be strung in a necklace. The symbolic meaning of the *magatama,* which were often placed in tombs, can only be guessed at.

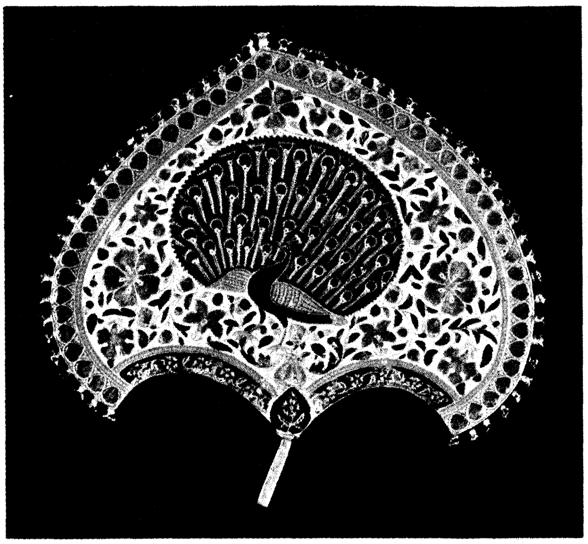
In historical times, traditional Japanese costume, male and female, has never allowed the use of ornaments of precious metal or stone, so that nothing in the history of Japanese craft and taste corresponds to the jeweller's work of the West. Hairpins with elaborate heads were increasingly used in the Tokugawa (Edo) period (1603–1868) by women of the geisha and courtesan classes but not by women of other classes. In the same period men were permitted the ostentation of the *inrō,* a small tiered box for tobacco, medicines, confections, and the like, which might be beautifully painted in lacquer and inlaid with mother-of-pearl or precious metal. The ivory girdle toggle called netsuke, always delicately and often intriguingly carved, was the only other personal ornament which usage allowed. (Ed.)

Indian. Extensive documentation on ancient jewels is provided by Buddhist statues and the cycles of wall paintings in the Ajantā caves (4th–5th century). The great variety of types of jewelry represented and the dominance of polychromy indicates the high degree of development attained by the art of jewelry-making in one of the most magnificent of early civilizations, the Indus. The lavish use of polychrome jewelry was possible because of the ancient practice of pearl-diving and because of the wealth and variety of deposits of precious and semiprecious stones to be found in India and the neighbouring countries of Ceylon, Thailand, and Burma. Indian women were thus the first to decorate themselves with huge quantities of jewels—so many that they were almost fully clothed without wearing any real garments. The clothing of these lovely Indian girls often consisted of tiaras, necklaces, earrings, armlets, bracelets, belts, "cache-sex" (serving the same function as a loincloth), and toe rings, worn on their bare skin and complemented by nothing more than precious veils and scarves.

Arnoldo Mondadori Editore–C.E.A.M.



Bird-shaped ornament of pearls, jade, and feathers, Chinese, Ch'ing dynasty, 18th century. In a private collection, Milan.



Gold and enamel turban ornament from Jāipur, Rājputana, India, 18th century. In the Victoria and Albert Museum. London

By courtesy of the Victoria and Albert Museum, London: photograph, John Webb

The important place given to polychromy in even such early jewelry, which contained only a negligible quantity of gold and other metal, is suggested by a woman's belt found in Harappā, one of the most ancient centres of Indus civilization, and dated from the end of the 3rd millennium BC. The belt was worn around the naked body of the woman, extending from the waist to the groin. The beads of which it is made are mainly red cornelian, green steatite, agate, jasper, amazonite from Gujarāt, jade from central Asia, lapis lazuli from Afghanistan, and a sort of faience.

In the Indus areas and in those under their influence, the setting, polishing, and piercing of precious and semiprecious stones underwent precocious evolution. Stone-cutting, however, was accepted very late in the Orient where, until only recently, it was considered preferable not to decrease the size of the stone. Over the centuries and the millennia, outside influence—Iranian, Greek, Buddhist, and Muslim—could not undermine the basic traditional characteristics of East Asian jewelry, which remained faithful to its preference for a many-hued rich effect that was less a form of artistic self-expression than a display of showy glitter aimed at astonishing the onlooker.

Kinds of Indi jewelry

If rich rajahs adorned themselves with jewels--on their turbans, on their ears, around their necks, inserted in their nostrils and even between their teeth—the precious objects worn by women were far more numerous, and each of them had its own specific name. For the head alone there were golden wreaths, large brooches, braids made from three bands of gold leaves with a star in the middle set with gems, braids to be placed along the part in the hair, lotus leaves made of gold sheet to be worn at the nape of the neck with bunches of gold flowers next to them, and tiaras in complicated shapes complete with many tinkling pendants. A variety of forms were used for the earrings, in which pearls, filigree, gems, and coral appeared in floral compositions based on the contrast between the different colours. Some Indian women embedded a jewel in the forehead or pierced the nose in order to wear a jewel in the left nostril. Necklaces were sometimes so long that they came below the navel, and different names were given to those made only of pearls and those of gold. The former were also distinguished according to the number of strings, of which there could be as many as several dozen. Some necklaces were made of a combination of precious stones and pearls, while others were made of amulets in various shapes. A very early type of Hindu amulet called a nauratan was made of a gold plaque with nine precious stones fastened above it. A series of *nauratans* could be used to form a necklace. Jewelled belts followed the shape of the body and often had extra pieces that reached up to the neck or down to the bracelets worn around the thigh. Ankle bracelets were often linked by tiny decorative chains running down the instep to the rings on the toes.

The magnificence of Indian jewelry is dominated by pearls, coloured stones, and enamel. Gaudy though small scale, the jewelry is made up of a combination of a series of linked components that break up the compositional unity. It depends mainly for its effect upon the chromatic play of the many different materials that have always abounded in the soil and sea of Indian territory and of nearby countries.

Southeast Asian. There is a long goldworking tradition among the peoples of Southeastern Asia, whose jewelry shows evidence of Tibetan, Chinese, and Indian stylistic influence. The areas in which personal ornamentation with precious objects underwent the greatest development were Burma, Cambodia, Laos, and Vietnam. Burmese jewels are outstanding for the beauty of their designs and for their technical artistry. Typical of them is the conical headdress, reflecting the traditional architectural form of the stupa (Buddhist shrine), and the bejewelled, rigid shoulder decorations with a raised line similar to that of pagoda roofs, worn by dancers in addition to arm and ankle bracelets, belts, and brooches made of gold and colored stones. Although it has its own distinct characteristics, Burmese jewelry was heavily influenced by Indian styles, especially in regard to a taste for great abundance; thus, each single jewel, rather than standing out, blends into the overall effect.

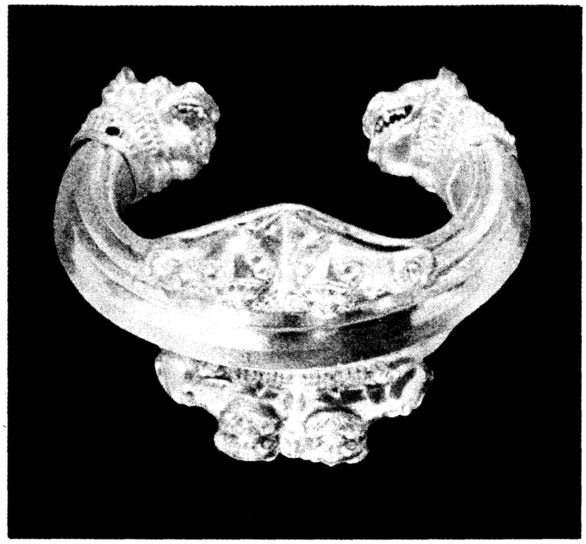
Burma, Cambodia, Laos, and Vietnam

In its jewelry forms Cambodia, Laos, and Vietnam were subject to greater Chinese influence because of their geographical position. In these territories, too, most documentation for the period when precious ornamentation had its most flourishing development is to be found in Buddhist sculpture. This means that the outstanding forms of expression in the art of jewelry were linked to religious rites, contributing to the glorification of the figures worshipped by the cult.

Scythian. It is to the Scythians, a seminomadic people from the Eurasian steppes who moved out from southern Russia into the territory between the Don and the Danube and then into Mesopotamia, that we owe a type of gold production, which, on the basis of its themes, is classified today as the animalistic style. During the early period (5th–4th century BC), this style appeared on shaped, pierced plaques made of gold and silver, which showed running or fighting animals (reindeer, lions, tigers, horses) alone or in pairs facing each other, embossed with powerful plasticity and free interpretation of the forms. The animalistic style had a strong influence in western Asia during the 7th century BC. Such ornaments as necklaces, bracelets, pectorals, diadems, and earrings making up the Ziwiye treasure (discovered in Iran near the border between Kurdistan and Azerbaijan) provide evidence of this Asiatic phase of Scythian goldworking art. The ornaments are characterized by highly developed expressive forms that are always based on animalistic themes. This Central Asian Scythian-Iranian style passed by way of Phoenician trading in the 8th century BC into the Mediterranean and into Western jewelry.

African. Personal decoration in African cultures usually consists of modest though showy material. The works with a relatively high degree of development come from those areas in which the influence of more advanced Mediterranean and Oriental cultures led to activities of some significance in the field of jewelry. Silver was the metal most commonly worked, especially in the northern coastal territories, and the forms used for ornaments were derived mainly from the art of Islām. Decoration that rarely surpassed the level of craftsmanship appears on objects such as bracelets, necklaces, rings, brooches, earrings, and belt buckles, and the techniques were usually limited to embossing, filigree, and the insertion of coins or semiprecious stones that had simply been polished.

Regions like Ethiopia, the Sudan, and the Bantu territory, partly because of their Egyptian–Nubian and Arabian origins and partly because they were the centres of a flourishing gold trade, developed a goldworking activity of fairly high quality, which was devoted mainly to the production of objects for the courts and for religious sites. These regions were also devoted to the production of personal ornaments such as embossed plaques, rings, necklaces, and tiaras.



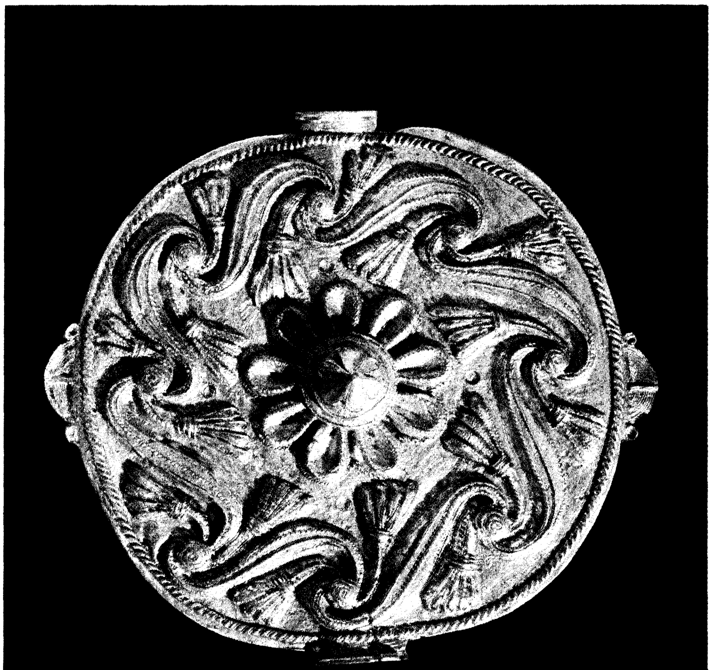
Scythian gold bracelet from the Ziwiye treasure, Persia, 7th century BC. In the Guennol Collection, New York.

By courtesy of the Guennol Collection; photograph. Metropolitan Museum of Art, New York

Jewelry of the Ashanti culture of Ghana

The same observations hold true for the Ashanti culture in Ghana, from which there is a large collection of gold jewelry in the British Museum in London. The local chieftain of each Ashanti tribe had a private workshop for gold jewelry in his small court. In the 18th and 19th centuries the most magnificent court was that of King Asantehene in Kumasi, the ancient Ashanti capital on the Gold Coast. A very widely used object was the emblem of the "bearer of souls," a decorated disk that, together with other insignia, was borne by the king's pages. On the back of the disk was a little tube through which a gold wire or cord was run. The decoration of these disks consisted of a mixture of separate and varied radial or spiral motifs, derived in an unorthodox manner from classic art. The mysterious presence of these ornamental motifs

By courtesy of the trustees of the British Museum; photograph, John Webb



Repoussé gold disk depicting the "bearer of souls," Ashanti (Ghana), 18th–19th century. In the British Museum.

in Ashanti jewelry can be explained only by the sporadic appearance of European goldsmiths in that area, probably during medieval times. Rigid necklaces were also in use, as were rings, which instead of the bezel had fully sculptured figures of animals.

In the past the sandy dunes of Senegal provided alluvial soil from which the natives obtained much gold. There, as in other parts of Africa, the metal that was not exported was used to make ornaments for the tribal chieftains. These were very elaborate objects with complicated decorative motifs worked in embossing or punched freehand. The objects were characterized by the repetition of the designs used and by protruding hemispheres that were smooth or decorated, according to their size.

In these regions, where the making of jewelry was directly dependent upon what was obtained from local deposits, gold was the only material used. Usually the type of decoration, taken from imported models or introduced directly by European goldsmiths, persisted as a repertoire was acquired, with a tendency toward ever great repetition. In other words, rather than an art form dominated by genuine native expression, this production has no relation to the local culture.

American Indian. Central and South American: *pre-*Columbian. Although precious metals were worked in some areas before the beginning of the Christian Era, the greatest development in this field was reached during the second half of the 1st millennium AD. The ancient peoples who, with no proven contact with the cultures of other continents, attained the highest degree of artistic evolution in this part of the world were located in the region near the Andes (Peru, Colombia, Ecuador) and in Central America (Mexico, Guatemala, El Salvador, and part of Honduras), while Panama, Costa Rica, and Nicaragua are considered areas of cultural exchange between South and Central America.

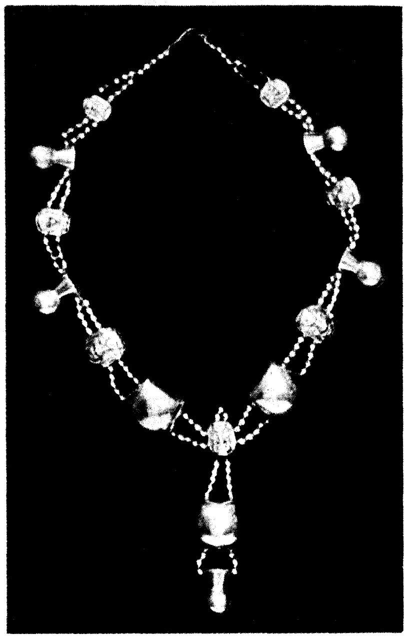
Andes region

In Peru the Chavin culture (c. 5th century AD) has supplied the oldest products, consisting of embossed gold sheets with ornamental motifs that suggest some Chinese influence, although this has not been proven. At the end of this period there was considerable development in techniques. This was followed by the Mochica culture (c. 3rd to 8th century), named for the Moche valley, where the discovery was made of objects decorated with filigree and set with gems. The Mochica culture produced large quantities of jewels in which animals were frequently represented, an example being a ring-shaped earring with a pendant in the form of a toucan (a tropical American bird). Contemporaneously, the Nazca culture flourished along the southern coast of Peru, producing gold ornaments of high technical quality in which anthropomorphic figures in high relief predominated. After centuries during which there was more or less a standstill because of a series of wars and political turmoil, the formation of the Inca empire in about 1300 encouraged the flourishing of local styles (Chimú, Chancay, Ica).

The richest gold and mineral deposits, which are still productive, were those in Colombia. It is not possible to establish definite dates for jewelry from Colombia and Ecuador, but an approximate chronology indicates the San Augustin zone as the oldest, followed by Chitcha. In the latter area, the "Quimbaya treasure" and objects from the upper Cauca River (Calima style) represent jewelry of the greatest importance and magnificence.

In the civilizations of the Andes, gold was lavishly used on clothes. About 13,000 pieces of gold were found sewn into a single Chimú poncho. On certain occasinos the priests wore tunics made entirely of braided gold sheet applied to the cloth. One of the commonest ornaments worn by important personages and warriors was the *nariguena*, a gold ornament that was hooked to the nostrils and might be in the shape of a simple ring, a laminated disk, or an upside-down fan decorated with pierced work. Anthropomorphic and zoomorphic figures appeared on pectorals which were embossed in high relief and decorated with overlapping cut and embossed pieces.

There are necklaces with a seashell motif in different shapes arranged one after the other and necklaces with other stylized zoomorphic forms that are all alike. One of the most outstanding of these necklaces is from Chimú (May 21, 1968, Christie sale). It is composed of a row of gold beads to which are attached eight similar figures of a deity in a ritual pose (1100–1200).



Gold necklace. Chimú, c. 1300. In the British Museum.

By courtesy of the trustees of the British Museum

Among the advanced techniques known by the people of the Andes were repoussé gold beating, cire perdue casting, and the making of gold alloys. The most common alloy, *tumboaga* (consisting of gold and copper), was also known to the Mexicans.

Meso-American region

The most outstanding artistic development during the pre-Columbian era took place, together with that in the region of the Andes, in that part of Central America that is approximately bordered on the north by the Mexican state of Sinaloa and on the south by Honduras and Nicaragua. Probably because the area was poor in gold, objects made of this material date from about 1,000 years after those from the Andes (c. 14th century BC). Mixtec graves have yielded outstanding examples of objects such as gold pendants, jewels combined with turquoise mosaic, and quartz ear spools. The few examples that remain from the Aztec period suggest the stylistic influence of the Andes region. Of the decorative animal motifs, the most frequent is the serpent; of the ornamental motifs, the spheroid, disk, and sphere.

It is thought that ornamental objects in precious materials from the pre-Columbian civilizations, especially the older ones, had some religious function in addition to being used in burial rites. Stylistically, pre-Columbian objects show an unusual amount of charming expressiveness. Symbolic concepts were transferred from stone and pottery to gold through transfigurations that enhanced the plasticity of the forms, displaying at the same time an awareness of structure and of compositional rhythms that forms the main appeal of these objects. Although this was not done frequently, at times the pre-Columbian Indians encrusted their jewels sparingly with rock crystal, lapis lazuli, emeralds, turquoise, pearls, and coral, the purpose nearly always being to increase the expressiveness of the object represented.

North American. The diverse forms taken by personal ornamentation are related to the type of life led by the numerous ethnic and tribal groups scattered throughout the vast American territory. The most highly developed tribes were those whose social organization permitted them to settle in one place for long periods of time, with the consequent evolution of religious and artistic activities.

On the basis of archaeological finds, Indian territory was divided into the following broad areas: the eastern forests, which includes the Great Lakes region and Florida, east of the Mississippi; the Great Plains, including the central part of North America between the eastern forests and the Rocky Mountains; the Southwest, which corresponds to what are now the states of Arizona, New Mexico, southeastern Utah and southern Colorado; the northwestern coast, from the bay of Yakutat in Alaska to the mouth of the Columbia River; and California, in the area included between the northwestern coast and the southwestern cultures.

The Great Plains and California produced no jewelry, the former area because its slow artistic evolution involved only the decoration of clothing with leather, and the latter because its tribes were economically at the pre-agricultural level and therefore lacking in any impressive, long-lasting form of artistic expression. Judged on the basis of the archaeological data that has come to light so far, the highest artistry was achieved by the southwestern cultures, followed by those of the eastern forests and of the northwest.

Personal ornamentation in all of the native cultures of North America shows no connection with the pre-Columbian cultures of Central and South America. One of the most striking differences between the two is that in North America copper was much more frequently used than gold. In some parts of North America this metal may have been used before its use became known in the Western world and at that far-off time it was valued like gold.

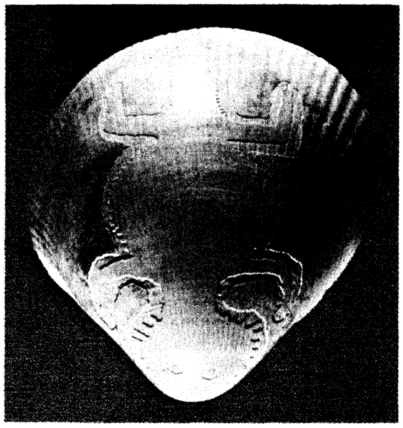
As far back as the archaic period, the practice of decorating shells with carving or champlevé enamel work was very widespread. Feathers and turquoise (used for mosaic) complete the list of precious materials available to the American Indians for personal ornamentation until the arrival of the white man.

On the whole, in their limited diversity, forms of artistic expression became traditional for particular cultures and were perpetuated by them. Even today, attempts are still being made to keep them alive.

Indians of the Southwest

In the southwest cultural area the first objects used for personal ornamentation go back to the first half of the 1st millennium AD and consist of bracelets made from a shell carved in the shape of a frog, exquisitely sculptured in miniature; zoomorphic subjects on auricular disks; rings with bird and snake motifs in pierced work; and other shell jewelry covered with turquoise mosaics.

By courtesy of the Arizona State Museum. Tucson; photograph, Helga Teiwes

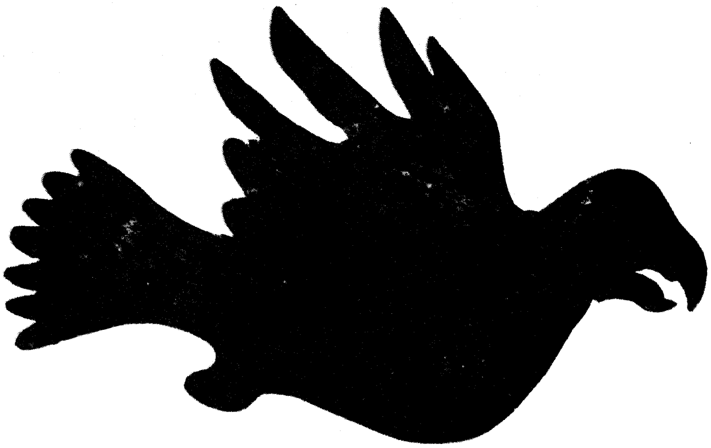


Shell etched in the horned toad motif, from the Snaketown excavations, Arizona; Hohokam. 900–1150 AD. In the Arizona State Museum, Tucson.

The Pueblo and Navajo tribes, which were part of the southwest cultural area, made beautiful necklaces and pendants from turquoise mosaics, shells, and coral. The Pueblo Bonito discoveries document this activity from pre-Columbian times. At the beginning of the second half of the 19th century, the Navajos learned to work silver from Mexican craftsmen and developed this skill with great ability, reworking motifs of Spanish American origin in their Indian traditional style.

In the Great Lakes region where the Forest culture was located, archaeological research has demonstrated the presence of copper ornaments as early as the 5th millennium BC. These consist of necklace beads formed of thin, narrow metal strips and of sheet metal in the shape of fish. The Hopewell finds include bobbin-shaped copper earrings and engraved sheets of silver, dated between 200 BC and AD 500, together with ornaments that were sewn into clothing or inserted in headdresses. From the Mississippi period there are pieces of embossed copper sheet and breastplates, disks, and plaques made of copper and shell with a wealth of engraved ornamental motifs, such as birds, sun symbols, isolated heads, human skulls, eagles, rattlesnakes, hands with outspread fingers and an eye designed on the palm, crosses, and figures of warriors.

By courtesy of the Ohio State Archaeological and Historical Society, Columbus



Flying peregrine falcon, copper repoussé, Hopewell, 1st century BC–1st century AD. In the collection of the Ohio State Archaeological and Historical Society, Columbus.

Beginning in the 17th century, the Seneca, Cayuga, Onondaga, and Iroquois tribes in the New York State region hammered, shaped, and cut European silver coins to be used for jewelry of all kinds. Also worthy of note among the Iroquois are bone combs with handles carved in zoomorphic shapes.

Indians of the northwest coast

In the culture of the Indians on the northwest coast, the influence of Arctic and even of Asiatic peoples can be observed. Persons of very high rank wore a characteristic type of headdress, which was made of wood, in a conical shape with wide brim, surmounted by sculptured human and animal figures. Another type was shaped like a crown or diadem with a rectangular plaque worked in relief placed in the middle of a leather forehead band from which ermine tails and bunches of sea-lion bristles stuck out. The sculpturing on these plaques is highly refined, and the rich shell inlay with which they are decorated makes them look like jewels. The engraving on combs is also outstanding.

The sculptural style peculiar to this culture is characterized by a conventional, formal naturalism that is extremely vigorous and dynamic. Often the same object combines parts that are fully sculptured with parts in low relief, and the depth of the carving may vary greatly.

Objects called copper coins, symbols of maximum power and wealth, were in the form of a shield made of copper sheet in a standardized shape (trapezoidal above and rectangular below). The upper half was taken up by a design such as a head worked in engraving or embossing.

During the 16th century, European conquest and rule of the American continent interrupted development of the arts among the natives, who were forced to live under conditions that were far from favourable to the continuation of traditional artistic activities.

BIBLIOGRAPHY. W.T. BAXTER, Jewelry, Gem Cutting *and Metalcraft,* 3rd ed. rev. (1950), is a manual for students learning the art of the jeweller; many pieces of jewelry, several instruments, and various methods that are used in this extraordinary work are illustrated. JOAN EVANS in *A* History of *Jewellery, 1100–1870,* 2nd ed. (1970), as in the author's earlier works—English Jewellery from the Fifth Century *A.D.*

to 1800 (1921) and *Magical Jewels of the Middle Ages and the Renaissance* (1922)—documents and makes a profound aesthetic and critical study of the style of jewels in the periods considered, giving examples in pictures and in sculpture. E.D.S. BRADFORD, *Four Centuries of European Jewellery* (1953, reprinted 1967), treats jewelry from the Renaissance up to the 20th century, describing the manufacture of the precious material and the technique that must be learned to transform that material into a work of jewelry. R.A. HIGGINS, *Greek and Roman Jewellery* (1961), is divided into two parts: a description of the technical processes and their historical evolution and a chronological listing of the jewels themselves. This scholarly work makes it particularly clear that, except for rare exceptions, the history of ancient jewelry is not merely a history of stylistic revolution but a combination of technical and creative development. ERICH STEINGRABER, *Alter Schmuck: Die Kunst d. europ. Schrnuckes* (1956; Ital. trans., *L'arte del gioiello in Europa dal Medioevo al liberty,* 1965), deals with the historical evolution of gem setting as an art form; *Der Goldschmied* (1966), is an exhaustive study of the history of the techniques of working with gold and silver. HENRI VEVER, *La Bijouterie française au XIX^e siècle (1800–1900),* 3 vol. (1906–08), documents the grand lines of goldsmiths and jewelers during the 19th century, evaluating the most authoritative and impartial sources of the time: official reports on the various expositions and articles published in journals and reviews. GUIDO GREGORIETTI, *Il Gioiello nei secoli* (1969; Eng. trans., *Jewelry Through the Ages,* 1969), treats the history of personal precious ornaments of all civilizations from the 3rd millennium BC to the present (also contains descriptions of the most important beds of celebrated diamonds, pearls, and corals). MARIA ACCASCINA, *L'oreficeria italiana* (1934), reviews the most important works of the Italian jewellers from the Teutonic era to the beginning of the 20th century, concentrating on the significant events influencing the current means of production. ANTONIO MORASSI, *Antica oreficeria italiana* (1936), the catalog of Milan's sixth triennial exhibition of ancient Italian jewelry, documents with examples the historical evolution of jewelry. FILIPPO ROSSI, *Capolavori di oreficeria italiana dall' XI al XVIII secolo* (1956), is a study that illustrates the important moments in Italian jewelry, beginning with the 11th century. HEINZ BATTKE, *Geschichte des Ringes in Beschreibung und Bildern* (1953), describes the history of rings, beginning with the Egyptian seal of the 18th dynasty (*c.* 1400 BC) and continuing up to the rings of the 19th century. The book presents 160 examples that give a complete panorama of the history of rings through the centuries. Other recommended studies of particular countries and periods include: ETIENNE COCHE DE LA FERTE, *Les Bijoux antiques* (1956; Eng. trans., *Antique Jewellery from the Second to the Eighth Century,* 1962); MARGARET FLOWER, *Victorian Jewellery,* rev. ed. (1967); R.G. CHANDRA, *Studies in the Development of Ornaments and Jewellery in Proto-historic India* (1964); GABRIEL MOUREY and A. VALLANCE, *European Art Nouveau Jewelry* (1969); GRAHAM HUGHES, *Modern Jewelry: An International Survey, 1890–1967,* rev. ed. (1968); E.D.S. BRADFORD, *Contemporary Jewellery arrd Silver Design* (1950); JOHN HAYCRAFT, *Finnish Jewellery and Silverware* (1962); GEORGES MARCAIS, *Les Bijoux musulmans de l'Afrique du Nord* (1958); MARY L. DAVIS and GRETA PACK, *Mexican Jewelry* (1963); LIDA VILIMOVA, *Egyptian Jewellery* (1969); and D.J. WILLCOX, *New Design in Jewelry* (1970).

(G.Gr.)

Jewish Mysticism

This article deals with the special nature and characteristics of Jewish mysticism, the main lines of its development, and its role in present-day religion and culture.

NATURE AND CHARACTERISTICS

The Judaic context. The term mysticism applies whenever a person is convinced that it is possible to establish *direct contact,* apart from sense perception and intellectual apprehension, with the divine—a reality undefinable by pure logic and believed to be the ultimate ground of being. Since mysticism springs from an aspiration to join and grasp that which falls outside ordinary experience, it is not easily restricted within precise limits. The boundary line that separates mysticism from metaphysics and cosmology (doctrines on the basic nature or structure of being and the world), from theosophies (systems of thought claiming special insights or revelation into the divine nature), and various forms of occultism (the study and control of supernatural powers), and from theurgy (the art of compelling or persuading divine powers) and even magic, often of the lowest kind, is not clear.

If mysticism is defined as the search for *direct* contact with the divine, however, it seems to be incompatible with Judaism. In its classical and normative form, Judaism appears as faith in a sole God who created the universe and who chose to reveal himself to a selected group by means of a rule of life he imposed on it—Torah ("Guidance" or "Teachings," incorrectly rendered as "Law"). The earthly destiny of the chosen nation, as well as the eternal salvation of the individual, in traditional Judaic beliefs, depends upon the observance of this rule of life, through which any relationship to God must take place. The fact is, however, that in the religious history of Judaism the quest for God goes beyond this relationship mediated by Torah, without ever dispensing with it (since that would take the seeker outside of Judaism) or pretending to reach the depths of the mystery of the divine, or still less to end in an ontological identification with it (where God and man are the same in nature and being).

Three types of Jewish mysticism. Three types of mysticism may be discerned in the history of Judaism: the ecstatic, the contemplative, and the esoteric. Though they are distinct types, in practice there are frequent overlappings and mixtures between them.

The first type is characterized by the quest for God—or, more precisely, for access to a supernatural realm, which is itself still infinitely remote from the inaccessible deity—by means of ecstatic experiences; this method is sometimes tainted by theurgy. The second follows the way of metaphysical meditation pushed to the limit, always bearing in its formulations the imprint of the cultural surroundings of the respective thinkers, who are exposed to influences from outside Judaism; this was the case with Philo of Alexandria (*c.* 15 BC–after *AD* 40) and a few of the Jewish thinkers of the Middle Ages, who drew their inspiration from Greco-Arabic Neoplatonism and sometimes also from Muslim mysticism.

Esoterism

The third type of mysticism claims an esoteric knowledge (hereafter called esoterism) that explores the divine life itself and its relationship to the extradivine level (the natural, finite realm) of being, a relationship that is subject to the "law of correspondences." From this perspective, the extradivine is a symbol of the divine; that is, a reality that reveals another, superior reality, whence reciprocal action of the one on the other (which corresponds to it) exists. This form of mysticism, akin to *gnosis*—the secret knowledge claimed by Gnosticism, a Hellenistic religiophilosophical movement—but purged, or almost purged, of the dualism that characterizes the latter, is what is commonly known as Kabbala (literally "tradition"). By extension, this term is also used to designate technical methods, used for highly diverse ends, ranging from the conditioning of the aspirant to ecstatic experiences to magical manipulations of a frankly superstitious character. If the concept of spiritual energy acting on matter and at a distance originally underlay these practices, it finally became unrecognizable and all that remained was a collection of "tricks of the trade."

The favour with which the doctrine of correspondences was regarded by ancient and medieval science, as well as the tendency in the three monotheistic religions (Judaism, Christianity, and Islām) to reconcile the results of rational reflection with the data of revelation, had the result of turning speculation on the origin and order of the universe toward mysticism.

Perennial human questions

It must also be noted that the quest for God implies the search for solutions to problems that go beyond those of religion in the narrow sense and that arise even when there is no interest in the relationship between man and supernatural powers. Man ponders the problems of his origins, his destiny, his happiness, his suffering—questions that arise outside of religion, as well as within nonmystical forms of religious life; the presence or absence of religious institutions or dogmas is of little importance when it comes to these questions. They were all formulated within nonmystical Judaism and served as the basis and framework for the setting and solution of problems in the various forms of Jewish mysticism. This mysticism, especially in its "Kabbalistic" form, brought about profound transformations in the concepts of the world, God and "last things" (resurrection, last judgment, messianic kingdom, etc.) set forth in biblical and rabbinical Judaism. Nevertheless, Jewish mysticism's own set of problems about the origins of the universe and of man, of evil and sin, of the meaning of history, of the afterlife and the end of time is rooted in the very ground of Judaism and cannot be conceived outside of an exegesis of revealed Scripture and rabbinical tradition.

MAIN LINES OF DEVELOPMENT

A study of the main lines of Jewish mysticism, following its actual historical development, reveals that during a very long period, from its origins in the 1st century-CE (Common Era, or AD) to the middle of the 12th century, only the first two of the three types outlined above existed. It was not until the second half of the 12th century that esoterism became clearly discernible; from then on it continued to develop in various forms up to very recent times.

Early stages to the 6th century CE. The centuries that followed the return from the Babylonian Exile in the 6th century BCE (Before the Common Era, or BC) witnessed the growth and intensification of reflection on the intermediary beings between man and God, of meditation on the divine appearances whose special place of occurrence had formerly been the most sacred part of the Jerusalem Temple, of speculation on the coming into being and organization of the universe and on the creation of man. None of these themes was absent from the Bible, which was held to be divinely revealed, but each had become the object of a constant ideological readjustment that also involved the infiltration of concepts from outside and reaction against them. The speculative taste of Jewish thinkers between the 2nd century BCE and the 1st century CE took them in many different directions: angelology (doctrine about angels) and its counterpart demonology (doctrine about devils); mythical geography and uranography, description of the heavens; speculation on the divine manifestations—which had as background the Jerusalem Temple worship and the visions of the moving "Throne" (the "Chariot," Merkava) in the prophecy of Ezekiel; on the double origin of man, a being formed of the earth but also the "image of God"; on the end of time; on resurrection (a concept that appeared only toward the end of the biblical period); and on rewards and punishments in the afterlife.

The literary crystallization of all this ferment was accomplished in writings, such as the book of Enoch, of which Pharisaic (rabbinical) Judaism—which became the normative Jewish [illegible] con[illegible] of Jerusalem [illegible]d the [illegible] of the [illegible] Temple (70 CE)—retained almost nothing and even the vestiges of which it tended to obliterate in its own writings; the Talmud and the Midrash (rabbinical legal and interpretative literature) touched these themes only with great reserve, often unwillingly and more often in a spirit of negative polemic.

As early as the 1st century CE, and probably even before the national calamity of 70, there were certainly sages or teachers recognized by the religious community for whom meditation on the Scriptures–especially the creation narrative, the public revelation of the Torah on Mount Sinai, the Merkava vision of Ezekiel, the Song of Solomon—and reflection on the end of time, resurrection, and the afterlife were not only a matter of exegesis and of attaching new ideas to texts recognized to be of divine origin but also a matter of inner experience. It was, however, probably in other circles that speculation on the invisible world was engaged in and where the search for the means of penetrating it was carried out. It is undeniable that there exists a certain continuity between the apocalyptic visions (*i.e.*, of the cataolysmic advent of God's Kingdom) and documents of certain sects (Dead Sea Scrolls) and the writings, preserved in Hebrew, of the "explorers of the supernatural world" (Yorde Merkava). The latter comprise ecstatic hymns, descriptions of the "dwellings" (hekhalot) located between the visible world and the ever-inaccessible divinity, whose transcendence is paradoxically expressed by anthropomorphic descriptions consisting of inordinate hyperboles (*Shiʿur qoma*, "Divine Dimensions"). In addition, a few documents have been preserved that attest to the existence of methods and practices having to do with the initiation of carefully chosen persons who were made to undergo tests and ordeals in accordance with psychosomatic criteria borrowed from physiognomy (art of determining character from physical, especially facial, traits). Some theurgic efficacy was attributed to these practices, and there was some contamination from Egyptian, Hellenistic, or Mesopotamian magic. (A curious document in this respect, rich in pagan material, is the Sefer ha-razim, the "Treatise on Mysteries," which was discovered in 1963.)

Two early sources of mystic writings

In this extrarational domain, there are many similarities between concepts reflected in unquestionably Jewish texts and the documents of contemporary non-Jewish esoterism, to the point that it becomes difficult, sometimes impossible, to distinguish the giver from the receiver. Two facts are certain however. On the one hand Gnosticism never ceases to exploit in its own way biblical themes (such as the tale of creation and speculation on angels and demons) that have passed through Judaism, whatever their original source may have been; on the other hand, though Jewish esoterism may borrow this or that motif fom ancient *gnosis* or syncretism (fusion of various faiths) and may even raise to a very high rank in the hierarchy of being a supernatural entity such as the angel Metatron, also known as "little Adonai" (*i.e.*, little Lord or God), it still remains inflexibly monotheistic and rejects the Gnostic concept of a bad or simply inferior demiurge who is responsible for the creation and governing of the visible world. Finally, it is noteworthy that during the centuries that separate the Talmudic period (2nd to 5th centuries AD) from the full resurgence of Jewish esoterism in the middle of the 12th century, the texts that have been preserved progressively lose their density and affective authenticity and become reduced to the level of literary exercises that are more grandiloquent than substantial.

Sejer yetzira. In the ancient esoteric literature of Judaism, a special place must be given to the Sefer yetzira ("Book of Creation"), which deals with cosmogony and cosmology (the origin and order of the universe). Creation, it affirms with a clearly anti-Gnostic insistence, is the work of the God of Israel and took place on two different levels: the ideal, immaterial level and the concrete level. This was done according to a complex process that brings in the ten numbers (sefirot, singular sefira) of decimal notation and the 22 letters of the Hebrew alphabet. The ten numbers are not to be taken merely as arithmetical symbols: they are cosmological factors, the first of which is the spirit of God—with all the ambiguities that this term ruah has in Hebrew—while the nine others seem to be the archetypes of the three elements (air, water, fire) and the spatial dimensions (up, down, and the four cardinal points). After having been manipulated either in their graphic representation or in combination, the letters of the alphabet, which are considered to be adequate transcriptions of the sounds of the language, are in turn instruments of creation.

The ontological value of speech

The basic idea of all this speculation is that speech (that is, language composed of words, which are in turn composed of letters/sounds) is not only a means of communication but also an operational agent destined to produce being—it has an ontological value. This value, however, does not extend to every form of language; it belongs to the Hebrew language alone.

The universe that is produced by means of the sefirot and the letters is constituted according to the law of correspondences between the astral world, the seasons that mark the rhythm of time, and man in his psychosomatic structure.

The "Book of Creation" certainly does not proceed entirely from biblical data and rabbinical reflection upon them; certain Greek influences are discernible, even in the vocabulary. What is important, however, is its influence on later Jewish thought, down to the present time: philosophers and esoterists have vied with one another in commentating it, pulling it in their own direction, and adjusting it to their respective ideologies. Even more important is the fact that Kabbala (see below The making of the Kabbala) borrowed a great part of its terminology from it (*sefira*, among others), naturally making semantic adaptations as required.

The speculation traced above developed during the first six centuries of the Common Era, both in Palestine and in Babylonia (later called Iraq); Babylonian Judaism had its own social and ideological characteristics, which put it in opposition to Palestinian Judaism in various aspects, including esoterism as well as other manifestations of the life of the spirit. The joint doctrinal influence of the two centres was to spread during the period from the mid-8th to 11th century among the Jews established in North Africa and Europe; mystical doctrines also filtered in, but very little is known about the circumstances and means of their penetration.

Greek philosophy and Islamic mysticism

The Arabic-Islāmic influence (7th–13th century). Arabic-Islāmic culture provided another important influence in Jewish mystical development. A considerable part of Jewry, which had fallen under Muslim domination in the 7th and 8th centuries, participated in the new Arabic-Islāmic civilization; the Jews of Asia, Africa. and Spain soon adopted Arabic, the prevailing language of culture and communication. By way of Arabic-language culture, elements of Greek philosophy and Islāmic mysticism penetrated Judaism and contributed to the deepening of certain theological concepts that were Jewish in origin but had become the common property of the three religions of the Book: affirming the divine unity, purging all anthropomorphism from the idea of God, and approaching the divine by progressing on a spiritual path that leads through an ascetic discipline (both physical and intellectual) to a detachment from this world and a freeing of the soul from all that distracts it from God. Greek philosophy and Islāmic mysticism, moreover, raised very serious questions that threatened many traditional beliefs, such as the creation of the world, the providential action of God, miracles, eschatology (doctrines about the resurrection of the body, rewards, and especially material punishments in the hereafter). Even in the Christian West, where cultural contacts between the majority society and the Jewish minority were far from reaching the breadth and intensity of the Judeo-Arab relations, Jewish intellectuals were unable to remain totally impervious to the incursions of the surrounding civilization. Moreover, at the beginning of the 12th century, if not earlier, European Judaism received part of the intellectual Arabic and Judeo-Arabic heritage through translations or adaptations into Hebrew, its only cultural language.

The making of the Kabbala (c. 1150–1250). It was in these circumstances that, starting around 1150, manifestations of markedly theosophic ideologies appeared in the south of present-day France (in the regions of Provence–Languedoc–Roussillon). Two types can be distinguished at the outset, which are very different as to their manner of appearance, their form, and their content.

Sefer *ha-bahir*. The first type is represented in fragmentary, poorly written, and badly assembled texts that began to circulate in Provence–Languedoc during the third quarter of the 12th century. Their inspiration, however, leaves no doubt as to the community of their origin. They were in the form of a Midrash; that is, an interpretation of Scripture with the help of a particular interpretative method, full of sayings attributed to ancient rabbinical authorities. This whole body of texts, probably imported from the Near East (Syria–PalestineIraq), is known as the "Midrash of Rabbi Nehunya ben Haqana" (from the name of a 2nd-century rabbi) or Sefer *ha-bahir*, "The Book of Brightness" (from a characteristic word of the first verse of Scripture to be elucidated in the work). The authorities cited are all inauthentic (as was often the case in late works), and the content of this Midrash, even its nonmystical content, is entirely Gnostic; a Gnosticism that tries nevertheless to escape any ontological dualism (and, as a matter of fact, succeeds).

Role of the divine powers

Its object is to present the origin of things and the course of history centred naturally on that of the chosen people, with the vicissitudes caused in turn by obedience to God and by sin, as bound and conditioned by the manifestation of divine powers. These "powers" are not "attributes" derived and defined by philosophical abstraction, although that is one of the terms used to designate it: they are hypostases (essences or substances). They are inseparable from God, but each one is clothed in its own personality, each operates in its own manner, in the leaning toward severity or mercy, in dynamic correspondence with the behaviour of man, especially of the Jew, in the visible world. They are ranked in a hierarchical order, which is not yet as fixed as it became starting with the second generation of Kabbalists in Languedoc and Catalonia (see below *The school of Gerona* [*Catalonia*]). The rich nomenclature used to designate the "powers" exploits the resources of both the Bible and rabbinical tradition, of the "Book of Creation," of some ritual observances, and also of the letters of the Hebrew alphabet and the signs that can be added to them to indicate the vowels. All of this combines to give a symbolical rendering of the myth, cosmology, sacred history, and eschatology through which an anonymous group of theosophists attempt to formulate their doctrine: a Gnostic myth, except for the adjustments that eliminate the radical depreciation of the visible world.

Thus, according to the Sefer *ha-bahir*, the universe is the manifestation of the hierarchically organized divine powers, and the one that is at the bottom of the hierarchical ladder has special charge of the visible world. This entity is highly complex. Undoubtedly there are survivals of Gnostic speculation on Sophia ("Wisdom"), who is involved, sometimes to her misfortune, in the material world. This power is also the divine "Presence" (*Shekhina*) of rabbinical theology but profoundly transformed: it has become a hypostasis; by a bold innovation, moreover, it is characterized as a feminine being and thus finds itself, while remaining an aspect of the divinity, in the position of a daughter or a wife, who owns nothing herself and receives all from the father or the husband. It is also identified with the "Community of Israel," another radical innovation. but facilitated by ancient speculation based on the allegorical interpretation of the Song of Solomon, which represents the relationship of God to the chosen nation in terms of the marriage bond. Thus a theosophical equality is established between the whole of the people chosen by God, constituted into a kind of mystical body, and an aspect of the divinity, whence the solidarity and linked destiny of the latter and the human group in question. As a matter of fact, a comparable relationship between the "Presence" and Israel was not totally foreign to ancient rabbinical theology. In this light, the obedience or disobedience of Israel to its particular vocation is a determining factor of cosmic harmony or disruption and extends to the inner life of the divinity. This is the essential and definitive contribution of the *Sefer* ha-bahir to Jewish theosophy. In the same document may be seen the resurgence of a notion fought against by the older theologians—that of *metensōmatōsis*, the reincarnation into several successive bodies of a soul that has not attained the required perfection in a previous existence.

School of Isaac the Blind. Parallel to the appearance of the Sefer *ha-bahir* but independent of it, another theosophic tendency unfolded in Languedoc, the second type referred to above. The two movements would take only about thirty years to converge, to constitute what may conveniently, though not quite precisely, be called classical Kabbala. The second school flourished in Languedoc during the last quarter of the 12th century and crossed the Pyrenees into Spain in the first years of the 13th century.

The most eminent spokesman of this school was Isaac ben Abraham, known as Isaac the Blind. For this theosophist, among whose extant works there is in particular a very obscure commentary of the "Book of Creation," the general vision of the universe proceeds, to use the words of Gershom G. Scholem (the eminent 20th-century Kabbala scholar), from the link he discovers between the hierarchical orders of the created world and the roots of all beings implanted in the world of the *sefirot.* One can already see a Neoplatonic influence in the reflections of Isaac: *e.g.*, the proceeding of things from the One and the corresponding return to the heart of the primordial undifferentiatedness, which is the fullness of being and at the same time every conceivable being. This return is not merely eschatological and cosmic but is in some way realized in the life of prayer of the contemplative mystic privileged to have supernatural inspirations, "appearances" of the prophet Elijah, by means of concentration, of orientation of action and thought (*kawwana*), and of "adhesion" (devequt), being-with-God, though not, indeed, a transforming union by which the human personality blends completely into the deity or becomes one with it.

Return to the undifferentiated One

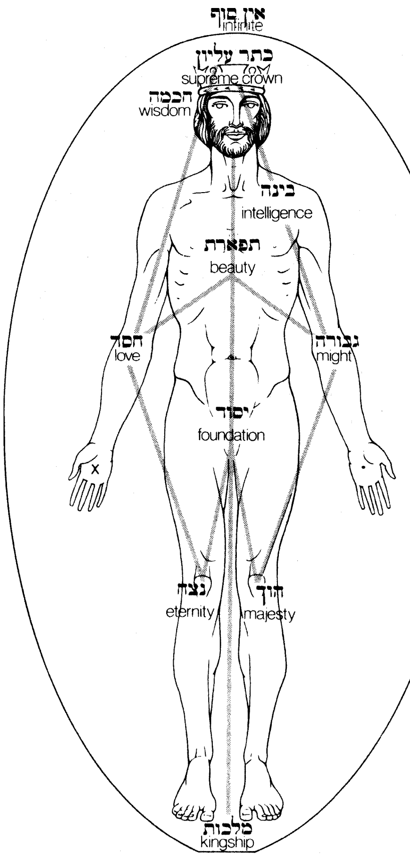
The synthesis of the themes of the bahir and the cosmology of the "Book of Creation," accomplished by Isaac or by others in the doctrinal environment inspired by his teachings, is and remains the foundation of Kabbala whatever enrichment, adjustments, even changes of orientation and sometimes radical modifications the composite may have undergone subsequently.

The ten sefirot. It is also in this environment that the nomenclature of the ten *sefirot* became more or less fixed; it is important to remember this, whatever variant terminologies and even divergent concepts as to the nature of these entities may exist elsewhere—*e.g.*, as internal powers of the divine organism (Gnostic point of view), as hierarchically ordered intermediaries between the infinite and the finite (Neoplatonic concept), or simply as instruments of the divine activity, neither partaking of the divine substance nor being outside it.

Here then is the classical list of the sefirot:

1. *keter ʿElyon*	The Supreme Crown (its identity or nonidentity with the Infinite, *En Sof*, the unknowable deity, remains problematical)
2. *ḥokhma*	Wisdom, the location of primordial ideas in God
3. *bina*	Intelligence, the organizing principle of the universe
4. *ḥesed*	Love, the attribute of goodness
5. *gevura*	Might, the attribute of severity
6. *tifʾeret*	Beauty, the mediating principle between the preceding two
7. *netzaḥ*	Eternity
8. hod	Majesty
9. yeson	Foundation of all the powers active in God
10. *malkhut*	Kingship, identified with the *Shekhina* ("Presence")

The School of Gerona (*Catalonia*). The double current of the gnosticizing theosophy of the *Sefer ha-bahir* and the contemplative mysticism of the masters of Languedoc became one in the elaborations it was subject to at the hands of the Kabbalists in Catalonia, where the Jewish community of Gerona was, during the first half of the 13th century, a veritable seat of esoterism. These elaborations followed the same overall lines, though they were at the same time highly diversified, depending on the personal inclinations of each writer. To the school of Gerona belong, among others, masters such as Ezra ben Solomon, Azriel of Gerona, Jacob ben Sheshet, Moses ben Naḥman (or Naḥmanides, *c.* 1195–1270, the famous Talmudist, biblical commentator, and mystical philosopher); their influence on the subsequent course of Jewish mysticism is of fundamental importance. None of them has left a complete synthesis of his theosophy; they expressed themselves, with more or less reserve, by means of commentaries, sermons, polemic or apologetic treatises or, at the most, brief summaries for the noninitiated. It is not impossible, however, to discover through these texts their vision of the world and compare it with the views of the Jewish thinkers who attempted to harmonize the biblical-rabbinical tradition with Greco-Arab philosophy, whether of Neoplatonic or Aristotelian inspiration.



Sefirot superimposed on the human body (see text).

From C D Ginsberg *Kabbalah, The Essenes;* Samuel Weiser Inc

At the base of the Kabbalistic view of the world there is an option of faith: it is by a voluntary decision that the unknowable deity—who is "nothing" or "nothingness" (nonfinite) because he is a fullness of being totally inaccessible to any human cogitation—set into motion the process that leads to the visible world. This concept radically separates Kabbala from the determinism from which the philosophy of the period could not, without internal contradictions, free the principle of being. In addition it offersa solution consistent with faith to the problem, highly embarassing for the philosophers, of creation ex *nihilo* (out of nothing): the paradoxical reinterpretation of the concept of the "nothing" eliminates the original matter coeternal with God and solves the opposition between divine transcendence (remoteness from the world) and immanence (presence in the world); issuing from the unfathomable depth of the deity and called to return to it, the world, visible as well as invisible, is radically separated from God, who is at the same time constantly present. The correspondence between the *sefirot*, which are modes of the divine manifestation, and all the degrees of being gives meaning to the structure of the world and to the history of humanity centred on the revelation especially given to the chosen people, a revelation that is a rule of life for this people and, consequently, the criterion of merit and sin, or good and evil. Thus, from the top to the bottom of the ladder, there are but corresponding realities that control one another; contrary to the opinion of the philosophers, evil is also a reality since it is the rupture of the universal harmony. It is also the consequence of this rupture, in the form of punishment, but it is repairable. From this perspective, scrupulous observance of the Torah, the revealed Law (both in the written text and the oral tradition), is the essential factor for the very maintenance of the universe. From that point on, the "rational" motivation of the commandments, which raises insurmountable difficulties for the theologians of philosophical orientation, is in the eyes of the Kabbalists but a false problem; the real problem is the fundamental nature of the Torah. Kabbala brings more than one solution to it, whereas philosophy is not even able to raise the question.

The divine "nothing"

It follows from this general concept that the Jewish faith, with its implications—the conviction of holding the undiluted truth, the faithful preservation of ritual practices, and the eschatological expectation—is safeguarded from all the doubts that either philosophical speculation or the rival religious doctrines of Christianity and Islām could evoke in the minds of Jewish believers. Considered from this point of view, Kabbala, already at the stage it had reached at Gerona, turns out to have been a significant factor in the survival of Judaism, which was exposed everywhere in medieval society to the perils that the history of the period reveals.

Besides the Gerona school and the doctrinal descendants of Isaac the Blind in Languedoc, there was another school of Jewish esoterism in southern Europe during the first half of the 13th century. This school—whose followers preferred to remain anonymous and therefore published their writings, such as the ***Sefer ha-Iyyum*** ("Book of Speculation"), either without giving any author's name or by attributing them to fictitious authorities—directed its speculation both to the highest levels of the divine world, where it discerned further aspects beyond the ten ***sefirot*** and attempted to give an idea of them by resorting to the symbolism of light, and to the primordial causes and the archetypes contained in the deity or directly issued from it. The sometimes striking similarity between these speculations and those of John Scotus Erigena a notable 9th-century Christian philosopher, seems to indicate not only a typological kinship of themes between this Kabbalistic current and Latin-language Christian Neoplatonism but also a concrete influence of the latter upon the former. The same may be true about Isaac the Blind and the school of Gerona, but certain knowledge is lacking.

The cycles of existence

Sefer ha-temuna. Still another current manifested itself at the same period; it found its literary expression in the *Sefer ha-temuna* ("The Book of the Image") of unknown authorship. This very obscure document claims to explain the figures of the letters of the Hebrew alphabet. In fact, the speculation of this treatise bears on two themes that were not foreign to the school of Gerona, but it develops them in a personal manner that decisively influenced the future of Jewish theosophy. On the one hand, it deals with a theory of different cycles through which the world must travel from the time of its emergence to its reabsorption into the primordial unity and, on the other hand, with various readings that correspond to these cycles in the divine manifestation that is constituted by the revealed Scriptures. In other words, the reading, thus the interpretation, and consequently the message of the Torah vary according to the cycles of existence; the passage to a cycle other than that under whose governance humanity is presently living could thus entail the modification, even the abrogation, of the rule of life to which the chosen people are presently subject, an explosive notion that opened the way to an overthrow of the traditional values of Judaism.

Medieval German (Ashkenazic) Hasidism. The period from *c.* 1150 to 1250, which witnessed the establishment of Kabbala in the south of France and in Spain, is no less important for the shaping of Jewish mysticism in the other branch of European Judaism, in northern France (and England) and in the Rhine and Danube regions of Germany. Unlike medieval Kabbala, which was to experience a broad and varied development starting in the second half of the 13th century, the movement designated somewhat summarily as German (or Ashkenazic, from a biblical place-name conventionally used to designate Germany) Ḥasidism (Pietism), would hardly survive as a living and independent current beyond the second quarter of the 13th century. There was undoubtedly within Franco-German Judaism a certain continuity of mystical tradition, based on the ***Sefer yetzira*** and the *hekhalot* (see above ***Sefer yetzira);*** certain elements of theurgy and magic of Babylonian origin had perhaps also reached it through Italy; and it would seem that the gnosticizing current crystallized in the ***Sefer** ha-bahir* did not pass without leaving traces in Germany. The intellectual atmosphere of Franco-German Judaism, however, differed greatly from that reigning in Spain or even Provence–Languedoc; it was characterized by an almost exclusively Talmudic culture, less intellectual contact with the non-Jewish environment than in the countries of Muslim civilization, and a very limited knowledge of the Jewish theology in Arabic of the Middle East, North Africa, and Spain. This situation would change only in the last third of the 12th century; until then, the "philosophical" equipment of the Franco-German Jewish scholar consisted essentially of a Hebrew paraphrase, dating perhaps from the 11th century, of the treatise ***Beliefs and Opinions*** by Sa'adia ben Joseph (the great 9th–10th-century Babylonian Jewish scholar and philosopher), and the commentary on the "Book of Creation," written directly in Hebrew (in 946) by the Italian Jew Shabbetai Donnolo. Even when the cultural influence of Spanish Judaism came to be felt more strongly in France–England and Germany, the speculative Kabbala noted above hardly penetrated there. Thinkers within Franco-German Judaism who inclined toward theological speculation had their own problems, which resulted in a mysticism strongly imbued with asceticism, a type of mysticism toward which the general situation of the Jews in those regions contributed, as, especially after the First Crusade, they were severely afflicted by bloody persecutions. The main speculative problem was that of the relationship between God in his pure transcendence and total unity and his manifestations in creation, as well as in revelation and communication with inspired men. Reflection on this problem led to the elaboration of various supernatural hierarchies between the inaccessible God and the created universe or the recipient of divine communication; data on angels taken from the Bible and rabbinical and mystical tradition, as well as speculation on the *Shekhina*, were used as material for these hierarchies and also gave a peculiar coloration to liturgical practice. The latter was marked, moreover, by a concern for spiritual concentration by means of fixing the attention on the words and even the letters of the synagogue prayers. Whatever the historical interest of these speculations, they had no great repercussions on the subsequent course of Jewish esoterism; the only exceptions are the mysticism of prayer and demonology, which was sometimes influenced by the beliefs of the Christian environment and fully developed in Hasidic circles. On the other hand, the ascetic morality of the movement, which found its literary expression in the work of Eleazar ben Judah of Worms (c. 1160–1238) and in the two recensions of the "Book of the Pious" ***(Sefer Ḥasidim)***, was to mark Jewish spirituality, esoteric or not, from then on.

Ashkenazic speculation and asceticism

The making of the Zohar *(c.* 1260–1492). Once the actually marginal episode of German Ḥasidism was finished, almost all of the creative activity in Jewish mysticism was to be situated or would originate in Spain, up to the expulsion of the Jews in 1492.

After the flowering of the schools described above came to an end, around the year 1260, two other currents appeared. The first, in its own manner, resumed relations with Gnosticism in that it placed the problem of evil at the centre of its reflection. The texts that reflect this tendency do not maintain evil in a state of dependence on the "attribute of judgement" within the structure of the ***sefirot*** set up by the previous Kabbalists but locate it outside the divinity, constructing a parallel system of "left-hand ***sefirot,"*** with a corresponding development of an exuberant demonology. The second movement, whose main representative was the 13th-century visionary-adventurer Abraham ben Samuel Abulafia, found its justification in inner experiences considered "prophetic" and encouraged by training methods akin to those of Yoga, the Byzantine Hesychasts (mystical, quietist monks), and the Muslim Ṣūfīs (mystics); moreover, an important place was given to speculations on the letters and vocalic signs of the Hebrew script. Unlike the protagonists of other mystical schools of Spain that until then had not sought to spread their ideas outside the circle of initiates, Abulafia applied himself in various places to propaganda and exhibitions that disaffected and worried the leaders of Judaism and caused their initiator to be pursued even by the non-Jewish authorities. The numerous writings that he left were later to stimulate a few minds among the Kabbalists.

The world of Moses de León

The work of Moses ben Shem Tov de León, in the last quarter of the 13th century, marked one of the most important turning points in the development of Jewish mysticism. Moses de León was born in the middle of the 13th century and died in 1305; he was the author of several esoteric works, which he signed with his own name. But at the same time, in order to better spread his ideas and to more effectively combat philosophy, which he considered a mortal danger to the Jewish faith, he turned to the composition of pseudepigrapha (writings ascribed to other authors, usually in past ages) in the form of Midrashim (plural of Midrash) on the Pentateuch, the Song of Solomon, Book of Ruth, and Lamentations, in which Talmudic authorities appeared, of whom only the names were even partially authentic, a procedure already used by the *Sefer. ha-bakir* (see above *Sefer ha-bahir);* in its most finished version (for there were several of them), the plot of the tales centred around Rabbi Simeon ben Yohai, a doctor of the 2nd century, about whom the Talmud already related some curious anecdotes, most of them semilegendary. Moses de León thus produced, over a period of about 30 years, first a work entitled *Midrash ha-ne'elam* ("The Mystical Midrash") whose method was largely allegorical and whose tongue was mainly Hebrew, and then a larger work, the *Sefer ha-zohar* ("The Book of Splendor"), or more briefly the *Zohar,* whose content is theosophic and which was written in artificial Aramaic. The book culminates in a long speech in which Simeon ben Yohai, on the day of his death, supposedly exposes the quintessence of his mystical doctrine. The literary hoax of Moses de León was not immediately accepted as authentic by all the esoterists and still less by scholars outside the theosophic movement; it took half a century or more for the *Zohar* and imitations of it to be recognized as authoritative ancient works, and even then it was not without some reluctance. The nearly contemporary imitations of the *Zohar* that were incorporated into it or appended to it were sometimes of a markedly different ideological orientation: the *Ra'ya Mehemana* ("Faithful Shepherd"—that is to say, Moses, who is the central figure of this composition, the particular subject of which is the interpretation and theosophic justification of the precepts of the Torah); and the *Tiqqune Zohar,* elaborations in the same vein bearing upon the first word of the book of Genesis (*Bereshit*, "In the beginning"). Although critics were never fully silenced and the authenticity of the *Zohar* was already questioned in the 15th century, the myth created by Moses de León and his imitators became a spiritual reality for the majority of believing Jews; it still retains this character among "Orthodox" Jews. The *Zohar,* believed to be based on supernatural revelations and reinterpreted in diverse ways, would serve as support and reference for all the Jewish theosophies in the centuries ahead.

Doctrine and symbolism of the *Zohar*

As to doctrine, the *Zohar* and its appendixes develop, amplify, and exaggerate speculation and tendencies that already existed, rather than offering any radical innovation. All of the ideas had already been accepted for a long time in Jewish theosophy: the springing forth of being from the depth of the divine "nothing"; the solidarity of the world of the *sefirot* (complicated by the introduction of four ontological levels at each one of which the schema of the ten *sefirot* is reproduced) with the visible world; the indispensable contribution of man (that is, of the Jew) who observes the biblical and rabbinical precepts in their slightest details, to universal harmony—these emphases remain the main lines of the *Zohar.* But all these themes (the speculations of the *Sefer ha-Temuna*, mentioned above, on the cosmic cycles and the "Prophetic Kabbala" of Abulafia being tacitly set aside) were largely organized and enhanced by the use, or rather the unscrupulous appropriation, of materials taken from rabbinical tradition and ancient esoterism as well as from more recent theological and philosophical currents of thought, despite the lack of esteem that the writers of the Zoharic corpus felt and sought to make others feel toward works created by gentiles.

The method of symbolic representation used by the writings of the Zoharic corpus was supported by a system of interpretation that made use of the originally Christian concept of the fourfold meaning of Scripture: literal, moral, allegorical (philosophical), and mystical. The symbolism thus set up boldly made use of an exuberant anthropomorphic and even erotic imagery whose function was to convey the manifestation of the levels of the sefirot to each other and to the extradivine world. The myth of the primordial man (Adam Qadmon), a virtually divine being, reappeared here under a new form, and it was to remain in the subsequent development of Kabbala.

The *Zohar* thus claims to provide a complete explanation of the world, man, history, and the situation of the Jew; on a higher level, to justify the biblical revelation and rabbinical tradition, down to the slightest detail, including the messianic expectation; and thereby to neutralize philosophy. But, while setting itself up as the defender of the traditional religion regulated by the Talmud and its commentaries, in a sense it places itself above tradition, by proclaiming boisterously the incomparable value of the theosophic teaching of "Rabbi Simeon ben Yohai" and the superiority of the esoteric doctrine over the Talmudic studies, which were open to all and which, along with the observance of the precepts, were, according to common opinion, supposed to constitute the basic justification of the life of the Jew. There is in this attitude —more accentuated in the *Ra'ya Mehernana* (see above) than in the *Zohar* proper—a revolutionary potentiality, a possible threat to the primacy of practice and study of Torah; the future would show that this danger was not completely unreal.

The Lurianic Kabbala. After the establishment of the Zoharic corpus, no major changes took place in Jewish esoterism until the middle of the 16th century, when in Safed (in Upper Galilee, Palestine; present-day Ẕefat, Israel) a religious centre of extreme importance for Judaism was established, which was mainly inspired by teachers coming from families expelled from Spain. Until the expulsion of the Jews from Spain (1492) and during the two generations that followed it, the Kabbalistic literary output had certainly been abundant, in Spain till the expulsion as well as in Italy and the Middle East; but it was primarily a matter of systematizing or even popularizing the *Zohar* or of extending the speculation already developed in the 13th century; there were also some attempts at reconciling philosophy and Kabbala. It should be noted that even the traditionalist theologians adopted a careful and rather reserved attitude toward Kabbala.

The teaching of Isaac Luria

The tragedy for Judaism of the expulsion from Spain and of the forced conversions to Christianity that preceded it by a century, and which would become even more extensive in Portugal shortly afterward, deeply marked the victims. These events, accentuating the already existing pessimism in response to the situation of the Jewish people dispersed among the nations, intensified the messianic expectation. This expectation does not seem to have been unrelated to the beginnings of the printed transmission of Kabbala—the first two printed editions of the *Zohar* date from 1558. All these factors, joined with certain internal developments of speculative Kabbala in the 15th century, prepared the ground for the new theosophy inaugurated by the teaching of Isaac ben Solomon Luria, who was born in Jerusalem in 1534, educated in Egypt, and died in Safed in 1572; although his teaching is traditionally associated with Safed, he spent only the last three years of his life there. Luria wrote very little; his doctrine has been transmitted, amplified, and probably somewhat distorted through the works of his disciples, of which the main one was Hayyim Vital (1543–1620), who wrote *'Etz Ḥayyim* ("Tree of Life"), the standard presentation of Lurianic Kabbala.

The theosophy of Luria, whose novelty was proclaimed by its creator and perfectly realized by the esoterists who held to the Zoharistic Kabbala (organized and codified precisely in Safed, during the lifetime of Luria, by Moses ben Jacob Cordovero, 1522–70), is of extreme complexity in its details, although basically it is but one more attempt to reconcile divine transcendence with immanence and to bring a solution to the problem of evil, which the believer in the divine unity can recognize neither as a power existing independently of God nor as an integral part of him.

The theosophic vision of Luria is expressed in a vast mythical construct, which is typologically akin to certain Gnostic and Manichaean (3rd-century dualistic) systems but which strives at all costs to avoid dualism. The essential elements of this myth are as follows: the withdrawal (*tzimtzum*) executed by the divine light, which originally filled all things, in order to make room for the extradivine; the sinking, as a result of a catastrophic event that occurred during this process, of luminous particles into matter (qelippof, "shells," a term already used in Kabbala to designate the evil powers); whence the necessity of saving these particles and returning them to their origin, by means of "repair" or "restoration" (*tiqqun*). This must be the work of the Jew who not only lives in complete conformity to the religious duties imposed on him by tradition but who also dedicates himself, in the framework of a strict asceticism, to a contemplative life founded on mystical prayer and the directed meditation (*kawwana*) of the liturgy, which is supposed to further the harmony (*yiḥud*, "unification") of the innumerable attributes within the divine life. The successive reincarnations of the soul, a constant theme of Kabbala that Lurianism developed and made more complex, are also invested with an important function in the work of "repair." In short, Lurianism proclaims the absolute requirement of an intense mystical life with, as its negative side, an unceasing struggle against the powers of evil. Thus it presents a myth that symbolizes the origin of the world, its fall, and its redemption; it gives meaning to the existence and to the hopes of the Jew, not merely exhorting him to a patient surrender to God but moving him to a redeeming activism, which is the measure of his sanctity. Obviously, such requirements make the ideal of Lurianism possible only for a small elite; ultimately, it is realizable only through the exceptional personage of the "just"-- the ideal holy Jew described above.

Shabbetai Tzevi, a 17th-century "messiah"

Shabbetaianism. During the 60 years that followed the death of the founder, the Kabbala linked to the name of Luria and overlaid with accretions from the other mysticisms of Safed spread through the Jewish Diaspora and deeply permeated its spiritual life, liturgy, and devotional practices. It emphasized the necessity of "repair" of a world in which the uneasiness of the Jew kept growing, for in spite of certain favourable factors—the relative tolerance of the Ottoman Empire and the peaceable establishment of an important Marrano (Iberian Jewish, or Sefardic) community in Amsterdam—there was no overall solution to the problem of the "conversos" (converts) who had remained in the Iberian Peninsula. The other half of the Jewish people, the Ashkenazim, also experienced a serious crisis: its most prosperous and dynamic section, the Jewish population of Poland, was sorely tried, almost totally ruined, and in large part forced to move back toward the west because of the massacres and the destruction that took place during the Cossack uprising of 1648. These ideological and historical data may provide the necessary context for understanding the astonishing though short-lived success of Rabbi Shabbetai Tzevi of Smyrna (1626–76), who proclaimed himself messiah in 1665. The "messiah" was forcibly converted to Islām in 1666 and ended his life in exile ten years later, but despite his failure he had faithful followers. A sect was thus born and survived largely thanks to the activity of Nathan of Gaza (*c.* 1644–90), an unwearying propagandist for the "messiah," who justified the actions of Shabbetai Tzevi, which were contrary to the Law, and his final apostasy by theories that were based on the Lurian theory of "repair": it had to be understood as the descent of the just into the abyss of the "shells" in order to liberate from it the captive particles of divine light. The Shabbetaian crisis lasted nearly a century, some of its aftereffects even longer. It led to the formation of sects whose members were externally converted to Islām—*e.g.*, the Dönme (Turkish, "apostates") of Salonika, whose descendants still live in Turkey—or to Roman Catholicism —*e.g.*, the Polish supporters of Jacob Frank (1726–91), the self-proclaimed messiah and Catholic convert. In Bohemia–Moravia, however, the Frankists outwardly remained Jews. This crisis did not discredit Kabbala, but it led the spiritual authorities of Judaism to watch over and severely curtail its spread and to exercise rigorous ideological control, by concrete acts of censorship and repression, over anyone, even a person of tested piety and recognized knowledge, who was suspected of Shabbetaian sympathies or of messianic pretensions.

Modern Ḥasidism. Though it is true that the messianic movement centred around Shabbetai Tzevi could only produce disillusionment and that if it had not been contained it could have led Judaism to its ruin, yet it answered not only the theosophic aspirations of a small number of visionary scholars but also an affective need of the Jewish masses that was left unsatisfied by the dry intellectualism of the Talmudists and the economic and social oppression of the ruling classes (both Jewish and non-Jewish). This was the case especially in Poland, which before the partition of the Polish kingdom (1772–95) included Lithuanian, Belorussian, and Ukrainian territories. It was there that the so-called Hasidic movement, in no way connected with medieval German Hasidism, originated around the middle of the 18th century—a movement in which the Lurian Kabbala, theoretically maintained as the basis of speculation, underwent adjustments and transformations that continue to the present day.

The Ba'al Shem Ṭov

If modern Ḥasidism may be regarded as a mass movement, having a minimum of organization, using the methods of propaganda and preaching, and forming groups of acknowledged members, then the legend is credible that traces it back to a single founder, Israel ben Eliezer. known as Ba'al Shem Tov (Master of the Good Name; that is, a possessor—he was not the only one of his kind--of the secret of the ineffable name of God, which bestows an infallible power to heal and perform other miraculous operations). This man was born around 1700 and died in 1760 in southern Poland. Though relatively untrained according to the norms of the rabbinical Judaism of his time, he was a spiritual personage of exceptional quality and was able to win to his ideas not only the common people but also many representatives of the intellectual elite. The mist of legend that surrounds him is too dense for it to be possible to reconstruct entirely his personal doctrine, which he probably never systematized. Drawing his inspiration from the methods of the itinerant preachers whose activity was becoming more intense in 18th-century eastern European Judaism, he delivered his teaching in the form of homiletic interpretations of sacred texts, having recourse to fables and parables borrowed from daily life and from folklore; this method remained constant in Hasidism, but it is undeniably an exaggeration and even an error in perspective to consider, as did Martin Buber (see below *Jewish mysticism* today), that the tale and the anecdote are the most authentic expression of the doctrine and the spirituality of Ḥasidism. It is indeed in the doctrinal works, most of them expressed in the form of sermons on the weekly sections of the Pentateuch and other liturgical lessons, that the thought of the Hasidic "rabbis" is expressed. It is very diversified thought, for there are as many bodies of doctrine in Ḥasidism as there were creative spirits during the first three generations of the movement. It is, nevertheless, possible to point to a few traits that are fundamental and common to Ḥasidism as a whole.

In theory, it remains rooted in the Lurianic Kabbala—and nothing essential separates it at this point from its most implacable adversaries in the traditional Judaism of eastern Europe. What is unique to it is to have made of devequt, "being-with-God," an object of aspiration and even a constant duty for all Jews and in all circumstances of life, even those seemingly most profane; in other words, it demands a total spiritualization of Jewish existence. This requirement entails a re-evaluation, less new in its principle than in its concrete application, of the speculative concepts of Kabbala: the emphasis is placed on the inner life of the believer, and it is at this level that the supercosmic drama is played (a drama whose stage was, according to bookish theosophy, in the universe of the *sefirot*); according to several teachers, the same emphasis on inwardness holds for messianic redemption. At the same time, Ḥasidism transforms into social reality a requirement that was also part of the Lurian doctrine of "repair," though it was unfortunately distorted by Shabbetaianism: Ḥasidism puts at the centre of the religious life and organization of the group, as an indispensible guide and unquestioned authority, the inspired leader, endowed with supernatural powers—the "just" ***(tzaddiq)***, the "miracle-working rabbi" ***(Wunder-rebbe)***. Ḥasidism thus produced, wherever it triumphed, an undeniable spiritual renewal; the reverse of the medal was the cult of personality, competition between "dynasties" of "rabbis," obstinacy in maintaining the Hasidic community apart from the surrounding society, with all the social and economic consequences that this will to isolation entailed and for which it would be false to lay all the blame on the environment, despite its definite hostility toward the Jews.

The *tzaddiq*

From its very beginnings, Ḥasidism was to encounter strong resistance on the part of the official Judaism of the period, which had been sensitized to the anarchism of the Shabbetaians and was at the same time solicitous for the prerogatives of the established community leaders and rabbis, the vigilant guardians over the traditional laws and their application, who were confined to the formal study of the Talmud and its commentaries. The behaviour of the followers of Hasidism, though irreproachable in its strict, even rigorous observance of ritual rules, displayed several traits that were distasteful to its adversaries (besides the unconditional submission to the *tzaddiq*, who often doubled as the rabbi of the official congregation): desertion of the general communal synagogues, meetings in small conventicles, modifications of the liturgy, casual dress during prayer, and preference given to mystical meditation rather than to the dialectical study of the Talmud, which requires instead serious intellectual concentration. Nevertheless, the conflict between the Ḥasidim and the "Opponents" (Mitnaggedim) did not finally degenerate into a schism; after three generations, a kind of tacit compromise was established between the two tendencies—Hasidic and Talmudic—without the consciousness of differences ever being erased. The compromise was rather to the advantage of Hasidism, but not without a few concessions on its part, notably on the question of education.

The strong organization of the Hasidic groups allowed them to survive the dislocation of eastern European Judaism as a result of the events of World War II, but its vital centres are today in the United States rather than in Palestine, in part because of economic reasons, in part because of the more or less reserved, and at sometimes frankly hostile, attitude of the Hasidic "rabbis" toward political Zionism and the State of Israel. The best known of the U.S.-based groups is the very active Lubavitchers (after Lyubavich, Russia, seat of a famous school of Ḥasidism), whose headquarters are in the Williamsburg district of Brooklyn, New York.

JEWISH MYSTICISM TODAY

The role played by Kabbala and Ḥasidism in the thought and spirituality of contemporary Judaism is far from being insignificant, though its importance is not as great as in former times. Of course, there is hardly any really living Kabbalistic and Hasidic literature, but the personal thought of religious writers such as Abraham Isaac Kook (*c.* 1865–1935), spiritual leader, mystic, and chief rabbi of Palestine, continues to exercise a marked influence. Furthermore, the renewal of religious thought in "westernized" Jewish circles between the two wars received a powerful impulse from the philosopher Martin Buber (1878–1965), whose work is in part devoted to the propagation of Hasidic ideology as he understood it. "Neo-Orthodoxy," founded in Germany by Samson Raphael Hirsch (1808–88), was quite indifferent to mysticism at the outset, but it too came to be influenced by it, especially after the rediscovery of living Judaism in Poland during World War I by Western Jewish thinkers. Also significant is the work of Abraham Joshua Heschel (1907–72), a Polish Jewish writer of distinguished Ḥasidic background and double culture—traditional and Western—who emigrated to the United States.

Christian and secular Kabbala

Jewish mysticism also has exercised some influence on thought outside the Jewish community. Kabbala, distorted and deflected from its own intentions, transcended the frontiers of Judaism and helped nourish and stimulate certain currents of thought in Christian society, from the Renaissance to the present: "Christian Kabbala," born in the 15th century under the impetus of Jewish converts from Spain and Italy, claimed to find in the Kabbalistic documents, touched up if necessary or even forged, arguments for the truths of the Christian faith. Thus a certain number of Christian humanist scholars became interested in Jewish mysticism, several of whom acquired a fairly extensive knowledge of it on the basis of authentic texts. Among them were Giovanni Pico della Mirandola (1463–94) and Gilles of Viterbo (Egidio da Viterbo; *c.* 1465–1532) in Italy, and Johannes Reuchlin (1455–1522), who was responsible for writing one of the principal expositions of Kabbala in a language accessible to the learned non-Jewish public ***(De arte Cabbalistica***, 1517), in Germany, while the visionary Guillaume Postel (1510–81) was attracting disciples in France. The occult philosophy of the 16th century, the "natural philosophy" of the 17th and 18th centuries, and the occult and theosophic theories cultivated even today and which have coloured the ideology of Freemasonry—all of these focus and continue to make borrowings from Kabbala, though they rarely grasp its spirit and meaning. The same is true of most of the books on Kabbala put out by publishers of occult and theosophic literature today.

The rigorous scholarly study of Jewish mysticism is a very recent phenomenon. The state of mind and the tendencies of the founders of the "science of Judaism" (the scholarly study of Jewish religion, literature, history, etc.) in Germany during the first half of the 19th century were too permeated with rationalism to be favourable to an investigation in depth of a movement judged to be obscurantist and retrograde. Granting some valuable earlier works, research on a large scale and application of the proved methods of philology and history of religions began only with the work of Gershom G. Scholem, professor of Kabbala at the Hebrew University, Jerusalem from 1923 to 1965, and has been continued by his disciples, both direct and indirect. This research touched all of the areas of Jewish mysticism that are briefly described in this article; however, the gaps in knowledge remain serious in every area. Critical editions of mystical texts are few in number; unpublished documents are cataloged in a very incomplete manner; and only a few monographs on writers and particular themes exist, though these are indispensible preliminaries to a detailed and thorough synthesis—the one outlined by Scholem in 1941, in his ***Major Trends in Jewish Mysticism***, though of exceptional value in its time, must be taken up again and completed.

BIBLIOGRAPHY

General works: G.G. SCHOLEM, ***Major Trends in Jewish Mysticism***, rev. ed. (1961), the standard survey of the subject, with a chapter-by-chapter bibliography, ***On the Kabbalah and its Symbolism*** (Eng. trans. 1965), several studies on some of the great themes of Jewish mysticism; A.E. WAITE, ***The Holy Kabbalah: A Study of Secret Tradition in Israel*** (1929), a theosophical view of Jewish mysticism.

Archaic Period: (Translations)—HUGO ODEBERG, ***3 Enoch; or the Hebrew Book of Enoch*** (1928); ISIDOR KALISCH, ***A Book on Creation*** (1877). (Studies)—J. ABELSON, ***The Immanence of God in Rabbinical Literature*** (1912); G.G. SCHOLEM, ***Jewish Gnosticism, Merkabah Mysticism, and Talmudic Tradition*** (1960); JOSHUA TRACHTENBERG, ***Jewish Magic and Superstition*** (1939, paperback 1961).

Classical Period: G.G. SCHOLEM, ***Ursprung and Anfänge der Kabbala*** (1962; French trans., ***Les Origines de la Kabbale***, 1966); ***Le commentaire d'Ezra de Gérone sur le Cantique des Cantiques***, trans. by G. VAJDA (1969); ***The Zohar***, trans. by H. SPERLING and M. SIMON, 5 vol. (1931–34; paperback, sel. and ed. by G.G. SCHOLEM, 1963).

Kabbala at Safed: MOSES CORDOVERO, ***The Palm Tree of Deborah***, trans. from the Hebrew by L. JACOBS (1960); R.J.Z. WERBLOWSKY, ***Joseph Karo: Lawyer and Mystic*** (1962).

Hasidism: (Translations)—L.I. NEWMAN, *The Hasidic Anthology* (Eng. trans. 1963); DOB BAER OF LUBAVITCH, *Tract on Ecstasy,* trans. from the Hebrew by L. JACOBS (1963). (Studies)—MARTIN BUBER, *The Origin and Meaning of Hasidism* (Eng. trans. 1960); s. ~ DRESNER, *The Zaddik: The Doctrine of the Zuddik According to the Writings of Rabbi Yaakov Yosef of Polnoy* (1960); S. POLL, *The Hasidic Community of Williamsburg* (1962).

Christian Kabbala: J.L. BLAU, *The Christian Interpretation of the Cabala in the Renaissance* (1944); F. SECRET, *Les Kabbalistes chrétiens de la Renaissance* (1964).

(G.V.)

Jewish Myth and Legend

Jewish myth and legend comprises a vast body of stories transmitted over the past 3,000 years in Hebrew and in vernacular dialects, such as Yiddish (Judeo-German) and Ladino (Judeo-Spanish), spoken by Jews in various parts of the world. These stories have played an important role in the history of Jewish religion and culture.

Significance and characteristics. Apart from their intrinsic appeal, Jewish myths and legends claim attention for three special reasons: (1) Those incorporated in the Old Testament now form part and parcel of the cultural heritage of the Western world and have exerted a profound influence on its literature and art. (2) During the Middle Ages Jews were among the principal transmitters of Oriental tales to the West, so that many familiar Eastern stories can be traced to Jewish compilations. (3) Since these stories have been accumulated through centuries of constant migration, they provide an unrivalled body of "clinical" material for studying the process by which popular tales in fact travel and are transformed.

Basic character and role

Not all of the stories are of Jewish origin; many can be readily paralleled elsewhere and are derived from tales the Jews picked up from their non-Jewish neighbours in the lands of their dispersion. Even what is borrowed, however, is usually impressed with a distinctive Jewish stamp, being adapted to point up some precept of the Jewish religion, to illustrate some facet of Jewish life, or to exemplify some trait of Jewish character and temperament. The dominant overall feature of the stories is, indeed, their religious and moral tone; most of them are, in fact, told specifically as part of the homiletic exposition of Scripture. Such stories are taught to Jews from early childhood as a regular part of their religious education. To the tradition-minded Jew, therefore, they are more than mere literary fancies and assume a kind of doctrinal complexion. Biblical characters and events present themselves to him more in the lineaments of later legend than in their original biblical form; while popular notions about heaven and hell, rewards and punishments, the coming of the Messiah, and the resurrection of the dead derive mainly from this source rather than from Scripture itself.

Virtually all the standard types of folktale are represented. Conspicuously absent, however, are pure fairy tales because fairies, elves, and the like are foreign to the Jewish imagination, which prefers to people the otherworld with angels and demons subservient to God.

A distinction must be made, of course, between myth and legend. In common parlance, a myth is a story about gods or otherworldly beings. Judaism, however, is a rigorously monotheistic religion; hence, in this narrower sense, there can be no original Jewish myths. Nevertheless, from the earliest times, Jews have not disdained to borrow those of their pagan neighbours and then adapt them to their own religious outlook. If, however, the term is interpreted in a larger sense, to mean the portrayal of continuous, transtemporal concerns in the context of particular and punctual events, myth is indeed one of the essential vehicles by which Judaism conveys its message; for it is only when historical happenings are translated into this wider dimension that they cease to be mere antiquarian data and acquire continuing relevance. In Judaism, for example, the Exodus from Egypt is projected mythically from something that happened at a particular time into something that is continually happening, and it thus comes to exemplify the situation and experience of all men everywhere—their emergence from the bondage of obscurantism, their individual revelations at their individual Sinais, their trek through a figurative wilderness, even their death in it so that their children or children's children may eventually reach the figurative "promised land." By the same token, the historical destruction of the Temple of Jerusalem is transformed by myth into a paradigm of the continuing mutual estrangement of God and man, their exile from one another.

Legend, on the other band, implies no more than a fanciful embroidering of purportedly historical fact. Unlike myth, it does not transcend the punctual and local.

Myth and legend in the Old Testament. The vast repertoire of Jewish myths and legends begins with the Old Testament. Their overall purpose in Scripture is to illustrate the ways of God with man, as exemplified both in historical events and in personal experience. The stories themselves are often derived from current popular lore and possess abundant parallels in other cultures, both ancient and modern. In each case, however, they are given a peculiar and distinctive twist.

Myths in Genesis

Myths. Old Testament myths are found mainly in the first 11 chapters of Genesis, the first book of the Bible. They are concerned with the creation of the world and of man, the origin of the continuing human condition, the primeval Deluge, the distribution of peoples, and the variation of languages.

The basic stories are derived from the popular lore of the ancient Near East and can be paralleled in the extant literature of the peoples of the area. The Mesopotamians, for instance, also knew of an earthly paradise such as Eden, and the figure of the cherubim—properly griffins rather than nightgowned angels—was known to the Canaanites. In the Bible, however, this mythical garden of the gods becomes the scene of man's fall and the background of a story designed to account for the natural limitations of human life. Similarly, the Babylonians, too, told of the formation of man from clay, but in the scriptural version his function is to bear rule over all other creatures, whereas in the pagan tale it is to serve as an earthly menial of the gods. Again, the story of the Deluge, including the elements of the ark and the dispatch of the raven and dove, appears already in the Babylonian myths of Gilgamesh and Atraḫasis. There, however, the hero is eventually made immortal, whereas in the Bible this detail is omitted because to the Israelite mind no child of woman could receive that status. Lastly, while the story of the Tower of Babel was told originally to account for the stepped temples (ziggurats) of Babylonia, to the Hebrew writer its purpose is simply to inculcate the moral lesson that man should not build beyond his assigned station.

Myths in the Prophets and Holy Writings

Scattered through the Prophets and Holy Writings (the two latter portions of the Hebrew Bible) are allusions to other ancient myths—*e.g.,* to that of a primordial combat between Yahweh and a monster variously named Leviathan (Wriggly), Rahab (Braggart), or simply Sir Sea or Dragon. The Babylonians told likewise of a fight between their god Marduk and the monster Tiamat; the Hittites told of a battle between the weather god and the dragon Illuyankas; while from Ras Shamra (ancient Ugarit), in north Syria, has come a Canaanite poem relating the discomfiture of Sir Sea by the deity Baal and the rout of an opponent named Leviathan. (Originally, this myth probably referred to the annual subjugation of the floods.)

Ancient myths are utilized also in the form of passing allusions or poetic "conceits," much as modern Westerners may speak of Cupid or the Muses. Thus, there are references in the prophetic books to a celestial upstart hurled to Earth on account of his brashness and to the imprisonment of certain rebellious constellations.

The prophets used such myths paradigmatically to illustrate the hand of God in contemporary events or to reinforce their forecasts. Thus, to Isaiah the primeval dragon becomes the symbol of that continuous force of chaos and evil that will again have to be vanquished before the Kingdom of God can be established on Earth. Similarly, for Ezekiel the celestial upstart serves as the prototype of the prince of Tyre, destined for an imminent fall; and Habakkuk sees in the impending rout of certain invaders a repetition on the stage of history of Yahweh's mythical sortie against the monster of the sea.

Legends and other tales. Old Testament legends often embellish the accounts of national heroes with standard motifs drawn from popular lore. Thus, the story (in Genesis) of Joseph and Potiphar's wife recurs substantially (with other characters) in an Egyptian papyrus of the 13th century BCE (before the Common Era, or BC). The depositing of the infant Moses in the bulrushes (in Exodus) has an earlier counterpart in a Babylonian tale about Sargon, king of Akkad (c. 2334–c. 2279 BCE), and is paralleled later in legends associated with the Persian Cyrus and with Tu-Kiieh, the fabled founder of the Turkish nation. Jephthah's rash vow (in Judges) whereby he is committed to sacrifice his daughter recalls the classical legend of Idomeneus of Crete, who had similarly to slay his own son. The motif of the letter whereby David engineers the death in battle of Bathsheba's husband recurs in Homer's story of Bellerophon and again in the episode of Rosencrantz and Guildenstern in *Hamlet*. The celebrated judgment of Solomon concerning the child claimed by two contending women is told, albeit with variations of detail, about Buddha, Confucius, and other Oriental sages; while the story of how Jonah was swallowed by a "great fish but subsequently disgorged intact finds a parallel in the Indian tale of the hero Saktideva, who experienced the same thing during his quest for the Golden City. On the other hand, it should be observed that many of the parallels commonly cited from the folklore of primitive peoples may be, in fact, mere playbacks of biblical material picked up from Christian missionaries.

Etiologic tales

Sometimes, worldwide folktales serve in the Old Testament to account for the names of places in Palestine or for the origins of traditional customs and institutions. Thus, the familiar story of the man who has to struggle with the personified current of a river before he can cross it is localized (in Genesis) at the ford of Jabbok simply because that name suggests the Hebrew word abk ("struggle"); and Samson's felling of 1,000 Philistines with the jawbone of an ass is placed at Ramath-lehi because *leḥi* is Hebrew for "jawbone." Similarly, a taboo against eating the sciatic nerve of an animal is validated (in Genesis) by the legend that Jacob was struck in the hip when he tussled with an otherworldly being at Penuel (Face of God); and the custom of annually bewailing the vanished spirit of fertility is rationalized (in Judges) as a lamentation for the hapless daughter of Jephthah.

Besides myths and legends the Old Testament also contains a few examples of fables (didactic tales in which animals or plants play human roles). Thus, the serpent in Eden talks to Eve, and Balaam's ass not only speaks but also "flairs" spirits; while in the celebrated parable of Jotham (in Judges) trees compete for kingship.

Finally, in the Book of Job (38:31) there are allusions to star myths concerning the binding of Orion (called the Fool) and the "chaining" of the Pleiades.

Present-day interpretations. The tendency to interpret biblical tales and legends as authentic historical records or as allegories, or as the relics of solar, lunar, and astral myths, is now a thing of the past. For the modern folklorist, their primary interest lies in the fact that they push back to remote antiquity several tales and motifs long known from later literature. For the theologian, however, they pose the deeper problem of distinguishing clearly between the permanent message of Scripture and the particular form in which it is conveyed. Such a process of "demythologization" is today one of the central concerns of religious thought. It involves recognition of the fact that the natural language of religious truth is myth so that the continuing relevance of ancient scriptures depends not on a total rejection of that vehicle but rather on a constant expansion and remodelling of it—*i.e.*, on remythologization rather than demythologization. In the final analysis, the traditional portrayal of God himself is simply a mythical representation of ultimate reality, but that reality transcends the particular images in which it happens to be expressed. At the same time, it must be clearly understood that expressions that can be reconciled with modern Western patterns of thought only if taken as metaphors were literal statements of fact to ancient and primitive peoples. Gods, for example, were not merely "personifications" of natural phenomena but rather the effective potencies of the phenomena themselves conceived from the start as personal beings, much as a modern child might conceive of a railroad engine as "Mr. Choochoo."

Angelology, demonology, and eschatology

Myth and legend in the Persian period. When, in 539 BCE, the Jews came under Persian domination, they absorbed a good deal of Iranian folklore about spirits and demons, the eventual dissolution of the world in a fiery ordeal, and its eventual renewal. This introduced a new element into Jewish myth and legend. Hierarchies of angels, archangels such as Michael, Gabriel, and Uriel (modelled loosely upon the Iranian *amesha spentas*), and the demonic figures of Satan, Belial, and Asmodeus (corresponding to the Iranian Angra Mainyu [Ahriman], Druj, and Aēshma daeva) now entered their popular mythology, and there was a preoccupation with apocalyptic visions of heaven and hell and of the Last Days. Unfortunately, no Jewish texts of this genre from the Persian period itself are extant so that these new elements can be recognized only inferentially from their survival in later times, notably in such products of the ensuing Hellenistic age as the Dead Sea Scrolls.

The principal monument of Jewish story in the Persian period is the biblical Book of Esther, and this is basically the Judaized version of a Persian novella about the shrewdness of harem queens. The story was adapted to account for a popular festival named Purim, but this is probably a transmogrification of the Persian New Year. Such leading elements of the tale as the parade of Mordecai through the streets dressed in royal robes, the fight between the Jews and their adversaries, and the hanging of Haman and his sons seem, indeed, to reflect customs associated with that occasion, viz., the ceremonial ride of a common citizen through the capital, the mock combat between two teams representing Old Year and New Year, and the execution of the Old Year in effigy.

BY courtesy of the Offentliche Kunstsammlung Basel, Switzerland

"Esther and Ahasuerus," tempera painting by Konrad Witz (c. 1400-c. 1445). In the Offentliche Kunstsammlung Basel. 85.5 X 79.5 cm.

Myth and legend in the Hellenistic period. Historiatett Bibles and legendary histories. When, in 330 BCE, Alexander the Great completed his conquest of the Near East, Judaism entered a new phase. The dominant features of the ensuing Hellenistic age were an increasing cosmopolitanism and a fusion of Oriental and Greek cultures. These found expression in Jewish myth and legend in the composition (in Greek) of stories designed to link the Bible with general history, to correlate biblical and Greek legends, and to claim for the Hebrew patriarchs a major role in the development of the arts and sciences. It was asserted, for instance, that Abraham had taught astrology to the king of Egypt; that his and Keturah's sons had aided Heracles against the giant Antaeus; and that Moses, blithely identified both with the semi-mythical Greek poet Musaeus and with the Egyptian Thoth, had been the teacher of Orpheus (putative founder of one of the then current "mystery cults") and the inventor of navigation, architecture, and the hieroglyphic script. Leading writers in this vein were Artapanus, Eupolemus, and Cleodemus (all c. 100 BCE), but their works are known to us only from stray quotations by Eusebius and Clement of Alexandria, early Church Fathers.

Novelistic versions of biblical figures

Furthermore, the Jews followed a current Greek literary fashion of retelling Homeric and other ancient legends in "modernized," novelistic versions, well seasoned with romantic elaborations. Among the Dead Sea Scrolls has been found a paraphrase of Genesis in which the biblical narrative is tricked out with several familiar folklore motifs. Thus, when Noah is born, the house is filled with light, just as it is said elsewhere to have been at the birth of the Roman king Servius Tullius, of Buddha, and (later) of several Christian saints. When Abraham's life is threatened he dreams of a cedar about to be felled—the same omen said to have presaged the deaths of Domitian and Severus Alexander. (True, the parallels are of later date, but they illustrate the persistence of age-old popular traditions.) The same trend toward fanciful elaboration of scriptural tales is manifested also in the *Testaments of the Twelve Patriarchs* ("testaments" meaning last wills), in which the virtues and weaknesses of the sons of Jacob are illustrated by moralistic legends. There is also a lengthy paraphrase of early biblical narratives, mistakenly attributed to Philo, the famous Alexandrian Jewish philosopher of the first century CE (Common Era, or AD).

Apocrypha and Pseudepigrapha. The principal monuments of Jewish literature during the Hellenistic period are the works known collectively as the Apocrypha and Pseudepigrapha. The former are certain later writings excluded by Jews from the canon of the Old Testament but found in the Greek Septuagint version. The latter are other late writings not included in any authorized version of the Scriptures and spuriously attributed to biblical personalities.

Judaized versions of foreign tales

The Apocrypha include several Judaized versions of tales well represented in other cultures. The book of Tobit, for instance, turns largely on the widespread motifs of "The Grateful Dead" and the "Demon in the Bridal Chamber." The former relates how a traveller who gives burial to a dishonoured corpse is subsequently aided by a chance companion who turns out to be the spirit of the deceased. The latter tells how a succession of bridegrooms die on the nuptial night through the presence of a demon beside the bridal bed. Similarly, in Bel and the Dragon (2nd century BCE) occurs the equally familiar motif that fraud (in this case perpetrated in a temple) is detected by the imprint of the culprit's foot on strewn ashes—a motif that reappears later in the French and Celtic romance of Tristan and Iseult. Again, *Susanna and the Elders* (also 2nd century BCE) revolves around the well-worn theme that a charge of unchastity levelled against a beautiful woman is refuted when a clever youngster ("Daniel come to judgment") points out discrepancies in the testimony of her accusers. The story has a close parallel in a Samaritan tale about the daughter of a high priest in the 1st century CE; while the motif of the clever youngster who surpasses seasoned judges recurs later in infancy gospels and in the tale of ʿAlī Khamājah in *The Thousand and One Nights.*

The most interesting folktale in the Pseudepigrapha is that contained in the *Martyrdom of Isaiah* (1st century CE?), which tells how the prophet, fleeing from King Manasseh, hid in a tree that opened miraculously and how he eventually perished when it was sawn asunder. A similar tale is related in the Talmud about a certain Isaac ben Joseph and (later) in the Persian epic *Shāh-nāmeh* (c. 1000 CE) about the hero Jamshid.

Myth and legend in Talmud and Midrash. *Midrash and Haggada.* Toward the end of the 1st century CE, through a process known as "canonization," certain traditional Hebrew writings came to be recognized as an authoritative corpus of divine revelation, later called the Hebrew Bible or Old Testament. The study of them became, henceforth, an essential element of the Jewish religion. This meant that the sacred text had to be subjected to a form of interpretation that would bring out its universal significance and permanent relevance. The process was known as Midrash (literally "searching the Scriptures"), and a leading constituent of it was the spicing of homiletic discourses with elaborative legends—a pedagogic device called Haggada ("storytelling"). Originally transmitted orally, the legends were eventually committed to writing in that vast sea of literature known as the Talmud (the authoritative compendium of early rabbinic law and lore), as well as in later compilations geared to particular books or sections of the Old Testament, to scriptural lessons read in the services of the synagogue, or to specific biblical characters or moral themes (see also TALMUD AND MIDRASH).

The range of Haggada is virtually inexhaustible; a few representative examples must suffice. In regard to biblical characters, both Moses and David were born circumcised; Cain had a twin sister; Abraham will sit at the gate of hell to reproach the damned on Judgment Day; Aaron once locked the angel of death in the tabernacle; Solomon understood the language of animals; King Hiram, who supplied materials for the Temple, entered paradise alive; the flesh of Leviathan will feed the righteous in the world to come.

Indebtedness of Haggada to non-Jewish sources

In such fanciful elaborations of Scriptures, Haggada does not disdain to draw on classical tales. The men of Sodom, it is said, subjected itinerant strangers to the ordeal of Procrustes' bed; the Earth opened to rescue newborn Hebrew males from the Pharaoh, as it did for Amphiaraus, the prophet of Argos, when he fled from Periclymenus after the attack on Thebes; Moses spoke at birth, as did Apollo; Solomon's ring, cast into the river, was retrieved from a fish that had swallowed it, as was that of Polycrates, the tyrant of Samos, in the story told by Herodotus; the Queen of Sheba had the feet of an ass, like the child-stealing witch (Onoskelis) of Greek folklore; no rain ever fell on the altar at Jerusalem, just as none was said to have fallen on Mt. Olympus.

Other familiar motifs also appear. Moses qualifies as a husband for Zipporah by alone being able to pluck a rod from Jethro's garden—a variant of the tale told later about the sword Excalibur in the Arthurian legend; David's harp is played at night by the wind, like that of Aeolus; Isaiah, like Achilles and Siegfried, has only one vulnerable spot in his body—his mouth; Job has a magic belt, which relieves his pains.

Legends are developed also from fanciful interpretations of scriptural verses. Thus, Adam is said to have fallen only a few hours after his creation because the Hebrew text of Ps. 49:12 can be literally rendered "Adam does not last the night in glory." Lamech slays the wandering Cain—a fanciful interpretation of his boast in Gen. 4:23–24. Melchizedek is immortal in view of Ps. 110:4: "You are a priest for ever after the order of Melchizedek." The first man is a hermaphrodite (this notion has analogues elsewhere) because Gen. 1:27 says of God's creation, "Male and female he created them."

Fables and animal stories. Midrash also uses fables paralleled in non-Jewish sources. Aesop's fable of the "Lion and the Crane" is quoted by a rabbi of the 1st century CE, and the tales of the "Fox in the Vineyard and of the "Camel Who Got Slit Ears for Wanting Horns" likewise make their appearance.

Sometimes, too, material is drawn from medieval bestiaries (manuals on animals, real or imaginary, with symbolic or moralistic interpretations). Bears, we are told, lack mother's milk; hares and hyenas can change sex; only one pair of unicorns exists at a time; there is a gigantic bird (*ziz*) that reaches from Earth to sky.

Contribution of Haggada to Christian and Islāmic legends. Several of the stories related in Haggadic literature were later adopted and adapted by Christian writers. Thus, the legend that Adam was created out of virgin soil was taken to prefigure the fact that the second Adam (*i.e.*, Jesus) was likewise born of a virgin; while the story that the soil in question was taken from the site of the future Temple was transformed into the claim that Adam had been molded out of the dust of Calvary. Similarly, the legend that, at the dedication of the Temple, the doors had swung open automatically to admit the ark of the Covenant was transferred to the consecration of a church by St. Basil; and the Talmudic tale that the bronze Nicanor gates of the Temple had floated to Jerusalem when cast overboard for ballast during their shipment from Alexandria was applied to the doors of a sacred edifice erected in honour of St. Giles.

Nor was it only the Christians who absorbed Haggadic legends. The Qur'ān, the sacred book of Islām, likewise incorporates a good deal of such material in its treatment of such biblical characters as Joseph, Moses, David, and Solomon.

Myth and legend in the medieval period. *Jewish contribution to diffusion of folktales.* The Middle Ages was a singularly productive period in the history of Jewish myth and legend. Jews now began to play a prominent role in the transmission of Oriental tales to the West and thereby enhanced their own repertoire with a goodly amount of secular material. Especially in Spain and Italy, Arabic versions of standard collections were translated into Hebrew and thence into Latin, thus spreading the stories to the Christian world. The Indic fables of Bidpai, for example, were rendered into Hebrew from the 8th-century Arabic version of 'Abd Allāh ibn al-Muqaffaʿ, and from this Hebrew rendering there subsequently developed, in the 12th century, John of Capua's *Directorium humanae vitae* ("Guide for Human Life"), one of the most celebrated repertoires of moralistic tales (*exempla*) used by Christian preachers. So, too, the famous *Senbād-nāmeh* ("Fables of Sinbad"; one of the sources, incidentally, of Boccaccio's *Decameron*) was rendered from Arabic into Hebrew and thence into Latin; while the renowned romance of *Barlaam and Josaphat*—itself a Christian adaptation of tales about the Buddha—found its Jewish counterpart in a compilation entitled *The Prince and the Dervish,* adapted, from an Arabic text, by Abraham ben Samuel ibn Hisdai, a leader of Spanish Jewry in the 13th century.

The *Alexander Romance,* Arthurian legend, and Aesop's fables

Hebrew versions of medieval romances. Here, too, however, the traffic moved in both directions: Hebrew translations were also made from Latin and other European languages. There are, for instance, several Hebrew adaptations of the *Alexander Romance,* based mainly (though not exclusively) on Leo of Naples' Latin rendering of the Greek original by Callisthenes. The central theme is, of course, the exploits of the great Macedonian conqueror, and the narrative is spiced with fanciful accounts of his adventures in foreign lands and of the outlandish peoples he encounters. There is likewise a Hebrew reworking of the Arthurian legend, in the form of a secular sermon in which Arthurian and biblical scenes are blithely mixed together. Finally, there is a Hebrew *Ysopet* (the common title for a medieval version of Aesop) that shares several of its fables with the famous collection made by Marie de France in the late 12th century.

Jewish contributions to Christian and Islāmic tales. Moreover, apart from these Hebrew translations of Oriental and European works, a good deal of earlier haggadic material is embodied in the *Disciplina clericalis* of Peter Alfonsi, a baptized Jew of Aragon originally known as Moses Sephardi. This book, composed in the 12th century, is the oldest European collection of novellas and served as a primary source for the celebrated *Gesta Romanorum* ("Deeds of the Romans") of the same period—a major quarry for European storytellers, poets, and dramatists for many centuries.

Haggadic material percolated also to Arabic writers during this period. Not only does the Qur'ān incorporate such material but also the Egyptian recension of *The Thousand and One Nights* seems to have drawn extensively on Jewish sources, as, for instance, in its tales of "The Sultan and His Three Sons," "The Angel of Death," "Alexander and the Pious Man," and the legend of Baliqiyah.

Major medieval Hebrew collections. Between the 11th and 13th century the tendency developed in Europe to compile, both for entertainment and edification, comprehensive collections of tales and fables; standard examples are the British *Gesta Romanorum,* the Spanish *El novellino,* and the aforementioned *Disciplina clericalis.* Among Jews similar collections were made, especially in Morocco as well as in Moorish Spain. Two of the most important are *The Book of Comfort* by Nissim ben Jacob ben Nissim of al-Qayrawān (11th century) and *The Book of Delight* by Joseph ben Meir ibn Zabara of Spain (end of the 12th century). The former, composed in Judeo-Arabic, is a collection of some 60 moralizing tales designed to comfort the author's father-in-law on the loss of a son. It belongs to a well-known genre of Arabic literature, derives mainly from Arabic sources, and is permeated by a preoccupation with divine justice, typical of the Mu'tazilite school of Islamic theology. It was later translated into Hebrew. *The Book of Delight* consists of 15 tales, largely about the wiles of women, exchanged between two travelling companions—a form of cadre, or "enclosing tale," adopted on a more extensive scale by Chaucer in his *Canterbury Tales,* which dates from the same period. Typical is the tale of the "Silversmith and His Wife," which relates how a craftsman, persuaded by his greedy wife to make a statue of a princess, gets his hands cut off by the king for violating the Islamic law against making images, while his wife reaps rich rewards from the flattered princess. Although most of the stories are taken from Arabic sources, some indeed find parallels in rabbinic literature. To the latter category belongs, for instance, the famous tale of the matron of Ephesus, who, while keeping vigil over her husband's tomb, at the same time engages in an intrigue with a guard posted nearby to watch over the corpses of certain crucified robbers. When, during one of their trysts, one of the corpses is stolen and her lover therefore faces punishment, the shrewd woman exhumes the body of her husband and substitutes it. This tale is found already in the *Satyricon* of Petronius and was later used by Voltaire in his *Zadig* and by the 20th-century English playwright Christopher Fry in his A *Phoenix Too Frequent.*

The story of the matron of Ephesus

Of the same genre but deriving mainly from west European rather than Arabic sources are the *Mishle shuʿalim* ("Fox Fables") of Berechiah ha-Nakdan (the Punctuator), who may have lived in England toward the end of the 12th century. About half of these tales recur in Marie de France's *Ysopet,* and only one of them is of specifically Jewish origin. Berechiah's work was translated into Latin and thence became a favourite repertoire of European storytellers.

Among anonymous compendiums of this type is *The Alphabet of Ben Sira,* extant in two recensions, probably of the 11th century. This is basically a collection of proverbs attributed to the famous sage of the apocryphal book Ecclesiasticus (Wisdom of Jesus the Son of Sirach). In one of the recensions they are illustrated by appropriate tales. The author is represented as an infant prodigy who performs much the same feats of sapience as are attributed to Jesus in some of the Infancy Gospels.

The *Josippon, Jashar,* and *Jerahmeel* collections

Medieval historiated Bibles and legendary histories. Two other developments mark the history of Jewish myth and legend during the Middle Ages. The first was a revival of the Hellenistic vogue of compiling large-scale compendiums in which the history of the Jews was "integrated," in legendary fashion, with that of the world in general and especially with classical traditions. Two major works of this kind, both composed (apparently) in Italy during the 9th century, are (1) *Josippon,* composed by a certain Ben Gorion, which presents a fanciful record from the creation onward and contains numerous references to foreign nations; and (2) the *Book of Jashar,* a colourful account from Adam to Joshua, named for the ancient book of heroic songs and sagas mentioned in the Bible (Josh. 10:13; II Sam. 1:18). There is also a voluminous *Chronicles of Jerahmeel,* written in the Rhineland in the 14th century. This draws largely on Pseudo-Philo's earlier compilation, mentioned above. and is of special interest because it includes Hebrew and Aramaic versions of certain books of the Apocrypha.

Medieval Haggadic compendiums. The other development was the gathering of Haggadic legends and tales into comprehensive, systematic compendiums. Works of this kind are (1) *Yalquṭ Shimeʿoni* ("The Collection of Simeon"), attributed to a certain Rabbi Simeon of Frankfurt am Main; (2) *Midrash ha-gad01* ("The Great Midrash"), composed after the death of Moses Maimonides (1204), whom it quotes; and (3) the *Midrash of David ha-Nagid,* grandson of Maimonides. About 100 years later appeared a similar work, *Yalquṭ ha-Makiri* ("The Collection of Makhir"), on the Prophets and Holy Writings, compiled by one Makhir ben Abba Mari in Spain (see TALMUD AND MIDRASH). It has been suggested that the compilation of such works was spurred by the necessity of providing "ammunition" for the public disputations with Christian ecclesiastics that the church forced upon Jewish scholars in this period.

Myth and legend in the modern period. *Kabbalistic tales.* In the 16th century, Jewish myth and legend took several new directions. The disappointment of messianic expectations through the dismal eclipse of the pretender Shabbetai Tzevi produced, by way of compensation, an Increased interest in occult speculation and in the mystical lore of the Kabbala (esoteric Jewish mysticism). Important schools of Kabbala arose in Italy and at Safed, in Palestine, and tales of the miraculous Faust-like powers of such masters as Isaac Luria and Hayyim Vital Calabrese began to circulate freely after their deaths.

The legend of Joseph della Reyna

Another reaction to the dashing of messianic hopes is represented by the beautiful story of the Kabbalist Joseph della Reyna and his five disciples, who go journeying through the world to oust Satan and prepare the way for the Deliverer. Warned by the spirits of such worthies as Rabbi Simeon ben Yohai and the prophet Elijah, they nevertheless succeed eventually in procuring their blessing and help and are sent on to the angel Metatron. The latter furnishes them with protective spells and spices and advises Joseph to inscribe the ineffable name of God on a metal plate. When, however, they reach the end of their journey Satan and his wife, Lilith, attack them in the form of huge dogs. When the dogs are subdued they beg for food. Moved to pity, Joseph gives them spices to revive them. At once they summon a host of devils. Two of the disciples die of terror; two go mad, and only Joseph and one disciple are left. The Messiah weeps in heaven, and Elijah hides the great horn of salvation. A voice rings out telling Joseph that it is vain to attempt to hasten the footsteps of the Redeemer.

The repertory of Jewish tales and legends was seasoned, however, by other elements. During the 16th century—the age of the great navigators—stories began to circulate about the discovery of the Ten Lost Tribes in remote parts of the world.

Judeo-German (Yiddish) tales. It was at the same period that Judeo-German (Yiddish) came increasingly to replace Hebrew as the language of Jewish tales and legends in Europe, a major factor in this development being the desire to render them accessible to women unschooled in the sacred tongue. Not only were the synagogal lessons from Scripture legendarily embellished in a so-called *Taitsh Humesh* ("Yiddish Pentateuch"), in the more fancifully titled *Tze'ena u-re'ena* ("Go Forth and See"; *cf.* S. of Sol. 3:11) by Jacob ben Isaac Ashkenazi, and in adaptations of the story of Esther designed for dramatic presentation on the feast of Purim, but the Hebrew *Chronicles of Josippon* also assumed Yiddish dress. More secular productions were a verse rendition of the Arthurian legend, entitled *Artus Hof* ("The Court of King Arthur"), based largely on Gravenberg's medieval *Wigaleis,* and the *Bove Buch* by Elijah Levita, which retold the romance of Sir Bevis of Southampton.

These "frivolous" productions were in time offset by collections of moral and ethical tales. The principal of these are (1) the *Brantspiegel,* attributed to a certain Moses Henoch (Prague 1572), and (2) the *Maʿaseh Buch* ("Story Book"), a compendium of 254 tales compiled by Jacob ben Abraham of Meseritz and first published at Basel in 1602. The latter was drawn mainly from the Talmud but was supplemented by later legends about medieval rabbis. Jewish legends also circulated in the form of ephemeral chapbooks, a large selection of which is preserved in the library of the Yiddish Scientific Institute in New York City.

Judeo-Persian and Judeo-Spanish (Ladino) tales. A similar development, though on a lesser scale, took place among Jews who spoke other vernacular dialects. Major monuments of Judeo-Persian literature are poetic embellishments of biblical narratives composed by a certain Shāhīn of Shīrāz in the 14th century and by Joseph ben Isaac Yahudi (*i.e.,* the Jew) some 300 years later. These. however, are exercises in virtuosity rather than in creative storytelling. In Judeo-Spanish (Ladino) there are versified elaborations of the story of Joseph, entitled *Coplas de Yoçef* ("Song of Joseph"), composed, in 1732, by Abraham de Toledo and embodying a certain amount of traditional haggadic material. From a revival of literary activity in the 18th century comes a comprehensive "legendary Bible" called Me-'am Lo'ez ("From a People of Strange Tongue"; *cf.* Ps. 114:1), begun by one Jacob Culi and continued by later writers, as well as several renderings of standard Hebrew collections and a number of Purim plays. Until the Nazi holocaust in the 1940s, Judeo-Spanish folktales were still current in Macedonia and Yugoslavia, but these leaned more on Balkan than on Jewish sources.

Hasidic tales. The rise of the Hasidic sect (a popular pietistic-mystical movement) in eastern Europe at the end of the 18th century begat a host of legends (circulated mainly through chapbooks) concerning the lives, wise sayings, and miracles of such *tzaddiqim,* or masters, as Israel ben Eliezer, "the Besht" (1700–60), and Dov Baer of Meseritz (died 1772) (see also JEWISH MYSTICISM). These, however, are anecdotes rather than formally structured stories and often borrow from non-Jewish sources.

Droll stories. To the popular creativity of the ghetto belong also the droll tales of the wise men of Chelm (in Poland)—Jewish counterparts of the German noodles (stupid people; hence "noodle stories") of Schildburg and of the more familiar English Wise Men of Gotham. These, too, were circulated mainly in Yiddish popular prints. Typical of them is the tale of the two "sages" who went for a walk, one carrying an umbrella and the other without one. Suddenly it began to rain. "Open your umbrella," said the one without one. "It won't help," answered the other, "it's full of holes." "Then why did you bring it?" rejoined his friend. "I didn't think it would rain," was the reply.

Modern Israeli folktales. The ingathering of Jews from many lands into the modern state of Israel has made that country a happy hunting ground for the student of Jewish folktales. Assiduous work in the collection of these has been undertaken by Dov Noy of Hebrew University in Jerusalem, aided by enthusiastic amateurs throughout the country. Mainly, however, the stories are mere retellings of traditional material.

BIBLIOGRAPHY. The following works provide a general coverage of the field for English-speaking readers.

Old Testament: T.H. GASTER, *Myth, Legend, and Custom in the Old Testament* (1969).

Apocrypha and Pseudepigrapha: R.H. CHARLES (ed.), *The Apocrypha and Pseudepigrapha of the Old Testament in English,* 2 vol. (1913, reprinted 1964); M.R. JAMES, *The Lost Apocrypha of the Old Testament, Their Titles and Fragments* (1920).

Hellenistic: W.N. STEARNS (ed.), *Fragments from Graeco-Jewish Writers* (1908); C.C. TORREY (ed. and trans.), *Lives of the Prophets* (1946); T.H. GASTER (trans.), *The Dead Sea Scriptures in English Translation,* pp. 256–267, 2nd ed., (1964).

Talmud and Midrash: The Legends are digested and annotated in LOUIS GINZBERG'S classic, *Legends of the Jews,* 7 vol. (1909–39), also available in a one-volume abridgment (1961). The *Midrash Rabbah* has been translated and edited by HARRY FREEDMAN and MAURICE SIMON, 10 vol. (1939); and

the *Midrash on Psalms,* by W.G. BRAUDE, 2 vol. (1959). Principal editions and translators are listed in H.L. STRACK, *Einleitung in Talmud und Midraš*, 5th ed. (1921; Eng. trans., *Introduction to the Talmud and Midrash*, 1931).

Medieval: The Book of Jashar, trans. by M.M. NOAH (1840); M.R. JAMES (trans.), *The Biblical Antiquities of Philo* (1917); MOSES GASTER (trans.), *The Chronicles of Jerahmeel* (1899); I.J. KAZIS (ed. and trans.), *The Book of the Gests of Alexander of Macedon* (in Hebrew) (1962); CURT LEVIANT (ed. and trans.), *King Artus* (1969); MOSES HADAS (trans.), *The Book of Delight* (1932). On the diffusion of medieval Jewish tales, see J. JACOBS, *Jewish Ideals, and Other Essays,* pp. 135–161 (1896); LOUIS GINZBERG, "Jewish Folklore: East and West," *Independence, Convergence, and Borrowing in Institutions, Thought, and Art,* Harvard Tercentenary Conference of Arts and Sciences, pp. 89–108 (1937).

Judeo-German (Yiddish): MOSES GASTER, *Ma'aseh Book,* 2 vol. (1934); N.C. GORE (ed. and trans.), *Tzeenah u-Reenah: A Jewish Commentary on the Book of Exodus* (1965).

Judeo-Spanish (Ladino): IGNACIO GONZALEZ LLUBERA (ed.), *Coplas de Yocef* (1935); C. CREWS, "Judaeo-Spanish Folktales in Macedonia," *Folk-Lore,* 43:193–225 (1932).

Hasidic: MARTIN BUBER (ed.), *Tales of the Hasidim,* 2 vol. (1947–48).

Israeli: DOV NOY (ed.), *Folktales of Israel* (1963); ZEV VILNAY, *Legends of Palestine* (1932; orig. pub. in Hebrew, 1929).

Comparative Studies: MOSES GASTER, *The Exempla of the Rabbis* (1924, reprinted 1968). A useful motif index of Haggadic legends to accompany STITH THOMPSON'S standard *Motif Index of Folk-Literature,* 6 vol., rev. and enl. ed. (1955–58), has been compiled by DOV NOY, *Motif Index of the Talmudic-Midrashic Literature* (1954), and is available on microfilm. For a convenient bibliographical survey of the entire field, see ABRAHAM BERGER, "The Literature of Jewish Folklore," *Journal of Jewish Bibliography,* 1:12–20, 40–49 (1938–39).

(T.H.G.)

Jewish Peoples, Arts of

There is no agreement on the definition of a Jew, much less Jewish art. One would be hard put to define in terms of its "Jewishness" the art, for example, of Gustav Mahler, Marcel Proust, or Amedeo Modigliani; yet they were all Jews. Since Jewish art is not satisfactorily defined simply as art produced by people who happen to be Jewish, and since no definition can subsume the diverse works of such artists as Proust, who was an assimilated Jew, and Sholem Aleichem, who was not, the definition assumed by this article is a limited one: art that is indisputably part of Jewish tradition, reflecting Jewish religion and Jewish customs, and literature written in Hebrew and Yiddish. Jewish artists, such as Mahler, Proust, and Modigliani, who have made major contributions to and are difficult to extricate from non-Jewish traditions, are treated more fully in other articles.

Literature

Literature in the Hebrew language has been produced continuously from the early 12th century BC. If, as is believed, the Canaanite population living in Palestine at the time of the Israelite conquest spoke in part an earlier form of Hebrew, then there is evidence of a Hebrew literature existing as early as the 14th century BC. During part of this literary history (c. 1200 BC–AD 200), Hebrew was a spoken language in Palestine, first in the dialect called biblical Hebrew and later in the form of Mishnaic Hebrew. During a second period (AD 200–1880), Hebrew was a literary language only, the writers of which spoke, and sometimes also wrote concurrently, a variety of other languages. Hebrew literary centres existed throughout the Mediterranean area and consecutively in northwestern Europe, central and eastern Europe, and, since mid-19th century, in America. The last period (1880 onward) coincides with the revival of Hebrew as a spoken language and saw the gradual shift of its centre to Palestine.

Hebrew literature is not synonymous with Jewish literature. Some Hebrew writing was produced by the Samaritans and in the 17th century by Protestant enthusiasts. Jews also produced important literatures in Greek, Aramaic, Arabic, Judeo-Spanish, Yiddish, and modern European languages. (Ed.)

Because Hebrew literature has had a great impact on Western literary tradition as a whole, it cannot easily be isolated from that tradition and is therefore discussed in the article LITERATURE, WESTERN. Early Hebrew literature, inextricably bound up with religion, is most extensively covered in BIBLICAL LITERATURE; JEWISH MYSTICISM; TALMUD AND MIDRASH; EXEGESIS AND HERMENEUTICS, BIBLICAL; and JEWISH MYTH AND LEGEND.

Origins of Yiddish literature

Yiddish is the main colloquial language of the majority of Ashkenazic (central and east European) Jews and one of their two literary languages (the other is Hebrew). Literature written in the Yiddish language began with the need of the Jews living in the Germanic Frankish lands of western Europe during the Middle Ages for a literature in their own vernacular: most of them no longer had enough knowledge of Hebrew, which had become the language of prayer and scholarship, to understand fully new works written in it. (Popular literature in Hebrew, intertwined with Yiddish traditions, developed much later.) Living among people speaking Middle High German, the Jews had adopted that language, but they continued to need a religious usage terminology, which had to be Hebrew. Thus there evolved an amalgam of the two languages called Jiidisch-Deutsch (Judeo-German), soon corrupted to Yiddish-Teitch; and finally called simply Yiddish. For a time there was little difference between Yiddish-Teitch and Middle High German, except for the infusion of Hebrew words; but eventually (possibly as early as the 9th century) Yiddish-Teitch became a separate language. With the Jewish migrations to eastern Europe after the massacres during the Crusades, the language acquired a strong Slavonic element and, cut off from its German source, developed on independent lines, flowering into a rich literary language. An oral and manuscript Yiddish literature first appeared in the 12th and 13th centuries, a printed literature in the 16th century.

Part of the cultural and spiritual transition of many Jews from the 18th-century ghetto to the modern 19th-century Western world was the adoption of modern European languages as literary vehicles. It is questionable whether—and in what ways—such literature can be characterized as Jewish.

This section of the article discusses literature written by Jews in Yiddish and, more briefly, in modern European languages.

IN YIDDISH

Early history. Yiddish, like many other languages, was for a long time looked upon as a vulgar tongue, good enough for the spoken but not the written word. The language of the scholars was Hebrew. When the first Yiddish books were written in the 12th and 13th centuries, they were intended only for women and the ignorant. These works were of two types: religious and liturgical—translations from the Bible, prayer books, and religious poems and secular writing, consisting mainly of adaptations from the German and Italian—the literature of the so-called *Spielmann* period, named after the German minstrels of the medieval period. Jewish folk singers and jesters sang or performed these texts at various entertainments. Most of them were verse romances, tales of knights and their ladies. Some even had erotic connotations. One of the most famous of these works, the *Bove-Buch*, was written in 1507 (and printed in 1541) by Elijah Bachur, more popularly known as Elijah Levita, a German Jew who emigrated to Italy in 1496, a time of trouble for Jews in his native country. Written in ottava rima (an eight-line stanza with a rhyme scheme of ABABABCC), it is a heroic tale adapted from an Italian romance (*Buovo d'Antona*) for its Jewish public. Another work by the same author is *Paris un Viene,* published posthumously in 1595. Both books are rhymed romances, written in the spirit of the Renaissance. They are at the same time fantastic, realistic, and highly dramatic. Another, earlier work, the *Shmuel Buch* (1544; "Samuel Book"), by an unknown writer, is a retelling of the Book of Samuel, with King David possessing all the characteristics of a medieval knight. Rabbis and pious Jews con-

Literature of the *Spielmann* period

demned these works as frivolous; but they tolerated them because they recognized that the Jewish woman of the day, who was as often as not breadwinner as well as wife and mother, needed some diversion.

As a result of the catastrophes and expulsions that the German Jews experienced in the wake of the Thirty Years' War (1618–48) in Germany and their subsequent emigration in large numbers to Poland, secular Yiddish literature disappeared for a long time. It did not appear again until the second half of the 18th century. The writings of the interim period were almost all of a religious, didactic nature. The most famous of them is the *Tzeno Ureno,* written by Jacob ben Isaac Ashkenazi of Janov, Poland. It appeared in Prague in 1608. It was a paraphrase of the Pentateuch (first five books of the Old Testament), with the addition of legends and commentaries from the Talmud (authoritative body of Jewish law and tradition), exegeses of Rashi (11th-century scholar and commentator), the Midrash (mystical commentary on the Pentateuch), and other holy books. The language used is much closer to that of modern Yiddish than to that of the earlier, 16th-century works. As late as the beginning of the 20th century it was still customary for women to read this book on the sabbath.

By the second half of the 18th century, Yiddish had become exclusively the language of eastern European Jews. With the coming of the Hasidic movement (a revolt, based on mystical enthusiasm, against the aridity of rabbinic intellectualism), a number of Hasidic storybooks began to appear in Yiddish. The greatest of them are the stories of Rabbi Nahman of Bratslav (then Russian). He was born in the south of Russia and was a great-grandson of Ba'al Shem Tov, the founder of Ḥasidism. He did not write his stories himself but dictated them to his disciple Rabbi Nathan Nemeirover. These tales, which have achieved fame in the German translation, by Martin Buber, are masterworks of Jewish folklore.

At about the same time in Poland, the *Haskala,* or Enlightenment (a moderate tendency toward Westernization), began to influence both Jewish life and literature. Most of the writers of the Enlightenment wrote either in Hebrew or in Russian, but some condescended to write in Yiddish. For the first time, works of popular science began to appear, as well as pamphlets directed against the Hasidim, the parodies of Hasidic literature. One of the earliest writers of this period was Israel Axenfeld. His stories and plays satirized the superstitions and impracticality of the eastern European Jews and criticized their total denial of secular education. Shlomo Ettinger was less bitter than Axenfeld and a superior artist as well. His play *Serkele* and other dramatic works and his epigrammatic poems and fables appeared after his death in 1856.

Importance of Mendele in modern Yiddish literature

Beginnings of modern literature. Modern prose in Yiddish began in the 1850s with the stories of Eisik Meir Dik; the satiric novel *Dos poilische Yingel* (1869; *The Polish Boy)* by Isaac Joel Linetzky; and the novels of Sholem Yakov Abramovitsch, better known as Mendele Mokher Sforim (Mendele the Itinerent Bookseller), "the grandfather" of modern Yiddish literature. Mendele also wrote in Hebrew and is a classic in both languages. Among his best Yiddish works are *Dos Vinschfingerl* (1865; "The Wishing Ring"), *Die Klatsche* (1873; "The Mare"), and *Fishke der Krumer* (1869; "Fishke, the Lame"). Mendele, like all of these writers, deplored and mocked the unworldliness of the Russian and Polish Jews, their exaggerated piety, and their obsolete system of educating their children. He also castigated the community elders who, for their private gains, exploited the poor Jews. By Mendele's time, great numbers of Jews already lived in the large cities of Russia, western Europe, and the United States. Jewish professors held high positions in the universities of the world. The revolutionary movement had reached even the Jews in the small villages of Poland and the Ukraine. Young men and women left their homes to rebuild Palestine. There was also an increasing emigration to the United States, South America, and elsewhere. But Mendele chose to write about the poorest and the most backward citizens of Kabtzansk (Pauper Village) and Glupsk (Village of Fools). From a purely literary point of view, Mendele himself was rather backward. His achievement was social: his themes did much to modernize the Jewish masses of eastern Europe.

Sholem Yakov Rabinovitsch, or Sholem Aleichem, considered himself a disciple of Mendele and his spiritual grandson. His style resembles that of Mendele, but he was a God-inspired master of description and one of the greatest humorists in world literature, to be rated only a few steps below the Russian writer Nikolay Gogol. Sholem Aleichem described the same muddy villages as Mendele had and even gave them similar satiric names; but he saw in the Russian Jews, restricted to the Pale, a heroic group. Seemingly condemned to extinction, they used all of their intelligence, wisdom, and shrewdness in order simply to survive and to insure a better lot for their children. All of Sholem Aleichem's heroes were possessed by the passion for making a living. He is perhaps the only writer in world literature whose main theme is the struggle to make ends meet. Sholem Aleichem was often only one step from sentimentality, but his artistic instinct seldom let him take that step. His greatest works are *Tevya der Milchiger* ("Tevya, the Milkman"), *Motel Peisy dem Chasen's* ("The Cantor's Motel Peisy"), and *Funem Yarid* ("Back from the Fair"), the first volume of an unfinished autobiographical novel. His language is rich, his descriptions alive and funny. Sholem Aleichem spent the last year or so of his life in the United States; and his descriptions of the Russian Jewish immigrants in New York are genuine, pertinent, and highly humorous.

Isaac Leib Peretz, who also wrote in Hebrew, was a contemporary of Sholem Aleichem and the first important romantic writer in Yiddish literature. His best works are *Chassidish* ("Hasidic") and *Folkstimliche Geshichten* ("Folktales") and a mystical drama, *Bei Nacht oifn alten Mark* ("At Night in the Old Marketplace"). Both stylistically and thematically, he was closer to modern European literature than his predecessors. It was Peretz who first described love in the Yiddish villages and cities. He dared to take a positive position toward religion, Ḥasidism and the Kabbala (Jewish mystical and theosophical doctrine). He wrote poetry, plays, and short stories. In contrast to the garrulous writers before him, his style was taut and concise. Peretz believed that it was possible for the Jews of Russia and Poland to create their own secular culture. He may rightly be called the father of Yiddishism, the spiritual founder of the movement that demanded cultural autonomy for the Jews of Russia. In contrast to Sholem Aleichem, he was an opponent of political Zionism. What he had in common with Mendele and Sholem Aleichem was that he was never able to make a proper living out of writing. For many years he served as an employee of the same Jewish community in Warsaw that he criticized in his writings. His house was often filled with young writers, whom he encouraged and helped. Although Peretz was potentially a greater artist than his fellow classic writers, he lacked the time and perhaps the patience necessary to create large canvases. He was a short story writer, a poet, a dramatist, a journalist, a speaker, an editor, and a cultural leader. He distinguished himself in all of these capacities.

The father of Yiddishism

Among the foremost essayists and men of letters of this period was David Frischmann, who wrote in Yiddish and in Hebrew. Frischmann in his essays preached "Europeanism." He was hypnotized by the scientific and cultural achievements of central and western Europe and considered the Yiddish and Hebrew literature of his day utterly provincial. He admired and translated the German philosopher Friedrich Nietzsche and kept his readers informed about "the new winds of Europe." He died in 1922, a lonely and disillusioned man in the Germany that he had so admired before World War I and the horror that came with it.

World War I to the present. Another essayist and man of letters, Hillel Zeitlin, was the first modern Yiddish writer who, after considerable soul-searching, came to believe that the Jewish people could not endure without religion; he propounded this belief in all of his writings. He fought bitterly against the Enlightenment and its idolatrous worship of worldliness. Zeitlin was a mystic and a

Kabbalist, and he translated parts of the Zohar (mystical commentary on the Pentateuch). He was also a sharp-tongued literary critic. He attacked especially the pro-Soviet writers, with their "proletarian" literature. His analysis of the precarious Judaism of the modern Jew is still valid today and may prove even more valid in the future. He was the father of Aaron Zeitlin, perhaps the greatest of the modern Yiddish poets, who shared his father's interest in mysticism and wrote books on psychic research in both Hebrew and Yiddish.

Sholem Asch, although he looked upon himself as a disciple of Peretz, had neither Peretz' scholarship nor his sense of style. He wrote his fiction on a broad scale and was one of the first Yiddish writers to create historical novels and plays. His books were translated into many languages and his play Got *fun* Nekomeh (1907; The *God* of Vengeance, 1918) was first produced by Max Reinhardt in Berlin (1910) and enjoyed considerable success. Among his best known earlier works are: *Kiddush Hashem* (1919), Die *Kishefmacherin fun* Kastilic (1921; "The Witch of Castile"), A *Shtetl* (1904; "A Village"), Farn Mabul (A Tale of Three Cities, 1933), Motke Gonef (1917; "Motke the Thief," 1935), and Onkl Moses (1918; *Uncle* Moses, 1938), a description of Jewish immigrant life in New York. Written in his later years, a trilogy — Der Man fur *Nazares* (1939; The *Nazarene*), Der Apostol (1943; The *Apostle*), and Mary (1949) — which deals with the New Testament, was the target of much controversy among Yiddish critics and readers. With Sholem Asch, an epoch of so-called "primitivism" was introduced into Yiddish literature. The new generation of Yiddish writers was no longer steeped in Hebrew letters and Jewish lore. Among the most talented writers of this group were Abraham Reisen, Isaac Meir Weisenberg, Abraham Moshe Fuchs, Joshua Perle, Ozer Warshafsky, Moses Stavsky, Rachel Feigenberg, Ephraim Kagonovsky, Israel Rabon, Sh. Berlinsky, Fishel Bimko, Sh. Horontchik, and Hertz Bergner.

Coexisting with the "primitives" were some writers who were still nourished by the old wells of Jewish knowledge: Micah Joseph Berdichevsky, who excelled in descriptions of Hasidic life, though he spent most of his literary life in Germany; Zalman Shneur, who wrote both in Hebrew and in Yiddish and was a highly gifted follower of Mendele; Hersh David Nomberg and Onoychi (pen name of Zalman Yitzchok Aaronson), who wrote Hasidic stories; and the brothers I.J. and I.B. Singer (see below).

To the roster of the most gifted of the Yiddish poets in Europe belong Abraham Reisen, Shimon Shmuel Frug, Aaron Einhorn, Isaac Katzenelson, Uri Tzvi Grinberg, Moshe Broderzon, Samuel Jacob Imber, Aaron Zeitlin, Itzik Manger, Melech Ravitch, Israel Stern, Jechiel Lehrer, Zusman Segalovitch, Kadia Maladovsky, Miriam Ulinover, and Rachel Korn.

Yiddish literature in the United States

Yiddish literature in the United States began to blossom after World War I. The writers before that period were mostly Socialist propagandists, such as Morris Wintehevsky, David Edelstadt, and Joseph Bovshover. Among the foremost Yiddish novelists and storytellers in the United States are: Lamed Shapiro, Der *Tseilem* (The Cross); Joseph Opatoshu, who wrote In poilische *Velder* (1921; In the Forest of *Poland*); David Pinski; the brilliant short story writer Yona Rosenfeld; O.B. Berkowitch, Boruch Glassman; Leon Kobrin, M. Chaimovitch; Isaac Raboy; Shaye Miller, I. Metzger; Benjamin Ressler; Shmuel Izban. Beniamin Demblin, Yitzchok Perlov, and Chaim Grade.

A writer who achieved considerable fame outside the Yiddish literary world and who, in 1933, emigrated to the United States, was Israel Joshua Singer. His novel *The* Brothers Ashkenazi (1936), which describes the growth of the industrial city of Łódź as mirrored in the rise and fall of the Ashkenazi family, is considered a classic work. His other works are *Yoshe* Kalb (1932; The Sinner, 1933), a description of Hasidic life in Galicia, and The Family Carnovsky (1969), the story of refugees from Germany before and during World War II. He also wrote short stories and plays that were performed in Yiddish both in Europe and the United States. He was a realist without a trace of the sentimentality characteristic of most of the Yiddish writers. His younger brother Isaac Bashevis Singer, who has lived in the United States since 1935, is the author of Satan *in Goray* (1955), The Family Moskat (1950), The Slave (1962), and The Manor (1967); he has also written numerous short stories, including stories for children. His books have been translated into a number of languages. Yiddish literature in the United States has excelled in its poetry: among the outstanding American-Yiddish poets are Morris Rosenfeld, whose poems about the sweat shops and the life of the immigrants gained world fame; H. Leivick, author of *The* Golem and Yehoash, who translated the Bible into Yiddish. Such poets as Moshe Leib Halpern, Moshe Nadir, Zisha Landau, and A. Lutzky were first-rate by any measure.

The two foremost Yiddish critics were Bal Machshoves (1873–1924), who lived in Russia, and Shmuel Niger (the pen name of Samuel Charney), who lived in the United States. Both of them influenced Yiddish literature to a high degree and nurtured whole generations of Yiddish writers.

Yiddish literature in the Soviet Union

The Bolshevik Revolution (1917–22) attracted a number of Yiddish writers and later liquidated most of them. Some of them were native Russians, others emigrated there after the Revolution. The greatest of these was David Bergelson, a novelist and short story writer. His best novels are *Noch Alemen* (1913; "After Everyone") and *Arum Vogsal* (1909; "By the Depot"). His novel *Penek* was a desperate and unsuccessful attempt to adapt himself to the demands of Soviet criticism. The same fate befell a highly talented poet from Vilna, Moishe Kulbak, and A. Nister, a mystic who tried his hand at Socialist realism. Peretz Markish, Itzik Feffer, and Izzy Charik wrote numberless odes to Stalin and his regime. All of these poets and many others were shot by the Soviet authorities in the early 1950s. Of the Jewish-Soviet writers of that time, the talented Asher Schwarzman and Leyb Kvitko had the good fortune to die natural deaths.

In the 1970s in the Soviet Union, Yiddish literature is practically nonexistent. The few writers who remain form a nucleus around the magazine *Sovietisch Heimland* ("Soviet Homeland"). In the United States no young generation of readers or writers is growing up. The situation is not much better in Israel, where the new generation knows only Hebrew. There the few Yiddish writers, such as Yehiel Hoffer, Joseph Papiernikov, A. Shomri, Yakov Friedman, Abraham Karpinovitch, and F. Siegelbaum, are concentrated around the magazine Die goldene Keit ("The Golden Chain"), edited by Abraham Sutzkever, a refugee from the Nazi holocaust, who is one of the best Yiddish poets, a bard of the destruction of his people.

Small centres of Yiddish literature also existed in other countries. In Romania there was the fabulist Eliezer Steinbarg; the poets Yankev Sternberg and Moishe Altman, the essayist Biekel. In France there are M. Shlevin and Mendel Mann, whose trilogy about the Red Army has been translated into a number of languages. The most important Yiddish writers in Argentina are P. Yannsovitch, Kehos Kliger, and Josoph Okrutny.

The Yiddish theatre naturally had its playwrights. The first and perhaps the greatest of them was Abraham Goldfaden, who lived in the 19th century and was the father of the Yiddish theatre. Jacob Gordin, also in the 19th century, was for many years the foremost Yiddish dramatist in the United States. There was S. Ansky, who wrote the world-famous Dybbuk, David Pinski, Peretz Hirshbein, and Harry Sackler.

Memoirs in Yiddish literature

The treasure of Yiddish literature is its huge number of memoirs. A memoirist of talent was Yehiel Yeshaia Trunk. His eight-volume work, Poilen (1944), is a monument to Jewish life in Poland. The greatest part of the memoirs about the Hitler era have been and are being written by ordinary men and women. These works constitute an accusation against humanity that has no counterpart in the history of world literature. The present generation of readers is still too close to the enormity of the holocaust to face it. Perhaps the future will be able to appraise this overwhelming literature.

IN MODERN EUROPEAN LANGUAGES

To define what Jewish literature is in any language other than Yiddish or Hebrew is no easy matter. Are Saul Bellow and Bernard Malamud Jewish writers or American writers who happen to be Jews? Was Franz Kafka a Jewish, German, or Czech writer? Is it enough to write on Jewish themes to be called a Jewish writer? No one considered Edna Ferber or Harold Pinter a Jewish writer because they were Jews. Was Ludwig Lewisohn a Jewish writer because he wrote on Jewish themes and in his later years fervently expounded his faith in his writings? Was Abraham Cahan (1860–1951), who wrote both in English (*Yerkl* and *The Rise of David Levinsky*) and in Yiddish (*Bletter fun mein Leben* ["Pages from My Life"]) and preached Socialism and atheism, a Jewish writer, a Yiddish writer, an American writer?

It is a fact that in Russia Isaac Babel was thought of as a Jewish writer, while Osip Mandelstam, and Ilya Ehrenburg were looked upon as Russian writers of Jewish extraction. The French speak of Schwarz-Bart, Memi, Elie Wiesel, and Piotr Rawicz as Jewish writers; but no Frenchman would dream of calling Marcel Proust one. In Germany and Austria it was only with the beginning of Nazism that such writers as Heinrich Heine, Jakob Wassermann, Arthur Schnitzler, Stefan and Arnold Zweig, Franz Werfel, and Elsa Lasker Schiller were called Jewish writers, while Soma Morgenstern and Nelly Sachs were altogether ignored. Actually the writer belongs to the language in which he writes and to the land in which he lives and which he describes. In truth, the classification "Jewish" writer is an arbitrary one and can no more be clarified than "Catholic" or "Protestant" writer.

It is, however, a fact that Jewish themes and Jewish culture and folklore have been increasingly accepted in recent times. In the United States there are, in addition to those writers mentioned above, Henry Roth, Chaim Potok, Maurice Samuel, Meyer Levin, Hugh Nissenson, Philip Roth, Herman Wouk, Karl Shapiro, Alfred Kazin, Leslie Fiedler, Daniel Fuchs, Arthur Miller, Charles Angoff, and many others. In Canada, there are Abraham Moses Klein and Mordecai Richler. In England there was the famous Israel Zangwill; and there are today Wolf Mankowitz, Chaim Raphael, Chaim Bermant, and Gerda Charles. South Africa has Dan Jacobson, Italy the novelist Natalia Ginzburg.

There is no doubt that the interest in Jewish themes and in Jewishness in general was increased tremendously on the one hand as a result of the holocaust that destroyed a great part of the Jewish people and on the other by the birth of the State of Israel, which elicited the astonishment and admiration of both Jews and Gentiles.

Common characteristics

If there is any common characteristic among the writers who deal with Jewish life in languages other than Yiddish and Hebrew, it is that they thrive on thin soil and have short cultural roots. Often both the authors and their protagonists are children of immigrants who are not yet truly acclimated to the land their parents adopted. Their literature lacks the richness that is the result of a generation-long way of life. With few exceptions, writers who write in a language that is comparatively new to them are inclined to either exaggerated admiration for their own kind or an exaggerated self-hatred. Immigration and assimilation are not the healthiest soil for literature. Literature must have a factual and cultural address.

The literary masters were all too deeply bound to the culture of their own people. They knew well their history, their habits, their idiosyncrasies, their weaknesses and strengths. They emerged only when their people were ripe in maturity, saturated with tradition, stably and solidly rooted. It may take a long time before the Jewish milieu in the United States, in England, in France, becomes sufficiently independent spiritually to produce full-blooded Jewish artists. Jews who speak Yiddish in the United States, on the other hand, are generally an uprooted group. Their language has become impoverished rather than enriched. They have abandoned, whether willingly or not, too many of the customs and traditions that lend a group its own special colour. The Yiddish-speaking writes is in fact living in the past, both in his language and his themes. The soil for Yiddish in the United States is growing ever thinner, and the hope for the emergence of great writers in Yiddish ever more doubtful. (I.B.S.)

Dance and theatre

Although there is biblical evidence in the Song of Solomon and the Book of Job of Hebraic awareness of drama, Jewish religious traditions prohibited theatricals except for song, dance, and processionals. Until the emergence of the modern, professional Yiddish theatre in 1876, the ritual pageantry of song, dance, and processional of traditional holidays provided the only Jewish "theatrical" experiences.

DANCE

Antedating the birth of Jewish theatre art, the dance among Jews flowered in arbitrarily designated eras: Ancient (biblical), Diaspora (*c.* 300 BC–*c.* AD 500), Middle Ages, Hasidic (18th–19th century), and modern (20th century).

Social and religious dance. *Ancient.* Derived possibly from the customs of neighbouring pastoral and agricultural Semites and Greeks, the early songs and processionals, associated with taboo rites and sacrifices, celebrated fertility and harvest festivals.

Sukkot dances

On the first day of Sukkot, a seven- to nine-day religious festival of thanksgiving celebrated originally as an autumn harvest festival, there was a ritual dance procession around the altar, followed on the second day by a torch dance. In synagogues to this day, the waving of the *lulav*, or festive palm branch, and the *etrog*, the fruit of the citron (which are carried and waved during the festival of Sukkot), culminates in seven processions on Hoshana Rabba, or the seventh day of Sukkot.

There were also ancient Hebrew dances performed on other occasions: dances to celebrate the return of Jephtha; dances that were part of war ceremonials; dances before the ark (the sacred chest in which the two tablets of the Law were kept); the vineyard dance, wherein maidens danced to woo prospective grooms; and dances in praise of God:

> Let them praise his name with dancing, making melody to him with timbrel and lyre!
>
> (Psalms 149:3.)

Women were the principal performers in these dances, but men danced in praise of God.

To celebrate the Water-Drawing Festival of Sukkot, a part of the harvest celebration, dances and acrobatics continued through the night; at dawn, with trumpets blasting, water was poured from precious pitchers upon the altar. To proclaim nuptials, escorts of a bride and a groom sang and danced. At weddings, professional dancers performed, and the officiating rabbi sometimes danced to honour the bride. From antiquity to the Middle Ages a Jewish bride would dance as she brandished a sword to symbolically ward off suitors other than her beloved or to drive away evil spirits. In the medieval times this custom became prevalent in the Middle East. To honour a celebrity, there was a circle dance, with prescribed hand movements, which was accompanied by the juggling of knives, glasses, or eggs.

Diaspora and Middle **Ages.** To avoid antagonizing local rulers, the rabbinate often forbade dancing, a prohibition that was relaxed for the irrepressible holiday spirit during Purim (a festival to commemorate the deliverance of the Jews from the machinations of Haman, a minister of the Persian king Ahasuerus). Pantomime, song, and dance culminated in a fire dance, perhaps derived from primitive people who thus destroyed ever-threatening demons. Because of the possible danger of fires from the torches, the Eastern Roman emperor Theodosius II in AD 408 prohibited the fire dance by Jews throughout the Byzantine Empire.

Social and folk dancing at festivals and weddings, forbidden in the synagogues, found a home in the *tanzhaus.* There, the dances blended Jewish traditional with local native dance. Popular dances included *maien tanz* (like a Polish polonaise), *umgehender* (a courtly minuet), *schpring tanz* (a vigorous peasant dance), Judentanz, *Adam harishon tanz*, Doctor Foist *tanz,* and *fish tanz* (the forms of the last four are not known).

Wedding dances, combining ritual with gaiety, introduced a *tanz fuhrer* (dance leader), the forerunner of the *badchen* (a witty master of ceremonies), who led the dances and sang rhymed, often ribald verses at the expense of the wedding principals. Among the wedding dances were the *bettlertanz*, dances by local beggars culminating in a dance with the bride (this dance provides a climactic scene in the drama *The Dybbuk),* and the *Totentanz,* or dance of death, which originated in the time of a plague and symbolized death and resurrection.

Mixed dancing, by stringent custom, was restricted to close relatives. During the mid-17th century, ecstatic followers of the false messiahs abandoned themselves to indiscriminate and licentious dancing between the sexes; thereafter, more decorous mixed social dancing became popular, although limited among the strictly Orthodox.

Kabbalistic dances

During the Middle Ages, Kabbalists (adherents of the esoteric doctrine or mysticism) in Safed (now Zefat, Israel) welcomed the sabbath with song and dance. (This may have been the origins of later Hasidic dances.) Similar dances take place in the 20th century at the Western Wall in Jerusalem on the sabbath eve.

Hasidic. The Hasidic cult (devoted to mysticism and opposed to secular studies and Jewish rationalism), which became popular among the Ashkenazi (eastern European Yiddish-speaking) Jews in the early 18th century, was distinguished by the religious fervour and joy of its adherents. Their devotion and exaltation was expressed in a spontaneous dance wherein each individual sought *devegut* (a unity with God). The Ḥasid sways in prayer, or dances either alone or in a circle with his brethren as they hold each other's shoulders or waists; the rabbi often dances alone or leads the others. The height of Hasidic dancing occurs in their synagogues on Simhat Torah ("Rejoicing of the Law"), when the rabbi leads his followers in a dance while he embraces the Torah (first five books of the Old Testament) close to his heart.

Modern. The Palestinian hora, frequently performed outdoors, resembles early harvest dances imported from eastern European folk dances and blends peasant vigour with Hasidic ecstasy and abandon. It has become a popular dance among Jewish youth the world over.

Jewish wedding dances have taken on folkways from other groups and are no longer choreographed from religious mysticism such as the sword dance by the bride to ward off evil spirits. At an Orthodox wedding, or at the request of an Orthodox guest at a Reform wedding, the celebration will feature dances such as the *sher,* a form of square dance; the *mitzwa* or *kosher-tanz*, either a mixed dance with handkerchief as the only contact between the man and woman in observance of the Orthodox custom that prohibited mixed dancing, or a dance performed by the bride with a guest or with her groom; *machutanim-tanz,* a circle dance by relations of the couple; *broyges-tanz*, a dance game with mixed partners, expressing resentment followed by reconciliation, as well as others derived from local nationalities, such as the *quadrille* and *lancelot, kutzki, bulgar, pas d'Espagne*, and *vingerka.* The *redl, karahod,* and *hopke* are circle dances performed by men only. The *koilich-tanz*, a woman dancing with salt and halla or twisted white bread, to ensure prosperity for the newlyweds; the *besem-tanz*, a man manipulating a broom as sword, rifle, or horse; the *flasch-tanz*, a male dancer balancing a bottle on his head while he dances; and the *bube-tanz*, the grandmother's dance, are wedding dances that are still featured.

Professional dance. Since their slavery in the Diaspora, when some were employed as dancers and entertainers, the Jews have enriched the art and lore of world dance. In 1480, Jews and Moors were obliged to perform their national dances on Corpus Christi day (Roman Catholic festival in honour of the Eucharist); and, in the 17th century, the sultans Ibrahim I and Mehmed had their troupes of Jewish fiddlers and dancers. *Danza de la muerte* (dance of death), with Talmudic implications, attributed to Rabbi Santob de Carrion, appeared in Spain late in the 14th century. In the 16th and 17th centuries, Jewish *letzim* (itinerant jesters or mummers) toured Germany to appear at weddings, where their performances included walking on stilts, dancing, acrobatic tumbling, and pantomime. Early in the 19th century, the parents of the French tragedienne Rachel were wandering Jewish dancers.

Jews have also been dance teachers and choreographers. In 1313, Rabbi Hacen ben Salomo taught Christians a ritual choral dance at St. Bartholomew's in Saragossa, Spain. The grand Florentine Lorenzo de' Medici had a Jewish dancing master, Guglielmo da Pesaro, who wrote the book *Treatise on the Art of Dancing.* Other Jewish dance teachers and choreographers were Sieur Isaac d'Orléans in the French court (1700); Greccion Azziz and Emanuel de Rabbi Jalmoasis of Ancona, whom the Pope granted the privilege to teach dance; the great *danseur noble* Arthur Saint-Léon, the son of the ballet master at Stuttgart's Theatre Royal, who danced and produced ballets in Paris and St. Petersburg in the 19th century and whose *Coppélia* is still a popular favourite in the repertoire of leading ballet companies; Dame Marie Rambert (Miriam Rambach), who taught eurythmics (the art of harmonious bodily movements) to the Russian dancer Nijinsky; and Emil May, a Moscow choreographer.

Jewish ballet dancers

Prominent among Jewish ballet dancers were Adolph Bolm, Alexandre Sakharov, Alexis Dolinoff, Louis Chalif, David Lichine; Shulamith and Asaf Messerer, both stars of the Bolshoi Ballet; Ida Rubinstein, Dame Alicia Markova (Lillian Alicia Marks), Annabelle Lyon, Miriam Golden, Nora (Koreff) Kaye, Benjamin Zemach; in Israel there are Rina Nikova, Baruch Agudati, and Gertrud Kraus; and in Poland, Judith Berg, Bela Katzova, and Sam Hyor. In the mid-20th century, the outstanding Jewish dancer is Beatrice Abramowicz in Bucharest.

In the U.S., dancers identified with modern and avant-garde dance have been Dvora Lapson, Lillian Shapero, Lazar Galpern, Pauline Koner, Miriam Blecher, Belle Didjah, Sophie Maslow, Marie Marchowsky, Thelma Babitz, Nina Fonoroff, Helen Tamiris (Helen Becker), Anna Sokolow, Felicia Sorel, Gluck-Sandor (Samuel Gluck), Blanche Evans, Hans Wiener, and Jerome Robbins. Prominent modern dancers from other countries have been enthusiastically received when they have appeared for recitals in Israel. Folk dancing has attained great popularity. Yemenite folk dances, which revert to biblical sources, have developed out of Yemenite and biblical traditions and have become the specialty of a singing and dancing troupe, the Inbal Ensemble. Under the guidance of its founder and leading spirit, Sara Levi-Tonai, this trained, professional group performs at camps, kibbutzim, and villages as well as in concert halls in cities.

THEATRE

Excluded from theatrical activity by their religion, by harassment at the hands of hostile princes and states throughout the Diaspora, the Jewish people found the roots of their theatre in their traditional and religious rituals such as the impressiveness of the reading of the Torah and the protocol of the Passover feasts. Out of the religious festival of Purim came the earliest Jewish theatrical presentation.

Purim-shpil. The *Purim-shpil* (Yiddish "Purim play"; the term was in use by the mid-16th century) formalized the Purim merrymaking with a plot based on the Book of Esther. Growing in popularity, and thereby demonstrating an affinity for theatre among the Jews, the *Purim-shpil* became the main Jewish theatrical entertainment until the birth of the modern professional Yiddish theatre, in 1876. The first Purim play to be published (1708, and in Hebrew in 1710) was the anonymous *Beautiful New Ahasuerus Play,* followed by *Mekhirat Yosef* ("The Sale of Joseph"), by the writer Bermann

(Lemberg, 1710), which featured the merry-andrew, or buffoon (*letz*), stock comic Pickelherring, later to be known as Leckerlaufer. Among the popular scripts that soon followed were *The Sacrifice of Isaac* and *David and Goliath*. First performed by amateurs and later by travelling mimes and jesters, the Purim play was presented either in the courtyard of the synagogue, in large homes of wealthy Jews, in large storerooms, or in community halls.

Yiddish theatre. *19th century.* Deploring the jargon of Yiddish, Jewish scholars wrote dramas in Hebrew during the 17th, 18th, and early 19th centuries. But because they were dedicated to the 18th- and 19th-century Haskala spirit of elevating the uneducated Jewish masses, eastern European Jewish poets, short-story writers, and novelists turned to the lingua franca of the lowly Jews, Yiddish. The origins of the professional Yiddish theatre were in the mid-19th century among the folk singers—individuals and groups—who travelled among the eastern European Jewish settlements and performed in Yiddish during that part of the year when the ever popular Purim play was not performed. The lone minstrel, Velvel Zbarsher (Wolf Ehrenkrantz), sang and acted throughout the provinces of Romania, where he popularized his own lyrics. His dedication to uplifting the uneducated and impoverished Jews was typical of the many educated *maskilim* (participants in the Haskala movement) who sought to emancipate the Jews from their circumscribed ghetto life. The most prominent among the travelling troupes who entertained in the Jewish enclaves in Yiddish was the Broder Singers (from Brod in Galicia), who appeared in cafes, wine cellars, and assembly halls throughout eastern Europe. Their repertoire included songs by Michael Gordon, Zbarsher, Eliokhim Zunser, and Abraham Goldfaden. When, in 1876, Goldfaden provided them with a continuity of dialogue and plot for their presentation in a wine cellar in Iaşi, Romania, the professional Yiddish theatre was born.

After hiding from threatened impressment into the Romanian army during the Russo-Turkish War (1877–78), Goldfaden's troupe appeared in Bucharest. Here the talented actor Sigmund Mogulesco, joining the company as the first professional Yiddish actor, established what was to become the conventional and popular role of a versatile comic. With a repertoire of Goldfaden operettas that were filled with Jewish lore and evoked the human richness of Jewish life, the troupe made successful tours of Russian cities that had Jewish centres.

The assassination of Tsar Alexander II (1818–81) intensified anti-Semitism until on September 14, 1883, Yiddish stage presentations were forbidden in Russia. The Yiddish theatre briefly continued in Warsaw as a "German" theatre; a second edict ended that subterfuge.

Yiddish theatre in the United States

The lonely immigrant and the Yiddish theatre found and sustained each other in America, where the Jews had greater economic opportunities; where they were influenced by radicals, a free press, and the intelligentsia to free themselves from religiosity and to assume a greater degree of worldliness; and where they were not subject to police harassment and official repression. In 1882, the young Boris Thomashevsky brought a few Jewish actors from London to New York to present the first professional Yiddish play in America. Soon, under the monopoly of the playwright-managers Joseph Lateiner and Moses Hurwitz, who produced vulgar melodramas and shoddy operettas, sentimental representations of inexact Jewish history, and crude imitations of Goldfaden's folk epics, a burgeoning Yiddish theatre in New York provided entertainment for the hordes of uneducated Jewish immigrants.

In the Haskala tradition, journalists, critics, actors, and artists sought to uplift the masses and agitated for better theatre. In 1892 Jacob Gordin, a Tolstoyan reformer, ushered in the first "golden epoch" of the Yiddish theatre with his play *Siberia,* produced by Jacob P. Adler, the most notable actor of the early Yiddish theatre. Gordin's prolific pen provided the Yiddish theatre with adaptations from the world's classic repertoire, as well as original plays concerned with social problems and reflecting Gordin's penchant for realism. For the ensuing decade, Gordin and his imitators dominated the Yiddish theatre in New York and Europe. Gordin's influence and heritage gave the Yiddish theatre higher standards of literature, acting, and directing, as well as a respect for the dramatists' written words that put an end to vulgar interpolations and extraneous improvisations by the actors.

The actors attempted to better their lot, organizing, in 1899, a Yiddish actors' union, a forerunner and guide to American Actors' Equity Association, which was formed two decades later. The union helped the actors but imposed restrictions upon productions that interfered with artistic realization; for example, the rules of seniority frequently required that an ancient actress be given an ingenue role, and new talent encountered discouraging obstacles in rules that limited membership.

20th century. Around the beginning of the 20th century, Russian authorities relented, and the Yiddish theatre burst forth with new life. Perez Hirshbein, considered the greatest Jewish playwright, inspired by the artistry of the Moscow Art Theatre, toured Russia and Poland with his company from 1908 to 1910. When its most notable actor, Jacob Ben-Ami, came to New York on the eve of World War I, he introduced the concepts of Konstantin Stanislavsky, director of the Moscow Art Theatre, to the American theatre for the first time via the Yiddish stage. In Warsaw, the Kaminska family established a distinguished Yiddish theatre that, except for a hiatus during World War II, has continued as the Jewish State Theatre of Poland even though its actor-manager-star, Ida Kaminska (the daughter of the original founder) fled from Poland's resurgent anti-Semitism in 1968.

The Yiddish Art Theatre

Although Gordin's influence continued to prevail, and great folk and poetic dramas by such dramatists as David Pinski, Leon Kobrin, and Ossip Dymov were produced by such prominent stars as Jacob P. Adler and David Kessler, the Lateiner and Hurwitz type of shoddy melodrama and tawdry operetta persisted. During World War I, a plethora of sordid melodramas about fallen women infested the Yiddish stage. In 1918 Maurice Schwartz gathered aspiring members of the New York Yiddish theatre into a new venture that was to become, two years later, the Yiddish Art Theatre. of which he became director and star for three decades. Early in its first season, at the urging of Jacob Ben-Ami, Schwartz's troupe presented Hirshbein's *Farvorfen Vinkel* ("Forgotten Village"). With its enthusiastic reception, the Yiddish Art Theatre movement—and the second "golden epoch" of the Yiddish theatre—was launched. Schwartz's troupe endured, both in New York and on triumphant world tours, for over three decades.

The determination to bring worthwhile literary dramas to Jewish communities and to develop talent for the professional stage resulted in the creation of literary and dramatic clubs that flourished in many Jewish areas. From these clubs were born important Yiddish theatre troupes of high artistic merit. Warsaw had the I.L. Peretz Theatre Society (founded in 1911) and the contemporary VYKT (Varschaver Yiddische Kunst Teater, or Warsaw Yiddish Art Theatre); and Vilna (now Vilnius, Lithuanian Soviet Socialist Republic) had FADA (Farband fun Yiddishe Dramatische Actyorin, or Association of Yiddish Dramatic Actors). When, in 1915, the occupying German forces determined to undermine Russian power by encouraging nationalistic groups, the dramatist Arnold Zweig and fellow German officers helped FADA form what was to become the world-renowned Vilna Troupe, whose fragments—discords and misunderstandings created splits within the group—later toured the world and set standards of unsurpassed theatre art. Also in 1915, in New York, two dramatic clubs combined to form the Folksbuehne, which, in the latter half of the 20th century, is the most notable Yiddish theatre still operating in the United States. Various theatre clubs combined in 1923 into the Folks Farband far Kunst Teater. In 1926, with the aid of many Jewish workers' organizations, this active studio group became the Arbeiter Teater Farband. When in 1927 it presented its first public production, the left-wing Artef group started its memorable, albeit one-decade, existence, during which it presented politically provocative and stylistically challenging productions in New York.

To revive and stimulate interest in Hebrew as a living language, Habima (Hebrew: the Stage) was organized in Bialystok, Poland, in 1912, under the guidance of Nahum Zemach. Encouraged by Stanislavsky, it established itself in Moscow during the turbulence of the Russian Revolution of 1917. Despite the new government's edict against Hebrew as a secular language for the Jews, the high artistic merit of this troupe, which performed in Hebrew, made it a cult in Moscow. Its sensational production of the Russian playwright S. Ansky's Dybbuk electrified theatregoers on both sides of the Atlantic. During the company's tour of Europe and the United States in 1926, defections by and schisms among the members led the remnants to regroup in Palestine where, in 1931, the Habima permanently established itself in Tel Aviv and later became the official state theatre of Israel.

Between the two World Wars, the Soviet government encouraged and supported Yiddish theatres, the most successful of which was the Kamerny in Moscow. As one scholar puts it, this theatre "sneered at old-fashioned Jewish folkways, caricatured such life, and satirized it with political interpretations; it exaggerated in order to ridicule the Jewish past." Thus, this Moscow Yiddish folk theatre distorted Jewish heritage in attempting to combine it with the ideology of the October Revolution.

Caught in the crossfire of the savage Nazi scourge and Stalin's purge, almost all Yiddish theatres, actors, and directors were eliminated from the Russian scene; moreover, the *shtetl* (Jewish small town or small town community in eastern Europe), the milieu out of which were distilled the Yiddish theatre arts and Yiddish dramatic poetry, also disappeared.

In the 1970s, the best organized and most ongoing Jewish theatres are in Israel and Romania. The first attempt at Hebrew theatre in Palestine, in 1894, was followed periodically with attempts by sporadic groups to establish themselves in either Hebrew or Yiddish productions. With the birth of the State of Israel in 1948, the Hebrew theatre came of age; by mid-20th century, it was firmly established as a vigorous and important part of world theatre culture. In addition to the state-subsidized Habima, there are the Ohel and Cameri theatres in Tel Aviv; Hateatron Haironi (the Municipal Theatre, under the direction of the renowned Josef Millo) in Haifa; and the Histadrut's (General Federation of Labour's) Telem (Theatre for New Immigrants), which has functioned chiefly as a tool to teach uneducated Jews from Arab countries. As an educational tool, its mobile stage offers plays at various reception centres for new immigrants as well as at outposts where newcomers are settled. In addition to independent Hebrew professional theatre groups that appear from time to time in the leading cities, visiting troupes, such as the Bucharest State Yiddish Theatre, have appeared in Israel. There are about 300 amateur theatrical groups in this small country. Playing an important role in the nation's culture, the theatres avoid plays from the Yiddish traditional theatre or ghetto genre, because the highly sophisticated theatre audiences are aware of the avant-garde theatres of other lands and are receptive to this eclectic fare. European and American successes are popular. Seldom are their original plays concerned with biblical subjects, local problems, the fight for independence, or the pioneers who settled the land. Supported by the highest per capita theatregoing population in the world, the Israeli theatre is forging its own theatric-dramaturgic traditions of a high order.

Decline of Yiddish theatre

The decline of Yiddish as a spoken and written language has had a devastating effect upon the languishing Yiddish theatre. Other than intermittent productions in New York and Buenos Aires and rare productions in other world capitals, there remains only a waning State Yiddish Theatre in Warsaw and a flourishing, highly artistic State Yiddish Theatre in Bucharest. Supported by a benign government, the Yiddish theatre in Bucharest operates its own theatre and tours the provinces with both traditional and avant-garde productions. Thus, it is in Romania, where the modern Yiddish theatre was born, that it still flourishes.

In its brief lifetime, the theatre of the Jews has brought forth talented performing artists and important dramatic poets, served as a nationalizing fulcrum, provided a bridge over which new theatrical art forms came from Europe to America, and created theatrical art forms that have enriched the world stage. (D.S.L.)

Visual arts

The Jewish attitude toward art has been cautious at times, although contrary to the general assumption there was no prohibition of artistic expression in biblical times, nor in most periods of Jewish history. The Second Commandment defines itself as a law against making idols for the worship of "other gods." No fear of artistic representation as idolatry can be detected in the detailed instructions given by God to Moses on how to build the Tabernacle and its implements, which, indeed, includes the specification to adorn the ark with two winged cherubs. The prohibition against idol worship, and sometimes the destruction of idols, did not deter artistic representations throughout the Israelite (c. 1020–586 BC) and Hasmonean (142–63 BC) periods.

The traditional Jewish "way of life" in accordance with the Halakha—*i.e.*, the section of the Talmud dealing with religion and civil law—which was established during the 3rd to 6th centuries AD, stressed verbal rather than visual expression: the Talmud is the great rabbinic compendium of Oral Law, lore, and commentary. Artistic representation was not prohibited, since there was no longer any fear of idolatry. It was mainly encouraged for educational purposes, or for *hiddur mitzwa* (that is, adornment of the implements involved in performing rituals).

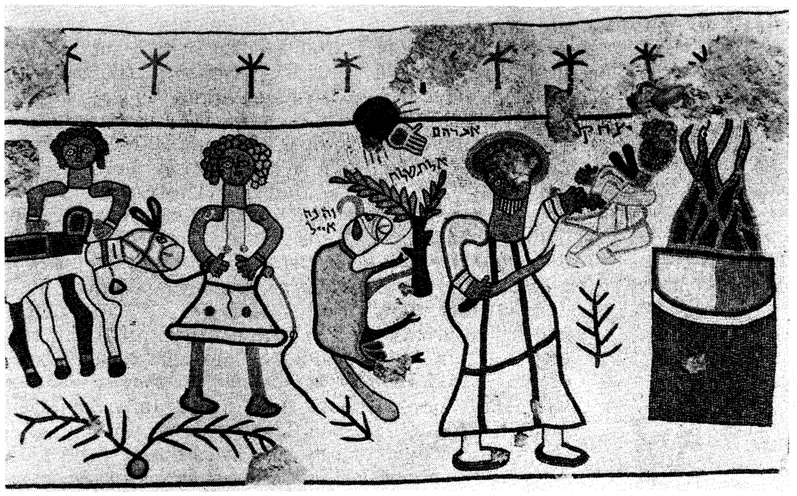
When Jewish antagonism to artistic expression did appear, it was limited to isolated areas and periods. Complete abstention from any adornment or from the use of decorated religious or secular objects pertained to small, ascetic Jewish sects called Hasidim, known from the Mishnaic period (450 BC to AD 200) and, later, from the 12th–13th centuries. Partial artistic abstention developed under the influence of external pressure prohibiting the depiction of the human form in sacred objects among the Jews living in Muslim countries and in the Byzantine Empire during the Christian Iconoclastic period, which prohibited the use of images in religious worship (725–843). During the 13th century, Jewish artists in the south of Germany introduced the human figure in distortion as a way to overcome this aversion.

Jewish art should be conceived as part of the arts of the Jewish people. There is no general Jewish style that can be traced throughout the 4,000 years of Jewish history. Because of the wide dispersion of the Jews, the style of Jewish art objects was usually similar to the contemporary style of their immediate environment. Nevertheless, some specific Jewish elements, motifs, forms, and styles were crystallized for limited periods in regional Jewish schools of artistic expression.

The arts of the Jews can be broadly divided into three major periods, each different in character from the other: the Israelite, the Jewish, and the modern periods.

ISRAELITE PERIOD

Very few art objects related directly to Israelite culture have survived from the Israelite period (see above). The Bible gives accounts of some major edifices, most of which were influenced by contemporary Egyptian, Canaanite, or Phoenician forms. These include the "Tent of Meeting" (Tabernacle) and its various appurtenances and Solomon's Temple in Jerusalem, built and decorated under the direction of the Zorite (Phoenician) artist Hiram. Solomonic public buildings were uncovered in Israel at Megiddo, Ḫaẕor and Gezer. Other royal buildings excavated are the palace of Ahab (*c.* 860 BC) at Samaria north of Nablus, and the palace of Jehoiakim (*c.* 608 BC) at Ramat Raḫel, south of Jerusalem; both were decorated with proto-Aeolian capitals that originated in



Sacrifice of Isaac, detail of a mosaic from the synagogue of Bet Alfa, Jezreel Valley, Israel, 6th century AD.

Picture from the photographic archive of the Jewish Theological Seminary of America, New York. Frank J. Darmstaedter

Archaic Greece. Ahab's palace was partly covered by surviving ivory panels that resemble the ivories found at Arslān Ṭāsh in northern Syria.

JEWISH PERIOD

Classical antiquity. The Jewish period begins with the classical phase, which spans the Hellenistic, Roman, and early Byzantine eras. Some elements introduced in Hellenistic times became characteristic throughout the Jewish period and lasted until the 19th century. The most important of these are the structure of the synagogue, the depiction of some biblical and legendary episodes, and the use of Temple symbols like the facade of the Temple, the ark of the Covenant, the *menora* (a seven-branched candelabrum), a shovel, the oil amphora, the shofar (ceremonial ram's horn), the *lulav* ("palm branch"), and the ***etrog*** ("citron"), all used in conjunction with the Sukkot festival.

Early development of the synagogue

Literature and archaeological finds of the period give evidence that synagogues existed both in Jerusalem and throughout the Diaspora (communities of Jews living outside Palestine), even before the destruction of the Temple in the year AD 70. The surviving synagogues from Roman times are of a basilican type of structure—*i.e.*, of an oblong construction ending in an apse, or a vaulted semicircular niche. Seats were built around the walls. The earliest example is the synagogue on Masada in the Judaean Desert, built by zealots during the revolt against Rome (AD 66–70). Most surviving synagogues of the late Roman and early Byzantine periods are located in the Galilee region of modern Israel and in Greece and Syria, and they are of two basic types. One type has its facade facing Jerusalem; the participants prayed within, facing the entrance. The 3rd-century synagogue at Kefar Nahum (Capernaum), in Israel, and the 4th-century synagogue in Ostia, near Rome, were of this type. The second type of synagogue had its back wall facing Jerusalem, and the Torah (scrolls of the Law) was kept permanently within a niche in the wall. Examples are the 6th-century-AD synagogue of Bet Alfa in the Jezreel Valley (Plain of Esdraelon) of Israel, and that in Dura Europos, on the Euphrates. Built in AD 245, it is now a part of the structure of the National Museum in Damascus, Syria.

Herod's Temple

The most important remaining building of this period of Jewish art is probably Herod's Temple, which was built on top of the Second Hasmonean Temple in Jerusalem. Although only the impressive supporting Western Wall (known as the Wailing Wall) and Southern Wall survive, they still give some indication of the magnificence of the building's scale. Herod's Temple was designed and decorated in the prevailing Roman style.

Synagogues were decorated by stone reliefs carved in the capitals of the columns and in friezes above them. The decorative motifs mainly consisted of acanthus leaves and floral ornamentation in addition to the traditional Temple symbols. The painted walls of Dura Europos bear witness to wall paintings mentioned in the Jerusalem Talmud (*c.* AD 400). They depict mostly biblical themes, with some Temple symbols. The style of the painting is Hellenistic in the Persian manner. Wall paintings, such as those of Dura, may have been modelled after illuminated biblical manuscripts, though no Hebrew manuscripts of the Classical period have been found.

During the 5th and 6th centuries, synagogues were decorated with floor mosaics in the Byzantine style. The motifs and themes used included Temple symbols, signs of the zodiac based on classical pagan models, and biblical scenes, of which the most famous are the sacrifice of Isaac in the synagogue of Bet Alfa in the Jezreel Valley, and the animals leaving Noah's ark in the Jarash synagogue in Jordan.

Jewish sepulchral art largely consisted of graffiti (crude drawings scratched onto a hard surface) and paintings of Temple symbols. These were executed on the interior walls of burial chambers and have been found at Jerusalem, at Bet She'arim in Israel, and in catacombs in Rome. Some burial caves in Israel had monuments built over them, such as the family tombs in the Qidron Valley and the public tombs around Jerusalem and in Bet She'arim.

Middle Ages and Renaissance. The Middle Ages introduced two distinct streams of culture into Jewish art, which were further developed during the Renaissance. These are the Sefardi tradition, originating in the East, and the Ashkenazi tradition of western Europe.

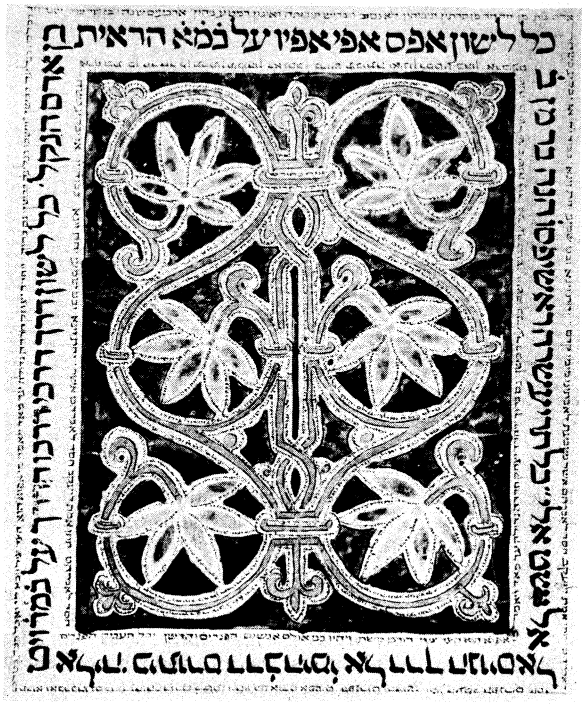
Sefardi art of the East

The Eastern stream was centred in the Muslim countries of the Near East and North Africa, as well as Yemen and Spain. After the expulsion of the Jews from Spain in 1492, a specific culture called Sefardi was created through the spread of Spanish Jews into Europe, the Americas, and the East.

Of the many large synagogues, schools, and other public buildings recorded in Jewish literature, only a few minor synagogues have survived in Egypt and Spain. These are mostly single halls with large Torah niches facing Jerusalem and a central bimah, an elevated platform from which the Torah is read. They were decorated with stucco sculpture and carved wood in the style of the Muslims, and the decorative motifs were Muslim as well, but Hebrew inscriptions were employed for ornamentation rather than Arabic ones. The 9th-century Ben-Ezra synagogue in Fusṭāṭ (Old Cairo) and the present church of El Transito (1357), Toledo, Spain, are examples of the single-hall synagogue. A more elaborate plan is that of the church of Santa Maria la Blanca in Toledo. Originally a synagogue built in 1205, it is divided into five aisles by four rows of horseshoe-shaped arcades.

The illumination of biblical manuscripts developed from the 9th century in the East, especially in Palestine, Egypt, and Persia. The carpet-like illumination of whole pages and sectional drawings followed motifs devised for illuminating the Qur'ān, the Muslim holy book. A special Jewish innovation was the outlining of drawings in micrography (tiny writing), using passages from the Masora (a collection of textual notes on the Hebrew Bible). Another typically Jewish element is the inclusion of a symbolic array of Temple implements as one of the carpet-like pages. This began in the East and developed mainly in Spain and the Provence region of France during the 14th century.

By courtesy of the Jewish National and University Library. Jerusalem



Carpet page with scrolls and leaves outlined by a micrographic inscription from the Masora. From the Damascus Keter, an illuminated Bible from Burgos, Spain, 1260. In the Jewish National and University Library, Jerusalem. 30.2 X 27 cm.

The 14th-century Spanish schools of illumination produced sumptuous Haggada manuscripts (book containing the text of the ritual service recited during the Passover *seder*). Full-page miniatures depicting biblical episodes, mainly from Genesis and Exodus, are found in manuscripts such as the "Golden Haggada" of Barcelona (*c.* 1320). The most important schools of Sefardi manuscript illumination in the 13th–14th centuries were those of northeastern Spain and southern France. In the 15th century, after the decline of the Castilian and Aragonese Jewish communities in central and northern Spain, the illumination schools of Galicia, in northwestern Spain and Portugal, became prominent.

Ashkenazi art of the West

The Ashkenazi Jewish culture of western Europe was centred in England (before the Jewish expulsion at the end of the 13th century), France (until the beginning of the 14th century when Jews were expelled), Italy, and Germany (where Jews were driven out during the 15th century).

The surviving Ashkenazi synagogues from the 12th century onward comprise mostly a single hall that is generally divided by one or two central pillars that support a vaulted ceiling. The Torah niche is in the east wall and the bimah in the centre. An example is the 14th-century Altneuschul (Old New Synagogue) in Prague, one of the very few medieval synagogues to survive Nazi destruction. Synagogue decorations were usually stone carvings, in the style of Gothic and Renaissance architectural ornamentation.

Ashkenazi illuminated manuscripts appear simultaneously in the 13th century in Germany, France, and Italy. To some extent the biblical manuscripts follow the oriental pattern of carpet pages, sectional divisions, and use of micrography. Ashkenazi manuscripts, however, tend to have more text illustrations, both on the text pages and as frontispieces. The Ashkenazi Haggadas, on the other hand, have illustrations within the text margins, rather than full-page miniatures. ***Mahzorim*** (festival prayer books) with marginal illustrations in text became fashionable in the Rhine region during the 13th–14th centuries. The Leipzig Mahzor (University Library, Leipzig, Germany) is one of the finest examples of illuminated *maḥzorim.* The existing works of Ashkenazi illumination produced in the 13th and 14th centuries in northeastern France, southern Germany, and central Italy were all directly related to the styles of Latin illumination in a given period or region. During the 15th century northern Italy became the most important centre of illuminated Hebrew manuscripts in the Renaissance style. Early printed books from Spain, Italy, and Germany also follow Christian Renaissance prototypes.

Some of the ritual objects introduced during the Middle Ages became established in the Renaissance for the subsequent generations of European Jews. The most common examples are the Torah ark, the Sefardi Torah case and the Ashkenazi Torah mantle, the crown and finials placed on top of the Torah, star-shaped, hanging sabbath lamps, tower-shaped spice containers for ceremonial use at the close of the sabbath, and lamps for Hanukka (Festival of Lights).

During the Middle Ages and the Renaissance, Jews excelled as artisans in many trades, some of which were known to be almost exclusively Jewish. Coin striking, silver engraving filled in with a black composition of metallic alloys called niello, sword production, and leather working were Jewish crafts. The making of astrolabes and drawing of maps were both considered Jewish arts in Majorca, Spain, Portugal, and North Africa.

Baroque age. The Jewish Baroque period of the 17th and 18th centuries was particularly important for introducing new decorative motifs, richer embellishment of ritual objects, and new design types for the building and arranging of synagogues.

In northern Italy, a "bipolar" arrangement developed inside the Sefardi synagogues, whereby a dramatically raised, stagelike bimah was placed along the wall opposite the ark of the Law. A lavish example is the 16th-century Sefardi synagogue in Venice, which was redesigned in the mid-17th century by the prominent Baroque architect Baldassare Longhena. In contrast, the Sefardi synagogues in The Netherlands, England, and the United States had the bimah in the centre, at varying distances from the Torah ark, as in the Touro Synagogue, designed in 1762 by the colonial American architect Peter Harrison for the Jewish community of Newport, Rhode Island.

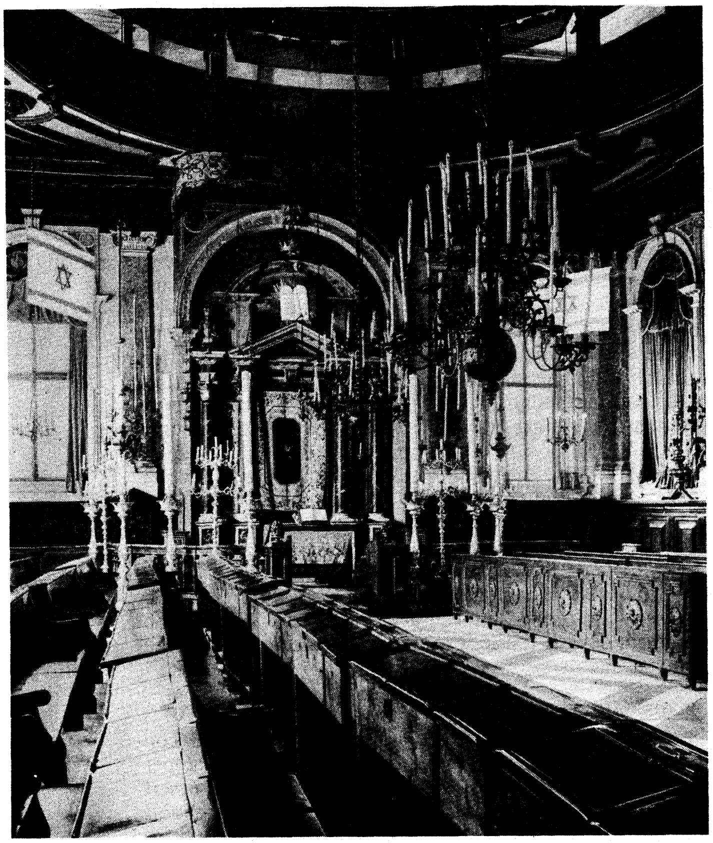
During the Baroque period two special types of synagogues were developed in eastern Europe: the four-pillared stone synagogue and the wooden synagogue. Most of these were destroyed in the 20th century during the Nazi occupation.

Innovations in synagogue design

In Polish synagogues the four pillars were erected around a central raised bimah. They generally supported a heavy canopy. The earliest known four-pillared synagogue was built in 1553 at Cracow. Some of these synagogues had a castle-like exterior to protect the building and the congregation from the attacks of Cossacks and Tatars. Interior wall paintings were sometimes used as decoration.

The timber synagogues of Poland were built with elaborately decorated ceilings. The bimah was usually in the centre, and some were of the four-pillar type. Most of the walls and ceilings of the wooden synagogues were painted with biblical, floral, and symbolic motifs. The earliest known timber synagogue was at Chodorbw (modern Khodorov, Ukrainian Soviet Socialist Republic) in Poland (1651), but many must have been built earlier.

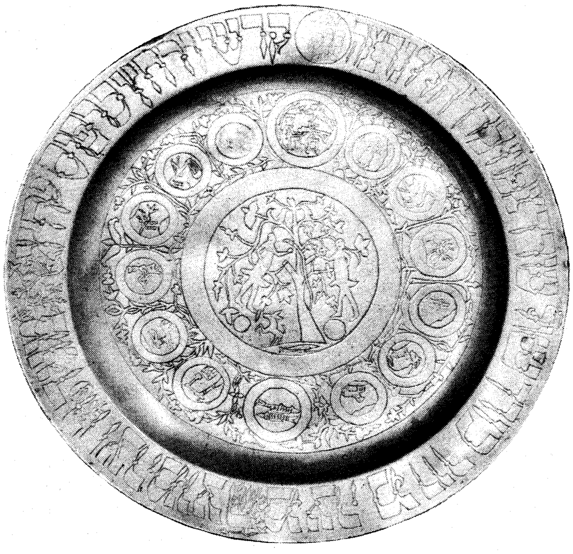
Objects of the Ashkenazi rite and the Italian rite (*rito italiano*) became more elaborately adorned with silver and gold during the Baroque period. New objects were introduced for the synagogue as for use in the home. The Torah ark, sometimes an entire structure of wood or



Tempio Israelitico Spagnola (Scuola Spagnola), Venice, a Baroque synagogue built in the 16th century and redesigned in the 17th by Baldassare Longhena.
Photo Giuseppe Bruno

marble, was adorned with an embroidered curtain and a valance to hang over the curtain. The Torah scroll was adorned with a breastplate to hang over its mantle and with a pointer to help in reading the scroll. A laver and basin for washing the hands of the *cohen* (priest), a standing Hanukka lamp, and the lamp of the eternal light became fashionable synagogue accessories. Ritual objects for the home were the work of folk artists who used both traditional forms and their imaginations to shape and to adorn them. The most favoured objects were Hanukka lamps, spice boxes, special dishes for the festivals of Passover and Purim, sabbath candlesticks, wine cups, and metal bindings for prayer books. In the 18th–19th centuries, schools of manuscript illuminators developed in Moravia and Bohemia and spread to Germany, The Netherlands, and England. Professional adornment of tombstones began in The Netherlands with biblical scenes, and spread to other areas of western and eastern Europe, where Jewish folk artists depicted no human forms.

Picture from the photographic archive of the Jewish Theological Seminary of America, New York. Frank J. Darmstaedter



Pewter Passover plate, German, 17th century. Adam and Eve are represented in the centre, surrounded by the signs of the zodiac and a depiction of the sacrifice of Isaac. The Hebrew lettering on the rim indicates the order of the Passover eve service. In the Jewish Museum, New York.

MODERN PERIOD

The emancipation of the Jews in the 19th century brought the art of the Jewish people out of its purely religious function and into the secular world, where it has remained. No Jewish national style has developed, and Jewish artists have integrated into the styles of the period. At the same time, art for ritual use has retained its traditional character.

Some Jewish 19th-century artists painted scenes of contemporary Jewish life. The most famous were Moritz Daniel Oppenheim in Germany, Solomon Alexander Hart in England, Mourycy Gottlieb in Poland, and Jozef Israels in The Netherlands. Of the Impressionistic school, the most famous Jewish painters were Camille Pissarro in France, Max Liebermann in Germany, and Isaak Ilich Levitan in Russia.

In the 20th century, several Jewish artists have become prominent, and although they may use Jewish subjects, the Jewishness of their style is debatable. Some of the best known are the painters Marc Chagall, Amedeo Modigliani, Chaim Soutine, and Jules Pascin; the sculptors Jacques Lipchitz, Bernard Reder, Naum Gabo, and Jacob Epstein; and the architects Erich Mendelsohn, Louis Kabn, and Moshe Safdie.

Jewish style of the 20th century

Attempts to create a Jewish national style occurred in 1906, when Boris Schatz founded the Bezalel School of Art in Jerusalem. His aim to combine Ashkenazi artistic forms with Yemenite style failed, and no distinctive Jewish national style has yet developed in Israel. Another attempt at the creation of a Jewish style was made in Russia immediately after the revolution of 1917, when a group of Jewish artists consciously studied Jewish folk art. Some of the late works of the painter Marc Chagall, such as his stained-glass windows installed at the Hadassah-Hebrew University Medical Center in Jerusalem, can be traced to this period when he worked on Jewish folk themes in the village of Vitebsk, in the Soviet Union. Twentieth-century synagogue architecture, sculpture, and ritual objects have been developed especially in the U.S., where they are characterized by formal monumentality and simplicity of decoration. (B.Ns.)

Music

History, geography, function, creators, and performers all contribute to the definition of Jewish music, a term that is most meaningful when applied to pure or stylized Hebrew musical tradition. In antiquity, when almost all Jews lived in one country, their music was largely vocal, instruments being used only for dance music and for the accompaniment of chant. Although there was non-Hebrew vocal music, most was in the Hebrew language, falling into the broad categories of religious and secular music. The religious stream included liturgical music, and, with the secular stream, it also contributed to popular music. The liturgical music of the Temple was performed by professional musicians, while synagogue music, originally performed by nonprofessionals, was influenced by both Temple liturgy and popular music.

MUSIC TRADITIONS

After the destruction of the Temple in Jerusalem (AD 70) by the Romans and the revolts against Rome (AD 135), the Jews were dispersed through North Africa and western Asia and migrated to Europe. Between AD 1 and 800, three main musical traditions emerged: the Mediterranean, or Sefardic, tradition, found at present in Spain, Italy, southern France, and the Balkan countries, northern Africa and the Near East, The Netherlands, the Caribbean, and in the former French colonies; the Ashkenazic musical tradition, its central European branch including German-speaking populations, its eastern branch encompassing Baltic and Slavic populations, both branches mingling in the Americas; and the Yemenite tradition, found in Yemen and Ethiopia. The three musical traditions and their subdivisions converge and can be heard in modern Israel.

Of the various properties of Jewish music, a few appear to be permanent and unaltered through the centuries: the preponderance of oral over written tradition; the importance of vocal rather than instrumental music; and the limitation to texts in Hebrew and its two most important derivatives, Yiddish, in the Ashkenazic orbit, and Ladino, or Judeo-Spanish, in the Spanish-Arabic. In all groups and all traditions, Hebrew chants stem from the oldest, possibly Hellenistic-Palestinian, origin. Because the earliest sources of written musical tradition come from the 12th century and are rare until the 18th century, the serious study of Jewish music must start with an examination of the problem of the continuity and authenticity of oral tradition.

Yemenite, Sefardic, and Ashkenazic traditions. Each of the three musical traditions differs from the others, so that they share only a few common elements. The Yemenite tradition is the oldest; although the Yemenites never came into contact with Christianity, they still sing tunes that have close parallels in the oldest strata of Gregorian chant. The Yemenites emigrated from Palestine during the rule of the last Hasmonean princes, and in Yemen, under later Islāmic rule, the Jews remained in strict isolation. Thus, their customs and melodies sometimes reach back into the pre-Christian epoch, from which also stem the oldest examples of Gregorian chant. The liturgical music has been preserved in a jealously guarded oral tradition.

Yemenite and ancient Christian chant

The continuity of the Sefardic tradition is better documented. Sefardic secular songs contain elements of medieval Spanish music and poetry, such as the ***Cantigas de Santa Maria,*** the ***El cantar de mío Cid,*** and the ***Ciclo de Roldán*** ("Cycle of Roland"), sung to old tunes in Ladino, which closely resembles Castilian Spanish. Many of these melodies are also used for liturgical chants.

The most recent of the traditions is that of Ashkenazic Jews. It begins in the Carolingian epoch, and it reflects best the historic transformations and the tragedies of European Jewry. Originating in northern Italy, the Ashkenazim settled in the German and French Rhineland, then branched off to Bavaria, Austria, Bohemia, and the other countries that constituted the Holy Roman Empire. After 1350, when the Jews were accused of spreading the plague and were massacred by the thousands, two or three waves of mass emigration reached Poland and Russia to the Ural Mountains and the Black Sea. Their tradition, called the east European, diverges from the central European only in those elements that originated after 1500. These show distinctly Slavic character.

Elements of musical style. There is some evidence that, in antiquity and in the Middle Ages, the various Jewish modes of chant corresponded to the expression of certain emotions, in accordance with the similar "doctrine of ethos" in ancient Greek music. Thus, laudatory prayers were differentiated from supplicatory ones by their musical modes. This approach later gave way to the belief in the harmony of the spheres and, in Ḥasidism (a pietistic movement arising in eastern Europe in the 18th century), to the doctrine of seven grades by which human chant may reach the throne of the Almighty.

After the destruction of the Second Temple, common law and oral tradition, rather than common language, linked the various groups. The differences between them are seen in elements of their chants. Their musical styles are represented in the criteria of mode versus scale; rhythm and metre; syllabic and melismatic tunes (see below); and monophonic and polyphonic music (see below). A scale is the result of the division of the octave, whereby the rising or descending steps fill the space of the octave. A mode, on the other hand, is in no way bound to an octave; it consists of a pattern of characteristic intervals. Thus, many individual melodies can be traced back to the same modal pattern. Modal patterns, such as the Dorian, Mixolydian, and others, continue to be used in Jewish tradition as well as in Christian Gregorian chant. The rhythm of the ancient chants was apparently free, although closely bound to textual accents. In the hymns that developed during the Middle Ages, melodies with regular metrical patterns also appear, although much more frequently in the Ashkenazic tradition than in the Sefardic. Most characteristic of the various rites of Jewish groups is the relationship of word and melody preferred by each group. Between the extremes of simple syllabic style (one note per syllable) and pure melismatic style (florid style with many notes per syllable), all shades and nuances occur. Thus, the east European Jews prefer a chant spiced with many melismatic passages, whereas the Yemenites use melismata only sparingly. The cantillation, or liturgical chanting, of Scripture also differs in the three main rites, although the so-called Masoretic accents, written signs regulating both cantillation and textual punctuation, are common to all groups. They date from the 9th century AD; the melodic motifs of cantillation, however, survive from antiquity. Jewish folk music does not make use of polyphony (music in several voice parts) but is monophonic (consisting of a single melodic line). In contrast, modern Israeli art music makes use of polyphony and all other techniques found in Western music. The works of 19th- and 20th-century Jewish composers do not represent Jewish music, except in compositions using traditional chants.

Word and melody

HISTORICAL DEVELOPMENT

Ancient and medieval music. The Old Testament contains many references to music. The practice of music permeates the wanderings of the Israelites through the desert, the dances of maidens, the outbursts of prophetic schools, the lyric prayers of young David, and the priestly praises in the First and Second Temples, recorded in Second Book of the Chronicles and in the book of Psalms. That most of these descriptions were exact has become apparent since the discovery of the Dead Sea Scrolls; for example, the scroll of ***The War of the Sons of Light Against the Sons of Darkness*** includes detailed description of the various trumpet signals. The accounts given by the Bible, the Jewish-Roman historian Josephus, and the Mishna (a section of the Talmud, the post-biblical legal code) stress the variety of the accompanying instruments: primitive clarinets, pipes, trumpets, lyres, harps, cymbals, hand drums, and even a rudimentary organ (or steam siren) were employed. Of these instruments, only the ritual and intentionally archaic ***shofar,*** or ram's horn, has remained for certain signals in the New Year's liturgy and the like. The ensemble must have sounded rather shrill because of the many high-pitched instruments.

Ancient instruments

Archaeological and literary evidence gives a rather clear picture of the Temple music; most of the instruments can be identified, and the unison singing of the Levites is established, as is the addition of boys' voices. But extant evidence yields no concrete idea of the Temple melodies and their nature.

The Talmudic and later rabbinic literature contains further descriptions of the Temple music, but these writings must be taken with considerable caution, for they all tend to exaggerate the number of participants and the quality of performance. In later history, only the European Jews had, by assimilation and acculturation, contact with Western music. The oldest notated documents of this musical symbiosis were written by Abdias the Proselyte (Obadya ha-Ger), an Italian monk of Norman extraction who had converted to Judaism after 1102. The three pieces notated by Abdias show resemblances both to Gregorian and to Sefardic-oriented chant, and similar phrases can be heard even today in Near Eastern synagogues.

The Ashkenazic musical tradition began in the Rhineland during the Carolingian period. Some ancient Ashkenazic psalmodies still form part of the daily morning prayers. The precentor (cantor, or ***hazzan)*** who carried the ancient, orally transmitted melodies emerged in Palestine and Babylonia in about the 7th or 8th century as an appointed official of the congregation. But his musical importance developed in Europe, where he was frequently singer, composer, and celebrant. The old Ashkenazic melodies, called Mi-Sinai tunes, especially those of the High Holidays, were composed by the cantors of the Rhineland and of southwest Germany during the period of the Crusades (1095–c. 1450). Their musical raw material blends ancient Jewish psalmodies with elements of the music of the trouvères and Minnesingers (medieval poet-composers of noble birth) and with street songs of the western German cities. There is some evidence that the transmitters from the Christian sphere were minstrels and vagrant singers.

Polyphonic compositions. During the Renaissance, the Italian Jews began to apply Western art music techniques to their synagogal music. Especially notable was Salomone Rossi, whose collection of Hebrew hymns, psalms, and other works for three to eight voices a cappella (Venice, 1622), contained a preface stating a revolutionary rabbinic decision that, in principle, allowed art music in the synagogue. Many French and Italian communities of the 17th and 18th centuries patronized minor Jewish and Gentile composers for festive cantatas. Among such works, the cantatas composed for the great Amsterdam Temple by the Sefardic composer Abraham Cáceres, with text by the Sefardic poet Moshe Hayyim Luzzatto, achieve a notable synthesis of Handel's aria style with liturgical chant. A much lower level of synthesis of operatic and traditional elements became typical for the general Ashkenazic development of cantorial songs, and by about 1820 the idiom of Haydn and Mozart had almost crowded out the traditional elements. Political emancipation and religious literalization saved the historical tradition of Ashkenazic music. The rejuvenator of this tradition was Salomon Sulzer, the chief cantor of Vienna. A friend of Schubert, and famous as a singer, organizer, and student of tradition, Sulzer set the example of style and performance for many European synagogues. His *Shir Zion* ("Song of Zion"; 1840, 1866) preserves traditional music, while making it accessible to organ accompaniment. The use of the organ became a controversy between the liberal and orthodox communities. Where the organ did not prevail, synagogue chant regressed to a rhapsodic popular style that has remained stagnant; where it did prevail, as in most Western and American communities, synagogue music began to rank with fine Protestant and Catholic church music. This was particularly evident in the United States, where, with the emerging Reform movement, the scientific study of Jewish musical traditions continued. Active 20th-century composers drawing on Jewish musical traditions include Abraham Binder, Leonard Bernstein, Lazar Weiner, Jacob Weinberg, Lazare Saminsky, Herbert Fromm, Isadore Treed, and Heinrich Schalit. The serious musical scholarship was initiated by the Hebrew Union College in Cincinnati and New York City.

Contributions of Salomon Sulzer

Israeli music

The rise of Jewish music had decidedly secular roots in Israel, to which the various waves of immigrants brought Yiddish and Ladino songs with them. But, from about 1910, immigrants also created working, dancing, and fighting songs in Hebrew. In the older dancing songs, some Hasidic elements are still noticeable, but they begin to vanish under the massive impact of Jewish immigrants from Arabic-speaking countries. Fine standards of European art music were set by Leo Kestenberg, an Israeli pianist and educator, in his work in musical education. The musicological departments of the universities of Jerusalem and Tel Aviv-Yafo have furthered the study of Jewish music, while the two music academies of these cities have specialized in the mass production of performing musicians. Meanwhile, there has emerged a generation of Israeli composers, among them Joseph Tal, Ödön Partos, Erich Walter Sternberg, and Paul Ben-Haim. The strongest stylistic trend blends the local colours of Israel with traditional elements of Jewish music within a Western framework that maintains close contact with European and American art music.

Research in Jewish music. The serious exploration of traditional Jewish music emerged about 1900. The first scientific studies were made by Eduard Birnbaum, a disciple of Sulzer. Abraham Zevi Idelsohn was the first to record thousands of songs in Palestine, between 1909 and 1920. His transcriptions are collected in the *Thesaurus of Hebrew Oriental Melodies* (10 vol., 1914–32). Curt Sachs, Eric Werner, and others who followed, specialized in historical or comparative studies. (E.We.)

BIBLIOGRAPHY

Literature: ABRAHAM ROBACK, *The Story of Yiddish Literature* (1940); MEYER WAXMAN, *A History of Jewish Literature*, 2nd ed., 6 vol. (1960); ISRAEL ZINBERG, *Toldot Sifrut Yisrael*, Russian, Hebrew, and Yiddish text, 6 vol. (1959–60), published, with additions, Yiddish and Hebrew texts, as *Di geshikhte fun literatur bay Yidn*, 10 vol. (1964–70); CHARLES A. MADISON, *Yiddish Literature: Its Scope and Major Writers* (1968); SOLOMON LIPTZIN, *The Flowering of Yiddish Literature* (1963) and *The Maturing of Yiddish Literature* (1970).

Music: ABRAHAM Z. IDELSOHN, *Thesaurus of Hebrew Oriental Melodies*, 10 vol. (1914–32), the standard collection of tunes according to their geographical origin, particularly important, as the last volume was printed before the mass-killing and mass-migration of European and Asiatic Jews; and *Jewish Music in Its Historical Development* (1929, reprinted 1967), a popular history, both its first and last thirds are dated, while the second third is still valuable; ALFRED SENDREY, *Bibliography of Jewish Music* (1951); PETER GRADENWITZ, *The Music of Israel* (1949), a more recent, stimulating book with the centre of gravity in Israel; ERIC WERNER, *The Sacred Bridge* (1959), a comprehensive work on the interdependence of Synagogue and Church for advanced students of liturgical music; a series of essays on Jewish music in the *Hebrew Union College Annual* (1940–66), studies that deal with aspects of Jewish music in antiquity and the early Middle Ages; and articles "Music" and "Musical Instruments," in *The Interpreter's Dictionary of the Bible* (1962), on the problems of music in the Bible according to the latest research, with full bibliographical list of references; E. GERSON-KIWI, "Jüdische Musik," in *Musik in Geschichte und Gegenwart* (1960), a reliable and comprehensive encyclopaedic article; ISRAEL ADLER, *La Pratique Musicale savante dans quelques communautés juives en Europe aux XVII^{e} et XVIII^{e} siècles*, 2 vol. (1966), the first fully-documented study on the attempts of Jewish communities to adopt Western art music before the Emancipation; CARL E. KRAELING, "Music in the Bible," and ERIC WERNER, "The Music of Post-Biblical Judaism," in *The New Oxford History of Music*, vol. 1, *Ancient and Oriental Music*, ed. by EGON WELLESZ (1957), with a useful introduction to scholarly studies of the subject.

Dance: DVORA LAPSON, "Modern Hebrew Dance," *American Dancer* (April 1, 1934), and her "Chassidic Dance," *Dance Observer* (November 1937), the best insight by a professional dancer; JULIAN MORGENSTERN, "The Etymological History of the Three Hebrew Synonyms for 'to dance' . . . ," *Journal of the American Oriental Society*, 36:321–332 (1917), an invaluable study of Biblical sources; ANNA DISLER, "La Danse Juive," *Archives Internationales de la Danse* (October 1935), an excellent monograph; LINCOLN KIRSTEIN, *Dance* (1936, reprinted 1969); JOHN MARTIN, *Introduction to the Dance* (1939); and CURT SACHS, *Eine Weltgeschichte des Tanzes* (1933; Eng. trans., *World History of the Dance*, 1937, reprinted 1963), three works that provide good historic perspective; ISRAEL ABRAHAMS, *Jewish Life in the Middle Ages*, rev. ed. (1932), reveals the importance of dance in the Diaspora.

Theatre: ZALMEN ZYLBERCWAIG and JACOB MESTEL, *Leksikon fun yidisn teater*, 2 vol. (1931–34); and ZALMEN ZYLBERCWAIG, *Leksikon fun yidisn teater*, 4 vol. (1959–69), the most complete directory and roster of the Yiddish theatre; SAMUEL J. CITRON, "Yiddish and Hebrew Drama," in BARRETT H. CLARK and GEORGE FREEDLEY (eds.), *A History of Modern Drama* (1947); DAVID S. LIFSON, *The Yiddish Theatre in America* (1965), and "The Yiddish Theatre," and "Israel and the Hebrew Theatre," in GEORGE FREEDLEY and JOHN A. REEVES (eds.), *A History of the Theatre*, 3rd rev. ed. (1968); and ZARA SHAKOW, *The Theatre in Israel* (1963), all comprehensive histories; ISAAC GOLDBERG, *The Drama of Transition* (1922), a most felicitous critique; HUTCHINS HAPGOOD, *The Spirit of the Ghetto* (1902, reprinted 1967), and MARK ZBOROWSKI and ELIZABETH HERZOG, *Life Is with People* (1952), fine evocations of the milieu that distilled the Yiddish theatre; DAVID BAER THEIKEL, *The Young Stage* (1940), indicates community sources of Yiddish talents.

Visual arts: Most items on Jewish visual arts can be found in LEO A. MAYER, *A Bibliography of Jewish Art* (1967). CECIL ROTH and BEZALEL NARKISS (eds.), *Jewish Art*, 2nd ed. (1971), is a general survey with 22 articles by different authors. A brief outline can be obtained from ERNEST NAMENYI, *L'Esprit* *de l'art juif* (1957; Eng. trans., ***The Essence of Jewish Art,*** (1960). On early and classical Jewish art, see E.E. URBACH, "The Rabbinical Laws of Idolatry in the 2nd and 3rd Centuries in the Light of Archaeological and Historical Facts," ***Israel Exploration Journal,*** 9:149–165, 229–245 (1959); CECIL ROTH, "Jewish Antecedents of Christian Art," ***Journal of the Warburg and Courtauld Institutes,*** 16:24–44 (1953); ERWIN R. GOODENOUGH, ***Jewish Symbols in the Greco-Roman Period,*** 13 vol. (1953–68); ELEAZAR L. SUKENIK, ***Ancient Synagogues in Palestine and Greece*** (1934); and CARL H. KRAELING, ***The Synagogue*** (1956). On medieval and modern European synagogues, see RACHEL WISCHNITZER, ***The Architecture of the European Synagogues*** (1964); and on American synagogues, the same author's ***Synagogue Architecture in the United States*** (1955); and A. KAMPF, ***Contemporary Synagogue Art*** (1966). On illuminated manuscripts, see JOSEPH GUTMANN, ***Images of the Jewish Past*** (1965); BEZALEL NARKISS, ***Hebrew Illuminated Manuscripts*** (1969), ***The Golden Haggadah*** (1970), and ***Machsor Lipsiae*** (1964); JACOB LEVEEN, ***The Hebrew Bible in Art*** (1944); and CECIL ROTH, ***The Kennicott Bible*** (1957) and ***The Sarajevo Haggada*** (1963). On ritual objects, see STEPHEN S. KAYSER (ed.), ***Jewish Ceremonial Art,*** 2nd ed. (1959); JOSEPH GUTMANN, ***Jewish Ceremonial Art*** (1964); and ABRAM KANOF, ***Jewish Ceremonial Art*** (1970). Modern art in Israel is discussed in BENJAMIN TAMUZ (ed.), ***Art in Israel*** (1965).

(I.B.S./D.S.L./B.Ns./E.We.)

Jewish Philosophy

The term Jewish philosophy refers to various kinds of reflective thought engaged in by persons identified as being Jews, in one sense or another. At times, as in the Middle Ages, this meant any methodical and disciplined thought, whether on general philosophical subjects or on specifically Judaic themes, pursued by Jews. In other eras, as in modern times, concentration on the latter has been considered a decisive criterion, so that philosophers who are Jewish but unconcerned with Judaism or the Jewish heritage and destiny in their thought are not ordinarily classified as Jewish philosophers.

PRE-HELLENISTIC AND HELLENISTIC THOUGHT

The concept of Wisdom

Bible and Apocrypha. Philosophy arose in Judaism under Greek influence; however, a kind of philosophical approach may be discerned in early Jewish religious works apparently subject to little or no Greek influence. The books of Job and Qohelet (Ecclesiastes) were favourite works of medieval philosophers, who took them as philosophical discussions untinged by theological preconceptions. The book of Proverbs introduces, in an apparently theological context, the concept of Wisdom (Hokhma), which was to have a primordial significance for Jewish philosophical and theological thought, and presents it as the first and favourite of God's creations. It is also praised in the book of the Wisdom of Jesus the Son of Sirach (Ecclesiasticus) as instilled by God into all his works and granted in abundance to those he loves. It is sometimes equated with fearing God and keeping the Law; however, in other passages piety seems to be regarded as superior to Wisdom. The Wisdom of Solomon, probably originally written in Greek, praises Wisdom, which is held to be an image of God's goodness and a reflection of the eternal light. God is said to have given the author knowledge of the composition of the world, of the powers, the elements, the nature of animals, the divisions of time, and the positions of the stars. In its vocabulary and perhaps in some of its doctrines, the work shows the influence of Greek philosophical conceptions. It has had considerable influence on Christian theology.

Philo Judaeus. The first systematic attempt to apply Greek philosophical concepts to Jewish doctrines was made by Philo Judaeus (Philo of Alexandria) in the 1st century CE (Common Era, which corresponds to AD). Philo, a scholar who combined Greek and Jewish learning, was influenced by Platonic and Stoic writings, and probably also by certain postbiblical Jewish beliefs and speculations. He apparently had some knowledge of the Oral Law, which was being evolved in his time, and he also knew of the Essenes, a contemporary rigorous sect, whom he praised highly.

The concept of Logos

Philo's main contribution was interpretative. He provided Jewish conceptions with the hallmark of intellectual and cultural respectability by stating them in Greek philosophical terms; he also showed that many Greek notions were consonant with Jewish doctrine, as he conceived it, and with the allegorical sense of biblical texts, as he read them. He had two schemes of reference—Jewish religious tradition and Greek philosophy—and the fact that he took care to stress the primacy of the former may have been more than mere lip service. It may be argued with some plausibility that in central points of his thought, such as his conception of the Logos (the divine Reason or Word), Philo used philosophical notions as trappings for an originally religious belief. A main function of the Logos in his thought is to serve as an intermediary between the transcendent, unknowable God and the world, a view that probably has a close connection with the view of his Jewish contemporaries concerning the world of God, by means of which he accomplishes his designs. On basic philosophical or theological problems, such as the creation of the world or the freedom of will, Philo's writings provide either vague or contradictory answers. He placed mystic ecstasy, of which he may have had personal experience, above philosophical and theological speculations.

Philo's approach, his method of interpretation, and his way of thinking, as well as some of his conceptions—primarily that of the Logos—exerted a considerable influence on early Christian thought but not, to any comparable extent, upon Jewish thought in that period. Later, in the Middle Ages, knowledge of Philo among Jews was either very slight or nonexistent. Not until modern times was his role in Jewish religious thought recognized.

Other ancient sources. Some traces of a knowledge of popular, mainly Stoic philosophy may be found in the Mishna, a codification of the Oral Law composed in Palestine in the 2nd century CE, and in the subsequent Talmudic literature set down in writing in Palestine and Babylonia. On the whole these traces are rather slight. Some scholars believe that the influence of Greek philosophy on Palestinian Jewry was far-reaching, but the case, to say the least, is not proven. Jewish theological and cosmological speculations occur in the Midrashim (plural of Midrash,), which, under the guise of interpreting biblical verses, propound allegorical interpretations, legends, and myths, and in the ***Sefer yetzira*** ("Book of Creation"), a work that is a combination of a cosmogony and a grammar and that was fictitiously attributed to Abraham. There is no clear evidence of the period in which it was written; both the 3rd century and the 6th or 7th century have been suggested. The book became a key work in later Jewish mysticism.

MEDIEVAL PHILOSOPHY

Islāmic background

In the 9th and 10th centuries, after a long hiatus, systematic philosophy and ideology reappeared among Jews, a phenomenon indicative of their accession to Islāmic civilization. The evolution of Islam in the 9th and 10th centuries showed that Greek scientific and philosophic lore could be separated, at least to some extent, from its pagan associations and could be transformed into another language and another culture; it also tended to show that a culture in which the sciences and philosophy or the sciences and theology or both of these combinations were an indispensable part could be based upon a monotheistic prophetic religion that in all relevant essentials, including adherence to a basic religious law, was closely akin to Judaism. The question of whether philosophy is compatible with religious law (the answer being sometimes negative) constituted the main theme of the foremost medieval Jewish thinkers. From approximately the 9th to the 13th centuries Jewish philosophical and theological thought participated in the evolution of Islamic philosophy and theology and manifested only in a limited sense a specifically Jewish character. Jewish philosophers showed no particular preference for philosophic texts written by Jewish authors over those composed by Muslims, and in many cases the significant works of Jewish thinkers constitute a reply or a reaction to the ideas of Islāmic philosophic and scientific writings.

Jewish kalām. Although several Jewish intellectuals in 9th–10th-century Babylonia were steeped in Greek philosophy, the most productive and influential Jewish thinkers of this period represented a very different tendency, that of the Mu'tazilite *kaldm*. *Kalām* (literally, "speech") is an Arabic term used both in Islāmic and in Jewish vocabulary to designate several theological schools that were ostensibly opposed to Greek, particularly Aristotelian, philosophy. The Aristotelians, both Islāmic and Jewish, regarded *kalām* theologians (called the Mutukallimiin) with a certain contempt, holding them to be mere apologists, watchdogs of religion, and indifferent to truth. Herein they did not do justice to their adversaries, for many representatives of the Mu'tazilite school of *kaldm*, formed in the 8th century, displayed a genuine speculative impulse. Its theology, forged in disputes with Zoroastrians, Manichaeans, and Christians, claimed to be based on reason.

Sa'adia ben Joseph. This belief in reason, as well as some of the tenets of Mu'tazilite theology, were taken over by Sa'adia ben Joseph, who was also influenced, either directly or through the intermediary of an Arabic philosopher, by the arguments of a Christian 6th-century philosopher, John Philoponus, against certain Aristotelian and Neoplatonic positions. Sa'adia's main theological work, *Kitdb al-amānāt wa al-iʿtiqādāt (Beliefs and Opinions),* is modelled on similar Mu'tazilite treatises and on a Mu'tazilite classification of theological subject matter known as the Five Principles. Like many Muʿtazilite authors, Sa'adia starts out by setting forth in his introduction a list and theory of the various sources of knowledge.

Sa'adia's four sources of knowledge

Sa'adia distinguished four sources: (1) the five senses, (2) the intellect, or reason, (3) necessary inferences, and (4) reliable information given by trustworthy persons. In Sa'adia's sense of the word, intellect, or reason (*al-ʿaql*), means first and foremost an immediate, a priori cognition, independent of sense experience. In *Beliefs and Opinions* the intellect is characterized as having immediate ethical cognitions—that is, as discerning what is good and what is evil—in opposition to the medieval Aristotelians, who did not regard even the most general ethical rules as a priori cognitions. The third source of knowledge comprises inferences of the type "if there is smoke, there is fire," which are based on data furnished by the first two sources of knowledge. The fourth source of knowledge is meant to validate the teachings of Scripture and of the religious tradition. Teachings of Scripture must be held true because of the trustworthiness of the men who propounded them. One of the main purposes of the work is to show that the knowledge deriving from the fourth source concords with that discovered by means of the other three, or, in other words, that religion and human reason agree.

Sa'adia opposed Aristotle's view that the natural order was eternal. He held, with other partisans of the Mu'tazilite *kaldm*, that the demonstration of the temporal creation of the world must precede and pave the way for the proof of the existence of God the Creator. Given the demonstrated truth that the world has a beginning in time, it can be proved that it could have been produced only through the action of a creator. It can further be proved that there can have been only one Creator.

Divine justice and human ethical judgment

The theology of Sa'adia, like that of the Mu'tazilites, hinges on two principles: the unity of God and the principle of justice. The latter principle takes issue with the view (widespread in Islām and present also in Judaism) that the definition of what is just and what is good depends solely on God's will, to which none of the moral criteria found among men is applicable; according to this view, a revelation from God can convert an action now generally recognized as evil into a good action. Against this way of thinking, Sa'adia and the Mu'tazilites believed that being good and just or evil and unjust are intrinsic characteristics of human actions and cannot be changed by divine decree. The notions of justice and of good, as conceived by man, are binding on God himself. Since, according to Sa'adia, man has a priori knowledge of good and evil, just and unjust, the fact that human ethical judgments are valid for God means that man's ethical cognitions are also those of the Deity.

The function of religious law—of central importance in traditional Judaism and Islām—is to impose on man the accomplishment of good actions and to prohibit bad ones. Because Sa'adia believed that man has a priori knowledge of good and evil and that this knowledge coincides with the principles underlying the most important portions of the revealed law, he was forced to ask the question whether this law is not superfluous. He could, however, point out that, whereas the human intellect recognizes that certain actions—for instance, murder or theft—are evil, it cannot by itself discover the best possible definition of what constitutes a particular transgression, nor can it, on its own, determine the punishment appropriate for a transgression. On both points. Sa'adia asserted, the commandments of religious law give the best possible answers.

The commandments that accord with the behests of the human intellect were designated by Sa'adia as the intellectual, or rational, commandments. According to him, they include the duty of manifesting gratitude to the Creator for the benefits he has bestowed upon man. Sa'adia recognized that a considerable number of commandments—for instance, those dealing with the prohibition of work on the sabbath—do not belong to this category. He held, however, that the obligation to obey them may be derived from the rational commandment that makes it incumbent upon man to be grateful to God, for such gratitude entails obedience to his orders.

The Karaites. Sa'adia's adoption of the rational Mu'tazilite theology was a part of his overall activity, directed toward the consolidation of rabbinical Judaism (based on the Mishna and Talmud), which was being attacked by the Karaites. This Jewish sect, founded by Anan ben David in the 8th century, rejected the authority of the Oral Law—that is, of the Mishna and the Talmud. In the 10th century and afterward, the Karaites accepted as their guides the Bible (Old Testament) and human reason, in the Mu'tazilite sense of the word. Their professed freedom from any involvement with postbiblical Jewish religious tradition facilitated a rational approach to theological doctrine. This approach led the Karaite authors to criticize their opponents, the adherents of rabbinical Judaism, for holding anthropomorphic beliefs based, in part, on texts of the Talmudic period. Karaite authors propounded, in conceptual terms, a theology of Jewish history in exile ***(galut).*** Life in exile is a diminished existence; nevertheless, the good or bad actions of the Jewish people (rather than their material strength or weakness) affect the course of history. Redemption may come when all Jews are converted to Karaism.

Karaite rationalism and atomism

The Karaites adopted, wholesale, Mu'tazilite *kalām*, including its atomism. The Mu'tazilite atomists held that everything that exists consists of minute, discrete parts. This applies not only to bodies but also to space, to time, to motion, and to the "accidents"—that is, qualities, such as colour—which the Islāmic and Jewish atomists regarded as being joined to the corporeal atoms (but not determined by them, as had been believed by the Greek Atomists). An instant of time or a unit of motion does not continue the preceding instant or unit. All apparent processes are discontinuous, and there is causal connection between their successive units of change. The fact that cotton put into fire generally burns does not mean that fire is a cause of burning; rather, it may be explained as a "habit," signifying that this sequence of what is often wrongly held to be cause and effect has no character of necessity. God's free will is the only agent of everything that occurs, with the exception of one category—human actions. These are causes that produce effects; for instance, a man who throws a stone at another man, who is then killed, directly brings about the latter's death. This inconsistency on the part of the theologians was necessitated by the principle of justice, for it would be unjust to punish a man for a murder that was a result not of his action but of God's.

This grudging admission that causality exists in certain strictly defined and circumscribed cases was occasioned by moral, not physical, considerations.

Jewish Neoplatonism. ***Isaac Israeli.*** Outside Babylonia, philosophical studies were pursued by Jews in the 9th and 10th centuries in Egypt and in the Maghrib (northwest Africa). The outstanding figure was Isaac ben Solomon Israeli, an Egyptian-born North African who has been called "the first Jewish Neoplatonist." In his philosophical works, such as the "Book of Elements" and the "Book of Five Substances," he drew largely upon a 9th-century Muslim popularizer of Greek philosophy, Abii Yiisuf Yaʿqūb al-Kindi, and also, in all probability, upon a lost pseudo-Aristotelian text. The peculiar form of Neoplatonic doctrine that seems to have been set forth in this text had, directly and indirectly, a considerable influence on medieval Jewish philosophy.

According to Israeli, God creates through his will and power. The two things that were created first were form, identified with wisdom, and matter, which is designated as the genus of genera (the classes of things) and which is the substratum of everything, not only of bodies, but also of incorporeal substances. This conception of matter seems to derive from the Greek Neoplatonists Plotinus and Proclus, particularly from the latter. In Proclus' opinion generality was one of the main criteria for determining the ontological priority of an entity (relative place in the levels of reality). Matter, because of its indeterminacy, obviously has a high degree of generality; consequently, it figures among the entities having ontological priority. According to the Neoplatonic view, which Israeli seems to have adopted, the conjunction of matter and form gives rise to the intellect. A light sent forth from the intellect produces the rational soul, and in its turn it gives rise to the vegetative soul.

Israeli's doctrine of prophecy seems to be the earliest Jewish philosophical theory attributing prophecy to the influence of the intellect on the imaginative faculty. According to Israeli, this faculty receives from the intellect spiritual forms that are intermediate between corporeality and spirituality. This explanation implies that these forms, "with which the prophets armed themselves," are inferior to purely intellectual cognitions.

Solomon ibn Gabirol. In essentials the schema of creation and emanation propounded by Isaac Israeli and his Neoplatonic source or sources was taken over by Solomon ibn Gabirol, a celebrated 11th-century Hebrew liturgical poet, who seems to have been the earliest Jewish philosopher of Spain. His chief philosophical work, "Fountain of Life," written in Arabic, has been preserved in full only in a 12th-century Latin translation entitled ***Fons vitae.*** This work, which makes no reference to Judaism or to specifically Jewish doctrines, is a didactic dialogue between a disciple and a master who teaches him true philosophical knowledge. Despite its prolixity and many contradictions, it is an impressive work. Few medieval texts so effectively communicate the Neoplatonic conception of the existence of a number of planes of being that differ according to their ontological priority, the derivative and inferior ones constituting a reflection in a grosser mode of existence of those that are prior and superior.

Ibn Gabirol's doctrine of divine will

A central conception in Ibn Gabirol's philosophy is concerned with the divine will, which appears to be both part of and separate from the divine essence. Infinite according to its essence, the will is finite in its action. It is described as pervading everything that exists and as being the intermediary between the divine essence and matter and form. Will was one of a number of traditional appellations applied in various medieval theologies to the entity intermediate between the transcendent Deity and the world or to the aspect of the Deity involved in creation. According to a statement in ***Fons vitae,*** matter derives from the divine essence, whereas form derives from the divine will. This suggests that the difference between matter and form has some counterpart in the Godhead and also that universal matter is superior to universal form. Some of Ibn Gabirol's statements seem to bear out the impression of superiority of universal matter; other passages, however, appear to imply a superiority of universal form.

Form and matter, whether they be universal or particular, exist only in conjunction. All things, with the sole exception of God, are constituted through the union of the two, the intellect no less than the corporeal substance. In fact, the intellect is the first being in which universal matter and form are conjoined. The intellect contains and encompasses all things. It is through the grasp of the various planes of being, through ascending in knowledge to the world of the intellect and apprehending what is above it—the divine will and the world of the Deity—that man may "escape death and reach "the source of life."

Judah ha-Levi. Judah ben Samuel ha-Levi (c. 1075–c. 1141), also from Spain and a celebrated Hebrew poet, was the first medieval Jewish thinker who consciously and consistently based his thought upon arguments drawn from Jewish history. His views are set forth in an Arabic dialogue, ***al-Hazari*** (Hebrew ***Sefer ha-Kuzari),*** the full title of which is translated as "The Book of Proof and Argument in Defense of the Despised Faith." This work is usually called ***Kuzari, i.e.,*** "the Khazar."

The ***Kuzari***

Basing his narrative on the historical fact that the Khazars (a Turkic-speaking people in Central Eurasia) were converted to Judaism (c. 740), ha-Levi relates that their King, a pious man who did not belong to any of the great monotheistic religions, dreamed of an angel, who said to him, "Your intentions are pleasing to the Creator, but your works are not." To find the correct way of pleasing God, the King seeks the guidance of a philosopher, a Christian, a Muslim, then, finally, after hesitating to have recourse to a representative of a people degraded by its historical misfortune, of a Jewish scholar who converts him to Judaism. The words of the angel heard in a dream may, in accordance with both religious and philosophical doctrine, be regarded as a kind of revelation. The use of this element of the story enabled ha-Levi to suggest that it is not the spontaneous activity of human reason that impels man to undertake the quest for the true religion; for this, one needs the gift of prophecy, or, at least, a touch of the prophetic faculty (or a knowledge of the revelations of the past).

The argument of the philosopher whose advice is sought by the King brings this point home. This disquisition is a brilliant piece of writing, for it lays bare the essential differences between the Aristotelian God, who is totally ignorant of and consequently wholly indifferent to human individuals, and the God of religion. Within the framework of philosophical doctrine, the angel's words are quite meaningless. Not only is the God of the philosophers, who is a pure intellect, not concerned with man's works, but the (cultural) activities, involving both mind and body, to which the angel clearly referred, cannot, from the philosophical point of view, either help or hinder man in the pursuance of the philosophers' supreme goal—the attainment of union with the active intellect, a "light" of the divine nature. This union was supposed to confer knowledge of all intelligible things. Thus, man's supreme goal was here held to be of a purely intellectual nature.

In opposition to the philosopher's faith, the religion of ha-Levi's Jewish scholar is based upon the fact that God may have a close, direct relationship with man, who is not conceived primarily as a being endowed with intellect. The postulate that God can have intercourse with a creature made of the disgusting materials that go into the composition of the human body is scandalous to the King and prevents his acceptance of the doctrine concerning prophecy, expounded by the Muslim sage (just as the extraordinary nature of the Christological dogmas deters him from adopting Christianity).

Knowledge of God through the contemplation of Jewish history

The Jewish scholar's position is that it is contemplation not of the cosmos but of Jewish history that procures knowledge of God. Ha-Levi was aware of the odium attaching to the doctrine of the superiority of one particular nation; he held, however, that only this doctrine explains God's dealing with mankind, which, like many other things, reason is unable to grasp. The controversies of the philosophers serve as proof of the failure of human intelligence to find valid solutions to the most important problems.

Ha-Levi's dialogue was also directed against the Karaites. He shows the necessity and celebrates the efficacy of a blind, unquestioning adhesion to tradition, which the Karaites rejected. Yet, he expounds a theology of Jewish exile that seems to have been influenced by Karaite doctrine. According to ha-Levi, even in exile the course of Jewish history is not determined like that of other nations by natural causes, such as material strength or weakness; the decisive factors are the religious observance or disobedience of the Jews. The advent of Christianity and of Islām prepares the other nations for conversion to Judaism, an event that will occur in the eschatological period (at the end of history).

Other Jewish thinkers c. 1050–c. 1150. During the period comprising the second half of the 11th century and the first half of the 12th century, many other Jewish thinkers appeared in Spain. Bahya ben Joseph ibn Pakuda wrote one of the most popular books of Jewish spiritual literature, *Kitāb al-hidāyah ilā farā'iḍ al-qulūb* ("Guidance to the Duties of the Heart"), which combines a theology influenced by Sa'adia with a moderate mysticism inspired by the teachings of the Muslim Siifis (mystics). The commandments of the heart—that is, those relating to men's thoughts and sentiments—are contrasted with the commandments of the limbs—that is, the Mosaic commandments enjoining or prohibiting certain actions. Bahya maintained that both sets of commandments should be observed (thus rejecting the antinomian position), but made clear that first and foremost he was interested in the commandments of the heart.

Abraham bar Hiyya, an outstanding mathematician, an astrologer, and a philosopher, outlined in *Megillat* ha-Megalle ("Scroll of the Revealer") a view of Jewish history that is rather reminiscent of ha-Levi but does not emphasize to the same degree the uniqueness of that history; it is also set forth in much less impressive fashion. Living in Barcelona under Christian rule, Bar Hiyya wrote his scientific and philosophical treatises not in Arabic but in Hebrew. Hebrew was also used by Abraham ibn Ezra (died *c.* 1167), a native of Spain who travelled extensively in Christian Europe. His commentaries on the Bible contributed to the diffusion among the Jews of Greek philosophical thought, to which Ibn Ezra made many, although as a rule disjointed, references. His astrological doctrine had a significant influence on some philosophers.

A borderline case: Abii al-Barakiit

The last outstanding Jewish philosopher of the Islāmic East, Abii al-Barakāt al-Baghdādī (who died as a very old man sometime after 1164), also belongs to this period. As a borderline case he illustrates a certain indeterminacy in the definition of a Jewish thinker. An inhabitant of Iraq, he was converted to Islām in his old age (for reasons of expediency, according to his biographers). His philosophy appears to have had a strong impact on Islāmic thought, whereas its influence upon Jewish philosophy and theology is very hard to pin down and may be practically nonexistent. His chief philosophical work, *Kitāb al-Muʿtabar* ("The Book of That Which Has Been Established by Personal Reflection"), has very few references to Jewish texts or topics. Abū al-Barakiit rejects Aristotelian physics completely; according to him, time is the measure of being, and not, as Aristotle taught, the measure of motion, and he replaces Aristotle's bidimensional concept of place with the tridimensional notion of space, the existence of which is independent of the existence of bodies.

Jewish Aristotelianism. With regard to the adoption of Aristotelianism, including systems that in many essentials stem from but also profoundly modify the pure Aristotelian doctrine, there is a considerable time lag between the Islāmic East, on the one hand, and Muslim Spain and the Maghrib, on the other.

Abraham ibn Dd'iid. Abraham ibn Dā'ūd (12th century), who is regarded as the first Jewish Aristotelian of Spain, was primarily a disciple of Avicenna, the great 11th-century Islāmic philosopher. According to a not unlikely hypothesis, he may have translated or helped to translate some of Avicenna's works into Latin, for Ibn Dā'ūd lived under Christian rule in Toledo, a town that in the 12th century was a centre for translators. His historical treatises, written in Hebrew, manifest his desire to familiarize his coreligionists with the historical tradition of the Latin world, which at that time was alien to most of them. But his philosophical work, Sefer *ha-emuna ha-rama* ("Book of Sublime Faith"), written in 1161 in Arabic, shows few, if any, signs of Christian influence.

The doctrine of emanation, set forth in the "Book of Sublime Religion," describes in the manner of Avicenna the procession of the ten incorporeal intellects, the first of which derives from God. This intellect produces the second intellect, and so on. Ibn-Dā'ūd questioned in a fairly explicit manner Avicenna's views on the way the second intellect is produced; his discipleship did not by any means spell total adherence. Ibn-Dā'ūd's psychology was also, and more distinctively, derived from Avicenna. The argumentation leading to a proof that the rational faculty is not corporeal attempts to derive the nature of the soul from the fact of immediate self-awareness. Like Avicenna, Ibn-Dii'iid tended to found psychology on a theory of consciousness.

Ibn-Dā'ūd often referred to the accord that, in his view, existed between philosophy and religious tradition. As he remarked, the "Book of Sublime Faith" was not meant to be read either by readers who, in their simplicity, are satisfied with what they know of religious tradition or by those who have a thorough knowledge of philosophy. It was intended for readers of one type only, those who, being, on the one hand, acquainted with the religious tradition and having, on the other, some rudiments of philosophy, are "perplexed." It was for the same kind of people that Maimonides wrote his Guide of the Perplexed.

The greatest medieval Jewish philosopher

Maimonides. Maimonides (Moses ben Maimon, 1135–1204), a native of Spain, is incontestably the greatest name in Jewish medieval philosophy, but his reputation is not derived from any outstanding originality in philosophical thought. Rather, the distinction of Maimonides, who is also the most eminent codifier of Jewish religious law, is to be found in the vast scope of his attempt, in the *Dalālat al-ḥā'irīn* (Guide of the Perplexed), to safeguard both religious law and philosophy (the public communication of which would be destructive of the law) without suppressing the issues between them and without trying to impose, on the theoretical plane, a final, universally binding solution of the conflict.

As Maimonides made clear in his introduction to the Guide, he regarded his self-imposed task as perilous, and he therefore had recourse to a whole system of precautions destined to conceal his true meaning from the people who, lacking the necessary qualifications, might misread the book and abandon observance of the law. According to Maimonides' explicit statement, these precautions include deliberately contradictory statements meant to mislead the undiscerning reader. The apparent or real contradictions that may be encountered in the Guide are perhaps most flagrant in Maimonides' doctrine concerning God. There seems to be no plausible hypothesis capable of explaining away the differences between the following three views:

1. God has an eternal will that is not bound by natural laws. Through an act of his will, he created the world in time and imposed on it the order of nature. This creation is the greatest of miracles; only if it is admitted can other miracles, which interfere with the causally determined concatenations of events, be regarded as possible. The philosophers' God, who is not free to cut the wings of a fly, is to be rejected. This conception is in keeping with the traditional religious view of God and is avowedly adopted by Maimonides because failure to do so would undermine religion.

2. Man is incapable of having any positive knowledge concerning God. No positive attributes—*e.g.*, wisdom or life—can be ascribed to God. Contrary to the attributes predicated of created beings, the divine attributes are strictly negative; they state what God is not: for instance,

he is not not-wise, and such a statement is not a positive assertion. Hence, only a negative theology is possible—saying what God is not. The way God acts can, however, be known. This knowledge is to be found in natural science.

3. God is an intellect. The formula current among medieval philosophers that maintains that in God the knowing subject, the object known, and the act of intellectual knowledge are identical derives from Aristotle's thesis that God knows only himself. Maimonides, however, in adopting the formula, interpreted it in the light of human psychology and epistemology (theory of knowledge), pointing out that, according to a theory of Aristotle, the act of human (not only of divine) cognition brings about an identity of the cognizing subject and cognized object. The parallel drawn by Maimonides between the human and the divine intellect quite evidently implies a certain similarity between the two; in other words, it is incompatible with the negative theology of other passages of the *Guide*. Nor can it be reconciled with his theological doctrine that the structure of the world—created in time—came into being through the action of God's will.

Maimonides' doctrine of prophecy

The enigma of the *Guide* would be nonexistent if Maimonides could be held to have believed that truth can be discovered in a suprarational way, through revelations vouchsafed to the prophets. This, however, is not the case. Maimonides held that the prophets (with the exception of Moses) combine great intellectual abilities, which qualify them to be philosophers, with a powerful imagination. The intellectual faculty of the philosophers and the prophets receives an overflow from the active intellect. In the case of the prophets, this overflow not only brings about intellectual activity but also passes over into the imaginative faculty, giving rise to visions and dreams. The fact that prophets have a strong imagination gives them no superiority in knowledge over philosophers, who do not have it. Moses, who belonged to a higher category than did the other prophets, did not have recourse to imagination.

The laws and religion as instituted by Moses are intended not only to ensure the bodily welfare and safety of the members of the community but also to facilitate the attainment of intellectual truths by individuals gifted enough to uncover the various hints embodied in religious laws and practices. This does not mean that all the beliefs inculcated by Judaism are true. Some indeed express philosophical truths, although in an inaccurate way, in a language suited to the intellectual capacity of the common people, who in general cannot grasp the import of the dogmas they are required to profess. Other beliefs, however, are false but necessary for the preservation of a public order upholding justice—*e.g.*, the belief that God is angry with wrongdoers.

As far as the Law—that is, the religious commandments—is concerned, two aspects of Maimonides' position may be distinguished. First, he maintained that it is unique in its excellence and valid for all time. This profession of faith, at least with regard to its assumptions about the future, lacked philosophical justification; however, it could be regarded as necessary for the survival of Judaism. Second, he asserted that certain precepts of the Mosaic Law were related to specific historical situations and the need to avoid too sharp a break with popular customs and practices, for instance, the commandments concerning sacrifice.

Influence of the *Guide of the Perplexed*

For at least four or five centuries the *Guide of the Perplexed* exercised a very strong influence in the European centres of Jewish thought; in the 13th century, when the *Guide* was twice translated into Hebrew, these centres were Spain, the south of France, and Italy. Rather paradoxically, in view of the unsystematic character of Maimonides' exposition, it was used as a standard textbook of philosophy and condemned as such when the teaching of philosophy came under attack. The performance of this function by the *Guide* was rendered possible, or at least facilitated by, the fact that from the 13th century onward the history of Jewish philosophy in European countries acquired a continuity it had never had before. This development seems to have resulted from the substitution of Hebrew for Arabic as the language of philosophical exposition. Because of the existence of a common and relatively homogeneous philosophical background—the Hebrew texts were much less numerous and less diverse than Arabic philosophical works—and the fact that Jewish philosophers reading and writing in Hebrew read the works of their contemporaries and immediate predecessors, something like a dialogue can be discerned. In striking contrast to the immediately preceding period, European Jewish philosophers in the 13th century and after frequently devoted a very considerable part of their treatises to discussions of the opinions of other Jewish philosophers. That many of the Jewish philosophers in question wrote commentaries on the *Guide* undoubtedly furthered this tendency.

Influence of Averroës and Christian Scholastics on medieval Jewish philosophy

Averroists. The influence of Maimonides' great Islāmic contemporary Averroes, many of whose commentaries and treatises were translated into Hebrew, was second only to that of Maimonides on Jewish intellectual development. Indeed, it may be argued that for philosophers, as distinct from the general reading public, it often came first. In certain cases commentators on the *Guide* tend, in spite of the frequent divergences between the two philosophers, to quote Averroes' opinions in order to clarify those of Maimonides.

The apparently significant influence of Christian Scholastic thought on Jewish philosophy was often not openly acknowledged by Jewish thinkers in the period beginning with the 13th century. Samuel ibn Tibbon, one of the translators of the *Guide* into Hebrew and a philosopher in his own right, remarked on the fact that the philosophical sciences were more widely known among Christians than among Muslims. Somewhat later, at the end of the 13th century and after, Jewish scholars in Italy translated into Hebrew varied texts of Thomas Aquinas and other Christian Scholastics; not infrequently, some of them acknowledged the debt they owed their Christian masters. In Spain and in the south of France, a different convention seems to have prevailed up to the second half of the 15th century. Whereas Jewish philosophers of these countries felt no reluctance about referring to Greek, Arabic, and other Jewish philosophers, as a rule they refrained from citing Christian thinkers whose views had, in all probability, influenced them. In the case of certain Jewish thinkers, this absence of reference to the Christian Scholastics served to disguise the fact that in many essentials they were representative of the philosophical trends, such as Latin Averroism, that were current among the Christian Scholastics of their time.

Quite evident is the resemblance between certain views professed by the Latin Averroists and the parallel opinions of Isaac Albalag, a Jewish philosopher who lived in the second half of the 13th century, probably in Catalonia, Spain, and who wrote a commentary in Hebrew on the *Tahāfut al-falāsifah* ("The Inconsistencies of the Philosophers"), an exposition of Avicenna's doctrine written by the Muslim philosopher al-Ghazālī. Albalag's assertion that both the teachings of the Bible and the truths demonstrated by reason must be believed even if they are contradictory clearly poses the question whether some historical connections exist between this view and the Latin Averroist doctrine that there are two sets of truths—the religious and the philosophical—and that these are not necessarily in accord. In most other points Albalag was a follower of the system of Averroës himself. This philosophical position may be exemplified by his rejection of the view that the world was created in time. He professed, it is true, to believe in what he called "absolute creation in time." This expression, however, merely signifies that at any given moment the continued existence of the world depends on God's existence, an opinion that is essentially in harmony with Averroes.

Joseph Caspi, a prolific 14th-century philosopher and exegetical commentator, maintained a somewhat unsystematic philosophical position that seems to have been influenced by Averroes. He expressed the opinion that knowledge of the future, including that possessed by God himself, is of a probabilistic nature. The prescience of the prophets is of the same nature. It is more than likely that Caspi's interest in this problem had some connection with the debate about future contingencies in which Christian Scholastics were engaged at that time.

Moses Narboni

Moses of Narbonne, or Moses Narboni, who lived in the south of France in the 14th century, was, like many other Jewish writers of this period, mainly a writer of commentaries. He wrote commentaries on biblical books, on treatises of Averroes, and on Maimonides' *Guide*. In his commentary on the *Guide,* Narboni often interprets the earlier Jewish philosopher's opinions by recourse to Averroes' views. Narboni also expounded and gave radical interpretations to certain conceptions that he understood as implied in the *Guide*. According to Narboni, God participates in all things, because he is the measure of all substances. God's existence appears to be bound up with that of the world, to which he has a relation analogous to that existing between a soul and its body (a comparison already made in the *Guide).*

Gersonides. Gersonides (Levi ben Gerson), another 14th-century Jewish philosopher born in the south of France, wrote the systematic philosophical work *Milḥamot Adonai* ("The Wars of the Lord") as well as many philosophical commentaries. Gersonides apparently never explicitly mentioned Christian Scholastic philosophers; he cited Greek, Arabic, and Jewish thinkers only, and in many ways his system appears to have stemmed from the doctrines of Maimonides or Averroes, regardless of whether he agreed with them or not. For example, he explicitly rejected Maimonides' doctrine of negative theology. A comparison of his opinions and of the particular problems that engaged his attention, with the views and debates found in Scholastic writings of his period, however, suggests that he was also influenced by the Latins on certain points.

Gersonides disagreed both with the Aristotelian philosophers who maintained the eternity of the world and with the religious partisans who believed in the creation of the world in time out of nothing. He maintained that God created the world in time out of a pre-existent body lacking all form. As conceived by Gersonides, this body seems to be similar to primal matter.

The problem of human freedom of action and a particular version of the problem of God's knowledge of future contingencies form an important part of Gersonides' doctrine. Gersonides, who, unlike the great Jewish and Muslim Aristotelians, believed in astrology, held that all happenings in the world except human actions are governed by a strict determinism. God's knowledge does not extend to individual human acts but embraces the general order of things; it grasps the laws of universal determinism but is incapable of apprehending events resulting from man's freedom. Thus, the object of God's knowledge is an ideal world order, which differs from the real world insofar as the latter is in some measure formed according to man's free will.

In political and social doctrine there is a fundamental difference between Maimonides and Gersonides. Gersonides does not appear to have assigned to the prophets any political function; according to him, their role consists in the prediction of future events. The providence exercised by the heavenly bodies ensures the existence in a given political society of men having an aptitude for and exercising the handicrafts and professions necessary for the survival of the community. He remarked that in this way the various human activities are distributed in a manner superior to that outlined in Plato's *Republic*. Thus, he rejected explicitly Plato's political philosophy, which, having been adapted to a society ruled through the laws promulgated by a prophet (Muhammad), had been an important element of Jewish philosophy in the Arabic period.

Hasdai Crescas. Hasdai ben Abraham Crescas (1340–1410), a Spanish-Jewish thinker, like Gersonides had thorough knowledge of Jewish philosophy and partial knowledge of Islāmic philosophy, and both seem to have been influenced by Christian Scholastic thought; moreover, in certain important respects Crescas was influenced by Gersonides himself. In Crescas' main work, *Or Adonai* ("The Light of the Lord"), however, one of his objectives, quite contrary to Gersonides, was to expose the weakness and insufficiency of Aristotelian philosophy. This attitude may be placed in the wider context of the return to religion itself, as opposed to the Aristotelian rationalization of religion, and the vogue of Kabbala (Jewish theosophical mysticism), both of which were characteristic features of Spanish Jewry in Crescas' time. This change in attitude has been regarded as a reaction to the increasing precariousness of the position of the Jewish community in Spain.

Anti-Aristotelianism

The low estimation of the certainties and the rationalistic arrogance of the medieval Aristotelians coincided chronologically with a certain disintegration of and disaffection toward classical Aristotelian Scholasticism. Relevant to this decline were the so-called voluntarism of Duns Scotus, the Nominalism of William of Ockham and other 13th–14th century Christian Scholastics, and the development, in the 14th century and after, of anti-Aristotelian physics at the University of Paris and elsewhere. Significantly there is a pronounced resemblance between Crescas' views and two of these trends, Scotism and the "new" physics.

Crescas accepted Gersonides' view that divine attributes cannot be negative, but unlike his predecessor he centred his explanation of the difference between the attributes of God and those of created existents on the antithesis between an infinite being and finite beings. It is through infinitude that God's essential attributes—wisdom, for instance—differ from the corresponding and otherwise similar attributes found in created beings. In Crescas', as in Spinoza's, doctrine (see below), God's attributes are also infinite in number. The central place assigned to the thesis of God's infinity in Crescas' system suggests the influence of Duns Scotus' theology, which is similarly founded upon the concept of divine infinity.

The problem of the infinite approached from an altogether different angle was one of the main themes of Crescas' critique of Maimonides' 25 propositions; these propositions, concerned mainly with Aristotelian physical doctrines, had been set forth in the *Guide* as the basis of Maimonides' proof for the existence of God. Crescas' declared purpose in criticizing and rejecting several of these propositions was to show that the traditional Aristotelian proofs (founded in the first place on physical doctrines) were not valid. In the course of his critique, Crescas attempted to disprove the Aristotelian thesis that the existence of an actual infinite is impossible. He held that space is not a limit but a tridimensional extension, that it is infinite, and that, contrary to Aristotle, the existence of a vacuum and of more worlds than one is possible. He also criticized as being impossible the thesis of the Aristotelian philosophers that there exists an infinite number of causes and effects, which have order and gradation. This thesis refers not to a temporal succession of causes and effects that have a similar ontological status but to a vertical series, descending from God to the lowest rung in creation. His attacks were likewise directed against the Aristotelians' conceptions of time and of matter.

Crescas' fundamental opposition to Aristotelianism is perhaps most evident in his rejection of the conception of intellectual activity as the supreme state of being for man and for God. Crescas' God is not first and foremost an intellect, and the supreme goal to which man can aspire is to love God with a love corresponding as far as possible to the infinite greatness of its object and to rejoice in the observance of his commandments. God, too, loves man, and his love, in spite of the lowliness of its object, is proportionate to his infinity.

Crescas attacked the separation of the intellect from the soul as conceived by the Aristotelians and attempted, perhaps in part under the influence of Judah ha-Levi, to refute the Aristotelian doctrine that the actualized intellect, in contradistinction to the soul, survives the death of the body. According to Crescas, the soul is a substance in its own right and can be separated from the body; it continues to subsist after the body's death.

Joseph Albo. Whereas Crescas unmistakably regarded the Aristotelian philosophers as adversaries to be criti-

Albo's mixture of religious tradition and rational philosophy

cized or combatted, the attitude of Joseph Albo (*c.* 1380–c. 1444), who regarded Crescas as his teacher, is much less clearly defined. Albo did not eschew self-contradiction, apparently considering it a legitimate precaution on the part of a philosophical or theological author; indeed, he indulged in it in a much more obvious way than did Maimonides. But, whereas the latter's fundamental philosophical position is fairly clear, the problem being how far he was prepared to deviate from Aristotelian doctrine in the interests of religion, there may be valid doubt whether Crescas and the Jewish religious tradition or Maimonides and Averroës were Albo's true masters. Mainly because of this perhaps deliberate failure to explain to the reader where he really stood, Albo has often been dismissed as an eclectic. He was strongly influenced not only by the authors just mentioned but also by Sa'adia. He seems to have had a considerable knowledge of Christian theology, even adopting for his own purposes certain Scholastic doctrines. He differs from Crescas and to some extent resembles Maimonides in having a marked interest in political theory.

The proclaimed theme of Albo's magnum opus, ***Sefer ha-'iqqarim*** ("The Book of Principles"), is the investigation of the theory of Jewish religious dogmas, the number of which Maimonides, in a nonphilosophical work, had set at 13, whereas Albo, following a doctrine that in the last analysis seems to go back to Averroes, would limit the number to three: the existence of God, divine providence in reward and punishment, and the Torah as divine revelation. One section, usually including the philosophical and the traditional religious interpretations side by side, is devoted to each of these dogmas. Albo's principal and relatively novel contribution to Jewish doctrinal evolution is the classification, in his introduction, of natural, conventional, and divine law.

Natural law (the universal moral law inherent in human nature) is necessary, because man, being political by nature, must belong to a community, which may be restricted in size to one town or may extend over the whole earth. Natural law preserves society by promoting right and repressing injustice; thus, it restrains men from stealing, robbing, and murdering. The positive laws instituted by wise men take into account the particular nature of the people for whose benefit they are instituted, as well as other circumstances. This means that they differ from the natural law in not being universally applicable. Neither natural law nor the more elaborate conventional laws, however, lead men toward true spiritual happiness; this is the function of divine laws instituted by a prophet, which teach men true theoretical opinions. Whereas Maimonides maintained that Judaism was the only divine law promulgated by a true prophet, Albo considered that the commandments given to Noah for all mankind also constitute divine law, which ensures, although to a lesser degree than does Judaism, the happiness of its adherents. This position justifies a certain universalism; in accordance with a Talmudic saying, Albo believed that the pious among the non-Jews—that is, those who observe Noah's laws—have a share in the world to come. But he rejected the pretensions of Christianity and Islām to be divine laws.

MODERN PHILOSOPHY

The Iberian-Dutch Philosophers. The expulsion of the Jews from Spain and Portugal produced a new centre of Jewish thought, Holland, where many of the exiled Jews found a new and safer domicile; the tolerance of the regime seemed to provide guarantees against external persecution. This did not prevent, and indeed may have furthered, the establishment of an oppressive internal orthodoxy that was prepared to chastise rebellious members of the community. This was evident in the cases of Uriel Acosta, or da Costa, and Benedict de Spinoza, two 17-century philosophers who rebelled against Jewish orthodoxy and who were excommunicated for their views (Acosta twice).

The two excommunicados: Acosta and Spinoza

Uriel Acosta. Acosta came to Amsterdam from Portugal, where, belonging to a family of Marranos (Jews who had converted to escape religious persecution), he had been brought up in the Catholic faith; his philosophical position was to a great extent determined by his antagonism to the orthodox Judaism that he encountered in Amsterdam. His growing estrangement from generally accepted Jewish doctrine is attested by his Portuguese treatise ***Sobre a Mortalidade da Alma*** ("On the Mortality of the Soul"). He considered that the belief in the immortality of the soul has had many evil effects, for it impels men to choose an ascetic way of life and even to seek death. According to him, nothing has tormented men more than the belief in an inner, spiritual good and evil. At this stage Acosta affirmed the authority of the Bible from which, according to him, the mortality of the soul can be proved.

In his autobiography, written in Latin and entitled ***Exemplar Humanae Vitae*** ("Example of a Human Life"), he takes a more radical position. He proclaims the supreme excellency of the natural moral law (which, when arguing before Jews, he seems to identify with the divine commandments to Noah, thus suggesting a correspondence with the view of Albo). Accordingly, he denies the validity of the argument that natural law is inferior to Judaism and Christianity, because he believes that both these religions teach the love of one's enemies, a precept that is not a part of natural law and is a manifest impossibility.

Benedict de Spinoza. Although modem philosophers of Jewish origin are not considered as belonging to the history of Jewish philosophy unless they deal with Judaic themes, this restriction may not apply to Spinoza for the following reasons. (1) It was through the study of Jewish philosophical texts that Spinoza was first initiated into philosophy. (2) Spinoza's system is in part a radicalization of, or perhaps a logical corollary to, medieval Jewish doctrines; and the impact of Maimonides and of Crescas is evident. (3) A considerable portion of Spinoza's ***Tractatus Theologico-Politicus*** deals with problems related to Judaism (see SPINOZA, BENEDICT DE).

Spinoza's view of prophecy

The first chapters of the *Tractatus* show that the doctrine of prophecy is of central importance to Spinoza's explanation of Judaism and that, in dealing with this subject, he used Maimonides' categories, although he applied them to different people or groups of people. Maimonides held that the prophets combined intellectual perfection, which made them philosophers, with perfection of the imaginative faculty. He also referred to a category of persons endowed with a strong imagination but possessing no extraordinary intellectual gifts; this category includes, for example, lawgivers and statesmen. Spinoza took over this last category but applied it to the prophets, whom he described as possessing vivid imaginations but as not necessarily having outstanding intellectual capacities. He denied that the biblical prophets were philosophers and used a philosophical and historical approach to the Scriptures to show that the contrary assertion is not borne out by the texts.

Spinoza also denied Maimonides' assertion that the prophecy of Moses was essentially different from that of the other prophets and that this was largely because Moses, in prophesying, had no recourse to the imaginative faculty. According to Spinoza, the distinctive fact about Moses' prophecy was that he heard the voice of God in a prophetic vision—that is, in a state in which his imagination was active. In this assertion Spinoza employed one of Maimonides' categories of prophecy, differentiated in the ***Guide*** according to certain characteristics of prophetic dreams and visions; however, Maimonides thought it improbable that the voice of God was ever heard in prophetic vision, and he held that this category is purely hypothetical. It seems evident that in his classification of Moses, Spinoza was concerned not with what really happened in history but with pigeonholing the biblical evidence into Maimonides' theoretical framework so that it fit in with his own theologico-political purpose: to show that there could be a religion superior to Judaism.

This purpose made it imperative to propound in the *Tractatus* ***Theologico-Politicus*** a theory concerning Jesus, whom Spinoza designates as Christus. The category and the status assigned to Jesus are by and large similar to those that Maimonides attributed to Moses. Thus, Jesus is referred to in the *Tractatus* as a religious teacher who makes recourse not to the imaginative faculty but solely to the intellect. His authority may be used to institute and strengthen the religion Spinoza called *religio catholica* ("universal religion"), which has little or nothing in common with any of the major manifestations of historic Christianity.

The difference between Judaism and Spinoza's *religio catholica* corresponds to the difference between Moses and Jesus. After leaving Egypt the Jews found themselves, in Spinoza's view, in the position of people who had no allegiance to any positive law; they had, as it were, reverted to a state of nature and were faced with the need to enter into a social pact. They were also an ignorant people and very prone to superstition. Moses, a man of outstanding ability, made use of the situation and the characteristics of the people in order to make them accept a social pact and a state founded upon it that, contrary to Spinoza's scheme for his ideal communities, were not based first and foremost upon utilitarian—that is, reasonable—consideration of the advantages of life in society over the state of nature.

The social pact concluded by the children of Israel in the desert was based upon a superstitious view of God as "King" and "Judge," to whom the children of Israel owed whatever political and military successes they obtained. It was to God rather than to the representatives of the popular will that the children of Israel transferred political sovereignty. In due course political sovereignty was vested in Moses, God's representative, and in his successors. It should be added that, in spite of Spinoza's insistence on the superstitious foundations of the ancient Israelite state, his account of its regime was not wholly unsympathetic. He believed, however, that it contained the seeds of its own destruction and that, with the extinction of this state, the social pact devised by Moses had lapsed and all the political and religious obligations incumbent upon the Jews had become null and void.

The rational state and its religion

It could be argued that, because the state conceived by Spinoza is based not on superstitious faith but on a social contract originating in rational, utilitarian considerations, it does not necessarily need to have its authority safeguarded and stabilized by means of religion. Nevertheless, Spinoza appears to have held the view—perhaps derived from a purely empirical knowledge of the behaviour of the common run of men—that there is a need for religion. In order to fulfill the need for some religion and to obviate the danger of harmful religions, he devised his *religio catholica,* the universal religion, which has the following distinctive traits: (1) Its main purpose, a practical one (which is furthered by recourse to the authority of Jesus), is to impel men to act in accordance with justice and charity. Such conduct is tantamount to obedience to the laws of the state and to the orders of the magistrates, in whom sovereignty is vested; for disobedience—even if it springs from compassionate motives—weakens the social pact, which safeguards the welfare of all the members of the community; in consequence, its evil effects outweigh whatever good it may produce. (2) Although religion, according to Spinoza, is not concerned with theoretical truth, in order to be effective the *religio catholica* requires dogmas, which he set forth in the *Tractatus.* These dogmas are formulated there in terms that can be interpreted in accordance both with the philosophical conception of God that Spinoza regarded as true and also with the superstitious ideas of ordinary people. It follows that if they are accepted as constituting by themselves the only creed that everybody is obliged to profess, people cannot be persecuted on account of their beliefs; Spinoza held that such a persecution may lead to civil war and may thus destroy the state. Philosophers are free to engage in the pursuit of truth and to attain, if they can, the supreme goal of man, freedom grounded in knowledge. There can be little doubt that the furtherance of the cause of tolerance for philosophical opinions was one of Spinoza's main objects in writing the *Tractatus.*

The relation between Spinoza's *Ethics,* his major philosophical work, and Jewish medieval philosophy is much more ambiguous than in the case of Spinoza's *Tractatus* Theologico-Philouophicus. In a way, Spinoza's metaphysical system, contained in the *Ethics*, can be regarded as being, in part, a spelling out of some extreme consequences, which could perhaps be legitimately drawn from medieval Aristotelianism; but the facts of the case are no doubt much more complicated than this.

German philosophers. *Moses Mendelssohn.* Moses Mendelssohn opened what may be called the German period of Jewish philosophy *(c. 1750–c.* 1830). This period, in which a considerable number of works on Jewish philosophy were written in German and often under the influence of German philosophy, is also marked by the emancipation of the Jews—that is, by the abrogation of discriminatory laws directed against them—and by their partial or complete assimilation. In this period in particular, the term Jewish philosophy applies especially to works the main purpose of which, or one of the main purposes of which, consists in proposing a definition of Judaism and a justification of its existence. The second task is often conceived as necessitating a confrontation of Judaism with Christianity rather than with philosophy, which served as a critical point of comparison for many medieval philosophers. This change seems to have been a result of the demarcation of the sphere of religion in such a way that, at least in the opinion of the philosophers, possible points of collision no longer existed between it and philosophy. This demarcation was largely furthered by the doctrine of Spinoza, from whom Mendelssohn and others took over and adopted for their own purposes certain fundamental ideas concerning Judaism. Like Spinoza, Mendelssohn held that it is not the task of Judaism to teach rational truths, although they may be referred to in the Bible. Contrary to what he called Athanasian Christianity—that is, the doctrine set forth in the Athanasian Creed—Judaism has no binding dogmas; it is centred on inculcating belief in certain historical events and on action—that is, observance of religious law (including the ceremonial commandments). Such observance is supposed to lead to happiness in this world and in the afterlife. Mendelssohn did not reject this view offhand, as Spinoza would have done; indeed, he seems to have been prepared to accept it, God's mysteries being inscrutable, and the radicalism and what may be called the consistency of Spinoza being the complete antithesis of Mendelssohn's apologetics. Non-Jews were supposed by Mendelssohn to owe allegiance to the natural moral law.

Solomon Formstecher. Whereas Mendelssohn continued the medieval tradition, at least to some extent, or adopted Spinoza's doctrine for his purposes, the Jewish philosophers of the first half of the 19th century may generally be regarded as disciples of the philosophers of their own time. In *Die Religion des Geistes* ("The Religion of the Spirit"), Solomon Formstecher (1808–89) may have been influenced by F.W.J. Schelling, an eminent German philosopher, in his conception of nature and spirit as manifestations of the divine. There are, in Formstecher's view, two types of religions that correspond to these manifestations: (1) the religion of nature, in which God is conceived as the principle of nature or as the world soul, and (2) the religion of the spirit, which conceives of God as an ethical being. According to the religion of the spirit, God has produced the world as his manifestation in full freedom and not, as the religion of nature tends to profess, because the world was necessary for his own existence.

The religions of nature and of spirit

The religion of the spirit, which corresponds to absolute religious truth, was first manifested in the Jewish people. The religious history of the world may be understood as a process of universalization of the Jewish religion. Thus, Christianity propagated Jewish conceptions among the nations; however, it combined them with pagan ideas. The pagan element is gradually being eliminated—Protestantism, for instance, in this respect, marks considerable progress. When at long last the Jewish element in Christianity is victorious, the Jews will be right to give up their isolation. The progress that will bring about this final religious union is already under way.

Samuel Hirsch. The main philosophical work of Samuel Hirsch, entitled ***Die Religionsphilosophie der Juden*** ("The Philosophy of Religion of the Jews"), was decisively influenced by G.W.F. Hegel. This influence is most evident in Hirsch's method and in the task that he assigned to the philosophy of religion—the transformation of religious consciousness into conceptual truth. Contrary to Hegel, however, he did not consider religious truth to be inadequate as compared to philosophical truth.

God revealed himself in the first stages of Jewish history by means of miracles and of prophecy. At present, he manifests himself in the miracle that is constituted by the existence of the Jewish people. At its beginning in the time of Jesus, Christianity was identical with Judaism. The decisive break between the two religions was caused by Paul. When the Pauline elements are eliminated from Christianity, it will be in all essentials in agreement with Judaism, which, however, will preserve its separate existence.

Nachman Krochmal. Nachman Krochmal, a native of Galicia (at that time, part of Austria), was the author of ***More nebukhe** ha-zman* ("Guide for the Perplexed for Our Time"), a treatise in Hebrew on the philosophy of history and on Jewish history that had a considerable influence. Krochmal's philosophical thought was centred on the notion of "spirit," Krochmal being mainly concerned with the "national spirit," the particular spirit that is proper to each people and that accounts for the peculiar characteristics differentiating one people from another in every domain of human activity. The national spirits of all peoples except the Jewish are, according to Krochmal, essentially particular. Hence, the national spirit either becomes extinct with the extinction of the nation or, if it is a powerful spirit, is assimilated by some other nation. The Jewish people have a special relation to the Universal Spirit, who is the God of Israel. This relation accounts for the perpetuity of the Jewish people.

Solomon Steinheim. Solomon Ludwig Steinheim, the author of *Die **Offenbarung nach dem** Lehrbegriff der Synagoge* ("The Revelation According to the Doctrine of the Synagogue"), was apparently influenced by the antirationalism of Friedrich Heinrich Jacobi, a German philosopher. His criticism of science is based on Jacobi's criticism, but he did not agree with Jacobi in opposing discursive reason to the intuitive knowledge of God—Steinheim contrasted human reason to divine revelation. The main point on which the revelation, vouchsafed to the prophets of Israel, is opposed to reason is to be found in the fact that the God posited by reason is subject to necessity, that he can act only in accordance with laws. Moreover, reason affirms that nothing can come from nothing. Accordingly, God is free to create not a good world, but only the best possible world. Revealed religion, on the other hand, affirms the freedom of God and the creation of the world out of nothing.

The Neo-Kantian philosopher of Judaism

Hermann Cohen. There seems to be little connection between the Jewish philosophers of the first half or two-thirds of the 19th century and Hermann Cohen (1842–1918), the head of the Neo-Kantian school centred at the University of Marburg. In a certain sense Cohen may be regarded as a rather unusual case among the philosophers of Judaism of his and the preceding generations because of the two aspects of his philosophical thought—the general and the Jewish—and the uneasy equilibrium between them. Judaism was by no means the only important theme of his philosophical system; it was one of several and not even his point of departure. There is no doubt that, for most of his life, Cohen was wholly committed to his brand of Kantianism, in the elaboration of which he displayed considerable originality—it has been maintained with some justification that his doctrine manifests a certain (unintentional) kinship with Hegel's. Cohen's idea of God, however, derives from an analysis and a development of certain conceptions of Immanuel Kant. In Cohen's view reason requires that nature be conceived of as conforming to one rational plan and that harmony exist between the domains of natural and of moral teleology (ultimate purposes or ends). These two requirements, in turn, necessitate the adoption of the idea of God—the word idea being used in the Kantian sense, which means that no assertion is made about the metaphysical reality of God.

Cohen seems to have changed his attitude in the last years of his life; although he did not explicitly renounce his previous positions, a considerable shift of emphasis can be discerned in his doctrines. The notion of the human individual—an individual who is weak and full of sin—comes to the fore, as well as the conception of a correlation, a relationship, between God and the individual. This relationship is one of love, the love of God for man and the love of man for God. It is difficult to reconcile the conception of God expounded in Cohen's work of his last period with his Kantian or Neo-Kantian attitude toward metaphysics.

Franz Rosenzweig. Franz Rosenzweig published his main philosophical work, ***Der Stern der Erlösung** (**The Star of Redemption***, 1971), in 1921. This work begins with a rejection of the traditional philosophical attitude that denies the fear of death, maintaining, instead, that this fear is the beginning of the cognition of the All. Man should continue to fear death, despite the indifference of philosophy and its predilection for accepting death. Traditional philosophy is interested exclusively in the universal, and it is monistic—its aim is to discover one principle from which everything can be derived. This tendency of philosophy, however, denatures human experience, which knows not one but three separate domains (which Kant had referred to in a different context), namely, God, the world, and man.

According to Rosenzweig, God (like the world and like man) is known through experience (the experience of revelation). In Greek paganism, the most perfect manifestation of paganism, every one of these domains subsists by itself: the gods, the cosmos, and man as the tragic, solitary, silent hero. Biblical religion is concerned with the relation between the three: the relation between God and the world, which is creation; the relation between God and man, which is revelation; and the relation between man and the world, which leads to salvation. The philosophy that renounces the ambition to find one principle for everything that exists and that follows biblical religion in centring on the connections between the three domains and between the words and acts that bring about and develop these connections, Rosenzweig termed the narrative philosophy; the term and the concept were taken over from Schelling, whose influence Rosenzweig repeatedly emphasized.

The biblical faith brought forth two valid religions—Christianity and Judaism. The first is described by Rosenzweig as the eternal way; the Christian peoples seek in the vicissitudes of time and history the way to salvation. In contradistinction to them, the existence of the stateless Jewish people is not concerned with time and history; it is—notwithstanding the hope for final salvation—already an eternal life, renewed again and again according to the rhythm of the liturgical Jewish year.

"I and Thou"

Martin Buber. Since the early years of the 20th century, Martin Buber has exercised a powerful influence on both Jews and non-Jews. In his early period Buber was led, partly through empathy with Jewish and non-Jewish mysticism, to stress unitive experience and knowledge, in which the difference between one man and another and between man and God tend to disappear. But in his final period he taught, following, as he claimed, a suggestion of Ludwig Feuerbach, a 19th-century German philosopher, that man can only realize himself as a human being in a relation with another, who may be another man or God. This conception of the "I and Thou" relationship leads to the formulation of Buber's view of the dialogical life—the mutual, responsive relation between man and others—and accounts for the importance that he attaches to the category of "encounter." (S.Pi.)

BIBLIOGRAPHY

General introductions: JULIUS GUTTMANN, ***Die Philosophie des** Judentums (1933;* Eng. trans., ***Philosophies of Judaism,*** 1964), the best general treatment of Jewish philosophy from

the ancient to the modern period, ending with Franz Rosenzweig; NATHAN ROTENSTREICH, *Jewish Philosophy in Modern Times: From Mendelssohn to Rosenzweig* (1968), a good philosophical discussion of the major thinkers in modern Jewish philosophy, especially perceptive on Hermann Cohen.

Hellenistic philosophy: NAHUM GLATZER (ed.), *The Essential Philo* (1971), lengthy selections from the major works of Philo, with notes: HARRY A. WOLFSON, *Philo*, rev, ed., 2 vol. (1962), the most comprehensive study of Philo in any language, with special emphasis upon Philo's influence upon later philosophy.

Medieval philosophy: ISAAC HUSIK, *A History of Medieval Jewish Philosophy (1940,* reprinted 1969), a thorough examination of each of the major medieval Jewish philosophers from Isaac Israeli through Joseph Albo, with good bibliography; *Three Jewish Philosophers: Philo, Saadya Gaon, Jehuda Halevi,* trans. and ed. by HANS LEWY, ALEXANDER ALTMANN, and ISAAK HEINEMANN (1965), a good introductory anthology containing representative selections from these classical figures with perceptive introductions and explanatory notes; GEORGES VAJDA, *Introduction à la pensée juive du moyen âge* (1947), a good general survey organized around the major philosophical traditions in medieval Jewish philosophy, with extensive bibliography.

Jewish Kalām: SAʿADIA BEN JOSEPH, *The Book of Reliefs and Opinions,* trans. by SAMUEL ROSENBLATT (1948), Sa'adia's important philosophical work dealing with the major topics in Jewish theology such as faith and reason, creation, God, and reward and punishment; *A Karaite Anthology,* trans. and ed. by LEON NEMOY (1952), an excellent collection of the more accessible Karaite materials, covering a wide range of themes, writers, and periods; ISRAEL EFROS, "Medieval Jewish Philosophy" (1967), a collection of essays (in Hebrew) containing a full-length study of Sa'adia's philosophy, in addition to short essays on Judah ha-Levi and Maimonides.

Jewish Neo-Platonism: ISAAC ISRAELI, *Works,* ed. and trans. by ALEXANDER ALTMANN and S.M. STERN (1958), contains complete translations of three of Israeli's works and excerpts from another, together with a comprehensive essay on his philosophy; SOLOMON IBN GABIROL, "Fountain of Life," trans. from the Latin into Hebrew by J. BLUWSTEIN (1950), in addition to a complete vocalized translation of the text this edition contains the Hebrew summary made by Shem Tov ibn Falaquera in the 13th century.

Judah ha-Levi: The Kuzari, trans. by HARTWIG HIRSCHFELD (1964), a complete English translation of the text and notes, with an introductory essay by HENRY SLONIMSKY; LEO STRAUSS, "The Law of Reason in the Kuzari," in *Persecution and the Art of Writing* (1952), in addition to this perceptive and provocative essay on ha-Levi's attitude towards philosophy and rationalistic ethics, the book contains important essays on Maimonides and Spinoza; BAHYA IBN PAQUDA, *The Duties of the Heart* (Hebrew ed. by A. ZIFRONI, *1928;* Eng. trans. by EDWIN COLLINS, 1904), one of the more widely read medieval classics of Jewish philosophy, concentrating upon ethics and personal piety; ABRAHAM BAR HIYYA, *The Meditation of the Sad Soul,* trans. with introduction by G. WIGODER (1971), one of the more accessible philosophical works of this important medieval astronomer and mathematician, devoted primarily to ethical and religious themes.

Jewish Aristotelianism: ABRAHAM IBN DAUD, *Das Buch Emunah Ramah,* ed. and trans. into German by S. WEIL (1852), one of the first manifestations of the Jewish assimilation of Aristotelian philosophy, containing a critique of Gabirol's neo-Platonism.

Maimonides: The Guide of the Perplexed, trans. with introductory essay by SHLOMO PINES (1963), an accurate and clear translation, with an excellent essay by Pines and a valuable prefatory essay by LEO STRAUSS; *A Maimonides Reader,* ed. with introduction and notes by I. TWERSKY (1972), a fine anthology containing important material from Maimonides' *Mishne Torah* and other legal writings, as well as from his shorter philosophical-theological essays and the *Guide;* HARRY A. WOLFSON, "Maimonides on Negative Attributes," in the *Louis Ginzberg Jubilee Volume,* pp. *411–446* (1945), the best historical and analytical study of Maimonides' doctrine of divine attributes; and "Halevi and Maimonides on Prophecy," *Jewish Quarterly Review,* n.s., 32:345–370, 33:49–82 (1942), an excellent historical analysis of the sources of doctrines of both philosophers on prophecy and a clear analysis of the differences between them.

Averroists: (Isaac Albalag): GEORGES VAJDA, *Isaac Albalag, Averroïste juif, traducteur et annotateur d'Al-Ghazâlî* (1960), a translation of and commentary on Albalag's main philosophical work. al-Ghazālī's "Inconsistencies of the Philosophers," containing valuable citations from the classical Arabic philosophical texts. *(Levi ben Gerson):* "The Wars of the Lord" (1866), Gersonides' major philosophical work (in Hebrew) dealing with the most difficult and controversial topics in medieval philosophy and science; SEYMOUR FELDMAN, "Gersonides' Proofs for the Creation of the Universe," *Proceedings of the American Academy for Jewish Research, 35:113–137* (1967), an analytical study of Gersonides' theory of creation and of his criticism of Aristotle's theory of eternity of the universe. *(Ḥasdai Crescas):* "The Light of the Lord" (1861), Crescas' most original critique (in Hebrew) of Aristotelian philosophy and his vigorous defense of traditional Judaism; HARRY A WOLFSON, *Crescas' Critique of Aristotle* (1929), the most important study of medieval Jewish and Arabic philosophy so far written, containing a translation and critical Hebrew text of part 1 of the treatise with comprehensive, detailed, and most valuable notes and introductory essay; JOSEPH ALBO, *Book of Principles,* trans. with critical text by ISAAC HUSIK, 4 vol. (1946), a fine translation with helpful notes of Albo's treatise in Jewish dogmatics.

Modern Jewish philosophy: (Iberian-Dutch philosophers): I.S. REVAH, *Spinoza et le dr. Juan de Prado* (1959), a study of the cultural background of Spinoza's Amsterdam, especially of the heterodox elements in Sefardic Judaism, containing valuable material pertaining to the excommunication of Spinoza and the ideas of Uriel da Costa; BARUCH SPINOZA, *Theologico-Political Treatise,* trans. by R.H.M. ELWES *(1883,* reprinted 1951), Spinoza's critique of the Bible and the Jewish religion; LEO STRAUSS, *Spinoza's Critique of Religion (1965),* an excellent philosophical study of Spinoza's *Treatise,* its relation to Maimonides, Uriel da Costa, and other Sefardic heterodox thinkers; HARRY A. WOLFSON, *The Philosophy of Spinoza,* 2 vol. *(1934,* reprinted 1969), a most detailed commentary on Spinoza's *Ethics,* containing valuable references to Spinoza's medieval sources such as Maimonides, Gersonides, and Crescas.

German philosophers: (Moses Mendelssohn): Jerusalem and Other Jewish Writings, trans. and ed. by ALFRED JOSPE (1969), a complete translation of *Jerusalem* and other miscellaneous writings, pertaining to questions on the Jewish religion, with a brief, informative, acute introduction by the editor; JULIUS GUTTMANN, "Mendelssohn's Jerusalem and Spinoza's *Theologico-Politico Treatise*" (in Hebrew) in his *Religion and Knowledge* (1956), a collection of Guttmann's essays that also includes important essays on ha-Levi, Maimonides, Gersonides, and Crescas, as well as on themes in modern philosophy of Judaism; NACHMAN KROCHMAL, "Collected Works" *(1961,* in Hebrew), containing his "Guide for the Perplexed of Our Times," ed. by S. RAWIDOWICZ and Rawidowicz' comprehensive analytical study of Krochmal's philosophy of Jewish history and its relationship to Hegelian philosophy. *(Hermann Cohen): Die Religion der Vernunft aus den Quellen des Judentums (1919,* reprinted 1966), Cohen's major work dealing with his philosophy of Judaism, organized according to the main themes of the Jewish religion; *Reason and Hope: Selections from the Jewish Writings of Hermann Cohen,* trans. and ed. by EVA JOSPE (1971), a useful collection from Cohen's miscellaneous writings on Jewish themes. *(Franz Rosenzweig): The Star of Redemption,* trans. by WILLIAM HALLO (1971), Rosenzweig's early but impressive statement of his "re-conversion" to Judaism, which has been influential in contemporary Jewish and non-Jewish theology; NAHUM GLATZER, *Franz Rosenzweig: His Life and Thought,* 2nd ed. (1961), an excellent anthology of Rosenzweig's various writings, containing many letters that reveal the more important and intimate episodes in Rosenzweig's career. *(Martin Buber):* See bibliography to article BUBER, MARTIN.

(Se.F.)

Jewish Religious Year

The term Jewish religious year as used in this article encompasses the cycle of sabbaths and holidays that are commonly observed by the Jewish religious community —and officially in Israel by the Jewish secular community as well. Dates here are listed as BCE (Before the Common Era = BC) and CE (Common Era = AD).

NATURE AND SIGNIFICANCE

The sabbath and festivals are bound to the Jewish calendar, reoccur at fixed intervals, and are celebrated at home and in the synagogue according to ritual set forth in Jewish law and hallowed by Jewish custom. According to Jewish teaching, the sabbath and festivals are, in the first instance, commemorative. The sabbath, for example, commemorates the Creation, and Passover commemorates the Exodus from Egypt over 3,000 years ago. The past is not merely recalled; it is also relived through the sabbath and festival observances. Creative physical activity ceases on the sabbath as it did, according to Genesis, when the Creation was completed; Jews leave their homes and reside in booths during the Sukkot festival as did their biblical ancestors. Moreover, sabbath and festival themes are considered to be perpetually significant, recurring and renewed in every generation. Thus the revelation of the Torah (the divine teaching or law) at Sinai, commemorated on Shavuot, is considered an ongoing process which recurs whenever a commitment is made to Torah study.

An important aspect of sabbath and festival observance is sanctification. The sabbath and festivals sanctified the Jews more than the Jews sanctified the sabbath and festivals. Mundane meals became sacred meals; joy and relaxation became sacred obligations (*mitzwot*). No less significant is the contribution of the sabbath and festivals toward communal awareness. Thus, neither sabbath nor festival can be properly observed in the synagogue according to the ancient tradition if fewer than ten male Jews are present. Again, a Jew prays on Rosh Hashana and mourns on Tisha be-Av not only for his own fate but for the fate of all Jews. The sense of social cohesiveness fostered by the sabbath and festival observances has stood the Jews well throughout their long, often tortuous history.

Through Christianity and Islām, the sabbath and festivals have had considerable impact on much of the inhabited world. The seven-day week, the notion of a weekly day of rest, and many Christian and Islāmic holiday observances owe their origins to the Jewish calendar. sabbath, and festivals.

THE JEWISH CALENDAR

Lunisolar structure. The Jewish calendar is lunisolar —*i.e.*, regulated by the positions of both the moon and the sun. It consists usually of 12 alternating lunar months of 29 and 30 days each (except for Heshvan and Kislev, which sometimes have either 29 or 30 days), and totals 353, 354, or 355 days per year. The average lunar year (354 days) is adjusted to the solar year (365¼ days) by the periodic introduction of leap years in order to assure that the major festivals fall in their proper season. The leap year consists of an additional 30-day month called First Adar, which always precedes the month of (Second) Adar. A leap year consists of either 383, 384, or 385 days and occurs seven times during every 19-year period (the so-called Metonic cycle). Among the consequences of the lunisolar structure are these: (1) The number of days in a year may vary considerably, from 353 to 385 days. (2) The first day of a month can fall on any day of the week, that day varying from year to year. Consequently, the days of the week upon which an annual Jewish festival falls vary from year to year despite the festival's fixed position in the Jewish month.

Leap year: Second Adar

Months and notable days. The months of the Jewish religious year, their approximate equivalent in the Western Gregorian calendar, and their notable days, are as follows:

Tishri (September–October)
1,2 Rosh Hashana (New Year)
3 Tzom Gedaliahu (Fast of Gedaliah)
10 Yom Kippur (Day of Atonement)
15–21 Sukkot (Tabernacles)
22 Shemini Atzeret (Eighth Day of the Solemn Assembly)
23 Simhat Torah (Rejoicing of the Law)
Heshvan, or Marḥeshvan (October–November)
Kislev (November–December)
25 Hanukka (Feast of Dedication) begins
Ṭevet (December–January)
2–3 Hanukka ends
10 ʿAsara be-Tevet (Fast of Tevet 10)
Shevaṭ (January–February)
15 Tu bi-Shevaṭ (15th of Shevaṭ: New Year for Trees)
Adar (February–March)
13 Ta'anit Esther (Fast of Esther)
14, 15 Purim (Feast of Lots)
Nisan (March–April)
15–22 Pesah (Passover)
Iyyar (April–May)
18 Lag ba-Omer (33rd Day of the Omer Counting)
Sivan (May–June)
6,7 Shavuot (Feast of Weeks, or Pentecost)
Tammuz (June–July)
17 Shiva' 'Asar be-Tammuz (Fast of Tammuz 17)
Av (July–August)
9 Tisha be-Av (Fast of Av 9)
Elul (August–September)

During leap years, the Adar holidays are postponed to Second Adar.

Since 1948 many Jewish calendars list Iyyar 5—Israel Independence Day—among the Jewish holidays.

Origin and development. The origin of the Jewish calendar can no longer be accurately traced. Some scholars suggest that a solar year prevailed in ancient Israel, but no convincing proofs have been offered, and it is more likely that a lunisolar calendar similar to that of ancient Babylonia prevailed in ancient Israel. In late Second Temple times (*i.e.*, 1st century BCE to 70 CE), calendrical matters were regulated by the Sanhedrin, or council of elders, at Jerusalem. The testimony of two witnesses who had observed the New Moon was ordinarily required to proclaim a new month. Leap years were proclaimed by a council of three or more rabbis with the approval of the nasi, or president, of the Sanhedrin. With the decline of the Sanhedrin, calendrical matters were decided by the Palestinian patriarchate (the official heads of the Jewish community under Roman rule). Jewish persecution under Constantius II (reigned 337–361) and advances in astronomical science led to the gradual replacement of observation by calculation. According to Hai ben Sherira (died 1038)—the head of a leading Talmudic academy in Babylonia—Hillel II, a Palestinian patriarch, introduced a fixed and continuous calendar in 359 CE. A summary of the regulations governing the present calendar is provided by Maimonides, the great medieval philosopher and legist, in his Code: *Sanctification of the New Moon*, chapters 6–10.

Regulation by the Sanhedrin and patriarchate

Fragments of writings discovered in a geniza (depository for sacred writings withdrawn from circulation) have brought to light a calendrical dispute between Aaron ben Meir, a 10th-century Palestinian descendant of the patriarchal (Hillel) family, and the Babylonian Jewish authorities, including Sa'adia ben Joseph—an eminent 10th-century philosopher and *gaon* (head of a talmudic academy). Ben Meir's calculations provided that Passover in 922 be celebrated two days earlier than the date fixed by the normative calendar. After a bitter exchange of letters, the controversy subsided in favour of the Babylonian authorities, whose hegemony in calendrical matters was never again challenged.

Calendars of various sectarian Jewish communities deviated considerably from the normative calendar described above. The Dead Sea (or Qumrān) community (made famous by the Dead Sea Scrolls discoveries) adopted the calendrical system of the noncanonical books of Jubilees and Enoch, which was essentially a solar calendar. Elements of this same calendar reappear among the Mishawites, a sect founded in the 9th century.

The Karaites, a sect founded in the 8th century, refused, with some exceptions, to recognize the normative fixed calendar and reintroduced observation of the New Moon. Leap years were determined by observing the maturation of the barley crop in Palestine. Consequently, Karaites often celebrated the festivals on dates different from those fixed by the rabbis. Later in medieval times, the Karaites adopted some of the normative calendrical practices, while rejecting others.

THE SABBATH

The Jewish sabbath (from Hebrew shavat, "to rest") is observed throughout the year on the seventh day of the week—Saturday. According to biblical tradition, it commemorates the original seventh day on which God rested after completing the creation.

Scholars have not succeeded in tracing the origin of the seven-day week, nor can they account for the origin of the sabbath. A seven-day week does not accord well with either a solar or lunar calendar. Some scholars, pointing to the Akkadian term shapattu, suggest a Babylonian origin for the seven-day week and the sabbath. But *shapattu*, which refers to the day of the Full Moon and is nowhere described as a day of rest, has little in common with the Jewish sabbath. It appears that the notion of the sabbath as a holy day of rest, linking God to his people and recurring every seventh day, was unique to ancient Israel.

Importance. The central significance of the sabbath for Judaism is reflected in the traditional commentative and interpretative literature called Talmud and Midrash (*e.g.*, "if you wish to destroy the Jewish people, abolish their sabbath first") and in numerous legends and adages from more recent literature (*e.g.*, "more than Israel kept the sabbath, the sabbath kept Israel"). Some of the basic teachings of Judaism affirmed by the sabbath are God's acts of creation, God's role in history, and God's covenant with Israel. Moreover, the sabbath is the only Jewish holiday the observance of which is enjoined by the Ten Commandments. Jews are obligated to sanctify the sabbath at home and in the synagogue by observing the sabbath laws and engaging in worship and study. The leisure hours afforded by the ban against work on the sabbath were put to good use by the rabbis, who used them to promote intellectual activity and spiritual regeneration among Jews. Other days of rest, such as the Christian Sunday and the Islamic Friday, owe their origins to the Jewish sabbath.

Worship, study, and abstinence from work

Observances. The biblical ban against work on the sabbath, while never clearly defined, includes such activities as baking and cooking, travelling, kindling fire, gathering wood, buying and selling, and bearing burdens from one domain into another. The Talmudic rabbis listed 39 major categories of prohibited work, including agricultural activity (*e.g.*, plowing and reaping), work entailed in the manufacture of cloth (*e.g.*, spinning and weaving), work entailed in preparing documents (*e.g.*, writing), and other forms of constructive work.

At home, the sabbath begins Friday evening some 20 minutes before sunset, with the kindling of the sabbath candles by the wife, or in her absence by the husband. In the synagogue, the sabbath is ushered in at sunset with the recital of selected psalms and the "Lekha Dodi," a 16th-century Kabbalistic (mystical) poem. The refrain of the latter goes: "Come, my beloved, to meet the bride," the "bride" being the sabbath. After the evening service, each Jewish household begins the first of three festive sabbath meals by reciting the kiddush ("sanctification" of the sabbath) over a cup of wine. This is followed by a ritual washing of the hands and the breaking of bread; two loaves of bread (commemorating the double portions of manna described in Exodus) being placed before the breaker of bread at each sabbath meal. After the festive meal, the remainder of the evening is devoted to study or relaxation. The distinctive features of the sabbath morning synagogue service include the public reading of the Torah, or Five Books of Moses (the portion read varies from week to week) and, generally, the sermon, both of which serve to educate the listeners. Following the service, the second sabbath meal begins, again preceded by kiddush (of lesser significance), and conforming for the most part to the first sabbath meal. The afternoon synagogue service is followed by the third festive meal (without kiddush). After the evening service, the sabbath comes to a close with the havdala ("distinction") ceremony, which consists of a benediction noting the distinction between sabbath and weekday, usually recited over a cup of wine accompanied by a spice box and candle.

The *havdala* ceremony

THE JEWISH HOLIDAYS

The major Jewish holidays are the Pilgrim Festivals: Pesaḥ (Passover), Shavuot (Feast of Weeks, or Pentecost), and Sukkot (Tabernacles); and the High Holidays: Rosh Hashana (New Year) and Yom Kippur (Day of Atonement). In common, their observance is required by the Torah and work is prohibited for the duration of the holiday (except on the intermediary days of the Pesah and Sukkot festivals, when work the neglect of which entails monetary loss is permitted). Purim (Feast of Lots) and Hanukka (Feast of Dedication), while not mentioned in the Torah (and therefore of lesser solemnity), were instituted by Jewish authorities in the Persian and Greco-Roman periods. Lacking the work restrictions characteristic of the major festivals, they are sometimes regarded as minor festivals. In addition, there are the five fasts: 'Asara be-Tevet (Fast of 10 Tevet), Shiva' 'Asar be-Tammuz (Fast of Tammuz 17), Tisha be-Av (Fast of Av 9), Tzom Gedaliahu (Fast of Gedaliah), and Ta ʿanit Esther (Fast of Esther); and the lesser holidays—*i.e.*, holidays the observances of which are few and not always clearly defined—such as Rosh Hodesh (First Day of the Month), Tu bi-Shevaṭ (New Year for Trees), and Lag ba-Omer (33rd Day of Omer Counting). The fasts and the lesser holidays also lack the work restrictions characteristic of the major festivals. Some of the fasts and Rosh Hodesh are mentioned in Scripture. but most of the details concerning their proper observance, as well as those concerning the other lesser holidays, are provided by the Talmudic and medieval rabbis.

Pilgrim festivals. In Temple times. all males were required to appear at the Temple three'times annually and actively participate in the festal offerings and celebrations. These were the joyous pilgrim festivals of Pesah, Shavuot, and Sukkot. Originally, they marked the major agricultural seasons in ancient Israel and commemorated Israel's early history; but after the destruction of the Second Temple in 70 CE, emphasis was almost exclusively placed on the commemorative aspect.

In modern Israel, Pesah, Shavuot, and Sukkot are celebrated for the number of days prescribed by Scripture, namely, seven days, one day, and eight days, respectively (with Shemini Atzeret added to Sukkot). Due to calendrical uncertainties which arose in Second Temple times (6th century BCE to 1st century CE), each festival is celebrated for an additional day in the Diaspora.

Pesah (Passover). Pesaḥ commemorates the Exodus from Egypt and the servitude that preceded it. As such, it is the most significant of the commemorative holidays, for it celebrates the very inception of the Jewish people —*i.e.*, the event which provided the basis for the covenant between God and Israel. The term pesah refers originally to the paschal (Passover) lamb sacrificed on the eve of the Exodus, the blood of which marked the Jewish homes to be spared from God's plague; its etymological significance, however, remains uncertain. The Hebrew root is usually rendered "passed over"—*i.e.*, God passed over the homes of the Israelites when inflicting the last plague on the Egyptians—hence the term Passover. The festival is also called Hag ha-Matzot ("Festival of Unleavened Bread"), for unleavened bread is the only kind of bread consumed during Passover.

The Festival of Unleavened Bread (matza)

Leaven (se'or) and foods containing leaven (*ḥametz*) are neither to be owned nor consumed during Pesah. Aside from meats, fresh fruits, and vegetables, it is customary to consume only those foods prepared under rabbinic supervision and labelled "kosher for Passover," warranting that they are completely free of contact with leaven. In many homes, special sets of crockery, cutlery, and cooking utensils are acquired for Passover use. On the evening preceding the 14th day of Nisan, the home is thoroughly searched for any trace of leaven (*bediqat* hametz). The following morning the remaining particles of leaven are destroyed by fire (bi'ur hametz). From then until after Pesah, no leaven is consumed. Many Jews sell their more valuable leaven products to non-Jews before Passover (mekhirat hametz), repurchasing the foodstuffs immediately after the holiday.

The unleavened bread (matza) consists entirely of flour and water, great care being taken to prevent any fermentation before baking. Hand-baked matza is flat, rounded, and perforated. Since the 19th century, many Jews have preferred the square-shaped, machine-made *matza*.

Passover eve is ushered in at the synagogue service on

The seder ceremonies

the evening before Passover, after which each family partakes of the seder ("order of service); *i.e.*, an elaborate festival meal in which every ritual is regulated by the rabbis. (In the Diaspora, the seder is also celebrated on the second evening of Passover.) The table is bedecked with an assortment of foods symbolizing the passage from slavery (*e.g.*, bitter herbs) into freedom (*e.g.*, wine). The Haggada (literally "narration"), a printed manual comprised of appropriate passages culled from Scripture, Talmud, and Midrash, accompanied by medieval hymns, serves as a guide for the ensuing ceremonies and is recited as the evening proceeds. The seder opens with the cup of sanctification (kiddush), the first of four cups of wine drunk by the celebrants. An invitation is extended to the needy to join the seder ceremonies, after which the youngest son asks four prescribed questions expressing his surprise at the many departures from usual mealtime procedure. ("How different this night is from all other nights!") The father then explains that the Jews were once slaves in Egypt, were then liberated by God, and now commemorate the servitude and freedom by means of the seder ceremonies. Special blessings are recited over the unleavened bread and the bitter herbs (*maror*), after which the main courses are served. The meal closes with a serving of *matza* recalling the paschal lamb, consumption of which concluded the meal in Temple times. The seder concludes with the joyous recital of hymns praising God's glorious acts in history and anticipating a messianic redemption to come.

The Passover liturgy is considerably expanded and includes the daily recitation of Psalms 113–118 (Hallel, "praise"), public readings from the Torah, and an additional service (*musaf*). On the first day of Pesah, a prayer for dew in the Holy Land is recited; on the last day, the memorial service for the departed (yizkor) is added.

Shavuot (Feast of Weeks, or Pentecost). Originally an agricultural festival marking the wheat harvest, Shavuot commemorates the revelation of the Torah at Sinai. Shavuot ("weeks") takes its name from the seven weeks of grain harvest separating Passover and Shavuot. The festival is also called Hag ha-Qazir (Harvest Festival) and Yom ha-Bikkurim (Day of First Fruits). Greek-speaking Jews called it *pentēkostē*, meaning "the fiftieth" day after the sheaf offering. In rabbinic literature, Shavuot is called atzeret ("cessation, conclusion"), perhaps because the cessation of work is one of its distinctive features, or possibly because it was viewed as concluding the Passover season. In liturgical texts it is described as the "season of the giving of our Torah." The association of Shavuot with the revelation at Sinai, while not attested to in Scripture, is alluded to in the Pseudepigrapha (a collection of noncanonical writings). In rabbinic literature the association first appears in 2nd-century materials. The association, probably an ancient one, was derived in part from the book of Exodus, which dates the revelation at Sinai to the third month (counting from Nisan), *i.e.*, Sivan.

Dating Shavuot

Scripture does not provide an absolute date for Shavuot. Instead, 50 days (or seven weeks) are reckoned from the day the sheaf offering (Omer) of the harvest was brought to the Temple, the 50th day being Shavuot. According to the Talmudic rabbis, the sheaf offering was brought on the 16th of Nisan; hence Shavuot always fell on or about the 6th of Sivan. Jewish sectarians, such as the Sadducees, rejected the rabbinic tradition concerning the date of the sheaf ceremony, preferring a later date, and celebrated Shavuot accordingly.

In Temple times, aside from the daily offerings, festival offerings, and first-fruit gifts, a special cereal offering consisting of two breads prepared from the new wheat crop was offered at the Temple. Since the destruction of the Second Temple, Shavuot observances have been dominated by its commemorative aspect. Many Jews spend the entire Shavuot night studying Torah, a custom first mentioned in the Zohar ("Book of Splendour"), a Kabbalistic work edited and published in the 13th–14th centuries. Some prefer to recite the *tiqqun lel* Shavuot ("Shavuot night service"), an anthology of passages from Scripture and the Oral Law (Mishna) compiled in the late medieval period. An expanded liturgy includes Hallel, public readings from the Torah, yizkor (in many congregations), and musaf. The Book of Ruth is read at the synagogue service, possibly because of its harvest-season setting.

Sukkot (Tabernacles). Sukkot ("booths"), an ancient harvest festival that commemorates the booths the Israelites resided in after the Exodus, was the most prominent of the three pilgrim festivals in ancient Israel. Also called Hag ha-Asif (Festival of Ingathering), it has retained its joyous, festive character through the ages. It begins on Tishre 15 and is celebrated for seven days. The concluding eighth day (plus a ninth day in the Diaspora), Shemini Atzeret, is a separate holiday. In Temple times, each day of Sukkot had its own prescribed number of sacrificial offerings. Other observances, recorded in the Mishna tractate Sukka, include the daily recitation of Hallel, daily circumambulation of the Temple altar, a daily water libation ceremony, and the nightly bet ha-sho'eva or she'uvah ("place of water drawing") festivities starting on the evening preceding the second day. The last mentioned featured torch dancing, flute playing, and other forms of musical and choral entertainment.

Ideally, Jews are to reside in booths—walled structures covered with thatched roofs—for the duration of the festival; in practice, most observant Jews take their meals in the sukka ("booth) but reside at home. A palm-tree branch (lulav), bound up together with myrtle (*hadas*) and willow ('arava) branches, is held together with a citron (etrog) and waved. Medieval exegetes provided ample (if not always persuasive) justification for the Bible's choice of these particular branches and fruit as symbols of rejoicing. The numerous regulations governing the sukka, lulav, and etrog comprise the major portion of the treatment of Sukkot in the codes of Jewish law. The daily Sukkot liturgy includes the recitation of Hallel, public readings from the Torah, the musaf service, and the circumambulation of the synagogue dais. On the last day of Sukkot, called Hoshana Rabba (Great Hoshana) after the first words of a prayer (hoshana, "save us") recited then, seven such circumambulations take place. Kabbalistic (mystical) teaching has virtually transformed Hoshana Rabba into a solemn day of judgment.

Simhat Torah

Hoshana Rabba is followed by Shemini Atzeret (Eighth Day of Solemn Assembly), which is celebrated on Tishri 22 (in the Diaspora also Tishri 23). None of the more distinctive Sukkot observances apply to Shemini Atzeret; but Hallel, public reading from the Torah, yizkor (in many congregations), musaf, and a prayer for rain in the Holy Land are included in its liturgy. Simhat Torah (Rejoicing of the Law) marks the annual completion of the cycle of public readings from the Torah. The festival originated shortly before the gaonic period (*c.* 600–1050 CE) in Babylon, where it was customary to conclude the public readings annually. In Palestine, where the public readings were concluded approximately every three years, Simhat Torah was not celebrated annually until after the gaonic period. Israeli Jews celebrate Simhat Torah and Shemini Atzeret on the same day; in the Diaspora, Simhat Torah is celebrated on the second day of Shemini Atzeret. Its joyous celebrations bring the Sukkot season to an appropriate close.

Ten Days of Penitence. The Ten Days of Penitence begin on Rosh Hashana and close with Yom Kippur. Already in Talmudic times they were viewed as forming an especially appropriate period of introspection and repentance. Penitential prayers (selihot) are recited prior to the daily morning service and, in general, during the period scrupulous observance of the Law is expected.

Rosh Hashana (New Year). According to Mishnaic teaching, the New Year festival ushers in the Days of Judgment for all of mankind. Despite its solemnity, the festive character of Rosh Hashana is in no way diminished. In Scripture it is called "a day when the horn is sounded"; in the liturgy "a day of remembrance." In the land of Israel and in the Diaspora, Rosh Hashana is celebrated on the first two days of Tishri. Originally celebrated by all Jews on Tishri 1, calendrical uncertainty led to its being celebrated an additional day in the Diaspora and, depending upon the circumstances, one or two days in Palestine. After the calendar was fixed in 359, it was regularly celebrated in Palestine on Tishri 1 until the 12th century, when Provençal scholars introduced the two-day observance. Considerable speculation in recent literature concerning the origin of the Jewish New Year festival proves mostly that its early history can only be conjectured, not reconstructed.

The most distinctive Rosh Hashana observance is the sounding of the ram's horn (shofar) at the synagogue service. Medieval commentators suggest that the blasts acclaim God as Ruler of the universe, recall the divine revelation at Sinai, and are a call for spiritual reawakening and repentance. An expanded New Year liturgy stresses God's sovereignty, his concern for man, and his readiness to forgive those who repent. On the first day of Rosh Hashana (except when it falls on the sabbath) it is customary for many to recite penitential prayers at a river, symbolically casting their sins into the river; this ceremony is called tashlikh ("thou wilt cast"). Other symbolic ceremonies, such as eating bread and apples dipped in honey, accompanied with prayers for a "sweet" and propitious year, are performed at the festive meals.

Yom Kippur (Day of Atonement). The most solemn of the Jewish festivals, Yom Kippur is a day when sins are confessed and expiated and man and God are reconciled. It is also the last of the Days of Judgment and the holiest day of the Jewish year. Celebrated on Tishri 10, it is marked by fasting, penitence, and prayer. Work, eating, drinking, washing, anointing one's body, sexual intercourse, and donning leather shoes are all forbidden.

The kol *nidre* prayer

In Temple times, Yom Kippur provided the only occasion for the entry of the high priest into the Holy of Holies; details of the expiatory rites performed by the high priest and others are recorded in the Mishna and recounted in the liturgy. Present-day observances begin with a festive meal shortly before Yom Kippur eve. The *kol* nidre prayer (recited before the evening service) is a legal formula which absolves Jews from fulfilling solemn vows, thus safeguarding them from accidentally violating a vow's stipulations. The formula first appears in gaonic sources (derived from the Babylonian Talmudic academies, 6th–11th centuries) but may be older; the haunting melody that accompanies it is of medieval origin. Virtually the entire day is spent in prayer at the synagogue, the closing service (ne'ila) concluding with the sounding of the ram's horn.

Minor festivals: Hanukka and Purim. Hanukka and Purim are joyous festivals lacking the work restrictions characteristic of the major festivals.

Hanukka (Feast of Dedication). Hanukka commemorates the Maccabean (or Hasmonean) victories over the forces of the Seleucid king Antiochus IV Epiphanes (reigned 175–164 BCE), and the rededication of the Temple Kislev 25, 164 BCE. Led by Mattathias and his son Judah Maccabee, the Maccabees were the first Jews who fought to defend their religious beliefs rather than their lives. Hanukka is celebrated for eight days beginning on Kislev 25. The Hanukka lamp or candelabrum (*menora*), which recalls the Temple lampstand, is kindled each evening. One candle is lit the first evening; an additional candle is lit each subsequent evening until eight candles are lit on the last evening. According to the Talmud (Shabbat 21b), the ritually pure oil available at the rededication of the Temple was sufficient for only one day's light but miraculously lasted for eight days, hence the eight-day celebration of Hanukka. Evidence from the Apocrypha (writings excluded from the Jewish canon but included in the Roman Catholic and Eastern Orthodox canons) and rabbinic literature shows an association between Sukkot and Hanukka, possibly accounting for the latter's eight-day duration. Hanukka joy is expressed in festive meals, song, games, and gifts to children. The liturgy includes Hallel, public readings from the Torah, and the ʿ*al ha-nissim* ("for the miracles") prayer. The Scroll of Antiochus, an early medieval account of Hanukka, is read in some synagogues and homes.

Biblical background of Purim

Purim (Feast of Lots). As recorded in the biblical Book of Esther, Purim commemorates the delivery of the Persian Jewish community from the plottings of Haman, Ahasuerus' (perhaps Xerxes, king of Persia, 486–465 BCE) prime minister. Mordecai and his cousin Esther, the King's Jewish wife, interceded on behalf of the Jewish community, rescinded the royal edict authorizing a pogrom against the Jews, and instituted the Purim festival. The historicity of the biblical account is questioned by many modern scholars. It is now generally conceded that the Book of Esther was written in the Persian period (it contains Persian but not Greek words) and reflects Persian custom. Except for the Book of Esther, the earliest mention of the Purim festival is from the 2nd–1st centuries BCE. The name of the festival was derived from the Akkadian *pûru*, meaning "lot."

In most Jewish communities, Purim is celebrated on Adar 14 (some also celebrate it on the 15th, others only on the 15th). On the evening preceding Purim, men, women, and children gather in the synagogue to hear the Book of Esther read from a scroll (*megilla*). The reading is repeated Purim morning. A festive meal during the day is accompanied by much song, wine, and merriment. Masquerades, Purim plays, and other forms of parody are common. Friends exchange gifts of foodstuffs and also present gifts to the poor. Aside from the Esther readings, the liturgy includes public reading from the Torah and recital of the Purim version of the ʿ*al ha-nissim* prayer.

The five fasts. The commemorative aspects of the fasts are bound up with their penitential aspects, all of which find expression in the liturgy. Thus the Jew not only relives the tragic history of his people with each fast, but is also afforded an opportunity to search within himself and focus on his own (and his people's) present and future. Penitential prayers (selihot) are recited on all fasts, and the Torah is read at the morning and afternoon services.

'Asara be-Tevet (Fast of Tevet 10) commemorates the beginning of the siege of Jerusalem by Nebuchadnezzar, king of Babylonia, in 588 BCE.

Shiva' 'Asar be-Tammuz (Fast of Tammuz 17) commemorates the first breach in the wall of Jerusalem by the Romans in 70 CE. It initiates three weeks of semi-mourning that culminate with Tisha be-Av.

Tisha be-Av (Fast of Av 9) commemorates the destruction of the First and Second Temples in 586 BCE and 70 CE. The most solemn of the five fasts, its self-denials are more rigorous than those prescribed for the others, and, like Yom Kippur, the fast begins at sunset. The book of Lamentations is read at the evening service, followed by poetic laments that are also recited Tisha be-Av morning.

Tzom Gedaliahu (Fast of Gedaliah) commemorates the slaying of Gedaliah, governor of Judah after the destruction of the First Temple.

Ta'anit Esther (Fast of Esther), which commemorates Esther's fast (cf. Esther 4:16), is first mentioned in gaonic literature.

The lesser holidays. Rosh Hodesh (First Day of the Month). A major festival in the biblical period, this day gradually lost most of its festive character. Since Talmudic times, it has been customary to recite Hallel on Rosh Hodesh. In the medieval period, aside from the liturgical practices carried over from the Talmudic period, it was celebrated with a festive meal. Always more diligently observed in Palestine than in the Diaspora, attempts to revive its full festive character are being made in modern Israel.

Tu *bi-Shevat* (New Year for Trees). First mentioned in the Mishna, where it marks the New Year for tithing purposes, it assumed a festive character in the gaonic period, and later in the medieval period it became customary to eat assorted fruits on the holiday. In modern times it is associated with the planting of trees in Israel.

Lag *ba-Omer* (*33rd* Day of the Omer Counting). It is a joyous interlude in the otherwise somber period of Omer counting (*i.e.*, of the 49 days to Shavuot), which is traditionally observed as a time of semi-mourning.

Usually celebrated as a school holiday with outings, it is first mentioned in medieval sources, which attribute its origin to the cessation of a plague that was decimating the students of Akiba, an influential rabbinic sage in the 2nd century, and to the anniversary of the death of another great rabbi, Simeon ben Yoḥai (died *c.* 170 CE).

THE SITUATION TODAY

Modern attitudes toward the sabbath and festivals vary considerably. Acculturated Jews under the sway of Western secularism often are ignorant of, or choose to neglect, traditional observances. Attitudes of committed Jews in the Western world are patterned mostly along the lines of accepted Orthodox, Conservative, and Reform practice. Thus for example, driving to synagogue services on the sabbath is unthinkable in Orthodox circles, a matter of dispute among Conservative rabbis, and normative practice for Reform Jews. Among Orthodox Jews, who best preserve the traditional observances, contemporary discussion centres mostly on technological advances and their effect on Halakhic practice (the behaviour laid down in the written and oral Torah). Whether or not hearing aids may be worn on the sabbath, and how crossing the international dateline affects observance of sabbaths and festivals typify the sort of problem raised in Orthodox *responsa* ("replies" to questions on law and observance). Recent (and often heated) discussion in conservative literature raises the possibility of abolishing the obligatory character of the additional festival days in the Diaspora (except for the second day of Rosh Hashana), thus unifying Jewish practice throughout the world. Reform Jews, the most innovative of the three groups, observe neither the additional festival days (including the second day of Rosh Hashana) nor the fasts and have introduced numerous modifications in the liturgy as well as in the observances. In recent years more radical Reform congregations have experimented freely with "psychedelic" sound and light effects and other novel forms of synagogue service.

In Israel sabbath is the national day of rest, and Jewish holidays are vacation periods. Municipal ordinances govern public observance of the sabbath and festivals; their enactment and enforcement vary with the political influence of the local Orthodox Jewish community. Attempts to interpret festivals along nationalistic lines are common; some kibbutzim (communal farms) stress the agricultural significance of the festivals. Independence Day is a national holiday; the preceding day, Remembrance Day, commemorates Israel's war dead. Yom Hashoa (Holocaust Day)—marking the systematic destruction of European Jewry between 1933 and 1945 and recalling the short-lived Ghetto uprisings—is commemorated officially on Nisan 27; many religious Israelis prefer to commemorate it on Tebet 10 (a fast day) now called *yom ha-qaddish* (day upon which the mourner's prayer is recited). Since the June 1967 war, Iyyar 28—Liberation of Jerusalem Day—is celebrated unofficially by many Israelis. Appropriate services are conducted on all the aforementioned holidays by most segments of Israel's religious community.

In Israel and the Diaspora, Jewish theologians often stress the timelessness and contemporaneity of holiday observances. Nevertheless, "revised" Passover Haggadot (plural of Haggada) in which contemporary issues are accorded a central position, appear regularly.

Scholarly research into the origin of the festivals, if unabated, has not advanced significantly in recent years, nor is it likely to unless new evidence is forthcoming. Attempts to trace the development and spread of festival observances have fared better, and studies such as A. Yaari's ***History of the Simhat Torah Festival*** (in Hebrew) bode well for the future.

BIBLIOGRAPHY. ROLAND DE VAUX, ***Les Institutions de l'Ancien Testament, 2*** vol. (1958–60; Eng. trans., ***Ancient Israel: Its Life and Institutions,*** 1961), summarizes the present state of biblical scholarship regarding the origin and development of the Jewish calendar, sabbath, and festivals. Other relevant works include: THEODOR H. GASTER, ***Festivals of the Jewish Year*** (1953), an anthropological, comparative, and often speculative approach to the sabbath and festivals; SHLOMO YOSEF ZEVIN, ***ha-Mo'adim ba-halakhah*** (1944), a modern classic (in Hebrew) treating talmudic and post-talmudic developments in the festival observances; and MENAHEM M. KASHER, ***Torah Shelemah,*** vol. *13* (1949), a comprehensive history of the Jewish calendar, also in Hebrew.

(S.Z.L.)

JimCnez de Cisneros, Francisco, Cardinal

Cardinal Francisco Jiménez de Cisneros, Spanish cardinal, grand inquisitor, and statesman, was the first of the great series of 16th-century ecclesiastics, both Catholic and Protestant, who, through sheer power of intellect and force of personality, rose from humble beginnings to commanding positions in the religious and political life of their countries.

By courtesy of the Hispanic Society of America

FUNDATOR. ✠. F. F. X. CARDINALIS HISP. ARCHIEPIS. TOL. COMPLVT. ACADEMIÆ

Jiménez de Cisneros, engraving after a bust by Felipe Vigarny (d. 1543).

Early life

Jiménez was born at Torrelaguna, in Castile, Spain, in 1436, the son of a poor hidalgo (lower nobility) tax collector, and was christened Gonzalo. He studied at the University of Salamanca and, after taking holy orders, spent a number of years in Rome (1459–66), where he disliked the Humanists at the papal court but was impressed by their learning. Pope Paul II gave him an "expective letter" for the first vacant benefice in the Archdiocese of Toledo. The archbishop, Alfonso de Carillo, refused to accept the letter and, in 1473, when Jiménez insisted on his rights, threw him into prison. Refusing release at the price of giving up his claims, Jiménez remained in prison until 1479, when Carillo gave way. In 1482 Cardinal Pedro González de Mendoza, impressed by Jiménez' ability and strength of character, made him vicar general of the Bishopric of Sigiienza. In 1484 Jiménez gave up this post and, apparently, a brilliant career and became a monk in the Franciscan monastery of San Juan de los Reyes at Toledo, taking the name of Fray (Brother) Francisco.

In 1492, on Mendoza's recommendation, Isabella I the Catholic, of Castile, appointed him her confessor. From then on his influence grew rapidly. In 1495 he succeeded Mendoza as archbishop of Toledo. This position gave JimCnez the opportunity to initiate the reform of the Spanish clergy. At synods of Alcalá (1497) and Talavera (1498) he promulgated a series of orders: the clergy had to give up the common practice of concubinage; they were required to reside in their parishes, to go to frequent confession, and to preach and explain the gospel to their parishioners every Sunday. A simple catechism was published with the decrees. The monks, first of Jiménez' own order of the Franciscans and then of the other orders, were required to observe their traditional rules. Aristocratic ecclesiastics resented this interference with their life style and appealed to Isabella and to Rome; 400 monks from Andalusia even fled to North Africa with their "wives" and became Muslims. But gradually the reforms became effective, at least in the monastic orders.

Against the advice of Hernando de Talavera, archbishop of Granada (who wanted to convert the Moors of Granada slowly by education), Jiménez introduced forced mass conversions. The Moriscos, although now nominally Christians, were neither willing to be assimilated to the Christian Spaniards nor were they accepted as equals by the latter. Jiménez' intervention was the direct cause of a Moorish revolt in 1499–1500, and he must be held largely responsible for making the Morisco problem insoluble. In 1609 the Moriscos were finally expelled from Spain.

Appointed grand inquisitor

Jiménez spent the last years of Isabella's reign mostly at her court as her principal religious and political adviser. Upon her death, in 1504, he supported the claims of Ferdinand II the Catholic, of Aragon, against his son-in-law, Philip of Burgundy, but helped to mediate the Agreement of Salamanca, which left Philip as king of Castile. On Philip's death (1506) Jiménez set up a regency government for Ferdinand, who was in Naples at the time, and stopped the intrigues of a group of high nobles who wanted to make over the regency lo the emperor Maximilian I. Ferdinand made him grand inquisitor and obtained for him the cardinal's hat in 1507. As grand inquisitor, Jiménez insisted that the inquisitors observe the inquisitorial rules strictly, but he also extended the authority of the Supreme Council of the Inquisition (the "Holy Office") over the local inquisitorial courts. It was said that he offered from his own resources to pay Ferdinand the sum of 600,000 ducats, which the *conversos* (converted Jews) had offered the King for the abolition of the Inquisition.

Jiménez was the guiding spirit behind the Spanish campaigns in North Africa (1505–10), which he helped finance from his archiepiscopal revenues. But, because of a greater interest in Italy, Ferdinand was content with the capture of Oran and other ports and refused to support Jiménez' plan for a crusade to conquer all North Africa.

Jimtnez knew that the reform of the morals and pastoral work of the clergy could not be effective without a parallel intellectual reform of the church. He thus began planning the foundation of a new university at Alcalá de Henares in 1498. It opened in 1508. Apart from the usual chairs in Thomist theology, Jiménez also established chairs in Scotist and Nominalist theology, as well as in Oriental languages. He attracted some of the finest contemporary scholars to Alcala (although Erasmus refused his invitation). These scholars cooperated in the production of the famous Complutensian Polyglot Bible (completed in 1517 and published *c.* 1522).

Regent of Castile

At the death of Ferdinand (1516) Jimknez once more became regent of Castile. 'The old antagonisms between the nobility and the cities and between the Castilians and the Aragonese broke out again. Some of Ferdinand's Castilian opponents had previously gone to the court in Brussels. They were now joined by Ferdinand's Aragonese ministers, who tried to secure their position with the prospective new ruler, Ferdinand's grandson, Charles of Burgundy (later Charles I of Spain and Emperor Charles V). They were bitterly resented by those who remained in Spain. A group of Castilian nobles plotted to place Charles' young brother Ferdinand on the throne, but Jiménez forestalled them and obtained the general recognition of Charles in Castile. It was due largely to the Cardinal's efforts that Charles could take over his new kingdom without open opposition (September 1517). But Jimknez died at Roa on November 8, 1517, without seeing the new king.

BIBLIOGRAPHY. In the absence of a fully satisfactory modern biography, the CONDE DE CEDILLO, *El Cardenal Cisneros, gobernador del reino* (1921); REGINALD MERTON, *Cardinal Ximenes and the Making of Spain* (1934); and WALTER F. STARKIE, *Grand Inquisitor* (1940), are all useful. There is an excellent short account of Jiménez' ecclesiastical reforms in MARCEL BATAILLON, *Érasme et l'Espagne*, ch. 1 (1937). Some of Jiménez' correspondence is published by DIEGO LOPEZ DE AYALA, *Cartas del Cardenal Don Fray Francisco Jiménez de Cisneros* (1867); and VINCENTE DE LA FUENTE, *Cartas de los secretarios del Cardenal D. Fr. Francisco Jiménez de Cisneros* (1876).

(H.G.K.)

Jinnah, Mohammed Ali

Mohammed Ali Jinnah, an Indian Muslim politician popularly known as Qāʾid-e A'zam, or the Great Leader, was the founder of the state of Pakistan.

Early years

He was born in Karāchi on December 25, 1876, the first of seven children of Jinnahbhai, a prosperous merchant. After being taught at home, Jinnah was sent to the Sind Madrasasah High School, in 1887. Later he attended the Mission High School, where, at the age of 16, he passed the matriculation examination of the University of Bombay. On the advice of an English friend, his father decided to send him to England to acquire business experience. Jinnah, however, had made up his mind to become a barrister. In keeping with the custom of the time. his parents arranged for an early marriage for him before he left for England.

BY courtesy of the Pakistan Embassy, Washington, D.C.

Jinnah.

In London he joined Lincoln's Inn, one of the legal societies that prepared students for the bar. In 1895, at the age of 19, he was called to the bar. While in London Jinnah suffered two severe bereavements—the deaths of his wife and his mother. Nevertheless, he completed his formal studies and also made a study of the British political system, frequently visiting the House of Commons. He was greatly influenced by the liberalism of William E. Gladstone, who had become prime minister for the fourth time in 1892, the year of Jinnah's arrival in London. Jinnah also took a keen interest in the affairs of India and in Indian students. When the Parsi leader Dadabhai Naoroji, a leading Indian nationalist, ran for the English Parliament, Jinnah and other Indian students worked day and night for him. Their efforts were crowned with success, and Naoroji became the first Indian to sit in the House of Commons.

When Jinnah returned to Karāchi in 1896, he found that his father's business had suffered losses and that he now had to depend on himself. He decided to start his legal practice in Bombay, but it took him years of work 'to establish himself as a lawyer.

It was nearly 10 years later that he turned toward active politics. A man without hobbies, his interest became divided between law and politics. Nor was he a religious zealot: he was a Muslim in a broad sense and had little to do with sects. His interest in women was also limited to Ruttenbai, the daughter of Sir Dinshaw Petit, a Bombay Parsi millionaire, whom he married over tremendous opposition from her parents and others. The marriage proved an unhappy one. It was his sister Fatima who gave him solace and company.

Jinnah first entered politics by participating in the 1906 Calcutta session of the Indian National Congress, the party that called for dominion status and later for independence for India. Four years later he was elected to the Imperial Legislative Council—the beginning of a long and distinguished parliamentary career. In Bombay he came to know, among other important Congress personalities, Gopal Krishna Gokhale, the eminent Marāthā leader. Greatly influenced by these nationalist politicians, Jinnah aspired during the early part of his political life to become "a Muslim Gokhale." Admiration for British political institutions and an eagerness to raise the status of India in the international community and to develop a sense of Indian nationhood among the peoples of India were the chief elements of his politics. At that time, he still looked upon Muslim interests in the context of Indian nationalism.

But by the beginning of the 20th century, the conviction had been growing among the Muslims that their interests demanded the preservation of their separate identity rather than amalgamation in the Indian nation that would for all practical purposes be Hindu. Largely to safeguard Muslim interests, the All-India Muslim League was founded in 1906. But Jinnah remained aloof from it. Only in 1913, when authoritatively assured that the league was as devoted as the Congress to the political emancipation of India, did Jinnah join the league. When the Indian Home Rule League was formed, he became its chief organizer in Bombay and was elected president of the Bombay branch.

"Ambassador of Hindu–Muslim unity"

Jinnah's endeavours to bring about the political union of Hindus and Muslims earned him the title of "the best ambassador of Hindu–Muslim unity," an epithet coined by Gokhale. It was largely through his efforts that the Congress and the Muslim League began to hold their annual sessions jointly, to facilitate mutual consultation and participation. In 1915 the two organizations held their meetings in Bombay and in 1916 in Lucknow, where the Lucknow Pact was concluded. Under the terms of the pact, the two organizations put their seal to a scheme of constitutional reform that became their joint demand vis-8-vis the British government. There was a good deal of give and take, but the Muslims obtained one important concession in the shape of separate electorates, already conceded to them by the government in 1909 but hitherto resisted by the Congress.

Meanwhile, a new force in Indian politics had appeared in the person of Gandhi. Both the Home Rule League and the Indian National Congress had come under his sway. Opposed to Gandhi's non-cooperation movement and his essentially Hindu approach to politics, Jinnah left both the Home Rule League and the Indian National Congress in 1920. For a few years he kept himself aloof from the main political movements. He continued to be a firm believer in Hindu–Muslim unity and constitutional methods for the achievement of political ends. After his withdrawal from the Congress, he used the Muslim League platform for the propagation of his views. But during the 1920s the Muslim League, and with it Jinnah, had been overshadowed by the Congress and the religiously oriented Muslim Khilāfat committee.

When the failure of the non-cooperation movement and the emergence of Hindu revivalist movements led to antagonism and riots between the Hindus and Muslims, the league gradually began to come into its own. Jinnah's problem during the following years was to convert the league into an enlightened political body prepared to cooperate with other organizations working for the good of India. In addition, he had to convince the Congress, as a prerequisite for political progress, of the necessity of settling the Hindu–Muslim conflict.

To bring about such a rapprochement was Jinnah's chief purpose during the late 1920s and early 1930s. He worked toward this end within the legislative assembly, at the Round Table Conferences in London (1930–32), and through his fourteen points, which included proposals for a federal form of government, greater rights for minorities, one-third representation for Muslims in the central legislature, separation of the predominantly Muslim Sind region from the rest of the Bombay province, and the introduction of reforms in the North-West Frontier Province. But he failed. His failure to bring about even minor amendments in the Nehru Committee proposals (1928) over the question of separate electorates and reservation of seats for Muslims in the legislatures frustrated him. He found himself in a peculiar position at this time; many Muslims thought that he was too nationalistic in his policy and that Muslim interests were not safe in his hands, while the Indian National Congress would not even meet the moderate Muslim demands halfway. Indeed, the Muslim League was a house divided against itself. The Punjab Muslim League repudiated Jinnah's leadership and organized itself separately. In disgust, Jinnah decided to settle in England. From 1930 to 1935 he remained in London, devoting himself to practice before the Privy Council. But when constitutional changes were in the offing, he was persuaded to return home.

Soon preparations started for the elections under the Government of India Act of 1935. Jinnah was still thinking in terms of cooperation between the Muslim League and the Hindu Congress and with coalition governments in the provinces. But the elections of 1937 proved to be a turning point in the relations between the two organizations. The Congress obtained an absolute majority in six provinces, and the league did not do particularly well. The Congress decided not to include the league in the formation of provincial governments and exclusive all-Congress governments were the result. Relations between Hindus and Muslims started to deteriorate, and soon Muslim discontent became boundless.

Creator of Pakistan

At this point, Jinnah emerged as the leader of a renascent Muslim nation. Events began to move fast. On December 22, 1939, soon after the resignation of the provincial Congress ministries, Jinnah called upon the Muslims to observe "the Day of Deliverance." Three months later, on March 22–23, 1940, in Lahore, the league adopted a resolution to form a separate Muslim state, Pakistan. The Pakistan idea was first ridiculed and then tenaciously opposed by the Congress. But it captured the imagination of the Muslims. Pitted against Jinnah were men of the stature of Gandhi and Nehru. And the British government seemed to be intent on maintaining the political unity of the Indian subcontinent. But Jinnah led his movement with such skill and tenacity that ultimately both the Congress and the British government had no option but to agree to the partitioning of India. Pakistan thus emerged as an independent state in 1947.

Jinnah became the first head of the new state. Faced with the serious problems of a young nation, he tackled Pakistan's problems with authority. He was regarded not as merely the governor general but was revered as the father of the nation. He worked hard until overpowered by age and disease, and he died in Karāchi, the place of his birth, on September 11, 1948.

BIBLIOGRAPHY. A definitive biography of Mohammed Ali Jinnah is yet to be written. But there are several readable accounts of his life and work both by Pakistani and foreign writers, the most recent of which is GULAM A. ALLANA, ***Quaid-e-Azam Jinnah: The Story of a Nation*** (1967). ***A*** more widely read book is HECTOR BOLITHO, ***Jinnah, Creator of Pakistan*** (1954). Among the older works MATLUBUL HASAN SAIYID, ***Mohammad Ali Jinnah: A Political Study*** (1945); and SAROJINI NAIDU, ***Mohamed Ali Jinnah, an Ambassador of Unity: His Speeches and Writings, 1912–1917*** (1920), are valuable. M.A.H. ISPAHANI, ***Qaid-e-Azam Jinnah As I Knew Him*** (1966), is written by one who worked with Jinnah in the Pakistan Movement and knew him well. JAMIL-UD-DIN AHMAD (ed.), ***Speeches and Writings of Mr. Jinnah, 2*** vol. (1960-64), ***Gandhi-Jinnah Talks*** (1944), and ***Letters of Iqbal to Jinnah*** (1963), throw light on important aspects of Jinnah's work. His role in the Pakistan Movement is dealt with in books dealing with the history of Pakistan, notably KEITH B. CALLARD, ***Pakistan*** (1957); and MOHAMMAD NOMAN, ***Muslim India*** (1942). The story of the events immediately preceding the creation of Pakistan is dealt with in CHAUDHRI MUHAMMED ALI, ***The Emergence of Pakistan*** (1967), which is an interpretation by a key figure among the Muslim officials; and V.P. MENON, ***The Transfer of Power in India*** (1957), by the most important Hindu official at the time. ALAN CAMPBELL-JOHNSON, ***Mission with Mountbatten*** (1951), is the version of one on the staff of Lord Mountbatten and is rather unsympathetic to Jinnah. A recent competent study of the events leading to Pakistan is H.V. HODSON, ***The Great Divide*** (1969).

(M.Hu.)

Joachim of Fiore

Mystic, theologian, biblical commentator, philosopher of history, and founder of an order of monks, Joachim of Fiore (also Flora, or Floris) attained great renown in his lifetime, and his ideas continued to be deeply influential on spiritual movements and thought in later times.

The known facts about the life of Joachim (Italian Gioacchino) of Fiore are few. He was born at Celico, Calabria, in southern Italy, *c*. 113011135. Legends about his parentage and youth are of little historical significance, but from an autobiographical reference it seems certain that he went on a pilgrimage to the Holy Land that reputedly had an effect on his conversion to the religious life. He became a Cistercian monk at Sambucina and in 1177 abbot of Corazzo (Sicily). About 1191 he broke away from the distracting duties of administration and retired into the mountains to follow the contemplative life. Although claimed as a fugitive by the Cistercians, he was allowed by Pope Celestine III to form the disciples who gathered around him at San Giovanni in Fiore (a town in present-day Cosenza province in Calabria) into the Order of San Giovanni in Fiore in 1196. He died in 1201/1202.

Spiritual experience and prophecy

Far more significant is the evidence for the inner development of a man who came to believe that spiritual understanding would be given to one who wrestled with the "letter" of the Scriptures to get at the "spirit." Three moments of special illumination are indicated, but the first is only known in legendary form, connected with either his pilgrimage or his novitiate at Sambucina. The second, recorded by himself, took place one Easter eve, after a period of frustrated study of the biblical book of Revelation when he felt himself "imprisoned" by difficulties. In the midnight silence, suddenly his mind was flooded with clarity, and his understanding released from prison. The third was an experience at Pentecost, when, after a time of agonizing doubt on the doctrine of the Trinity, Joachim had a vision of a psaltery with ten strings, in a triangular form, that clarified the mystery through a visual symbol and called forth paeans of praise from him. He expresses this experience of illumination given after mental striving in terms of the city seen intermittently by the approaching pilgrim or of the spirit breaking through the hard rind of the letter.

He was summoned by Pope Lucius III in 1184 and urged to press on with the biblical exegesis he had begun. This probably refers to the ***Liber concordie Novi ac Veteris Testamenti,*** in which Joachim worked out his philosophy of history, primarily in a pattern of "twos"—the concords between the two great dispensations (or Testaments) of history, the Old and the New. But already Joachim's spiritual experience was creating in his mind his truly original "pattern of threes." If the ***spiritualis intellectus*** springs from the letter of the Old and New Testaments, then history itself must culminate in a final age of the spirit that proceeds from both the previous ages. Thus was born his trinitarian philosophy of history in which the three Persons are, as it were, built into the time structure in the three ages or ***status*** of the Father, Son, and Holy Spirit.

The third ***status*** was to be won by the church only after arduous pilgrimage and great tribulation, like the Israelites marching through the wilderness and crossing the Jordan River into the Promised Land. As guides through this crucial stage, Joachim prophesied the advent of two new orders of spiritual men, one of hermits to agonize for the world on the mountaintop and one a mediating order to lead men on to the new spiritual plane. Although the third age belongs ***par excellence*** to contemplatives, secular clergy and laymen are not shut out of it. In a strange diagram, a "ground plan" of the New Jerusalem, various categories of monks are grouped around the seat of God, but below, secular clergy and tertiaries (lay members) live according to their rule.

In the ***Expositio in Apocalypsim*** Joachim seeks to probe the imminent crisis of evil, as pictured in the apocalyptic symbols of Antichrist, and the life of the spirit to follow. His third main work, the ***Psalterium decem chordarum,*** expounds his doctrine of the Trinity through the symbol of his vision of the ten-stringed psaltery. Here and in a lost tract he attacked the doctrine of "quaternity" (an overemphasis on the "one essence" of the Godhead that seems to separate it from the three Persons of the Trinity and so create a fourth), which he attributed to Peter Lombard, a 12th-century theologian. Besides this trilogy, written concurrently, Joachim left minor tracts and one uncompleted major work, the ***Tractatus super quattuor Evangelia.***

Joachim was a poet and artist. His lyricism breaks through the biblical exegesis that he chose as his medium and the turgid Latin of his style. Above all, his visual imagination is expressed in the unique ***Liber figurarum*** (discovered in 1937), a book of drawings and figures thought to be a genuine work by most Joachim scholars today. Here his vision of the culminating age of history is embodied in trees that flower and bear fruit luxuriantly at the top; his doctrine of the Trinity is expressed in remarkable geometric figures; his kaleidoscopic vision fuses images in some strange shapes, such as the tree that becomes an eagle, which may have influenced Dante. Joachim's figures probably carried his ideas in exciting and popular form far more widely than his indigestible writings.

Posthumous reputation and influence

In his lifetime Joachim was acclaimed as a prophet, gifted with divine illumination, and this is how he was seen by the first chroniclers after his death. The condemnation of his tract against Peter Lombard by the fourth Lateran Council in 1215 dimmed his reputation for a time, but the appearance of the Franciscan and Dominican mendicant orders, hailed as Joachim's new spiritual men, re-established him as a prophet. The Spiritual Franciscans at mid-13th century and various other friars, monks, and sects down to the 16th century appropriated his prophecy of a third age. But Joachim has always had a double reputation, as saint and as heretic, for cautious Christian thinkers and leaders have seen his writings as highly dangerous. The debate as to whether he was orthodox or heretic continues today.

BIBLIOGRAPHY

Chief works of Joachim: ***Liber concordie Novi ac Veteris Testamenti; Expositio in Apocalypsim*** (1527); ***Psalterium decem chordarum*** (1527); ***Tractatus super quattuor Evangelia,*** ed. by E. BUONAIUTI (1930); ***Liber figurarum,*** ed. by L. TONDELLI, M. REEVES, and B. HIRSCH-REICH (1954).

Secondary works: E. BUONAIUTI, ***Gioacchino da Fiore, i tempi, la vita, il messaggio*** (1931), good for background, some disputed views; A. CROCCO, ***Gioacchino da Fiore*** (1960), excellent on some aspects of Joachim's theology; H. GRUNDMANN, ***Studien iiber Joachim von Floris*** (1927), a fundamental work from which all modern research starts; ***Neue Forschungen iiber Joachim von Fiore*** (1950), surveys recent work to date; M. REEVES, ***The Influence of Prophecy in the Later Middle Ages: A Study in Joachimism*** (1969), attempts to trace Joachim's influence after his death; M. REEVES and B. HIRSCH-REICH, ***The*** Figurae ***of Joachim of Fiore*** (1972); L. TONDELLI, ***Il Libro delle Figure dell'Abate Gioacchino da Fiore,*** vol. 1 (1953), states the case for the authenticity of the ***Liber figurarum*** and examines its influence, especially on Dante; F. RUSSO, ***Bibliografia gioachimita*** (1954), useful but full of errors; K. LOWITH, ***Meaning in History,*** ch. 8 (1949), a consideration of the effects of Joachim's prophecy on modern political developments and ideology.

Articles: M.W. BLOOMFIELD, "Joachim of Flora: **A** Critical Survey of His Canon, Teachings, Sources, Biography and Influence," ***Traditio,*** 13:249–311 (1957); M. REEVES, "The ***Liber figurarum*** of Joachim of Fiore," ***Mediaeval and Renaissance Studies,*** 2:57–81 (1950), a short study with some illustrations.

(M.E.R.)

Joan of Arc, Saint

St. Joan of Arc was a peasant girl who, believing that she was acting under divine guidance, led the French army in repulsing an English attempt to conquer their country during the Hundred Years' War in a momentous victory at Orléans and, captured a year afterward, was burned by the English as a heretic. She became the greatest national heroine of her compatriots. Her achievement was a decisive factor in the revival of French determination to oust the English from France and in the awakening of French national consciousness.

St. Joan, equestrian miniature from the manuscript La *Vie des femmes célèbres* by Antoine Dufour, c. 1505. In the Musée Archéologique Thomas Dobrée, Nantes, France.
Giraudon

Joan was born the daughter of a plowman, Jacques of Arc, and baptized c. 1412 at Domremy, a village in the French part of the Duchy of Bar, between Champagne and Lorraine. She had three brothers, Jacquemin, Pierre, and Jean, and one sister, Catherine, who died while young. Joan, a peasant girl without formal learning, acquired her faith and her prayers from her mother. She was later remembered by her friends and neighbours as a hardworking, good, simple, and exceptionally pious child. At 13 she began to hear a voice from God that taught her self-discipline; later, the voice told her to go to France, where she would raise the siege of Orleans. She declared that she frequently saw St. Michael, St. Catherine, and St. Margaret, the patron saints of her country. These "voices" were to guide Joan throughout her life.

JOAN'S MISSION

The crown of France at the time was in dispute between the dauphin Charles, the son and heir of Charles VI, and the English king Henry VI, whose armies, in alliance with the Burgundians, were occupying nearly all the northern part of the kingdom. The apparent hopelessness of the Dauphin's cause at the end of 1427 was increased by the fact that, five years after his father's death, he had still not been crowned at Reims, the traditional place for the investiture of French kings. Reims was well within the territory held by his enemies. As long as the Dauphin remained unconsecrated, the rightfulness of his claim to be king of France was open to grave question in the minds of the people.

Led by her "voices," Joan, in May 1428, travelled from Domremy to Vaucouleurs, the nearest stronghold still loyal to the Dauphin, where she asked a captain for permission to join the Dauphin. He did not take the 16-year-old girl and her visions seriously, and she returned to Domremy. Joan went to Vaucouleurs again in January 1429; this time her quiet firmness and piety gained her the respect of the people, and the captain, persuaded that she was neither a witch nor feeble-minded, allowed her to go to the Dauphin at Chinon. She left Vaucouleurs about February 13, 1429, dressed in men's clothes and accompanied by six men-at-arms. Crossing territory held by the enemy, and after travelling for 11 days, she reached Chinon.

Meeting with the Dauphin

Joan at once went to the castle occupied by the Dauphin. He was uncertain whether to receive her and took counsel, receiving conflicting advice; but two days later Joan was granted an audience. The Dauphin had hidden himself among his courtiers, but she made straight for him and told him that she wished to go and make war against the English and that she would have him crowned at Reims. On the Dauphin's orders she was immediately interrogated by ecclesiastical authorities in the presence of Jean, Duc d'Alençon, a relative of Charles, who showed himself well-disposed toward Joan. For three weeks she was further questioned by eminent theologians at Poitiers. These examinations were occasioned by the ever-present fear of heresy. Joan told the ecclesiastics that it was not at Poitiers but at Orleans that she would give proof of her mission; and forthwith she dictated letters of defiance to the English. In their report the churchmen suggested that in view of the desperate situation of Orléans, which had been besieged by the English for several months, the Dauphin would be well-advised to make use of her.

Joan returned to Chinon. At Tours, during April, the Dauphin provided her with a military household of several men; Jean d'Aulon became her squire. She was joined by her brothers Jean and Pierre. She had her standard painted and a banner made for her; when the question of a sword was brought up, she declared that it would be found in the church of Sainte-Catherine-de-Fierbois, and one was in fact discovered there.

The troops, numbering several hundreds of men, were mustered at Blois and on April 27, 1429, they set out for Orléans. The city, besieged since October 12, 1428, was almost totally surrounded by a series of English strongholds. When Joan and La Hire, one of the French commanders, entered with supplies on April 29, she was told that action must be deferred until further reinforcements could be brought in.

The relief of Orléans

On the evening of May 4, when she was resting, Joan suddenly sprang up, apparently inspired, and announced that she must go and attack the English. Having herself armed, she hurried out to the east of the city toward an English fort, where, indeed, an engagement of which she had not been told was taking place. Her arrival aroused the French, who took the fort. The next day Joan addressed another of her letters of defiance to the English. On the morning of May 6 she crossed to the south bank of the river and advanced toward another fort; the English immediately evacuated it in order to defend a stronger position nearby, but Joan and La Hire attacked them there and took it by storm. Very early on May 7 the French advanced against the fort of Les Tourelles. Joan was wounded but quickly returned to the fight and it was thanks to her tenacity that the French commanders maintained the attack until the English capitulated. Next day the English were seen to be retreating, but, because it was a Sunday, Joan refused to allow any pursuit.

Joan left Orleans on May 9, 1429, and met the Dauphin at Tours. She urged him to make haste to Reims for his coronation. Though he hesitated because some of his counsellors were advising him to undertake the conquest of Normandy, Joan's importunity ultimately carried the day. It was decided, however, first to clear the English out of the other towns along the Loire River. Having met her friend the Duc d'Alençon, whom the Dauphin had made lieutenant general of the French armies, they moved off together and took a town and an important bridge. They then attacked Beaugency, whereupon the English retreated into the castle. Despite the opposition of the Dauphin and Georges de La Trémoille, one of his favourites, and the reserve of Alençon, Joan received the Constable de Richemont, who was under suspicion at the French court. After making him swear fidelity she accepted his help.

Battle of Patay

The French and English armies came face to face at Patay on June 18, 1429. Joan promised success to the French, saying that Charles would that day win a greater victory than any he had won so far. The victory was, indeed, complete: the English army was routed and with it, finally, its reputation for invincibility. A contemporary chronicler, although a partisan of Burgundy, freely acknowledged that Joan had now won so much honour and renown that all felt that the king's enemies had no longer any power against her.

Instead of pressing home their advantage by a bold attack upon Paris, Joan and the French commanders turned back to rejoin the Dauphin, who was staying with La Trémoille at Sully-sur-Loire. Accompanying the vacillating Charles on his perambulations among the towns along the Loire, she argued all the time to overcome his hesitancy and the advice of hostile counsellors. She was not unaware of the danger and difficulties but declared them of no account, insisting on the necessity of going swiftly to Reims. Ultimately she won Charles to her view.

Coronation of the Dauphin

From June 25, 1429, the army began to assemble at Gien; it was from there that the Dauphin sent out the customary letters of summons to the coronation. Joan wrote two letters: one to the people of Tournai, always loyal to Charles, the other to Philip the Good, the duke of Burgundy. She and the Dauphin set out on the march to Reims on June 29. Before arriving at Troyes, Joan wrote to the inhabitants, promising them pardon if they would submit. They countered by sending a friar, the popular preacher Brother Richard, to take stock of her; but though he returned full of enthusiasm for the Maid and her mission, the townsfolk decided after all to remain loyal to the English. The Dauphin held a council and Joan proposed that the town should be attacked; the next morning she began the assault and the inhabitants at once asked for terms. The royal army then marched on to Chalons; despite an earlier decision to resist, the count-bishop handed its keys to Charles. On July 16 the royal army reached Reims, which opened its gates to Charles. His coronation took place on July 17. Joan was present at the consecration, standing with her banner not far from the altar. After the ceremony she knelt before Charles, calling him her king for the first time. That very day she wrote to the Duke of Burgundy, adjuring him to make his peace with the King and to withdraw his garrisons from the royal fortresses.

The King left Reims on July 20, 1429, and for a month the army paraded through Champagne and the Île-de-France. On August 2 Charles decided on a retreat to the Loire, which implied abandoning the plan to attack Paris. The loyal towns that would thus have been left to the enemy's mercy expressed some alarm; Joan, who was opposed to Charles's decision, wrote to reassure the citizens of Reims on August 5, saying that the Duke of Burgundy, now in possession of Paris, had made a fortnight's truce, after which it was hoped that he would yield Paris to the King. In fact, on August 6, English troops prevented the royal army from crossing the Seine at Bray, so that it fell back on Provins, much to the delight of Joan and the commanders. Everywhere acclaimed, Joan was now, according to a contemporary chronicler, the idol of the French. For her part, she was enchanted with the people and the land and expressed the wish that, if it pleased God, she might, when she had finished her days, be buried in that district. She confided to Renaud de Chartres, the archbishop of Reims, that she would now like to go back to tending her parents' flocks.

Near Senlis, on August 14, 1429, the French and English armies again came face to face as at Patay; but this time only skirmishes took place, neither side daring to start a battle, though Joan carried her standard up to the enemy's earthworks and openly challenged them. Meanwhile Compiègne, Beauvais—whence the bishop, Pierre Cauchon, who was later to be one of Joan's two judges, fled precipitately—Senlis, and other towns north of Paris surrendered to the King; soon afterward, on August 28, a four months' truce for all the territory north of the Seine was concluded with the Burgundians.

Attack on Paris

Joan, however, was becoming more and more impatient; she thought it essential to take Paris. She and Alençon were at Saint-Denis on the northern outskirts of Paris on August 26, and the Parisians began to organize their defenses: Charles arrived on September 7 and an attack was launched on September 8, directed between the gates of Saint Honoré and Saint-Denis. The Parisians could be in no doubt of Joan's presence among the besiegers: she stood forward on the earthworks, calling on them to surrender their city to the King of France. Wounded, she continued to encourage the soldiers until she had to abandon the attack. When next day she and Alençon sought to renew the assault, they were ordered to retreat. Charles VII retired to the Loire, Joan following him. At Gien, which they reached on September 22, 1429, the army was disbanded and Alençon and the other captains went home. Only Joan remained with the King. Later on, when Alençon was planning a campaign in Normandy, he asked the King to let Joan rejoin him, but La Trémoille and other courtiers dissuaded him, and Joan went with the King to Bourges, where many years later her goodness and her generosity to the poor were recalled. Nearby Joan met a clairvoyant, who told Joan that she wished to effect a reconciliation between the Duke of Burgundy and the King of France. Joan replied that in her view no peace could be found save at the lance's point; and she told Charles that the woman's pronouncements were foolish and empty. In October 1429 she was sent against Saint-Pierre-le-Moutier; it was through her courageous assault, with only a few men, that the town was taken. Laying siege to La Charité-sur-Loire, and short of munitions, they appealed to neighbouring towns for help. The supplies arrived too late and after a month they had to raise the siege.

Battle of Compiègne

Joan then rejoined the King, who was spending the winter in towns along the Loire. Late in December 1429, Charles issued letters patent ennobling Joan, her parents, and her brothers. Early in 1430 the Duke of Burgundy began to threaten Brie and Champagne; the citizens of Reims became alarmed and Joan wrote to them in March to assure them of the King's concern and that she would come to their defense. When the Duke of Burgundy moved up to attack Compikgne, the townsfolk determined to resist and, in late March or early April, Joan left the King and set out to their aid, accompanied only by her brother Pierre, her squire Jean d'Aulon, and a small troop of men-at-arms. She arrived at Melun in the middle of April, and it was no doubt her presence that stimulated the townsfolk to declare themselves for Charles VII.

Joan was at Compikgne by May 14, 1430. There she found Renaud de Chartres, the archbishop of Reims, and Louis I de Bourbon, comte de Vendôme, a relative of the King's. With them she went on to Soissons, where the townsfolk there refused them entry. Renaud and Vendôme therefore decided to return south of the Marne and Seine rivers; but Joan refused to accompany them, preferring to return to "her good friends" in Compiègne.

CAPTURE, TRIAL, AND EXECUTION

On her way back, Joan heard that John of Luxembourg, the leader of a Burgundian band, had laid siege to Compiègne. Hurrying on, she entered Compikgne under cover of darkness, and the next afternoon she led a sortie and twice repelled the Burgundians, but was eventually outflanked by English reinforcements and compelled to retreat. Remaining till the last in order to protect the rear guard while they crossed the Oise River, she was unhorsed and could not remount. She gave herself up and, with her brother Pierre and Jean d'Aulon, was taken to Margny, where the Duke of Burgundy came to see her. In telling the people of Reims of Joan's capture, Renaud de Chartres accused her of rejecting all counsel and having acted willfully.

Imprisonment by John of Luxembourg

John of Luxembourg sent Joan and Jean d'Aulon to his castle in Vermandois. When she tried to escape in order to return to Compikgne, he sent her to a more distant of his castles. There, though she was kindly treated, she became more and more distressed at the predicament of Compikgne; her desire to escape became so great that she jumped from the top of a tower, falling unconscious into the moat. She was not seriously hurt, and when she had recovered she was taken to Arras, a town belonging to the Duke of Burgundy.

News of Joan's capture had reached Paris on May 25, 1430. Next day, the University of Paris, which had taken the English side, requested the Duke of Burgundy to hand Joan over for judgment either to the chief inquisitor or to the bishop of Beauvais, Pierre Cauchon, in whose diocese she had been captured. The university wrote also, to the same effect, to John of Luxembourg; and on July 14 the Bishop of Beauvais presented himself before the Duke of

Burgundy asking, on his own behalf and in the name of the English king, that the Maid be handed over in return for a payment of 10,000 francs. The Duke passed on the demand to John of Luxembourg, and by January 3, 1431, Joan was in the Bishop's hands. The trial was fixed to take place at Rouen. Joan was then imprisoned in a tower of the castle of Bouvreuil, which was occupied by the Earl of Warwick, the English commander at Rouen. Joan's offenses against the English king were common knowledge. But she was brought to trial before a church court because it would not have served English interests to do away either by imprisonment or execution with so popular a figure unless they had first undermined the people's faith in her, and, through her, in the recently crowned king. The purpose of the trial was to show that Charles VII was a fool in the hands of a witch, or at least an unbeliever. Her judges were to be the Bishop of Beauvais and Jean Lemaitre, the vice inquisitor of France.

Interrogation

The trial. From Jan. 13, 1431, statements taken in Lorraine and elsewhere were read before the bishop and his assessors and would provide the framework for Joan's interrogation. Summoned to appear before her judges on February 21, Joan asked for permission to attend mass beforehand, but it was refused on account of the heinousness of the crimes with which she was charged. She was ordered to swear to tell the truth and did so swear; but she always refused to reveal the things she had said to Charles. Cauchon forbade her to leave her prison, but Joan insisted that she was morally free to attempt escape. Guards were then assigned to remain always inside the cell with her; and she was chained to a wooden block and sometimes put in irons. Between February 21 and March 24 she was interrogated nearly a dozen times; on every occasion she was required to swear anew to tell the truth, but she always made it clear that she would not necessarily divulge everything to her judges since they were the enemies of King Charles. The report of this preliminary interrogation was read to her on March 24; apart from two points she admitted its accuracy.

When the trial proper began a day or so later, it took two days for Joan to answer the 70 charges that had been drawn up against her. These were based mainly on the contention that her whole attitude and behaviour showed blasphemous presumption: in particular, that she claimed for her pronouncements the authority of divine revelation; prophesied the future; endorsed her letters with the names of Jesus and Mary; claimed to be assured of salvation; immodestly wore men's clothing; and dared to claim that her saint spoke in French and not in English. Perhaps the most serious charge was of preferring what she assumed to be the direct commands of God to those of the church.

On March 31 she was questioned again on several points about which she had been evasive, notably on the question of her submission to the church. In her position, obedience to the court that was trying her would inevitably be made the test of such submission. She did her best to avoid this trap, saying she knew well that the church could not err; but it was to God that she held herself answerable for her words and actions. The trial continued: the 70 charges were reduced to 12, which were sent for consideration to many eminent theologians in Paris as well as at Rouen.

Meanwhile, Joan fell ill in prison; she was attended by two doctors, since it did not suit the Earl of Warwick that she should die with the trial unresolved. She received a visit on April 18 from Cauchon and his assistants, who exhorted her to heed their advice. Joan, who was seriously ill and obviously thought she was dying, begged to be allowed to go to confession and receive Holy Communion and to be buried in consecrated ground. But they continued to badger her with their admonitions until May 2, receiving only her constant reply, "I am relying on our Lord," "I hold to what I have already said." They became more insistent on May 9, threatening her with torture if she did not clarify certain points. She answered that even if they tortured her to death she would not reply differently, and added that in any case she would afterward maintain that any statement had been extorted from her by force. In face of this common-sense fortitude her enemies, by a majority of ten to three, decided that torture would be useless. Joan was informed on May 23 of the decision of the University of Paris that if she persisted in her errors she would be handed over to the secular authorities for punishment (*i.e.*, execution). A young theologian attempted once more to change her mind, but she maintained the truth of what she had said at the trial.

Abjuration, relapse, and execution. Nothing further, therefore, could be done. She was taken out of prison for the first time in four months on May 24 and was taken to the cemetery of the Church of St. Ouen where the sentence was to be read out. First she had to listen to a sermon by one of the theologians, who, during his peroration, violently attacked Charles VII, provoking Joan to interrupt him, because she thought he had no right to attack the King, a "good Christian," and should confine his strictures to her. After the sermon was ended she asked that all the evidence on her words and deed should be sent to Rome. But her judges ignored her appeal to the pope, to whom, under God, she would be answerable, and began to read out the sentence abandoning her to the secular power. Hearing this dreadful pronouncement, Joan quailed, saying she would do all that the church required of her. She was, therefore, presented with a form of abjuration, which must already have been prepared. She hesitated in signing it, eventually doing so "on condition it was pleasing to Our Lord." She was then condemned to perpetual imprisonment or, as some maintain, to imprisonment in a place habitually used as a prison.

Revocation of abjuration

In any case, the English insisted on her returning to her former prison; she obeyed the order of the vice inquisitor that she should wear women's clothes. But two or three days later, when the judges and others visited her, and found her again in male attire, she said she had made the change of her own free will, preferring men's clothes. They then pressed other questions, to which she answered that St. Catherine and St. Margaret had censured her "treason" in making an abjuration. These admissions were taken to signify relapse, and on May 29 the judges and 39 assessors agreed unanimously that she must be handed over to the secular authorities. The next morning, May 30, 1431, Joan received from Cauchon permission, unprecedented for a relapsed heretic, to make her confession and receive Communion. Accompanied by two Dominicans she was then led to the Place du Vieux-Marché. There she endured one more sermon, and the sentence abandoning her to the secular arm, that is, to the English, was then read out in the presence of her judges and of a great crowd. The executioner seized her and led her to the stake, lighting the pyre. A Dominican consoled Joan, who asked him to hold high a crucifix for her to see and to shout out the assurances of salvation so loudly that she should hear him above the roar of the flames. To the last she maintained that her voices were sent of God and had not deceived her.

Few witnesses of her death seem to have doubted her salvation; and 26 years later, at the rehabilitation, all witnesses agreed that she died a faithful Christian. The executioner had never before found his task so fearful; he afterward declared that Joan's heart would not burn and that he found it intact amid the ashes. A few days later the English king and the University of Paris formally published the news of Joan's execution.

Rehabilitation

Almost 20 years afterward, on his entry into Rouen in 1450, Charles VII ordered an inquiry into the trial. Two years later the cardinal legate Gillaume d'Estoutville made a much more thorough investigation. Finally, on the order of Pope Calixtus III, proceedings were inaugurated that revoked and annulled the sentence of 1431. Joan was canonized by Pope Benedict XV on May 16, 1920, and the church celebrates her feast on May 30. The French parliament, on June 24, 1920, decreed a national festival in her honour; this is held on the second Sunday in May.

CHARACTER AND IMPORTANCE

Joan's place in history is assured. Her relief of Orléans undoubtedly proved a turning point in the Hundred

Years' War; and her achievement in reviving the courage of the French people and their faith in the Valois dynasty is in no way belittled by acknowledging that the reconciliation between Charles VII and the Duke of Burgundy in 1435 was yet to provide the practical basis for a French recovery that could in any case hardly have been averted once England in its turn began to be rent by civil war. Beyond this, the story of Joan's mission will probably be a source of controversy for historians and psychologists. Innumerable facts about her campaigns and about the motives and actions of her supporters and enemies are subject to dispute: for instance, the number and dates of her visits to Vaucouleurs; how she was able to convince the Dauphin at their first meeting at Chinon; whether Charles's perambulations after Reims represent triumphant progress or scandalous indecision; what Cauchon meant by "perpetual imprisonment"; whether Joan, after her recantation, resumed men's clothes really of her own free will and at the bidding of her voices, or, as one story has it, because they were forced upon her by her English jailers.

Assessment

Assessments of Joan's character and mission can be equally varied; some will agree with the church that she was a saint, others dub her a deluded hysteric. Joan liked gay clothes and undoubtedly enjoyed the acclamations of the crowds. Her sturdy piety must at times have been infuriating: she constantly ordered the entire royalist army to go to confession. and she lost valuable tactical advantage by refusing battle on Sundays or festivals, saying that God would make up to them and they would win other victories. Her answer to her judges, demonstrating the pointlessness of employing torture, shows her practicality; she could shrewdly turn off stupid questions, replying to a query as to whether St. Michael had long hair by asking why it should have been cut off.

But essentially Joan exhibits the basic characteristic and stamp of the genuine prophets and saints, utterly subordinated to a particular task that they believe has been inexorably laid upon them by divine command; Joan differs from many of them only in that she seems never, except for a few days near the end, to have swerved from her appointed path.

BIBLIOGRAPHY. There is an extremely large literature on Joan of Arc, but no truly definitive biography. Important works include: J. QUICHERAT, *Procès de condamnation et de rehabilitation du Jeanne d'Arc*, 5 vol. (1841-49); *Le Prods de condamnation de Jeanne d'Arc, 2* vol., trans. and annotated by P. TISSET (1960–70); *La Réhabilitation de Jeanne la Pucelle,* 3 vol., trans. and annotated by P. DONCOEUR and Y. LANHERS (1956–61), containing the recorded testimony from the trial of more than 100 witnesses who had known Joan of Arc; and LANERY D'ARC, *Les Mémoires et consultations en faveur de Jeanne d'Arc* (1889), dealing with the memoirs of jurists and theologians, intended to facilitate her rehabilitation. In English, W.P. BARRETT, *The Trial of Jeanne d'Arc* (1931); D.S. RANKIN and C. QUINTAL, *The First Biography of Joan of Arc with the Chronicle Record of a Contemporary Account* (1964); and R. PERNOUD, *The Retrial of Joan of Arc* (Eng. trans. 1955), are particularly recommended.

(Y.L.)

Johanan ben Zakkai

Of all the Palestinian Jewish sages of the 1st century AD, none apparently proved so fundamentally influential in his own time and for subsequent generations of scholars and spiritual leaders as Johanan ben Zakkai. Although this seems to have been true even before 70 CE (Common Era = AD), when the Jerusalem Temple—the principal cultic and religious institution of the Jews—was destroyed and the country was devastated, it was certainly the case in the last third of the 1st century.

Sources of knowledge about Johanan

As is the case with all the Talmudic teachers (the rabbis who interpreted and applied the Oral Law), little strictly biographical information about Johanan ben Zakkai has been preserved: Talmudic and Midrashic sources (commentative and interpretative writings) are principally devoted to the teachings of the sages and of what they came to represent. Thus, what can be reported essentially about Johanan is this: even before 70 he acted as a leading representative of the Pharisees in debate with priestly and Sadducean authorities. (The Pharisees stressed rigorous observance of the Law, inclusion of the oral tradition as normative, and an interpretative adaptation of traditional precepts to new situations; the Sadducees, an elitist conservative group, accepted only the Written Law as authoritative and were more literalist and static in their interpretation.) Johanan's school was apparently famous, and one in search of learning would go to extremes, if need be, to be admitted there. Furthermore, Johanan was opposed to the policy of those who were determined on war with Rome at all costs. By quitting beleaguered Jerusalem according to most accounts in 70 (though it is possible that he left as early as 68) and being brought to the Roman camp, he somehow succeeded in getting permission to set up an academy in Jamnia (Jabneh), near the Judaean coast, and there he was joined by a number of his favourite disciples. Two of them, Eliezer ben Hyrcanus and Joshua ben Hananiah, who are credited with having smuggled their master out of Jerusalem in a coffin, were to become, by the end of the century and the beginning of the following one, the leading teachers of their generation and had a profound influence on the greatest scholars of the next generation. It is therefore hardly excessive to say that Johanan's teachings are to be traced not merely to the relatively few statements specifically attributed to him but to many views that become articulate during the 2nd century: for example, that acts of loving-kindness atone no less effectively than the former Temple sacrificial ritual and are indeed at the core of the universe since its creation; that the study of Torah (the divine instruction or Law) is a central purpose of man and a paramount form of serving God; that a number of ceremonies and regulations once confined to the Temple were to be adopted even outside the Temple complex "to serve as memorials of the Sanctuary"; at the same time, despite the unique sanctity of Jerusalem, basic decisions regarding practice and instruction were now to be permitted to the authorized scholars wherever circumstances compelled them to sit in session. Such views, truly radical in origin, became normative rabbinical teaching and permanent components of Judaism.

Thus, it may be said that, by establishing in Jamnia a major academy and authoritative rabbinic body, Johanan fixed the conditions of continuity of basic traditions; and that, by his lively sense of the need for reinterpreting inherited concepts lest they become irrelevant in new circumstances, he laid the foundations on which Talmudic and rabbinic Judaism built their structure.

Work on the Law (Halakha)

The chief preoccupation of Johanan and his students was the study and continuing development of the Law (Halakha). He and they also engaged in the study of nonlegal subjects (Aggada), especially in connection with biblical exegesis (Midrash), explanation and interpretation of the biblical contents. In addition, he was interested in esoteric themes related to the subject of creation and the visions of the Merkavah (the divine chariot of Ezek. 1), discourses on which were even delivered by some of his disciples. And, at least before the destruction of the Temple, if not thereafter as well, he seems to have held occasional sessions when certain ethical–philosophical questions, typical of Hellenistic–Roman popular philosophical discussion, were raised and explored. His homiletical interpretations of scripture often unite the symbolic with the rationalistic in a remarkable way. Why were not hewn stones permitted in the building of the altar? Because iron is for weapons of destruction, and the altar of God is intended to bring peace, he answers. Why is the ear of one who prefers servitude to have a hole bored in it? Because we are God's servants, and man heard at Sinai with his own ears. Let the unlistening ear be bored. Such are typical comments by Johanan. Although he had discouraged what must have seemed to him unwarranted messianic proclamations, a saying attributed to him in his last illness suggests that messianic speculation was not alien to him.

In the history of Talmudic literature and thought, Johanan is rightly seen as continuing the Hillelite tradition, although this should not be interpreted to mean that he inherited only Hillel's teachings (see HILLEL).

BIBLIOGRAPHY. The most important modern study of Johanan ben Zakkai is that by G. ALON, *Studies in Jewish History,* vol. 1, pp. 219–273 (1957), in Hebrew. In addition to the summary description in the general Jewish histories such as H.H. GRAETZ, *Geschichte der Juden,* vol. 2 (1897; Eng. trans., *History of the Jews, from the Earliest Times to the Present Day,* vol. 2, pp. 321–359, 1891), see W. BACHER, *Die Agada der Tannaiten,* vol. 1, pp. 25–46 (1884); J. GOLDIN, "On the Academy of Johanan ben Zakkai," in AMERICAN ACADEMY FOR JEWISH RESEARCH, *Harry Austryn Wolfson: Jubilee Volume on the Occasion of His Seventy-Fifth Birthday,* vol. 3, pp. 69–92 (1965) in Hebrew; and J. NEUSNER, *A Life of Yohanan ben Zakkai, ca. 1–80 C.E.,* 2nd ed. rev. (1970).

(J.Gol.)

Johannesburg

Johannesburg, the largest city of the Republic of South Africa and the second largest city on the African continent, is South Africa's economic metropolis. It stands at an altitude of 5,750 feet (1,753 metres) on the southern slopes of the Witwatersrand—commonly called the Rand—a rocky watershed of east–west ridges surrounded by the Transvaal highveld. The greater metropolitan area has a total population of about 1,500,000; loosely known as "Greater Johannesburg," it consists of a municipal and a magisterial area.

Johannesburg grew at a remarkable speed after the discovery of gold on the Rand in 1856. Today it stands at the centre of the country's gold-mining industry. Johannesburg is the only city of its size that is not situated on a coast, a lakeshore, or on a river. (For historical aspects see SOUTHERN AFRICA, HISTORY OF.)

The discovery of gold

History. Johannesburg was founded as a result of the discovery of gold in the Witwatersrand area of the Transvaal in 1886. The farms from Driefontein in the east to Roodepoort in the west were proclaimed as public diggings, and gold-mining operations began. The centre of contemporary Johannesburg now stands on land that was then government ground and was known as Randjeslaagte or Rantjeslaagte. There has been much controversy about the origin of the name Johannesburg; the earliest known official statement about the Johannes, for whom the city was named, dated from 1896, when it was recorded that it was called after Johann Rissik, acting surveyor general, and Christian Johannes Joubert, head of the mines department of the Zuid-Afrikaansche Republiek (South African Republic), in whose territory the goldfields were located. The two men had been appointed as commissioners to investigate the mining situation. By November 1886, according to the provisions of the act known as the Gold Law, the first members of a Goldfield's Diggers Committee were elected and became the city's first local government authority. This body was later called the Sanitary Committee but by 1897 had become a Stadsraad, or Town Council. In 1887, the year after the discovery of gold, the mining syndicates—forerunners of the mining companies of today—had begun to operate in Johannesburg; and houses, offices, churches, and shops had begun to spring up.

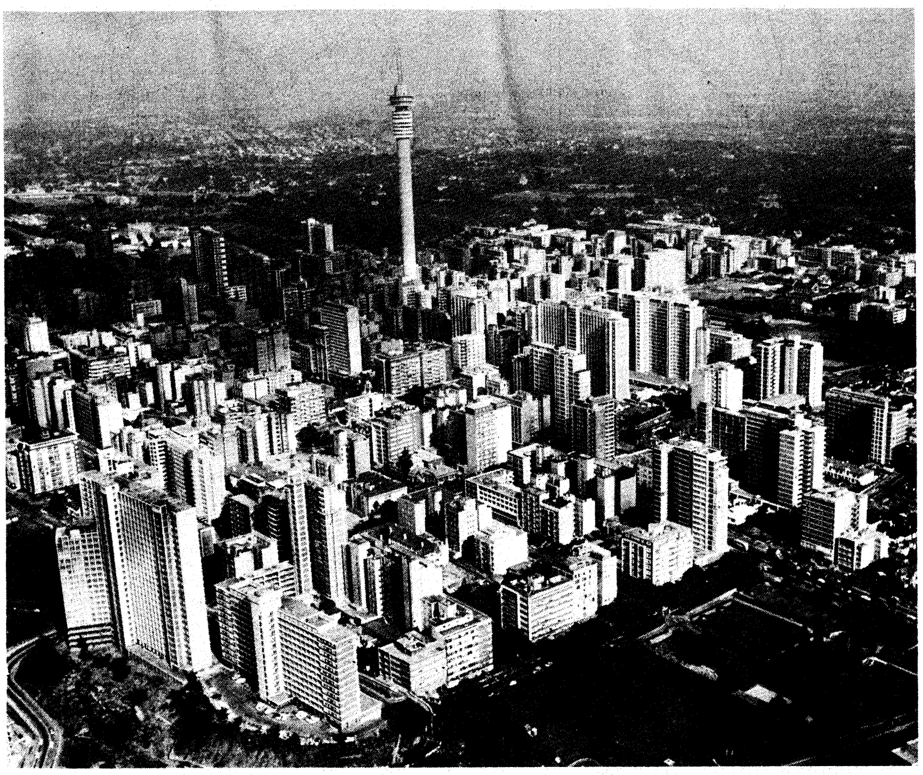
In 1889 the new city's expansion was halted when the miners reached a zone of pyrite (an ore of sulfur, associated with gold and sometimes called fool's gold) from which it was not known how to extract the gold. As a result, a number of mines closed down before three Scotsmen from Glasgow, the brothers R.W. Forrest and W. Forrest and J.S. MacArthur, discovered the cyanide method of gold extraction. This process saved the Rand, and Johannesburg's growth was resumed.

Although postal facilities of a sort were available from the outset and a telegraph office was opened in 1887, telephone service was not installed until 1894. Goods had to be carried in by ox wagon, and passengers were transported by coach and horses; the only railroad in the locality, from Johannesburg to nearby Boksburg, was primarily used for the transport of coal to the mines. In 1892, however, the city was linked by rail to the port of Cape Town. By 1896 the inhabitants of Johannesburg, most of whom were English-speaking and did not have the franchise, came to wish for representation in the government of the country. As a result, differences between the rulers of the country in Pretoria and the mining men in Johannesburg became acute. When the South African War between the United Kingdom and the Zuid-Afrikaansche Republiek broke out in 1899, many residents left Johannesburg. By agreement, the mines were handed over undamaged when the British took the city in 1900. At the conclusion of the war in 1902, by which time the city's population numbered 100,000, labour for the mines was so scarce that Chinese workers were imported. Due to political pressures, however, the Chinese were all repatriated by 1910, the year in which Johannesburg, together with the remainder of the Transvaal, became a part of the Union of South Africa. After World War I, labour unrest occasioned various strikes in the mines. In a strike of 1922, white miners opposed the use of Africans for semiskilled work. The strike was suppressed but at a cost of more than 200 lives. In 1928 Johannesburg was declared to be a city; it celebrated its 50th anniversary in 1936. World War II checked further development from 1939 to 1945, after which a building boom occurred, with urban expansion taking place to the north and northwest. By 1970 the city's municipal boundaries were extended to include an area of 104 square miles.

Expansion of the city after World War II

Physical description. Johannesburg is situated in the most thickly populated part of southern Africa. The mines, many of which were almost worked out in the early 1970s, lie to the south and southwest of the city. In September 1970 it was estimated that of the restricted municipal population about 483,000 were whites, about 804,000 were black Africans, 39,000 were Asians, and about 83,000 were Coloureds (of mixed white and nonwhite descent). The city's height above sea level, which as already mentioned is 5,750 feet, is measured at Joubert Park near the centre of the city; the highest point, however, is to be found near the suburb of Northcliff, which stands at an altitude of 5,928 feet. There are about 400 suburbs altogether; most of the city's workers live either in the suburbs or in the surrounding dormitory towns. The population is segregated by colour, in accordance with South Africa's official policy of apartheid (literally "apartness"), so that the nonwhite groups—black Africans, Coloureds, Asians—are each restricted to residence in certain areas, which are located particularly in the west and southwest. The major buildings in the city are grouped around Eloff Street, the main shopping area that leads south from the railway station, with Von Brandis, Joubert, and Rissik streets parallel to it. Other streets in the vicinity include Commissioner, Market, Fox, and Main streets (which constitute the commercial area) and Jeppe and Bree streets. Joubert Park, Milner Park, George Harrison Park (named for the Australian who first discovered gold-bearing quartz in the area in 1886), Turffontein Race Course, sports grounds, and mine dumps are all in the area. Johannesburg is the meeting place of roads from major towns in South Africa, such as Pretoria, Bloemfontein, Cape Town, and Durban. Not included in the municipal area, though administered by the municipality, is a group of towns inhabited by Africans (Bantu) that is known as Soweto (southwestern townships) and that covers an area of 26 square miles; it is linked with Johannesburg by road and rail. By 1961 the city council had provided subsidized housing in Soweto for 57,000 families and 26,000 single men—a total of about 370,000 persons. Water, sanitation, and lighting are supplied, and sports fields, schools, and other facilities are provided by the city council. All shopkeepers in the area are black Africans. Welfare and employment services are staffed by about 400 whites and about 3,000 black Africans. Other townships include Coronationville and Westbury, reserved for Coloured people, and—about 20 miles away—Lenasia, a township for Asians.

Climate. Johannesburg has an attractive climate; in July the mean temperature is about 50° F (10" C); in December it is 68" F (20" C). Rainfall averages about 30 inches a year. The sun shines for an average of nine hours a day in summer, and more than seven hours in winter.



Downtown Johannesburg skyline dominated by a television transmission tower.

Georg Gerster—Rapho Guillumette

Little remains of the original plant and animal life, except in the Melville Koppies Reserve to the northwest of the city. Suburban gardens flourish and are often elegant. The city's air and water are relatively pure in most parts; water shortages occur periodically, especially during the winter months.

City planning. The streets of the centre of the city, laid out on a grid plan established in 1886, are narrow, and city blocks are short. More imaginative street planning is in evidence in the suburbs, where many of the streets are tree lined.

Transport. Motorways are being built in many parts of the city to accommodate the large amount of traffic that enters and leaves the city each day. In the suburbs local public transport is provided by bus services, whereas the railroad, used by 160,000 passengers daily, serves commuters living outside municipal boundaries to the east, west, and south. Jan Smuts (international) Airport, from which scheduled domestic flights also leave, is situated 14 miles northeast of the city; charter planes fly from Rand Airport, nine miles to the east; and Baragwanath Aerodrome, seven miles to the southeast, has facilities for gliding and other aerial sports.

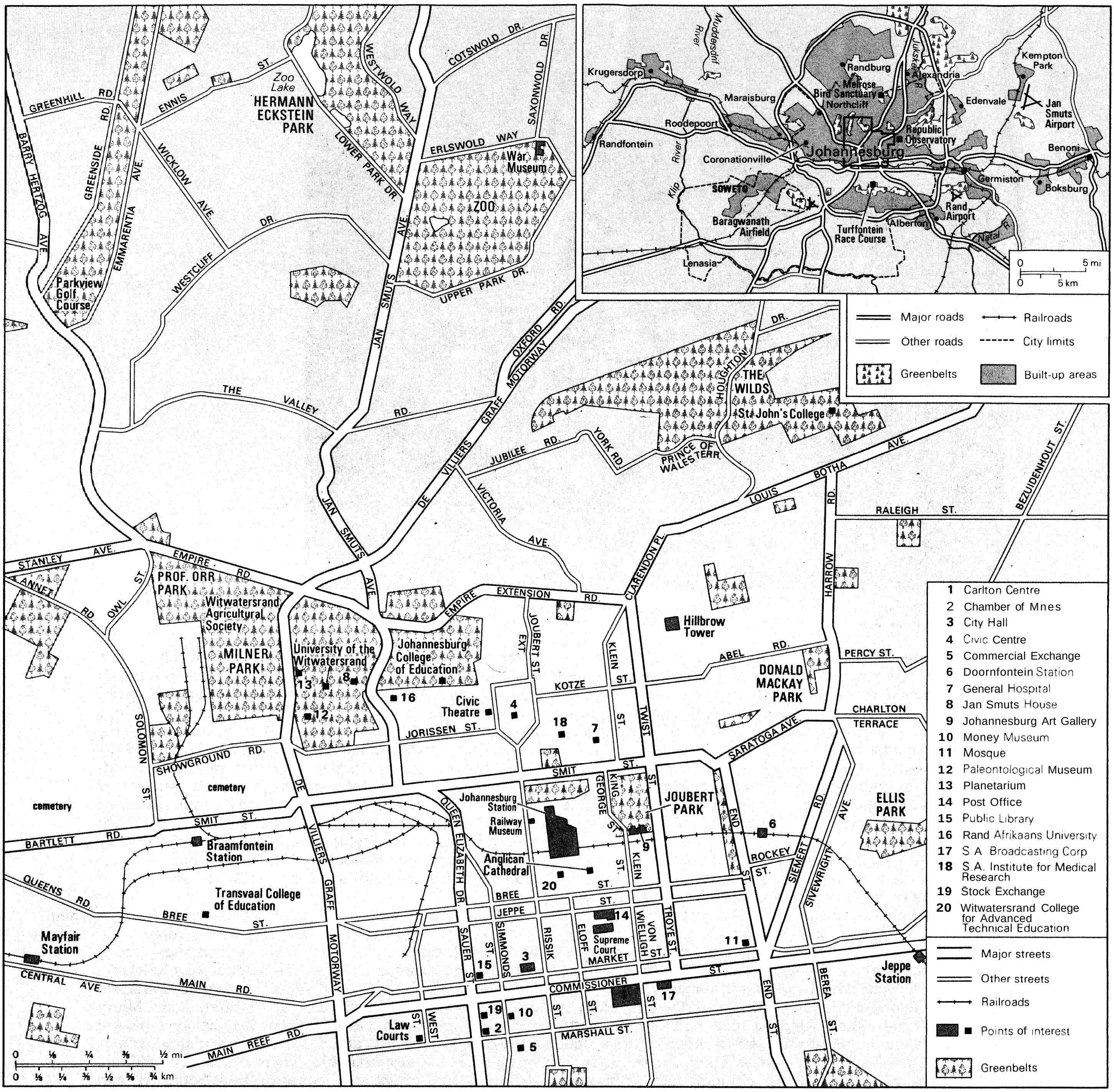
Demography. The city is cosmopolitan. Although English- and Afrikaans-speaking groups dominate the white section of the population, minority groups of Germans, Dutch, Hungarians, Italians, French, Scandinavians, Swiss, Poles, and others are also to be found, while Jews from all of these groups have also made Johannesburg their home. Africans speaking Zulu, Xhosa, Pedi, Venda, and Tswana also inhabit the city, while Asian groups include Japanese, Chinese, and Indians, speaking various languages and adhering to various faiths. The Coloured people of the city represent an intermixture of many groups, both white and nonwhite; they speak English or Afrikaans or both, and belong to Christian churches.

Housing. It is customary in Johannesburg to demolish and rebuild houses, rather than to follow the European practice of renovating or altering existing buildings. There are numerous blocks of flats or apartments for all types of income groups, and many South Africans have been able to maintain houses with gardens. Architectural styles of all kinds are in evidence. Buildings of the style developed from 18th-century Cape architecture are noticeable in residential areas, whereas business buildings and many houses are derived from modem European or United States models. Public buildings are modem adaptations of older European styles, particularly English, Dutch, and Renaissance Italian. By the early 1970s it had become the practice to construct apartment buildings (called flats in Johannesburg) so that residential density can be increased; it was still the custom, however, to build single-story homes for married nonwhites. Subsidized housing is available for eligible (*i.e.*, low-income) groups of the population, and large building projects are features of Soweto and other areas to the south of the city.

South Africa's business centre

Business and industry. Because Johannesburg is the centre of the South African business world, more well-to-do people are to be found in this city than elsewhere in the country. Gold mining is now conducted elsewhere, but most of the business and administrative headquarters of the gold-mining companies are located here. Many secondary industries, particularly of the heavier type, are to be found in the city, and opportunities for employment of many kinds are available. There are many banking, industrial, and commercial concerns, both South African and foreign; the city is also the home of the South African Stock Exchange, which was founded in 1887. Because of the city's importance, it has a number of branch offices of governmental institutions, as well as consular offices and other institutions (such as banks, building societies, and insurance companies) usually located only in capital cities.

Government utilities and services. Johannesburg is a municipally controlled city. It has a city council with 47 councillors (1971), elected by popular vote every five years; a mayor is chosen by the councillors from among themselves each year. The city council is the local government unit and is responsible to the Transvaal provincial council, which is in turn responsible to the central government. City councillors represent the same political parties that are represented in the national Parliament. Its local transport, electricity and gas utilities, fire-fighting services, and sanitation are run by the municipality. Water is provided by the municipality through the agency of the Rand Water Board. The rail-



Central Johannesburg and (inset) its metropolitan area.

ways are government owned, with the city's large railway station constituting the heart of the nation's rail system. Electricity for the mines and railways is provided by a public utility with headquarters in the city.

Police services are provided by the nationally administered South African Police Service, but traffic is regulated by the municipality's traffic officers. The city operates produce and livestock markets and the abattoir. The Witwatersrand Agricultural Society organizes a major show in Milner Park every year. It attracts agricultural and industrial exhibitors from throughout the country as well as from overseas. Trade is promoted by the Johannesburg Chamber of Commerce; the Commercial Exchange, a unique institution, owes its existence to commerce generated by the mining industry. Mining operations themselves are regulated by the South African Chamber of Mines, which also has its headquarters in the city.

Health and education. In addition to Johannesburg General Hospital, medical facilities for whites include many private nursing homes and institutions for sick children, maternity cases, and mine workers. Medical facilities for nonwhites are also extensive, and include clinical, curative, midwifery, and dental services; most mines also provide hospitalization for employees. Medical research is pursued at the South African Institute for Medical Research as well as at the medical school of the University of the Witwatersrand.

The University of the Witwatersrand itself, with an enrollment of about 8,000 students, is mainly intended for English-speaking students; higher education in Afrikaans is provided at the Rand Afrikaans University (founded in 1966). The Witwatersrand College for Advanced Technical Education specializes in technological training related to mining and other subjects. In addition, there are throughout the city **a** great number of schools,

Teacher-training colleges

both private and governmental, for pupils of all ages, from kindergarten to school-leaving age. Teacher-training courses are given at the Johannesburg College of Education (for English-speaking students), at the Goudstadse Onderwyskollege (for Afrikaans-speakers), at the Transvaal College of Education (for Asians), and at the Rand College of Education (for Coloured students).

Cultural institutions. Cultural institutions include the Johannesburg Public Library and a number of museums with collections relating to South African history, military history, medicine, archaeology, geology, costumes, transport, railroads, and Judaica. The Johannesburg Art Gallery exhibits modern European paintings as well as the work of South African artists. Commercial art galleries and antique shops are scattered throughout the city centre and in the northern districts. Branch library services are available in the suburbs and townships.

The media. Radio and press services are available for all sections of the population, but there is no television service. Domestic radio services are broadcast not only in the official languages, English and Afrikaans. but also in some African languages (such as Zulu, South Sotho, Tonga, and Venda). Overseas radio programs are broadcast from near Johannesburg in nine languages (including French, German, Portuguese, and Dutch, and such Bantu languages as Swahili, Chewa, and Tsonga). The commercial service of the South African Broadcasting Corporation has its main studios in the city. Daily newspapers are published in English and Afrikaans, in addition to which a number of publications for nonwhite readership are also produced. Newspapers for minority groups are also published in various languages, including Chinese. There are numerous learned and popular journals. Many book publishers are established in the city.

Recreation. There are a great number of sporting facilities, including golf courses, tennis courts, rugby and soccer grounds, swimming baths, and cricket grounds, as well as privately owned sports clubs of various kinds and commercial establishments providing facilities for ice hockey, horse racing, and ice skating. The Zoological Gardens have a large collection of wild animals, some fine gardens, and a lake. The larger parks constitute popular picnic areas, and there are numerous open areas in the suburbs. South African wild flowers are to be seen in the Wilds, which consists of 45 acres in the northern suburbs; wild birds may be seen at the Melrose Bird Sanctuary. The open space known as the Civic Centre has some fine examples of sculpture; the Civic Theatre is the home of opera, ballet, music, and drama. Throughout the city, many cinemas, halls, and theatres are to be found, as well as a number of recreation centres. There are also a planetarium and some observatories, including the Republic Observatory, which is of international importance. Kelvin House is the headquarters of many of South Africa's scientific and technical societies. A number of societies cater to the interests of photographers, philatelists, chess players, horticulturists, students of Jewish affairs, archaeologists, and other specialized groups.

BIBLIOGRAPHY. A.H. SMITH (ed.), ***Pictorial History of Johannesburg*** (1956); H.A. CHILVERS, ***Out of the Crucible: Being the Romantic Story of the Witwatersrand Goldfields and of the Great City Which Arose in Their Midst*** and an ***Epilogue, 1929–48: The Incredible City,*** by ALEXANDER CAMPBELL (1948); JOHANNESBURG, CITY COUNCIL, ***The City of Johannesburg: Official Guide,*** 3rd ed. (1962), fairly complete, but somewhat dated; JAMES GRAY, ***Payable Gold . . . An Intimate Record of the History of the Discovery of the Payable Witwatersrand Goldfields and of Johannesburg in 1886 and 1887*** (1937), and with ETHEL L. GRAY, ***A History of the Discovery of the Witwatersrand Goldfields*** (1940), a sequel to ***Payable Gold . . . ;*** JOHANNESBURG, CITY TREASURER'S DEPARTMENT, *Vade-Mecum,* 40th ed. (1969), gives official factual information, largely municipal, arranged alphabetically by subject; L.E. NEAME, ***City Built on Gold*** (1970), a historical approach largely based on information from the daily newspaper ***The Star,*** which dates from 1887; ERIC ROSENTHAL, ***Gold! Gold! Gold!: The Johannesburg Gold Rush*** (1970), a journalist's account of the history and development of the city and its mines.

(A.H.S.)

John XXII, Pope

The second pope to rule from Avignon in southern France, John XXII (Jacques Duese) is remembered today for his conflicts with the Franciscan order and with the Holy Roman emperor Louis IV, as well as for his unorthodox theological views. He also greatly increased the wealth of the papacy and centralized church administration.

Leonard von Matt—EB Inc.

John XXII, contemporary silver coin. In the coin collection of the Vatican Library.

Born of a wealthy bourgeois family at Cahors in southwestern France, he studied canon and civil law at Paris and Orléans. In 1309 he became chancellor to Charles II of Naples and was made a cardinal three years later. On August 7, 1316, he was elected pope at Lyons, France, succeeding Clement V, and proceeded to establish the papal court at Avignon on a permanent basis.

Early in his pontificate, John intervened in a long-standing conflict between two factions in the Franciscan order —the Spirituals, who favoured strict adherence to St. Francis' rule of poverty, and the Conventuals, who held to a broader interpretation. He supported the Conventuals and persecuted Spirituals who resisted his decision. He later condemned the whole Franciscan theory of evangelical poverty in two decretals (letters): ***Ad Conditorem Canonum*** (1322) and ***Cum Inter** Nonnullos* (1323), asserting scriptural evidence to show that Christ and the Apostles had owned property.

Conflict with Louis the Bavarian

John also intervened in a quarrel over the crown of the Holy Roman Empire between Louis the Bavarian (Emperor Louis IV) and Frederick of Austria. Louis defeated Frederick in 1322, but John forbade him to exercise imperial authority until he, as pope, settled the dispute. Louis's answer was the Sachsenhausen Appellation (May 22, 1324), in which he denied papal authority over imperial elections and attacked John's condemnation of the Spiritual Franciscans. At the same time Louis received at his court the political philosophers Marsilius of Padua and John of Jandun, who, in their work ***Defensor** pacis* ("Defender of the Peace"), had declared the authority of an ecumenical council superior to that of the pope. John retaliated by excommunicating Louis, but, on April 18, 1328, the Emperor had John deposed at Rome. (His condemnation of the Spirituals was held to conflict with the pronouncement of Nicholas III.) The Franciscan Peter of Corbara (Pietro Rainalducci) was elected antipope as Nicholas V, and Michael of Cesena, general of the Franciscan order, appealed to the authority of a church council against John. John thereupon excommunicated Peter and deposed Michael. When Louis returned to Germany in 1329, Peter submitted to John and was subsequently imprisoned at Avignon. The Emperor attempted, without success, to effect a reconciliation with the Pope, and thereafter the Franciscans and their philosopher ally Marsilius continued to carry on a vigorous antipapal propaganda from the imperial court at Munich.

New accusations of heresy were provoked by John's ideas about the experience of God by the souls of the blessed in the afterlife (the Beatific Vision), which he expressed in four sermons delivered in the winter of 1331–32. Most theologians held that the saints in heaven were immediately admitted to a full vision of the Godhead. John disagreed, holding that the fullness of the Beatific Vision would be delayed until the Resurrection of the dead and the Last Judgment at the end of the world. Thomas Wallensis, an English Dominican, was imprisoned for publicly disputing the Pope's position, which was subsequently condemned by a committee of doctors of the University of Paris. A pro-imperial cardinal, Napoleone Orsini, began secret negotiations with the Emperor for the convocation of an ecumenical council to judge John. John tried to reconcile his views with those of his opponents before his death on December 4, 1334.

During his pontificate, John had promoted missionary activity in Asia, establishing Catholic bishoprics in Anatolia, Armenia, Iran, and India. At Avignon he founded a papal library; and at Cahors, a university. Like most of the Avignon popes, he showed favouritism to his relatives and fellow countrymen. Of 28 cardinals created by him, 20 were from southern France, and three were his nephews. In appearance he was small, thin, and pale; in character, impetuous, astute, obstinate, and autocratic, though simple and approachable.

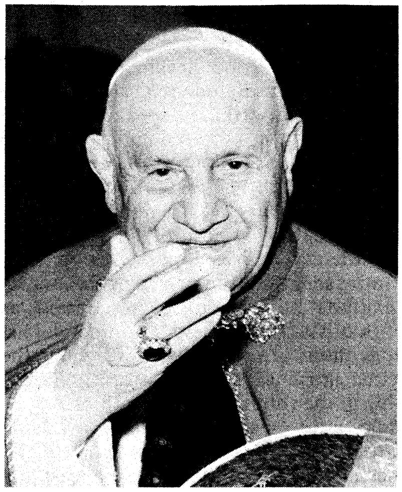
Assessment

His most lasting achievements were in the sphere of law and finance. He added to the body of church law the canons (decrees) of his predecessor, Clement V, and many of his own canons were later added. These were the last additions to canon law until the 16th century. The papal treasury, greatly depleted at the time of his accession, was greatly increased when he died. By the bulls (solemn documents) ***Execrabilis*** (1317) and ***Ex Debito*** (1319), he increased papal control over the distribution of church offices and over the fees paid by their recipients. He also compiled a new tax book, fixing fees for 145 documents issued by the papal chancery, which was not revised for the next two centuries.

BIBLIOGRAPHY. CUILLAUME MOLLAT, *Les Papes d'Avignon (1305–1378)*, 10th ed. (1964; Eng. trans., ***The Popes at Avignon, 1305–1378,*** 1963), is the standard and best general account of the Avignon papacy. YVES RENOUARD, ***La Papautf*** a *Avignon*, 3rd ed. (1954; Eng. trans., ***The Avigon Papacy, 1305–1403,*** 1970), is also very good, especially on the administrative side, and draws attention to the geographical importance of Avignon in the development of papal government. NOEL VALOIS, "Jacques Dubse, pape sous le nom de Jean XXII," ***Histoire littfraire de la France,*** 34:391–630 (1914), gives the best account of John's relations with the Franciscans and of his views on the Beatific Vision and his contributions to missionary activity and to the liturgy. See also S. BALUZE, ***Vitae Paparum Avenionensium,*** 2 vol. (1693), rev. by GUILLAUME MOLLAT (1917), contemporary lives with extensive notes by the chief modern authority; GUILLAUME MOLLAT, AUGUSTE COULON, and SUZANNE CLEMENCET, ***Les Registres de Jean XXII*** (issued in fascicles), papal letters with an invaluable introduction on papal administration; and DECIMA DOUIE, ***The Nature and the Effect of the Heresy of the Fraticelli*** (1932), somewhat out of date but still the fullest account in English of John XXII's relations with the Franciscans.

John XXIII, Pope

Although Pope John XXIII did not reach the papacy until he was almost 77 years old and held the office less than five years, he inaugurated changes in Roman Catholicism so far-reaching that his reign must be considered the beginning of a new era in the history of the Roman Catholic Church. There was nothing in the earlier life of Angelo Giuseppe Roncalli to suggest that his reign as pope would be anything other than the decorous, tradition-bound, static pontificate the Sacred College of Cardinals was counting on when they chose him as an "interim" pope—a papa di passaggio—in October 1958. His .brilliant, strong-willed predecessor, Eugenio Pacelli, as Pius XII, had shaped a defensive policy for dealing with the modern world that the cardinals thought would serve the needs of the church for another century, and they had every reason to believe, from his past behaviour, that the elderly Cardinal Roncalli would maintain the status quo for the few years he had to live. Then, according to the plan, the church would be turned over to a younger prelate who was cast in the Pacelli mold. As pope, however, Roncalli put into effect so many of the ideas he had entertained privately during a half century of obediently serving others that the church was never the same again.



John XXIII, 1963.
Keystone

Angelo Roncalli might have lived and died as an obscure parish priest had he been less of a conformist throughout all but the last few years of his life. The child of a peasant family, he began his career in the church with no connections of any significance and no powerful patron to guide him through the maze of ecclesiastical politics. His steady climb was due above all to his readiness to subdue his own preferences, follow orders, and adjust without complaint to the will of his superiors. Such was the clerical ideal proposed at his seminary in Bergamo, Italy, and it guided him throughout his life.

Early life and career. Angelo was born on November 25, 1881, one of 13 children born to Giovanni Roncalli, a tenant farmer of Sotto il Monte, a tiny village seven miles from the Lombard city of Bergamo. The Roncallis were poor but not anywhere as destitute as some of the later legends would have had it. "We had the necessities of life," the Pope used to say testily, "and we were strong and healthy." Though Angelo, the third child and oldest son of the family, went off to prepare for the priesthood as a child of 11, he continued to spend vacations with the family and remained close to them throughout his life.

Relations with his family

No matter how powerful he became, however, he never helped any of the other Roncallis to advance in the world. "The world is only interested in making money. . . ," he wrote to his brother Xaverio after he became pope. "A great honour has come to our family," he acknowledged, but he urged Xaverio and the others to remain humble, seek no honours or preferments, and take no material advantage of their relationship to the Pope. "At my own death," he concluded, "I shall not be denied the praise which did so much honour to the holiness of Pius X: born poor, he died poor." In his last will he bequeathed each of the living members of the family a legacy of less than $20—his total personal fortune.

Angelo, the churchman, however, was destined to spend most of his life among the powerful and cultivated. Inevitably, that set him apart from the other Roncallis almost from the beginning. While still a seminarian, he was already beginning to feel out of place in Sotto il Monte, where he was charged with priggishness and "putting on airs." "Only three days of the holidays have passed and already I am weary of them," he recorded in the diary that he allowed to be published after his death. He was eager, he wrote, to return to Bergamo, with its orderly life of study, prayer, and genteel fellowship, and to escape from the petty gossip, suspicion, and jealousies of the village.

Roncalli was not an especially brilliant student. He did well enough, however, to be sent to Rome for theological studies, in 1900. After only one year at the Seminario Romano, his education was interrupted when he was drafted into military service and assigned to an infantry company conveniently stationed at Bergamo. Later, as a priest, he returned to the army during World War I. Again he served in Bergamo, first as a hospital orderly and later as a military chaplain with the rank of lieutenant.

Even with the time out for military service, he was not yet 23 when he was ordained a priest, in Rome on August 10, 1904. The next day he said his first mass at St. Peter's Basilica. After a visit to the family in Sotto il Monte, he returned to the Seminario Romano for further study. Eventually he received a doctorate in canon law.

As a graduate student in Rome, Roncalli took the first step that was to lead to the papacy a half century later. Simply because he was a priest of Bergamo, he was asked by the reigning pope, Pius X, to assist in the ceremony of consecration for a new bishop, Giacomo Radini-Tedeschi, who had been appointed to take over the diocese of Bergamo. The new bishop, a member of the Italian nobility, was much taken by the young priest and asked him to serve as his secretary.

Back in Bergamo, Don Angelo, in addition to his secretarial duties, was assigned to the faculty of the diocesan seminary. For the next nine years he served as a professor of theology and spiritual director of the young men preparing for ordination. As the Bishop's right-hand man, he gained a wide range of pastoral experience.

Roncalli had great esteem for his superior, who was known as the most progressive prelate in Italy. After the Bishop's death, in 1914, he wrote an appreciative biography of the prelate and sent a copy to the pope, Benedict XV, who had been one of Radini-Tedeschi's personal friends.

Call to service under Pope Benedict XV

With the war behind him, Roncalli was summoned to wider service in the church. In 1920, Pope Benedict, recalling Radini-Tedeschi's biographer, named him a director of the Italian organization for the support of foreign missions. The position was not notably significant in a church top-heavy with dignitaries, but it brought him into personal contact with a number of important clerical figures throughout Europe, and his name became recognizable in ecclesiastical Rome. He also gained some attention because of his work as a part-time historian who specialized in some of the minor activities of Saint Charles Borromeo, a cardinal of Milan who played an important role in the 16th-century Counter-Reformation. It was Roncalli's researches for this project that first brought him into contact with Msgr. Achille Ratti, the Milanese librarian who would become Pius XI.

Service as a Vatican diplomat. Pius XI later remembered the Bergamo priest's gift for personal dealings and brought him into the Vatican's diplomatic service. Roncalli was appointed apostolic visitor to Bulgaria in March 1925. In keeping with custom, he was made an archbishop before he left Rome. He spent the next ten years in that obscure but delicate post, where he was expected to protect the interests of a small Roman Catholic community in a country overwhelmingly Eastern Orthodox. His diary reveals that he was often lonely and discouraged in Bulgaria, but he carried out the assignment with tact, patience, and notable good humour. Still, he was not deemed to be among the best qualified clerics in the papal diplomatic corps.

His next assignment was equally unpromising. He was appointed apostolic delegate to Greece, which was combined with naming him head of the Vatican diplomatic mission to Turkey. Again he was called upon to represent powerless Catholic minorities in an Eastern Orthodox nation, Greece, and a Muslim nation, Turkey. He made his home in Istanbul, where he was generally ignored by both the Turkish government and the Vatican but was warmly appreciated in the diplomatic colony as an amiable host and affable dinner companion.

None of these posts loomed large in the Western-oriented Vatican, and the Archbishop had good reason to believe that his career had reached a dead end. Later, he confessed that he was stunned by the announcement, at the end of 1944, that he had been named papal nuncio to Charles de Gaulle's newly liberated France; his first thought was that a clerical error must have been made in Rome.

Papal nuncio to France

The French post was particularly delicate at the time. Roncalli's predecessor, Msgr. Valerio Valeri, had been close to the collaborationist Gen. Philippe Pétain during the German occupation, and de Gaulle made it clear to the Vatican that, since Valeri had become ***persona non grata*** to the French people, he would have to be replaced immediately. France was still seething with a spirit of vengeance against former collaborators. It would be the new nuncio's obligation to deal with the ill will created by his predecessor and by the bishops who had cooperated with the hated Vichy government. Someone in the Vatican remembered the genial Archbishop languishing in the Near East, and it was decided that, though he was not noted for his political astuteness, perhaps he had precisely the qualifications needed under the circumstances. Roncalli was told that he would be expected to cool the atmosphere, re-establish the independence of the church, and gain the release of a number of German seminarians who were being held as prisoners of war. In addition, be had to deal with an outburst of radicalism among the younger French clergy, which the conservative forces in the Vatican Curia found highly disturbing.

His success in carrying out the assignment was acknowledged by the papacy when Archbishop Roncalli was named a cardinal by Pius XII. In January 1953 the red hat, the symbol of a cardinal, was conferred on him by the Socialist president of France, Vincent Auriol.

Reign as pope. As a cardinal, Roncalli immediately became eligible for one of the major Italian archbishoprics. Appointed patriarch of Venice at the age of 71, he had cause once more to believe that he had reached the end of the line. Thus, perhaps no one was more surprised than he when, after the death of Pius XII on October 9. 1958, he was elected pope on the 12th ballot—clearly a compromise candidate acceptable to all parties only because of his advanced years.

The second Vatican Council

Perhaps a younger pontiff would have been less daring and innovative than John XXIII turned out to be. Soon after his coronation, he announced almost casually that he was summoning an ecumenical council—a general meeting of the bishops of the church—the first in almost a century. He said the idea came to him in a sudden inspiration. His purpose was to "bring the church up to date" ***(aggiornamento)*** and to work for its spiritual regeneration. He was the first pope since the Reformation who acknowledged frankly that Catholicism stood in need of reinvigoration and reform.

It was long a truism among church historians that councils are followed by upheaval and disorder in the church. The Pope's decision, consequently, was received coolly by his conservative Curia, who were convinced that the church had prospered under Pius XII's leadership and who saw no good reason for the changes John envisioned. Some of the Vatican cardinals in fact did everything in their power to delay the council until the old man had passed from the scene and the project could be quietly dropped. But the Pope pushed on with his plan and lived long enough to preside over the first session of the second Vatican Council in the fall of 1962.

In keeping with his wishes, the council fathers pledged that they would be consistently positive. No condemnations or anathemas were to be made; political hostilities were to be ignored; and the church above all was to recognize that it was not the master but the servant of humanity. The Pope made it clear that the second Vatican Council was convened as a pastoral council. No new dogmas were to be pronounced, though old doctrines and disciplines were to be re-examined. What John sought, he said, was a "New Pentecost," a new outpouring of the Holy Spirit.

The council, according to John's design, would make a new start toward achieving Christian unity by putting aside the hostilities of the past and acknowledging the

Catholics' share of responsibility for the scandal of a divided Christianity. With his long experiences among the Eastern Orthodox, John's interest in Christian ecumenism seemed natural enough, but no one in Rome was quite prepared for the extent of his openness. He received Eastern Orthodox, Anglican, and Protestant religious leaders with extreme cordiality and made sure they were invited to send observers to the Vatican Council. He removed certain words offensive to Jews from the official liturgy of the church. On one notable occasion, he introduced himself to a group of Jewish visitors with the biblical words, "I am Joseph your brother," referring to the Old Testament story of the meeting of the sons of the patriarch Jacob at the court of Egypt.

He travelled around Rome freely, breaking with the tradition that the pope, deprived of his former temporal power, was a "prisoner of the Vatican." In an attempt to depoliticize the church, he played down his position as ruler of the Vatican and emphasized his role as "servant of the servants of God," a traditional title of the pope. In that spirit he called on the President of Italy and cordially received the son-in-law of the Soviet premier, Nikita Khrushchev, in private audience. Among his other visitors were the Archbishop of Canterbury — the first such meeting since the 14th century — the Moderator of the Scottish Kirk, and a Shintō (the religious cult of Japan) high priest — the first such official in history to be received at the Vatican.

The *Pacem in Terris* encyclical

During the Cuban missile crisis of 1962, the Pope publicly urged both the United States and the Soviet Union to exercise caution and restraint and won the appreciation of both Pres. John F. Kennedy and Premier Khrushchev. His major encyclical *Pacem in Terris* ("Peace on Earth"), addressed to all mankind, was received warmly throughout the world and praised by politicians as well as churchmen. Straightforward and frankly optimistic, it avoided the language of diplomacy and set forth the requirements for world peace in profoundly human terms. Distinguishing between the philosophy of Marxism and actual governments to which it gave birth, John suggested that peaceful coexistence between the West and the Communist East was not only desirable but was actually necessary if mankind was to survive. He thereby diluted the religious energy that had been poured into the Cold War as a result of the militant policies shaped by his predecessor.

John saw himself as a reconciler. In statement after statement he emphasized the church's significance as a suprapolitical spiritual force in the world. His greatest claim on the world's affection, however, rested on the warmth of his personality rather than on any of his formal statements. He remained simple and unaffected, in spite of the baroque setting in which he found himself, and instinctively appealed directly to human values that everyone could understand.

"Since you could not come to me, I came to you," he told the inmates of a Roman prison. When Mrs. John F. Kennedy, then wife of the President of the United States, came to call, he rehearsed "Mrs. Kennedy, Madame Kennedy" in his poor English. Then, when she appeared he spontaneously spread open his arms and cried out, "Jacqueline!" He once told a Communist diplomat, "I know you are an atheist, but won't you accept an old man's blessing?" When a shabby peasant woman reached up to touch him as he was being carried through St. Peter's, he stopped to clasp her hand. "There is no reason why you shouldn't get as close as the King of Jordan did," he said.

The roly-poly Pontiff — he was short of stature and never overcame a tendency toward corpulence — gradually became a kind of father figure for the world. When he died on June 3, 1963, it was generally recognized that he had become one of the best loved men on earth and one of the most popular popes of all time. His successor, Paul VI, instituted formal proceedings that could lead to his canonization as a saint. Had the ancient custom of popular canonization still been in effect in 1963, however, that honour would probably have been given to him immediately by the tearful crowd who were gathered in St. Peter's Square when his death was announced.

BIBLIOGRAPHY. Pope John's impact on the church, along with a biographical sketch, is dealt with in most of the books about the second Vatican Council and the new developments in contemporary Roman Catholicism. Among the best of these in English are XAVIER RYNNE, *Letters from Vatican City* (1963); GARY MACEOIN, *What Happened at Rome? The Council and Its Implications for the Modern World* (1966); and in translation, HANS KUNG, *Konzil und Wiedervereinigung* (1960; *The Council and Reunion,* 1961). Biographical works include E.Y. HALES, *Pope John and His Revolution* (1965); A LAZZARINI, *Giovanni XXIII* (1958; Eng. trans., *Pope John XXIII,* 1959); and MERIOL TREVOR, *Pope John* (1967). The Pontiff's diary, which he began writing as a seminarian and finished when he was pope, was published as *Il giornale dell' anima,* 4th ed. (1965; Eng. trans., *The Journal of a Soul,* 1965). His encyclicals and major public addresses may be found in *The Encyclicals and Other Messages of John XXIII* (1964).

(Jo.C.)

John of England

One of the most unpopular monarchs in his country's history, John, king of England from 1199 to 1216, aroused widespread hostility among barons, prelates, and commoners alike. His reign ended in a baronial rebellion that forced him to issue Magna Carta, the most important constitutional instrument in English history. Yet he was one of the most astute politicians and energetic administrators of the Plantagenet line of kings. A subject of controversy throughout his life, he was born at Oxford on December 24, 1167, the youngest son of Henry II and Eleanor of Aquitaine. Henry's plan (1173) to assign to his favourite son (whom he had nicknamed Lackland) extensive lands upon his marriage with the daughter of Humbert III, count of Maurienne (Savoy), was defeated by the rebellion the proposal provoked among John's elder brothers. Various provisions were made for him in England (1174–76), including the succession to the earldom of Gloucester. He was also granted the lordship of Ireland (1177), which he visited from April to late 1185, committing youthful political indiscretions from which he acquired a reputation for reckless irresponsibility. Henry's continued favour to him contributed to the rebellion of his eldest surviving son, Richard I (later called Coeur de Lion), in June 1189. For obscure reasons John deserted Henry for Richard.

First attempt to seize government

On Richard's accession in July 1189, John was made count of Mortain (a title that became his usual style), was confirmed as lord of Ireland, was granted lands and revenues in England worth £6,000 a year, and was married to Isabella, heiress to the earldom of Gloucester. He also had to promise (March 1190) not to enter England during Richard's absence on his crusade. But John's actions were now dominated by the problem of the succession, in which his nephew, the three-year-old Arthur I, duke of Brittany, the son of his deceased elder brother Geoffrey, was his only serious rival. When Richard recognized Arthur as his heir (October 1190), John immediately broke his oath and returned to England, where he led the opposition to Richard's dictatorial chancellor, William Longchamp. On receiving the news in January 1193 that Richard on his way back from the crusade had been imprisoned in Germany, John allied himself with King Philip 11 Augustus of France and attempted unsuccessfully to seize control of England. In April 1193 he was forced to accept a truce but made further arrangements with Philip for the division of Richard's possessions and for rebellion in England. On Richard's return, early in 1194, John was banished and deprived of all his lands. He was reconciled to Richard in May and recovered some of his estates, including Mortain and Ireland, in 1195, but his full rehabilitation came only after the Bretons had surrendered Arthur to Philip II in 1196. This led Richard to recognize John as his heir.

In 1199 the doctrine of representative succession, which would have given the throne to Arthur, was not yet generally accepted, and following Richard's death in April 1199 John was invested as duke of Normandy and crowned king of England in May. Arthur, backed by Philip II, was recognized as Richard's successor in Anjou and Maine, and it was only a year later, in the Treaty of

John, tomb effigy in Worcester Cathedral.

BY courtesy of the National Monuments Record. London

Le Goulet, that John was recognized as successor in all Richard's French possessions, in return for financial and territorial concessions to Philip. Meanwhile, John successfully resisted Scottish claims to the three northern counties and conducted a vigorous overhaul of local administration in England.

Second marriage

The renewal of war in France was triggered off by John's second marriage. His first wife, Isabella of Gloucester, was never crowned, and in 1199 the marriage was dissolved on grounds of consanguinity, both parties being great-grandchildren of Henry I. John then intervened in the stormy politics of his county of Poitou and, while trying to settle the differences between the rival families of Lusignan and Angoulême, himself married Isabella (August 1200), the heiress to Angoulême, who had been betrothed to Hugh IX de Lusignan. This politically conceived marriage provoked the Lusignans into rebellion the next year; they appealed to Philip II, who summoned John to appear before his court. In the general war that followed his failure to answer this summons, John had a temporary success at Mirebeau in August 1202, when Arthur of Brittany was captured, but Normandy was quickly lost (1204). By 1206, Anjou, Maine, and parts of Poitou had also gone over to King Philip.

These failures, foreshadowed under Henry II and Richard, were brought about by the superiority of French resources and the increasing strain on those of England and Normandy. Nevertheless. they were a damaging blow to John's prestige, and, equally important, they meant that John resided now almost permanently in England. This factor, coinciding with the death (1205) of the chancellor and Archbishop of Canterbury, Hubert Walter, gave his government a much more personal stamp, which was accentuated by the promotion of members of his household to important office. His determination to reverse the continental failure bore fruit in ruthlessly efficient financial administration, marked by taxation on revenues, investigations into the royal forests, taxation of the Jews, a great inquiry into feudal tenures, and the increasingly severe exploitation of his feudal prerogatives. These measures provided the material basis for the charges of tyranny later brought against him.

Quarrel with the church

John's attention was diverted and his prestige disastrously affected by relations with the papacy. In the disputed election to the see of Canterbury following the death of Hubert Walter, Pope Innocent III quashed the election of John's nominee in procuring the election of Stephen Langton (December 1206). John, taking his ground on the traditional rights of the English crown in episcopal elections, refused to accept Langton. In March 1208, Innocent laid an interdict on England and excommunicated John (November 1209). The quarrel continued until 1213, by which time John had amassed more than $100,000 from the revenues of vacant or appropriated sees and abbeys. But such a dispute was a dangerous hindrance to John's intention to recover his continental lands. In November 1212 he agreed to accept Langton and the pope's terms. Apparently at his own behest, he surrendered his kingdom to the papal nuncio at Ewell, near Dover, on May 15, 1213, receiving it back as a vassal rendering a tribute of 1,000 marks ($666 13s. 4d.) a year. He was absolved from excommunication by Langton in July 1213, and the interdict was finally relaxed a year later. John thus succeeded in his aim to secure the papacy as a firm ally in the fight with Philip and in the struggle already pending with his own baronage. But his treatment of the church during the interdict, although arousing little if any opposition among the laity at the time, angered monastic chroniclers, who henceforth loaded him with charges of tyranny, cruelty, and, with less reason, of sacrilege and irreligion.

Baronial rebellion and Magna Carta

In August 1212, recurrent baronial discontent had come to a head in an unsuccessful plot to murder or desert John during a campaign planned against the Welsh. Pope Innocent's terms had included the restoration of two of those involved, Eustace de Vesci and Robert Fitzwalter, and, although the barons soon lost papal support, they retained the protection of Stephen Langton. John, skillfully isolating the malcontents, was able to launch his long-planned campaign against the French, landing at La Rochelle in February 1214. He achieved nothing decisive and was forced to accept a truce lasting until 1220. Returning to England in October 1214, he now had to face much more widespread discontent, centred mainly on the northern, East Anglian, and home counties. After lengthy negotiations in which both sides appealed to the Pope, civil war broke out in May 1215. John was compelled to negotiate once more when London went over to the rebels in May, and on June 19 at Runnymede he accepted the baronial terms embodied in Magna Carta, which ensured feudal rights and restated English law. This settlement was soon rendered unworkable by the more intransigent barons and John's almost immediate appeal to Pope Innocent against it. Innocent took the King's side, and in the ensuing civil war John captured Rochester castle and laid waste the northern counties and the Scottish border. But his cause was weakened by the arrival of Prince Louis (later Louis VIII) of France, who invaded England at the barons' request. John continued to wage war vigorously but died at Newark on October 18–19, 1216, leaving the issues undecided. His death made possible a compromise peace, including the restoration of the rebels, the succession of his son Henry III, and the withdrawal of Louis.

John's reputation, bad at his death, was further depressed by writers of the next generation. Of all centuries prior to the present, only the 16th, mindful of his quarrel with Rome, recognized some of his quality. He was suspicious, vengeful, and treacherous; Arthur I of Brittany was probably murdered in captivity, and Matilda de Braose, the wife of a recalcitrant Marcher baron, was starved to death with her son in a royal prison. But John was cultured and literate. Conventional in his religion rather than devout, he was remembered for his benefactions to the church of Coventry, to Reading Abbey, and to Worcester, where he was buried and where his effigy still survives. He was extraordinarily active, with a great love of hunting and a readiness to travel that gave him a knowledge of England matched by few other monarchs. He took a personal interest in judicial and financial administration, and his reign saw important advances at the exchequer, in the administration of justice, in the importance of the privy seal and the royal household, in methods of taxation and military organization, and in the grant of chartered privileges to towns. If his character was unreliable, his political judgment was acute. In 1215 many barons, including some of the most distinguished, fought on his side.

BIBLIOGRAPHY. KATE NORGATE, *John Lackland* (1902), the most exhaustive scholarly biography, is still valuable.

W.L. WARREN, *King John* (1961; rev. paperback ed., 1966), is a lively, modern account. J.C. HOLT, *King John* (1963), discusses both the medieval and modern assessment of the King. For accounts of the reign, see SIDNEY PAINTER, *The Reign of King John* (1949), which is thorough and extensive; J.C. HOLT, *The Northerners* (1961), a study of the baronial opposition, and *Magna Carta* (1965), for the origins and course of John's quarrel with the barons. Among more general works, A.L. POOLE, *From Domesday Book to Magna Carta, 1087–1216*, 2nd ed. (1955), is an authoritative survey of English history in this period. D.M. STENTON, *English Justice Between the Norman Conquest and the Great Charter, 1066–1215* (1964), has an important essay on King John and the courts of justice. R.V. TURNER, *The King and His Courts: The Role of John and Henry III in the Administration of Justice, 1199–1240* (1968), a general survey of the King's legal powers and functions, is also valuable. F.M. POWICKE, *The Loss of Normandy, 1189–1204*, 2nd ed. (1961), a definitive study of John's defeat on the Continent, is also useful on financial and military administration.

(J.C.H.)

John III Sobieski, of Poland

Of the ten elective kings who ruled Poland between 1573 and 1795, only three were Poles by birth. The most famous and perhaps the most important of them was John III Sobieski (reigned 1674–96), the last warrior on the Polish throne, the subduer of the Ottoman Turks, and a writer whose letters to his wife are famous. The kingdom of Poland-Lithuania, weakened by internal revolt and by wars with Russia, Sweden, Brandenburg, Transylvania, and the Ottoman Empire in the mid-17th century, was restored to greatness for the last time under Sobieski's leadership but proved unable to regain its previous preponderance in eastern Europe. Though a genuine patriot, Sobieski also knew how to further his own and his family's interests. His greatness as a king has been disputed, but he was certainly one of the greatest commanders in chief Poland has ever had.

By courtesy of the trustees of the British Museum; photograph. J. R. Freeman & Co. Ltd.

John III Sobieski, engraving by Carel Allardt (1648–1709).

Sobieski's ancestors were of the lesser nobility, but one of his great-grandfathers was the famous grand-hetman (military commander) St. Żółkiewski, and when he was born in Olesko, August 17, 1629, his father, James (Jakób) (1588–1646), had already taken a step to the higher ranks, sharing an office on the royal court. At the end of his life, the father even became castellan of Cracow, an office that secured him the highest rank among the members of the Polish Senate, or first chamber of the parliament. John was well-educated and toured western Europe in his youth, as was usual for a Polish noble of his class. When the Swedes invaded Poland in 1655, he joined them in opposition to the Polish king John Casimir. The following year he changed sides again and became one of the leaders in the fight to expel the Swedes. In 1665, through the influence of his patroness, Queen Maria Louisa (Ludwika), he was appointed to the prestigious office of grand marshal. In 1666 he became field commander of the Polish army. In October 1667 he defeated the Tatars and the Cossacks near Podhajce (now in the U.S.S.R.), and in the spring of 1668, when he triumphantly returned to Warsaw, he was named commander in chief. In 1665 he had married an ambitious young French widow, Marie-Casmire de la Grange d'Arquien (Marysieńka). Marysienka planned to have John elected king after King John Casimir's resignation in 1668. When this plan failed—the nobility elected Michael Wilniowiecki in 1669—she began working to obtain support from Louis XIV of France for her husband's advancement. Since they were often separated—the husband on the front, his wife on journeys to France—Sobieski wrote long letters to Marysieńka, which are now a highly interesting and important historical source. Her letters have not been preserved. During the short reign of King Michael (1669–1673), Sobieski distinguished himself by further victories over the Cossacks, and simultaneously he tried to undermine Michael, whose policies favoured the Habsburgs against France. Michael died in November 1673, and almost on the same day Sobieski won a splendid victory over the Turks under Hussein Paşa near Chocim (Hoţin). Although this victory did not alter the disastrous conditions of the Peace of Buczacz concluded in 1672 (Poland had to cede territory to the Turks and to pay a considerable indemnity), Sobieski's reputation was so great that in May 1674 he was elected king in preference to the candidate backed by the Habsburgs. At first Sobieski followed a pro-French policy. He tried to end the Turkish war by French mediation and concluded the secret treaty of Jaworów with France (June 1675), in which he promised to fight the Holy Roman (Habsburg) Emperor after the conclusion of peace with the Turks. In fact, only an armistice with them was concluded at Żórawno (October 1676), and the conditions were only slightly more favourable than those of Buczacz.

Sobieski's marriage

Sobieski's hopes of compensating for losses to the Turks in the southeast by using French and Swedish support to make territorial gains from Prussia in the northwest were also disappointed. Furthermore, Louis XIV was neither ready to recognize Marysieńka's French relatives as members of a royal family nor willing to support the succession of Sobieski's son James (Jakób) to the Polish throne. The great nobles, especially those from Lithuania, were opposed to the French alliance because they feared that Sobieski was striving to attain absolute power with the help of France. It was becoming clear, moreover, that it was impossible to reconcile the interests of Poland and those of Louis, whose aim was to use Sobieski as an obedient vassal against the Habsburgs. Poland, for its part, had no differences with Habsburgs and, after a series of Turkish attacks, came to regard the Ottomans, the allies of France, as its deadliest enemies. Sobieski, therefore, though always an admirer of France, shifted away from the French alliance and concluded a treaty with the Holy Roman emperor Leopold I against the Turks (April 1, 1683). By the terms of the treaty, each ally had to support the other with all his might if the other's capital were to be besieged. Thus, when a great Turkish army approached Vienna late in the summer of 1683, Sobieski himself rushed there with about 25,000 men. Because he had the highest rank of all military leaders gathered to relieve Vienna, he took command of the entire relief force (about 75,000 men) and achieved a brilliant victory over the Turks at the Kahlenberg (September 12, 1683), in one of the decisive battles of European history.

The siege of Vienna

In the campaign that followed in Hungary (in the autumn of 1683), however, Sobieski was less successful, and his relations with the emperor Leopold deteriorated because of differences in temperament and conflicting political plans. Sobieski's idea was to liberate Moldavia and Walachia (present-day Romania) from Ottoman rule and to expand Poland's influence to the shores of the Black Sea. But his advances into Moldavia, undertaken between 1684 and 1691, were mostly failures, and during the last one he was even in danger of being captured. Despite his previous victories, he was thus not able to achieve his objective. Only after his death, in 1699, were the territories that had been lost in 1672 recovered.

In the last years of his life, from 1691 until his death in 1696 (June 17), Sobieski was often seriously ill and had to face quarrels with the nobles and within his own family. His eldest son, James, was bitterly opposed to the Queen and the younger princes. All of Sobieski's sons were interested in succeeding to the throne and tried to obtain help, either from the Emperor or from France. The marriage of Sobieski's daughter Kunegunda to Maximilian II Emanuel, elector of Bavaria (1694), was the only bright spot in these rather gloomy years.

Although the second half of the reign was much less brilliant than the first, the personal wealth of the royal couple continued to grow because they knew how to obtain money in exchange for offices and signs of royal favour. Thus, the King left a considerable fortune when he died.

Sobieski also spent large sums on his residences in Żółkiew and Jaworów and especially on the palace of Wilanów near Warsaw, a fine example of Baroque architecture. He was also a patron of poets and painters. Of all the Polish rulers of the 17th century, he was the best educated and took the greatest interest in literature and cultural life.

The struggle against Ottoman power in Europe was the keystone of Sobieski's foreign policy, with which all other foreign relations were closely connected. When the Russians, traditionally Poland's enemies, showed willingness to join the league against the Turks, Sobieski concluded with them the "Eternal" Peace of 1686 (the Grzymułtowski Peace). In this treaty, Kiev, which had been under temporary Russian rule since 1667, was permanently ceded by Poland. But despite all the failures and disappointments he experienced after 1683, Sobieski was able to deliver southeast Poland from the threat of Ottoman and Tatar attack.

His failures

In domestic policy Sobieski was least successful. All his endeavours to strengthen the position of the crown and stabilize the army failed completely, and his own sons opposed him. The nobles showed little interest in defending the country after the great victory of 1683 had been won, and the Lithuanian magnates fought each other rather than the Turks. Thus, John Sobieski, although a brilliant general and organizer, was unable to prevent rebellion in his family and the dissension among his subjects that finally led to Poland's downfall in the 18th century. This tends to make him, in the final reckoning, a somewhat tragic figure.

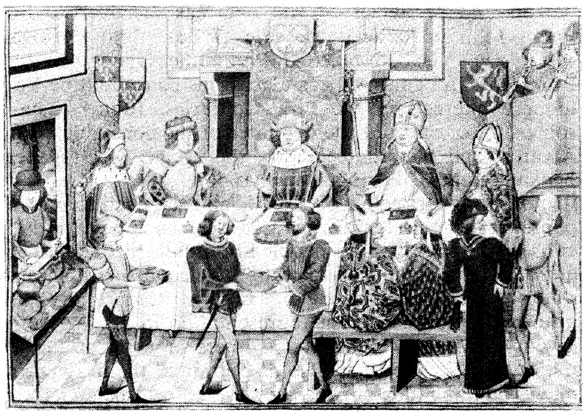
BIBLIOGRAPHY. For further information, see O. FORST DE BATTAGLU, "Jan Sobieski, 1674–96," in The *Cambridge History* of *Poland,* vol. 1, pp. 532–556 (1950).

(G.K.S.R.)

John I of Portugal

The first ruler of the House of Aviz, John I, king of Portugal from 1385 to 1433, consolidated his own position and preserved his country's independence from Castile by the great victory of Aljubarrota in 1385. His conquest of Ceuta in northwest Africa (1415) marked the beginning of the Portuguese and European age of overseas expansion and exploration, and his treaty with England (1386) raised the Anglo-Portuguese connection to the status of a firm and permanent alliance.

Born at Lisbon on April 11, 1357, John was the illegitimate son of King Pedro I and Teresa Lourenço. At age six he was made master of the military Order of Aviz; he received an ecclesiastical and military education, probably at Aviz in Alentejo. On his father's death, in 1367, his half brother Ferdinand became king and embarked on a calamitous rivalry with the new ruler of Castile, Henry II, who finally forced Ferdinand to accept a Castilian marriage for his infant heiress, Beatriz, thus compromising the future independence of Portugal.



John I of Portugal (centre) entertaining John of Gaunt (on his right), miniature from a late-15th-century manuscript by Jehan de Wavrin, published as Anchiennes Croniques *d'Engelterre* (3 vol., 1858–63). In the British Museum (Royal MS. 14 E.iv).

BY courtesy of the trustees of the British Museum

When Ferdinand died, in 1383, his widow, Queen Leonor, submitted to the demand of her Castilian son-in-law, John I, that he be recognized as king of Portugal. John of Aviz, who had hitherto remained carefully in the background, though he had been arrested for a time in 1382, was now persuaded by a group of young nationalists, led by Nuno Álvares Pereira, to murder Queen Leonor's favourite and adviser, the Galician João Fernandes Andeiro, conde de Ourém. Popular support was at once stirred up for John, and Queen Leonor fled from Lisbon, appealing to Castile for help. In May 1384 Castilian armies besieged John in Lisbon until the outbreak of plague forced them to withdraw (September).

John had been named defender of the realm; but in April 1385 representatives of the three estates met in the Cortes (assembly) of Coimbra, and, after it had been demonstrated that King Pedro's elder surviving sons had not been legitimized, John was elected king. The cities of Lisbon and Porto and the merchants and trade guilds enthusiastically backed him, but much of the older nobility still favoured the Castilian succession. John and Nuno Alvares, now his constable, marched northward and obtained the submission of the chief places but returned on hearing that the Castilians were preparing a major invasion. As Spanish forces entered central Portugal, John and Nuno Alvares advanced to bar the road to Lisbon and won the famous Battle of Aljubarrota (August 14, 1385). This victory assured Portugal's independence and made John a desirable ally. He had already received some English aid, and a small party of English archers had fought for him at Aljubarrota; and he now concluded the Treaty of Windsor (May 9, 1386), which became the cornerstone of the Anglo-Portuguese alliance. In consequence, John of Gaunt, duke of Lancaster, arrived in Galicia hoping, through his second marriage, with the daughter of King Pedro I of Castile, to realize his claim to the Castilian crown. John of Aviz sealed the alliance by marrying, in February 1387, Gaunt's daughter, Philippa of Lancaster. A joint invasion of León was unsuccessful, and Lancaster withdrew. John of Aviz made a ten-year truce with Castile in 1389, but frontier warfare with Castile was intermittent thereafter until peace was finally made in 1411.

Elected king

John's elder sons had now reached the age at which they could become knights, and it was ostensibly on their behalf that he organized an expedition against Ceuta, which fell in a day (August 24, 1415). He had probably hoped to advance into Morocco and tap the African caravan routes, but, instead, Ceuta became a beleaguered outpost supplied from the Portuguese Algarve. This stimulated the maritime explorations, beginning with the rediscovery and settlement of the Madeira Islands and the Azores.

The children of John and Philippa

John's court now became a centre of culture, influenced through Queen Philippa by English traditions. Of their sons, the *inclita geração* ("illustrious generation"), the eldest, Edward, administered the kingdom under his father and was later king; the second, Pedro, travelled through Europe and was regent after Edward's death; and the third, Henry, known as the Navigator, was the patron and organizer of the overseas discoveries. For them, John introduced the title of duke into Portugal. Their sister, Isabel, married Philip the Good, duke of Burgundy and count of Flanders, thus consolidating Portuguese interests in the Low Countries. John's bastard son, Afonso, married the daughter of the constable Nuno Álvares, and their descendants, the House of Braganca, became kings of Portugal from 1640.

John's long struggle with Castile and the need to recompense a new aristocracy caused serious financial difficulties, but he rallied his people around his throne and acquired a reputation as a cautious leader and shrewd statesman. He rewarded the faithful trade guilds by granting them permanent representation in the House of Twenty-four, in which two members of each of the 12 major guilds were to sit. He also granted them a special magistrate, the judge of the people. At Porto he ended the unpopular civil jurisdiction exercised over the city by the Bishop. He displayed his devotion by building near Leiria the great Abbey of Batalha to commemorate his victory and serve as a pantheon. He was the author of a work on hunting (*Livro da montaria*), his favourite pastime.

He died in Lisbon on August 14, 1433. There is a portrait of John in the Museum of Ancient Art, Lisbon, and recumbent statues of John and Philippa over their tomb at Batalha.

BIBLIOGRAPHY. John's reign, like those of his predecessors, D. Pedro and Fernando, was chronicled by FERNAO LOPES. There are several modern editions in Portuguese, but only a short excerpt has been translated by E. PRESTAGE in *The Chronicles of Fernão Lopes and Gomes Eannes de Zurara* (1928). The chronicle of Nuno Alvares Pereira is perhaps also by Lopes. GOMES EANNES DE ZURARA wrote the *Crónica da Tomada de Ceuta. A* modern biography in Portuguese is DAMIAO PERES, *D. João I* (1917). The events of the beginning of the reign are dealt with fully by S. DIAS ARNAUT, *A crise nacional dos fins do século XIV* (1960). *An* excellent account of the international politics and military enterprises of the period is in P.E. RUSSELL, *English Intervention in Spain and Portugal in the Time of Edward III and Richard II* (1955).

(H.V.L.)

John II of Portugal

King of Portugal from 1481 to 1495, John II is regarded as one of the greatest Portuguese rulers, chiefly because of his ruthless assertion of royal authority over the great nobles and of his resumption of the exploration of Africa and the quest for India.

John was the great-grandson of the founder of the House of Aviz, John I, and only surviving son of Afonso V by his queen and cousin, Isabella. Born in Lisbon in 1455, he was educated by the humanists of the court and was married to his cousin Leonor in 1471. He participated in his father's conquest of Arzila in Morocco, where he was knighted, and was given a separate household at Beja in southern Portugal. In 1474 his father entrusted him with the "trade of Guinea" and the African explorations. When Afonso V claimed the Castilian throne in opposition to Isabella I, plunging Portugal into war, he appointed John his regent (April 1475). The prince mobilized an army and marched to support his father, but the Battle of Toro (March 1476) checked the Portuguese intrusion into Castile. Afonso V departed for France in a fruitless search for an alliance while John defended the frontier and parried a Spanish counterattack. Afonso's lack of success caused him to announce his abdication. John was proclaimed king, but his father returned and resumed his reign, concluding the disadvantageous Treaty of Alcáçovas before his death in August 1481.

The oath of obedience

At John II's accession, this peace treaty had obliged him to place his young children under Spanish guardianship near the frontier as a pledge of their marriage to Castilians. Afonso had been limited in authority by the ambitious House of Braganca, the wealthiest family in Portugal. John summoned the Cortes (assembly) at Évora (November 1481) and imposed a drastic oath of obedience on his vassals. He also reasserted the *beneplacet*, requiring papal letters to be submitted to him before publication. He successfully negotiated a revision of the treaty with Spain, by which his children were restored to him. He then accused the Duke of Bragança of treason and had him tried and beheaded at Bvora (June 1483). Although evidence was produced that the Braganças had intrigued with Castile, it seems clear that John sought vengeance for their having caused the death of his maternal grandfather, the regent Dom Pedro. He confiscated the Braganças' vast estates and appointed royal judges in what had been private jurisdictions of the nobility. When a second conspiracy sought to remove him and bestow the crown on his wife's brother, the Duke of Viseu, John killed his rival with his own hand (August 1484).

In Africa, Afonso V had preferred crusading in Morocco to trade and discovery of the west coast. John II himself never returned to Africa after Arzila but supported the development of commerce and exploration.

Encouragement of African exploration

In December 1481 he sent Diogo de Azambuja to build the fortress of St. George at Mina (the "gold mine"), near Benin, a powerful native kingdom in the territory of modern Nigeria. Gold currency had been restored by his father, and the new -trade now doubled the royal revenues, and in 1485 John assumed the title of Lord of Guinea. He had already sent Diogo Cão to search for the seaway to India, and Cão had discovered the Congo. Christopher Columbus tried to interest him in his plan to reach India by a western route. John rejected this but licensed Fernão Dulmo to search for new islands, apparently without result. In 1485 he sent Cão on a second voyage that reached southwest Africa but failed to find the Cape. When in 1486 merchants in Benin heard news of a native potentate far to the east who was thought to be the legendary Christian ruler, Prester John, the King sent Pero da Covilhã and Afonso Paiva to visit India and Ethiopia by an overland route. He also sent Bartolomeu Dias to take over Cão's task of finding the southern extremity of Africa: Dias' return in December 1488 demonstrated that Africa could be rounded and India reached by sea, but it was only after John's death that Vasco da Gama's successful expedition to India was launched. In 1490 a mission was sent to consolidate relations with the kingdom of Congo. In 1493 Columbus arrived in Lisbon with tidings of, as he supposed, islands off Asia, and the Pope awarded these discoveries to the Spanish crown. But John II protested and began negotiations leading to the celebrated Treaty of Tordesillas (June 1494), which gave to Spain all lands west of a line 370 leagues to the west of the Cape Verde Islands. This line, however, reserved Brazil (still apparently unknown) for Portugal.

John had previously negotiated with the papacy, ceding the *beneplacet* but retaining the right to have ecclesiastical cases settled in Portugal and obtaining permission for a crusade against the Moors. He bent a minor expedition against Anafé and obtained tribute from Safi and Azemmour in Morocco, but in 1489 his attempt to build a fortress at the mouth of the Loukkos was prevented by the ruler of Fez.

Policy toward the Jews

When in May 1492 Ferdinand and Isabella of Spain decided to expel the Jewish population, John received a delegation of Spanish Jews who offered 60,000 cruzados for the permanent admission of 600 wealthy families to Portugal, together with a fee of eight cruzados a head for the temporary admission of others, who would be allowed to remain eight months, after which John would supply ships for them to leave. In fact, ships were provided only for Tangier and Arzila; some Jewish children were sent to settle the island of São Tomé.

John's son, Afonso, was married to the eldest daughter of the Spanish rulers, but soon after the Prince was thrown from a horse and killed (July 1491). John was deeply affected. He thought of legitimizing his other son Jorge but had already promised the succession to his wife's brother Manuel. He suffered a long illness and died in October 1495 at the small castle of Alvor in the province of Algarve. John's exercise of personal power, particularly against the nobles, explains the epithet, the "Perfect Prince," which owes its origin to Lope de Vega's play about him.

BIBLIOGRAPHY. Sources are the contemporary chronicler RUI DE PINA, *Croniqua Delrey Dom Joham II* (1792; modern edition by ALBERTO MARTINS DE CARVALHO, 1950), whose version is borrowed and amplified by GARCIA DE RESENDE. The only biography in English is that of ELAINE SANCEAU, *The* Perfect *Prince* (1959).

(H.V.L.)

John the Apostle, Saint

St. John has since ancient times been regarded as the author of three letters, the fourth Gospel, and Revelation in the New Testament.

Life with Jesus

The son of Zebedee, a Galilean fisherman, and Salome, John and his brother James were among the first disciples called by Jesus. In the Gospel of Mark he is always mentioned after James and was no doubt the younger brother. His mother, Salome, was among those women who ministered to the circle of disciples. James and John were called by Jesus "Boanerges," or "sons of thunder," perhaps because of some character trait such as the zeal exemplified in Mark 9:38 and Luke 9:54 when John and James wanted to call down fire from heaven to punish the Samaritan towns that did not accept Jesus. John and his brother, together with Simon Peter, formed an inner nucleus of intimate disciples. In the Fourth Gospel, ascribed by early tradition to John, the sons of Zebedee are mentioned only once, as being at the shores of the lake of Tiherias when the risen Lord appeared; whether the "disciple whom Jesus loved" (who is never named) mentioned in this Gospel is to be identified with John (also not named) is not clear from the text. After the Resurrection, John continued to play a leading role in the church at Jerusalem, together with Peter and James. His authoritative position is shown by his visit with Peter to Samaria to lay hands on the new converts there. It is to Peter, James (not the brother of John but a brother of Jesus), and John that Paul successfully submitted his Gospel for recognition. What position John held in the controversy concerning the admission of the Gentiles to the church is not known; the evidence is insufficient for a theory that the Johannine school was anti-Pauline—*i.e.*, opposed to granting Gentiles membership in the church. The authorship of the Gospel and letters of John and of the book of Revelation is discussed in BIBLICAL LITERATURE.

Obscurity of later life

John's subsequent history is obscure and passes into the uncertain mists of legend. At the end of the 2nd century, Polycrates, bishop of Ephesus, claims that John's tomb is at Ephesus, identifies him with the beloved disciple, and adds that he "was a priest, wearing the sacerdotal plate, both martyr and teacher." That John died in Ephesus is also stated by Irenaeus, bishop of Lyons c. AD 180, who says John wrote his Gospel and letters at Ephesus and Revelation at Patmos. During the 3rd century, two rival sites at Ephesus claimed the honour of being the Apostle's grave. One eventually achieved official recognition, becoming a shrine in the 4th century. In the 6th century, the healing power of dust from John's tomb was famous (it is mentioned by the Prankish historian Gregory of Tours); at this time also, the church of Ephesus claimed to possess the autograph of the Fourth Gospel.

By courtesy of the Stadelsches Kunstinstitut, Frankfurt

St. John the Apostle, on Patmos, writing the book of Revelation. Painting by an unknown artist, 14th century. In the Städelsches Kunstinstitut, Frankfurt, West Germany.

Legend was also active in the West, being especially stimulated by the passage in Mark 10:39, with its hints of John's martyrdom. Tertullian, the 2nd-century North African theologian, reports that John was plunged into boiling oil from which he miraculously escaped unscathed. During the 7th century, this scene was portrayed in the Lateran basilica and located in Rome by the Latin Gate; it is still annually commemorated on May 6. John's feast day otherwise is December 27. This belief that John did not die is based on an early tradition. In the original form of the apocryphal Acts of John (second half of the 2nd century) the Apostle dies; but in later traditions he is assumed to have ascended to heaven like Enoch and Elijah. A popular tradition known to Augustine declared that the earth over his grave heaved as if the Apostle were still breathing.

The legends that contributed most to medieval iconography are mainly derived from the apocryphal Acts of John. These Acts are also the source of the notion that John became a disciple as a very young man. Iconographically, the young, beardless type is early (as in a 4th-century sarcophagus from Rome), and this type came to be preferred (though not exclusively) in the medieval West. In the Byzantine world, the evangelist is portrayed as old, with long, white beard and hair, usually carrying his Gospel. His symbol as an evangelist is an eagle. On account of the inspired visions of the book of Revelation the Byzantine churches entitled him "the Theologian"; and the title appears in Byzantine manuscripts of Revelation, but not in manuscripts of the Gospel.

BIBLIOGRAPHY. A scholarly study of St. John's life may be found in commentaries on the Gospel, Epistles, and Apocalypse; especially commentaries on the Gospel by R.E. BROWN (1966–70), J.N. SANDERS (1968), and R. SCHNACKENBURG (1968). See also the entry under December 27 in *Butler's* Lives of *the* Saints, vol. *4* (1956). For an appreciation of Johannine theology, see F.M. BRAUN, Jean *le théologien*, 3 vol. (1959–67).

(H.Cha.)

John the Baptist, Saint

St. John the Baptist was a Jewish prophet of priestly origin, who, at the beginning of the Christian Era, preached the imminence of God's Final Judgment and baptized those who repented in self-preparation for it. In the Gospels of the New Testament, his baptism of Jesus of Nazareth marks the start of Jesus' public activity. John's principal importance for history lies not in any sect that he founded—though "Baptist" groups developed after his death—but in the fact that his message was taken up by the early Christian Church, in which his role as the forerunner of Christ was treasured, many of his disciples were later found, and many of his sayings were transmitted. His rite of baptism was adapted and adopted by the church and became one of its principal sacraments. Nevertheless, scholars differ widely in assessing John and determining the relationship that existed between him and Jesus.

Sources of information about John

The primary sources for information about John's life and activity are the four Gospels (Matthew, Mark, Luke, and John), the Acts of the Apostles, and the Jewish historian Josephus' Antiquities of the Jews. In using these works for historical reconstruction, allowances must be made for the known tendencies of each writer. All four Gospels recognize in John the start of the Christian Era, and each in its own way tries to reconcile John's precedence in time and Jesus' acceptance of his message and of a baptism of repentance from his hands (elements suggesting subordination to John), with the author's belief in Jesus as Messiah and Son of God. The Gospel According to Mark presents Jesus as the hidden Messiah, known only to a narrow circle, and John as the prophet Elijah returned and as the one who had to "come first to restore all things" but who also remained hidden and suffered death with little acknowledgment of his true status (Mark 9). An early collection of sayings of Jesus, known to biblical scholars simply as Q, similarly represents the Baptist as the herald of the Coming One and of the imminent Kingdom, yet tries to stress his preparatory character, and so his subordination to Jesus. Matthew and Luke develop these two sources. The Gospel According to Matthew emphatically identifies John as the returning Elijah, herald of the Kingdom of God (Matt. 3). For Matthew, John's death, like that of Jesus, illustrates the old Israel's hostility to God's offer of salvation. Luke, in his Gospel and in the Acts of the Apostles, neglects the identification with Elijah but describes John as Jesus' forerunner and as inaugurator of the time of fulfillment of prophecy. Luke's account of the infancy of John and of Jesus does not necessarily derive from an underlying document composed in a "Baptist" sect, although it uses material perhaps transmitted by former disciples of the Baptist. It depicts the birth of Jesus and John in two parallel series of scenes, each with an angelic annunciation, conception, marvellous birth, circumcision, hymns greeting the children and predicting their destiny, and infancy. Even in his mother's womb John recognizes Jesus—also still in his mother's womb—as his Lord. The Gospel According to John reduces the Baptist from an Elijah to a model Christian preacher, a mere voice; it omits any description of Jesus' baptism. Its tendency has often been labelled polemic against a continuing group of disciples of John, but it is more plausibly explained by the evangelist's desire that this ideal witness recognize the full character of the Christ as he is presented in the Gospel According to John and as a necessary consequence of the tension between the highly developed understanding of Christ in this Gospel and those details in early Christian tradition that suggested Jesus' subordination to John. The Gospels are thus primarily interested in the relations between John and Jesus.

Alinari

St. John the Baptist, painting by Pinturicchio (c. 1454–1513). In the Cathedral of Siena, Italy.

Josephus sought to present Jewish religious phenomena in Hellenistic categories and to de-emphasize any political elements unfavourable to Roman imperial control.

Life and work

After allowances are made for the tendencies of each of these sources, the following items about John appear relatively trustworthy. He was born somewhere in Judaea (localized at 'En Kerem from at least AD 530) to Zechariah, a priest of the order of Abijah, and his wife, Elizabeth, perhaps a relative of Mary, the mother of Jesus. His formative years were spent in the Judaean desert, where monastic communities, such as the Essenes (a strict Jewish sect that existed from about the 2nd century BC to the end of the 1st century AD), and individual hermits often educated the young in their own ideals.

In 27/28 or 28/29 John attained public notice, not as a priest but as a prophet. He was active in the region of the lower Jordan Valley, from "Aenon near Salim" (near-modern Nablus) to a point east of Jericho. His dress of an austere camel's hair garment was the traditional garb of the prophets, and his diet of locusts and wild honey represented either strict adherence to Jewish purity laws or the ascetic conduct of a Nazirite (a Jew especially vowed to God's service). His mission was addressed to all ranks and stations of Jewish society. His message was that God's wrathful judgment on the world was imminent and that, to prepare for this judgment, the people should repent their sins, be baptized, and produce appropriate fruits of repentance. Certain problems about, the meaning of John's message continue to be debated: In Matt. 3, John says, "he who is coming after me is mightier than I"; this might refer to God himself, a human messiah, or a transcendent divine being. He also says, "I baptize you with water . . . ; he will baptize you with the Holy Spirit and with fire"; this second baptism might symbolize the judgment the one coming would carry out. John's followers were characterized by penitent fasting, beyond the demands of Jewish Law, and special prayers. John's ethical call for justice and charity in Luke 3 requires righteousness from everyone in his own situation.

Although, like earlier prophets, John had an inner circle of disciples, baptism was not an admission rite into this group. It was a rite (immersion in running water) that symbolized repentance in preparation for the coming world judgment and was to be accompanied, before and afterward, by a righteous life. It was hardly conceived as a sacrament, in the Christian sense, conveying forgiveness, or as superseding Judaism and marking off a new people, including both Jews and Gentiles, prepared for God's final Kingdom; nor is a hypothesis that it symbolized a new Israel's crossing of the Red Sea toward a new national deliverance demonstrable. Equally unprovable is that it was a rite symbolizing man's reunion with divinity and return to his heavenly home—a sacrament of salvation and rebirth. The Jewish rite of baptism of converts differs fundamentally and is not its source. There were several other baptizing groups found about the same time and place, but none of these various and little-known baptisms can be shown to have inspired John's. It may have resembled in parts the initiatory baptism of the Essenes, though their other baptisms were more concerned with maintaining their community's ritual purity. John's baptism probably symbolized not so much anticipated entrance into the Kingdom of God as an anticipatory submission to the coming world judgment, which was represented as a coming second "baptism" by the Holy Spirit in a river of fire.

Possible relationship with the Essenes

The discovery of the Dead Sea Scrolls has drawn attention to the numerous parallels between John's mission and that of the Essenes, with whom John may have received some of his religious training. Both were priestly in origin, ascetic, and with intense and, in many respects, similar expectations about the end of the world. But John neither belonged to nor intended to found any organized community; he did not stress study of the Mosaic Law; and his message was more widely directed (to the poor, to sinners) than was that of the Essenes.

Jesus, who was baptized by John, saw in John the last and greatest of the prophets, the one who prepared for the coming of God's Kingdom (Mark 9, Matt. 11, Luke 7), and in many ways his ministry continued and developed John's. Whether John, who probably expected a divine Son of Man, recognized him in Jesus is not clear, but many of his disciples later followed Jesus.

Some time after baptizing Jesus, John was imprisoned by Herod Antipas, ruler of Galilee and central Transjordan. His crime was hardly the innocuous moral message Josephus presents, nor would his message, as found in the Gospels, have had much more immediate political bite. Herod had married (illegally, by Jewish Law) Herodias, the former wife of his half brother, after divorcing his first wife, the daughter of King Aretas IV of the Nabataeans, an adjacent Arab people. John's denunciation of this marriage doubtless presented Herod with the danger that his Jewish subjects would combine with his semi-Arab subjects in opposition to him. John's execution certainly preceded Aretas' victory over Herod in 35–36, a defeat popularly considered to have been divine vengeance on Herod for killing John. According to the Gospels, John's death preceded Jesus'; any greater chronological precision depends on the dates of Jesus' ministry and death. It is probable that John's followers recovered and buried his body and revered his tomb. The traditional burial site, at Sebaste (originally Samaria), near "Aenon by Salim," is attested from 360 onward. John's chief feast days are June 24 (nativity) and August 29 (beheading).

BIBLIOGRAPHY. On the ancient sources, see the standard commentaries on the Gospels, the Acts of the Apostles, and Josephus ("Loeb Classical Library"). Among recent monographs, CARL H. KRAELING, *John the Baptist* (1951), is primarily a historical investigation, while WALTER WINK, *John the Baptist in the Gospel Tradition* (1968), studies the Gospels' tendencies and their differing representations of John. For analogous contemporary baptizing groups, see J. THOMAS, *Le Mouvement baptiste en Palestine et Syrie . . .* (1935), and more recent literature on the community that produced the Dead Sea Scrolls. Most encyclopaedias of religion or biblical studies contain articles on John. Representative are: P. VIELHAUER in *Die Religion in Geschichte und Gegenwart*, 3rd ed., vol. 3, pp. 804–808 (1959), who offers the mainline view of critical New Testament scholarship; and W.R. FARMER in *The Interpreter's Dictionary of the Bible*, vol. 2, pp. 955–962 (1962), who stresses the political aspects of John's activity.

(Jo.St.)

John the Fearless, Duke of Burgundy

Second of the four dukes of Burgundy of the French royal house of Valois, who ruled over a large part of the Netherlands as well as the duchy and county of Burgundy between 1384 and 1477, John the Fearless exercised a more decisive influence on the internal history of France than any other Burgundian duke. Although his reign as duke was relatively short, 1404–19, his role in maintaining Burgundian power in Europe and in improving the administration of the scattered Burgundian territories was of crucial significance for the future of the Burgundian state.

The son of Philip the Bold, duke of Burgundy, and Margaret of Flanders, John the Fearless was born on May 28, 1371, outside Dijon, Burgundy, in the ducal castle of Rouvres, where he spent the greater part of his childhood. In 1385 he married Margaret of Bavaria, and, in the following decade, his father initiated him into the arts of government and warfare, though he was not given any post of responsibility. Even in 1396, at the age of 24, when he became leader of the Burgundian crusade against the Ottoman Turks in defense of Hungary, his leadership was only nominal. The actual conduct of the expedition, which ended in the disastrous defeat of the crusaders on the battlefield of Nicopolis and the capture of John by the Turks (an adventure that earned him the epithet the Fearless), was entrusted to a group of councillors and military advisers appointed by Philip the Bold. John evidently benefitted from the blunders of these commanders, for his subsequent career showed that he was the only one of the Valois rulers of Burgundy who knew how to handle an army.

John the Fearless, portrait by an unknown master of the southern Netherlands, c. 1415. In the Musée Royal des Beaux-Arts, Antwerp.

BY courtesy of the Musee Royal des Beaux-Arts, Antwerp; photograph, A.C.L. Brussels

Succession

When John at last succeeded his father in 1404 as duke of Burgundy and count of Burgundy, Flanders, and Artois, he was 33 years old.

John the Fearless spent most of his time and his considerable political and military energies in France, Paris being his normal place of residence and seat of government. His only significant personal participation as duke of Burgundy in major events outside France took place in 1408, when he led a Burgundian army to aid his beleaguered brother-in-law, the Bishop of Liège, John of Bavaria, against the citizens of Liège, who were in open revolt. On the field of Othée, on September 23, 1408, the men of Liège were decisively defeated, and Burgundian influence was extended over the city and over the bishopric of Liège. From the start, then, John found himself involved in French affairs and was in part responsible for provoking a civil war in France with a rival house, headed by his first cousin, the King's younger brother, Louis, duc d'Orléans. Each man sought control of the mad king Charles VI and his queen and of the capital Paris. While the notorious murder by Duke John of his cousin by hired assassins in 1407 enabled John to subdue Paris and the crown, the opposition to the Burgundians by Louis's followers and heirs continued. Their faction was named after its main supporter, Bernard VII, comte d'Armagnac. During the five years between 1413 and 1418, in which the Armagnacs succeeded in driving the Burgundians out of Paris, the internal situation in France was further complicated by a new English invasion led by the ambitious king, Henry V. Duke John was one of those French princes who, while pretending to do his best to reach the battlefield of Agincourt to give battle to the English (1415), was unaccountably delayed on the way. His intermittent negotiations with King Henry V did not, however, lead to a firm Anglo-Burgundian alliance, and in the autumn of 1419 John turned instead to the Armagnacs, in the hopes of arranging a truce or even making a firm peace settlement with their youthful leader, the dauphin Charles (the future Charles VII), in an alliance against the English. The two princes, each with ten companions, met on the bridge of Montereau, some 50 miles southeast of Paris. On September 10, 1419, as the diplomatic parley began, John the Fearless was struck down and killed during a dispute started by the Armagnacs, a political assassination that contemporary evidence shows was almost certainly carefully premeditated.

Reputation

John the Fearless pursued aims similar to those of the other rulers of his day: the consolidation and extension of his own and his family's power. In spite of his lapses into violence, his love of intrigue, his hypocrisy, and his rashness, he was a successful diplomat and military leader; he was more dynamic and more of a reformer than his son Philip the Good and more cunning, though less scrupulous, than his father. Yet he has received less attention from historians than either of them. In the eye of history, especially French history, he has long been regarded as a traitor and assassin. There was, perhaps, a dark and sinister element in his character, but he lived in an age when vice, tyranny, and murder were the common properties of every ruler. If he wrought destruction in France, he also brought peace and prosperity to his own Burgundian lands.

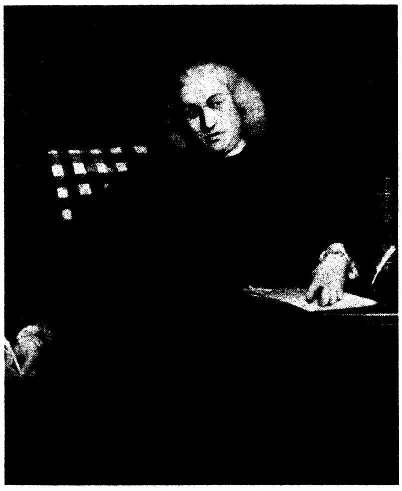
BIBLIOGRAPHY. J. D'AVOUT, ***La Querelle des Armagnacs et des Bourguignons*** **(1943),** the only detailed study of French internal affairs in John the Fearless' time; J. CALMETTE, ***Les Grands Ducs de Bourgogne*** (1956; Eng. trans., ***The Golden Age of Burgundy: The Magnificent Dukes and Their Courts,*** **1962),** the best general survey of the Valois dukes of Burgundy; R. VAUGHAN, ***John the Fearless: The Growth of Burgundian Power*** **(1966),** deals with the history of the Burgundian state under Duke John and lists relevant literature.

(R.Va.)

Johnson, Samuel

The English poet, critic, essayist, and lexicographer Samuel Johnson became famous not only for his writings but also for his forceful, witty conversation. After Shakespeare, Johnson is possibly the best known figure and the most frequently quoted in the whole range of English literature.

BY courtesy of the National Portrait Gallery, London



Dr. Johnson, oil painting by Sir Joshua Reynolds, 1765. In the National Portrait Gallery, London.

EARLY LIFE AND INFLUENCES

He was born at Lichfield, Staffordshire, on September 18 (new style; September 7, old style), 1709. His father, Michael Johnson, was a prominent citizen of Lichfield and was sheriff of the city at the time of Samuel's birth. As a bookseller, he conducted a substantial but not very profitable business. Samuel was not a healthy child. His eyes were weak and he was the victim of a tubercular infection in the glands of the neck, commonly known then as "the King's Evil." In the hope that the cure for this disease lay in the royal touch, Mrs. Johnson travelled to London with Samuel in March 1712, and the boy was duly touched by Queen Anne. His memories of the ceremony were naturally slight, but he retained "a sort of solemn recollection of a lady in diamonds and a long black hood." It is not recorded that he derived any physical benefit from the royal touch, but the gold amulet that the Queen hung round his neck remained there until his death.

Michael Johnson was a high churchman with Jacobite sympathies, favouring the Stuart rather than the Hanoverian succession. His wife, Sarah Ford, was a devout woman with leanings toward Calvinism. She taught her son to learn the collect for the day by heart and expounded to him the contrast between heaven and hell. In 1717 Johnson entered Lichfield grammar school and began the study of Latin. One of his schoolfellows, Edmund Hector, who was to become his lifelong friend, recalling in later years Johnson's "uncommon abilities for learning," wrote: "His ambition to excel was great, though his application to books . . . was very trifling . . . his dislike to business was so great that he would procrastinate his exercises to the last hour." It was a disposition that remained with Johnson throughout his long literary career.

Friends and influences at Oxford

When he was promoted to the upper school, Johnson came under the discipline of the headmaster, a scholar but a tyrant who beat his boys indiscriminately—"to save them from the gallows." As Johnson later said, "My master whipt me very well. Without that, Sir, I should have done nothing." After a brief period at the grammar school at Stourbridge, where he was a student and also took some part in the teaching of the younger boys, Johnson helped his father in the bookshop. Rambling along his father's shelves, he read widely and with the instinct of an incipient scholar—"not voyages and travels, but all literature, Sir, all ancient writers, all manly . . ." Thus, by the time that he entered Pembroke College, Oxford, in 1728, he was familiar with many works unknown at the universities; and when he was first introduced to William Jorden, tutor of the college, his own contribution to the conversation was a quotation from the Latin grammarian and philosopher Macrobius. Johnson thought little of Jorden's scholarship but praised his goodness of heart. William Adams, fellow and later master of Pembroke, meant more to Johnson and remained his friend through many years. He had some good friends, too, among the undergraduates of his own generation. He was near to his old schoolfellow, John Taylor of Ashbourne, at Christ Church, and among Pembroke men he had the reputation of being gay and frolicsome. But it was a forced gaiety. Frustrated and embittered by poverty, he defied authority and "thought to fight his way by his literature and his wit." One example of his scholastic facility survives from his undergraduate days—a Latin translation of Alexander Pope's "Messiah," which was included in a *Miscellany* published at Oxford in 1731.

How the impecunious Michael Johnson was enabled to send his son to college is not wholly clear. Possibly a small legacy received by Mrs. Johnson may have helped. But, in any event, Johnson was compelled to leave Oxford in December 1729 after a residence of 13 months. His prospects were poor. He had no degree or other qualification; his father's business was declining; an application for an ushership (assistant teacher) at Stourbridge was unsuccessful. His father died at the end of 1731, and in the following year Samuel accepted a post as undermaster in the grammar school at Market Bosworth. The work brought him neither health nor happiness, and he resented the arrogance with which he was treated by Sir Wolstan Dixie, in whose house he lived. Through the influence of his old friend Edmund Hector, who had become a surgeon in Birmingham, Johnson secured the task of translating into English the French version of ***A Voyage*** to ***Abyssinia*** by Father Jerome Lobo. Johnson's preface gives an early indication of his instinctive sympathy with the natives of an invaded country. The invaders of Abyssinia were missionaries who preached the gospel with swords in their hands. The preface also gives an authentic foretaste of Johnson's prose style:

> The Reader . . . will discover, what will always be discover'd by a diligent and impartial Enquirer, that wherever Human Nature is to be found, there is a mixture of Vice and Virtue, a contest of Passion and Reason, and that the Creator doth not appear Partial in his Distributions, but has balanced in most Countries their particular Inconveniences by particular Favours.

Marriage and early literary endeavours

Johnson's sojourn in Birmingham brought him something more than a fee of five guineas for his first book—it brought him a wife in the person of Elizabeth, widow of Harry Porter, a mercer. She was 20 years older than Johnson, and it is not easy to determine the grounds of mutual attachment. Johnson, in later years, used to speak of his wife's beautiful blonde hair; and whatever the lady thought of Johnson's looks ("lean and lank . . . the scars of the scrofula deeply visible") or of his "convulsive starts and odd gesticulations," she at least appreciated the vigour and good sense of his conversation. The marriage took place at Derby in 1735, and the bride brought with her a reputed dowry of £700. On the strength of this and with the encouragement of a friend who was registrar of the ecclesiastical court of Lichfield, Johnson decided to set up at Edial, near Lichfield, a school on his own at which young gentlemen could be boarded and taught the Latin and Greek languages. He prepared an elaborate curriculum, but only a few young gentlemen, among whom was David Garrick, came as pupils, and at the end of two years Johnson had to admit failure.

But in spite of ill health, of melancholia, of his lack of a degree, and of the collapse of his school, the desire to be known as a scholar and a writer remained clear in his mind. While still a schoolboy he had written of

> . . . the young Authour, panting after fame,
> And the long honours of a lasting name

and in the year before his marriage he had written to the founder of the *Gentleman's Magazine*, offering to fill a column, on reasonable terms, with poetry, dissertations, critical remarks on ancient and modern authors, and other material. Furthermore, with plenty of time on his hands at Edial, he embarked, again with the encouragement of friends, on the writing of a tragedy, based on the story of Sultan Mahomet (Mehmed) II and the beautiful Greek maiden Irene, as told by Richard Knolles in his *General History of the Turks*. But the work was not finished, and Johnson, facing the fact that he must write something for which an editor or a bookseller would pay, decided to seek his fortune in London. So, in company with David Garrick, he rode to London in March 1737. Arthur Murphy, in "An Essay on the Life and Genius of Samuel Johnson LL.D." (1792), wrote:

> Two such candidates for fame perhaps never, before that day, entered the metropolis together. . . . They brought with them genius, and powers of mind, peculiarly formed by nature for the different vocations to which each of them felt himself inclined. . . . In three or four years afterwards Garrick came forth with talents that astonished the publick. . . . Johnson was left to toil in the humble walks of literature.

Contributions to the *Gentleman's Magazine*

The *Gentleman's Magazine* offered him the first opportunity of humble toil. Edward Cave, the publisher, quickly recognized his journalistic ability, and Johnson contributed a number of pieces in prose and verse—odes, epigrams, reviews, as well as a series of concise biographies. In the later part of 1737 he returned to Lichfield, finished his tragedy *Irene*, and brought his wife to London. Meanwhile, his ambition to be a writer of something more than ephemeral pieces for periodicals remained, and in 1738 his first substantial poem, *London*, was published. It was written in imitation of the third *Satire* of Juvenal, and in it Johnson embodied his protest against political corruption:

> Here let those reign, whom pensions can incite
> To vote a patriot black, a courtier white;

against the dangers of the London streets:

> Their ambush here relentless ruffians lay,
> And here the fell attorney prowls for prey;
> Here falling houses thunder on your head,
> And here a female Atheist talks you dead.

and, with more acutely personal feeling, against the miseries of the unknown and impecunious author:

> This mournful truth is ev'ry where confess'd,
> SLOW RISES WORTH, BY POVERTY DEPRESS'D.

London, published anonymously, had an immediate success. It went quickly into three editions and won high praise from Alexander Pope. But Johnson's fee was only ten guineas, and again he thought of schoolmastering as an alternative to being "starved to death in translating for booksellers." The headmastership of Appleby School in Leicestershire was offered to him, subject to his obtaining the degree of M.A., but negotiations for its conferment broke down both at Oxford and at Dublin; similarly, his lack of a degree in law frustrated his application for permission to practice as an advocate. To this period of embittered disappointment belong his two most violent and satirical strictures upon the government of Walpole: the first was *Marmor Norfolciense* (1739), an essay upon a Latin rhyme supposed to have been discovered in Walpole's county, Norfolk; and the second *A Compleat Vindication of the Licensers of the Stage* (1739), an ironical defense of the suppression of Henry Brooke's play *Gustavus Vasa*. Both satires are the protests of an angry young man rebelling against authority and striving, as he had striven at Oxford, to fight his way out by his wit. Johnson was no sentimental Jacobite, but his scorn of the Hanoverian government was never more bitterly explicit than in *Marmor Norfolciense:*

> Then o'er the World shall Discord stretch her wings;
> Kings change their Laws, and Kingdoms change their Kings.

Satires against the Walpole ministry

The story that a warrant was issued for Johnson's arrest has never been verified, but, meanwhile, as a member of Cave's staff, he was required to treat contemporary politics in more sober style. Reports of parliamentary debates had been a feature of the *Gentleman's Magazine* since 1732, but shortly after Johnson's arrival in London the House of Commons forbade publication of its proceedings, Cave, however, contrived to continue the publication as "Debates in the Senate of Magna Lilliputia"; and in the first instance Johnson was employed to assist in editing and expanding the reports, but from 1740 to 1743 the "Debates" were entirely Johnson's own work. He was not a reporter in the modern sense and was only once inside the House of Commons. Sometimes he had a few notes supplied by other reporters, sometimes nothing more than the subject of debate and the names of the speakers; sometimes, he confessed, the "Debates" were the mere coinage of his own imagination, and in later years he had some prickings of conscience about his freedom of invention. As a journalistic feat, they were a remarkable tour de force. He would shut himself up in a room at Cave's headquarters in St. John's Gate and deliver three columns for the *Magazine* in an hour. His often-quoted remark that he took care that the Whig dogs—members of the political party opposed to the Tories—should not have the best of it gives a false impression of what he was doing. His reports have been well described as leading articles on both sides of the question, rather than records of the cut and thrust of debate. The short speech put into the mouth of Walpole at the end of the debate on the motion for his removal from office is a remarkably objective record of a minister speaking with dignity and restraint in his own defense.

Friendship with Richard Savage

One of Johnson's closest companions in his early years in London was Richard Savage. Savage had in his time been actor, playwright, and poet. He claimed to have been nobly born and had had many friends among the great whose hospitality he persistently abused. Two qualities at least he shared with Johnson—poverty and patriotic indignation against the Walpole administration. When their fortunes were at their lowest, they would spend whole nights in "a perambulation round the squares of Westminster . . . when all the money they could both raise was less than sufficient to purchase for them the shelter and sordid comforts of a night cellar." When Savage died, Johnson lost no time in commemorating his friend. Written con amore and at a white heat, the *Account of the Life of Mr Richard Savage*, published anonymously in 1744, was the first of Johnson's prose works to captivate the public. The novelist Henry Fielding, who would publish *Tom Jones* five years later, called it the best treatise in the language on the excellencies and defects of human nature; Joshua Reynolds, then only 21 but destined to become one of England's greatest painters, could not put the book down until he had finished it.

RECOGNITION AND MATURE CAREER

The theatre. In addition to a great variety of hack work, Johnson was now turning his attention to Shakespeare. In 1745 he published his *Miscellaneous Observations on the Tragedy of Macbeth,* coupled with preliminary proposals for his own edition of the plays, but he was deflected from further Shakespearean work by the suggestion that he should compile a dictionary of the English language. That a syndicate of booksellers should have chosen Johnson for so gigantic a task is a tribute to the position he had made for himself during his nine years in London.

In those same years David Garrick had made more rapid and more brilliant progress. Deserting the law for the theatre, he had made his mark as an actor in 1741 and by 1747 had become the manager of Drury Lane Theatre. For the first performances under Garrick's management Johnson wrote a prologue, and, although he commonly spoke with contempt of actors and their profession, he was willing, for friendship's sake, to plead their cause:

> Ah let not Censure term our Fate our Choice,
> The Stage but echoes back the publick Voice,
> The Drama's Laws the Drama's Patrons give,
> For we that live to please, must please to live.

At this time, moreover, Johnson had a personal interest in dramatic production. His tragedy *Irene* had long lacked a publisher or a producer. Garrick, when he came into power, agreed to do his best for his old master, and *Irene* was produced "with a display of Eastern magnificence" in 1749. It ran for nine nights. It was essentially a moralist's play, and although it was said to be universally admired, it has never been revived.

The Vanity of Human Wishes

More permanent was Johnson's second didactic poem, *The Vanity of Human Wishes* (1749), in which the careers of Galileo, Wolsey, Charles XII of Sweden, and others are shown to illustrate the hazards of political ambition, the futility of military conquest, and the miseries of authorship:

> There mark what ills the scholar's life assail,
> Toil, envy, want, the patron, and the jail.
> See nations, slowly wise and meanly just,
> To buried merit raise the tardy bust.
> If dreams yet flatter, once again attend,
> Hear Lydiat's life, and Galileo's end.

Though at the time of its publication *The Vanity of Human Wishes* had a much less rapid sale than *London,* it is Johnson's greatest poem. Its manner is that of its period. But Johnson's panorama of the rise and fall of scholars and philosophers, of statesmen and kings, in the modern as well as the ancient world, is inspired by that "high seriousness" that endows the poem with universality. But the record of disillusionment is not the end.

> Must helpless man, in ignorance sedate,
> Roll darkling down the torrent of his fate?

No, says Johnson, he must pray for the love and patience and faith of the Christian.

The "Rambler." The contract for A *Dictionary* of *the English Language* had been signed in 1746, and Johnson published his *Plan of a Dictionary of the English Language* the following year. He made no complaint of what the booksellers offered him, but it was not enough to keep the wolf from the door. So, in 1750, he embarked upon the *Rambler,* a twopenny sheet published twice a week and containing a single anonymous essay. The *Rambler* is of fundamental importance in any estimate of Johnson's approach to literature. Before he embarked on the work he prayed that he might promote the glory of God and the salvation of himself and others. The Rambler, in short, was not an entertainer but an instructor, and 19th-century critics tended to dismiss his essays as lay sermons.

Johnson had recently founded his first club (the Ivy Lane Club) and from his tavern chair would take a prominent part in contributing to what he called "colloquial entertainment." But a printed essay demanded "more accurate thought and more laboured beauties." Talking for victory was legitimate; with the printed word came the moralist's responsibility, and in Johnson's literary creed the basic article of belief was that it was always a writer's duty to make the world better. This sense of responsibility determined the style of the essays. He created no character comparable to that of Addison and Steele's Sir Roger de Coverley in the *Spectator,* and the essays bear little relation to current events or current literature. On the other hand, they frequently reflect the social and literary conditions of the time: there is scorn for the virtuoso and the overdomesticated hostess; an objective picture of the prostitute's life; a vigorous protest against the death sentence for robbery; and, inevitably, a grimly humorous presentation of the journalist's lot. Moreover, they show a remarkable understanding of human frustration and blocked wills. As has recently been pointed out, in the writings of Johnson there is the closest anticipation of the theories of Freud, if not his language, before the 20th century.

Mrs. Johnson's death

The *Rambler* appeared twice a week for two years (1750–52). A few days after the issue of the last number, Johnson's constitutional melancholy was deepened by the death of his wife. The full story of his married life must necessarily be conjectural. None of Johnson's friends in his early London period appears to have met Mrs. Johnson, and there are but few contemporary references to her. That the marriage was based on mutual admiration and affection is reasonably clear. From the beginning Mrs. Johnson had relished the quality of her husband's conversation, and she had a special admiration for the *Rambler;* Johnson, on his part, appreciated her intelligent reading of comedy. But, domestically, they were not well suited. Johnson was as insistent about the quality of his food as he was careless and untidy about the house. What was more serious was that in her later years Mrs. Johnson became addicted to strong liquor and to drugs and was unwilling to satisfy her husband's physical desires. Something of this may be read between the lines of Johnson's prayers and meditations about her death. Nevertheless, Johnson's affection, tinged with some remorse, remained sincere; and its sincerity is not impaired by the discovery that on April 22, 1753, he purposed "to try . . . to seek a new wife without any derogation from dear Tetty's memory." The precise direction of his search is not known, but it is interesting that he contemplated an adventure that he was later to describe as the triumph of hope over experience. "Formosae, cultae, ingeniosae, piae" ("a woman of beauty, elegance, ingenuity, and piety")—so Johnson described his Tetty on the gravestone that he placed in Bromley Parish Church more than 30 years later. In lapidary inscriptions, as he said, a man is not upon oath.

In 1752 Johnson was a lonely man, but he had made some good friends. At the Ivy Lane Club he was delighted "to pass those hours in a free and unrestricted interchange of sentiments, which otherwise had been spent at home in painful reflection." There were ten members, including Sir John Hawkins, who lived to be Johnson's executor and biographer; John Hawkesworth, editor of the *Adventurer,* to which Johnson himself contributed a number of essays; John Ryland, one of the few of Johnson's early friends who lived to attend his funeral; John Payne, publisher of the *Rambler;* and Richard Bathurst, the physician whom Johnson loved better than any other creature. It was at the Ivy Lane Club that Johnson celebrated the publication of the first book of his friend Mrs. Charlotte Lennox by "a whole night spent in festivity." About five in the morning, according to Hawkins, "Johnson's face shone with meridian splendour, though his drink had been only lemonade."

The dictionary. In 1747 Johnson published his *Plan of a Dictionary of the English Language.* At the suggestion of Robert Dodsley, the author and bookseller, it was dedicated to Lord Chesterfield. At first Chesterfield showed some interest and made some suggestions for revision, but when he paid no further attention to the work, Johnson would not be obsequious—and Johnson did not forget. In April 1753 he was beginning work upon the second volume of the *Dictionary,* and he had still to write the preface, the grammar, and the history.

Two years later the work was finished. It had occupied Johnson and his amanuenses for eight and one-half years, and its accomplishment was described by Sir James Mur-

Character and reputation of the *Dictionary*

ray, editor in chief of the *Oxford English Dictionary*, as a marvellous one. It surpassed earlier dictionaries not in bulk but in precision of definition and in literary illustration. Its 40,000 words were, in fact, rather less than those in the work of his predecessor, Nathaniel Bailey; but what distinguished Johnson's work was the range of reading by which he exemplified the different shades of the meaning of a particular word. "I applied myself," he wrote in his preface, "to the perusal of our writers. . . noting whatever might be of use to ascertain or illustrate any word or phrase." Certain books (Bacon's *Essays,* South's *Sermons,* and others) have survived with Johnson's underlinings and indications of the passages chosen for quotation. His assistants copied the quotations on separate slips, and these were pasted below Johnson's own definitions. At the beginning Johnson had looked forward to hours that he would revel away in feasts of literature and ransack the obscure recesses of northern learning. But he soon realized that these were "the dreams of a poet doomed at last to wake a lexicographer." To the weariness of copying was added the vexation of expunging, and Johnson, having set practical limits to his work, was acutely conscious of its imperfections. His orthography was admittedly controvertible and his etymology uncertain; and, even while the dictionary was hastening to publication, some words, as he said, were budding and some were falling away. Nevertheless, the claim made in the final paragraph of his preface—one of the finest examples of his prose style—was abundantly justified:

> In this work, when it shall be found that much is omitted, let it not be forgotten that much likewise is performed; and though no book was ever spared out of tenderness to the author, and the world is little solicitous to know whence proceeded the faults of that which it condemns; yet it may gratify curiosity to inform it, that the *English Dictionary* was written with Little assistance of the learned, and without any patronage of the great; not in the soft obscurities of retirement, or under the shelter of academick bowers, but amidst inconvenience and distraction, in sickness and in sorrow.

Today the ordinary reader may tend to remember the few "wild blunders"; the "risible absurdities"; and the frankly personal prejudices shown in the definitions of "oats" ("A grain, which in England is generally given to horses, but in Scotland supports the people"), "excise," "pension," and other well-known examples. But Johnson's enrichment of the wordbooks of his predecessors by judicious definition and by linguistic illustration from the whole range of English literature was an enduring monument in the history of lexicography.

On the title page of the *Dictionary,* Johnson had the satisfaction of describing himself as a Master of Arts of the University of Oxford, and it is significant that the award was made in consideration of the religious and moral value of his essays. The degree, in short, was conferred upon the Rambler.

The *Dictionary* was well received, and opposition came only from "the Criticks of the coffeehouse whose outcries are soon dispersed in the air and are thought on no more." But there had been one outcry of enthusiastic praise that Johnson did not allow to be dispersed. Shortly before the work appeared, in two papers in the *World,* Lord Chesterfield, seeking to make amends for his previous neglect, hailed Johnson as the supreme dictator of the English language and so provoked the most famous of all Johnson's letters:

> The notice which you have been pleased to take of my labours, had it been early, had been kind; but it has been delayed till I am indifferent, and cannot enjoy it; till I am solitary, and cannot impart it; till I am known, and do not want it. . . .

Journalism. Johnson was soon at work upon an abridged edition of the *Dictionary*, but fame as a lexicographer had not relieved him of pecuniary distress. In March 1756 he was under arrest for the sum of £5 18s. Normally he would have appealed to one of the printers for whom he worked, but they were not available, and it was the printer-novelist Samuel Richardson who sent him six guineas. So, of necessity, his activity as a journalist continued. He edited Sir Thomas Browne's *Christian Morals* (1756), wrote prefaces to William Payne's *Introduction to the Game of Draughts* (1756) and Richard Rolt's *New Dictionary of Trade and Commerce* (1756), and contributed many articles to journals. In the *Literary Magazine,* Jonas Hanway's "Essay on Tea" provoked Johnson's description of himself as a "hardened and shameless tea-drinker" who had drunk of it for 20 years without hurt and therefore believed it not to be poison.

A more important review was that of Soame Jenyns' *Free Inquiry into the Nature and Origin of Evil* (1757), in which, with massive irony, Johnson demolished the conjecture of a superior race of beings deceiving and tormenting men for their own pleasure:

> The only end of writing is to enable the readers better to enjoy life, or better to endure it: and how will either of those be put more in our power by him who tells us, that we are puppets, of which some creature not much wiser than ourselves manages the wires.

Nor could Johnson swallow the bland assertion that poverty was generally compensated by better health and a more exquisite relish of the smallest enjoyments. "Life," he retorted, "must be seen before it can be known." These were topics on which he wrote with peculiar depth of feeling, and, as always when he felt deeply, his style became more simple.

Political writings

Neither Johnson's interests nor his writings were confined to the problems of literature and ethics. To the *Literary Magazine* of 1756 he contributed two articles on the political situation—"An Introduction to the Political State of Great Britain" and "Observations on the present State of Affairs." In both he writes with scorn of the power politics inherent in both English and French colonialism. The dispute between the two countries in America was "the quarrel of two robbers for the spoil of a passenger," and even those who had settled in the New World on the fairest terms had no other merit than that of "a scrivener who ruins in silence over a plunderer that seizes by force." Nor was he afraid of asserting that the French sent out better governors and that they treated the natives better than did the English. It was ridiculous, he wrote, to imagine that the friendship of nations, whether civil or barbarous, could be gained or kept but by kind treatment; and it was this basic mistrust of the motives and methods of colonizers that provoked his later and better known outburst against colonial claims. Meanwhile, he had projects of his own in hand or in view. He issued his proposals for his edition of Shakespeare in 1756, emphasizing, not for the first time, that his motive was not desire of fame but want of money. In April 1757 he wrote to his old friend Edmund Hector, in Birmingham, that the subscriptions, if slightly disappointing, were satisfactory, but early in 1758 he was obliged to borrow £40 from a bookseller.

"The Idler" essays

About the same time he undertook to contribute a weekly essay, to be entitled "The Idler," to the *Universal Chronicle.* Of these essays it may be said that the reader who has been nurtured in the tradition that while Johnson's talk is magnificent, his writings (other than the *Lives of the English Poets)* are unreadable, would be well advised to turn his attention to "The Idler." Though the moralist and the social reformer (especially on such topics as debtors' prisons and vivisection) are still evident, he is willing to turn aside to the human comedy: for instance, to the female bargain hunter ("whatever she thinks cheap, she holds it the duty of an economist to buy") or to the publisher ("Some never dealt with authors; others had their hands full; some had never known such a dead time; others had lost by all that they had published for the last twelvemonth"). There are also some character sketches, of which the most memorable is that of "Mr. Sober":

> Mr. Sober's chief pleasure is conversation; there is no end of his talk or his attention; to speak or to hear is equally pleasing; for he still fancies that he is teaching or learning something and is free for the time from his own reproaches.
>
> But there is one time at night when he must go home, that his friends may sleep; and another time in the morning, when all the world agrees to shut out interruption. These are the moments of which poor Sober trembles at the thought.

It is a rare and convincing piece of self-portraiture. Every week for two years, with a few exceptions, Johnson delivered his "Idler" essay to the printer.

Rasselas. In the midst of his work and worry came news of his mother's illness. On January 13, 1759, he contrived to send her 12 guineas. But he knew that he would need more, and on the same day that he wrote his last tender tribute to his mother (January 20), he implored Strahan, the printer, to let him have £30 on account of "a thing he was preparing for the press." The thing was to be entitled *The Prince of Abissinia,* better known as *Rasselas.* Written in the evenings of a week with the impending expenses of his mother's funeral in mind, it explores and exposes the vanity of the human search for happiness. The setting was no doubt prompted by Johnson's recollection of Jerome Lobo's *Voyage to Abyssinia,* and the work is addressed to those who "listen with credulity to the whispers of fancy and pursue with eagerness the phantoms of hope." Impelled by such eagerness, Rasselas, with his sister, leaves his happy valley because its pleasure has ceased to please and because he is fired with a desire to do something. He meets with men of varied occupations and interests and earnestly explores their manner of life—scholars, astronomers, shepherds, hermits, poets. With Imlac, the poet, he ranges over many of the basic problems of art and life but gains little satisfaction from the answers he receives. At his first entry into city life, Rasselas meets everywhere with gaiety and happiness, but closer association reveals a picture of levity and intemperance. From the extravagance of youth he turns to an inspiring lecture on morality, only to find that the lecturer's philosophy collapses at the first stroke of personal misfortune. When he passes hopefully to scenes of pastoral simplicity, he finds the shepherds cankered with discontent, since they are condemned to labour for the luxury of the rich; when hc seeks advice from a hermit, he is disappointed to be told that "the life of a solitary man will be certainly miserable, but not certainly devout." So the journey of disillusionment continues, until Imlac protests: "While you are making the choice of life, you neglect to live"—which is, perhaps, the most important moral to be drawn from the tale.

The moral of *Rasselas*

Of all Johnson's writings, none is more intensely characteristic than *Rasselas.* It is a remarkable example of his fluent productivity when he had a definite object and a definite date before him, but speed did not debase either the solemnity of his subject or the dignity of its treatment. Johnson allowed his imagination to wander into Ethiopia and Egypt, but fundamentally *Rasselas* is a spiritual autobiography. Furthermore, it was the one prose work of Johnson's that obtained an immediate and wide popularity in his lifetime. It satisfied the taste of the 18th-century reader for "impressive truth in splendid fiction drest"—and not only of the English reader. In a variety of translations its fame spread over Europe and beyond.

Johnson's circle. In spite of his struggles and sorrows, Johnson was, by this time, no longer a lonely hack writer. His writings, though they did not make his fortune, had brought him many friends. Joshua Reynolds had been enthralled by Johnson's *Life of Savage* and was no less enthusiastic about his conversation and his counsel; Charles Burney, a leading English music historian of his time, warmly commended the *Dictionary* and encouraged the edition of Shakespeare; Bennet Langton, a Lincolnshire gentleman and scholar, came to London at an early age for the express purpose of meeting the author of the *Rambler.* Johnson also cherished his Oxford friendships. He spent some time there in 1754 and again in 1755 and 1759 and, after the conferment of his degree, took a keen pleasure in wearing his gown. It was at Trinity in 1759 that Bennet Langton introduced him to Topham Beauclerk, who was descended from Charles II and Nell Gwyn.

Introductions, reviews, and dedications

Even so, Johnson was still at the beck and call of authors, editors, and friends for the writing of introductions, reviews, and, especially, dedications. Of James Bennet's edition of Roger Ascham's works he was virtually the editor and contributed a life of Ascham, a famous scholar and Latin secretary to Elizabeth I; he wrote the loyal address presented to George III on his accession by the painters, sculptors, and architects; he revised a pamphlet on the proper route for the coronation procession; he reviewed a work on Mary, queen of Scots, and wrote a dedication for Giuseppe Baretti's *Italian Dictionary.* Meanwhile his edition of Shakespeare tarried. In 1762 the unexpected happened. He was informed of the gracious intention of his majesty the King to confer on him a pension of £300 a year. Could he in decency take it, after defining "pension" in his *Dictionary* as "pay given to a state hireling for treason to his country"? At least, he felt, he must consult his friends. But Reynolds reassured him, and Lord Bute, the prime minister, told him that the award was made not for anything he might do but for what he had already done. So, from a full heart, Johnson thanked his lordship for sparing him "the shame of solicitation, and the anxiety of suspense."

Meeting with James Boswell. The following year was to provide yet another landmark in Johnson's life, for on May 16, 1763, his accidental meeting with James Boswell in the back parlour of Thomas Davies' bookshop in Covent Garden inaugurated one of the most famous companionships in history. Boswell, the eldest son of Lord Auchinleck, a Scottish judge, had studied law at Edinburgh and Glasgow and had a passionate desire to taste the felicities of London life. His mind was dominated by two ambitions—to meet famous men and to be a famous author himself. Staying in a country house in the previous year, he had read aloud, with enthusiastic comment, some of the *Rambler* essays; and, in the list of celebrities whom he wished to meet, the author of the *Rambler* stood high. A little daunted by Johnson's brusque rejoinders at their first meeting, Boswell was nevertheless encouraged by Davies to persevere; and a week later he waited, with some apologies, upon Johnson. Johnson cut him short. "I am obliged," he said, "to any man who visits me." So Boswell stayed and listened in rapture. "His conversation," he wrote in his diary, "is as great as his writing." He could give no higher praise. To Johnson new friendships were always welcome, and especially with young people (Johnson was then 53 and Boswell only 22), for such friendships lasted longest. To the charm of Boswell's enthusiasm he quickly succumbed. Together they enjoyed suppers at the Mitre Tavern and excursions on the Thames, but the signal mark of Johnson's favour was his offer to accompany Boswell in the coach to Harwich, whence Boswell was to embark on his continental tour. Johnson gave him much sound advice. In particular, he urged him to keep a diary, and it is to Boswell's diary that the world owes its intimate knowledge of Johnson. Many others left valuable records of his life and character, but it was Boswell who Johnsonized the land.

Foundation of The Club

In 1764 Johnson was happy to concur with Joshua Reynolds' suggestion for the foundation of what is still the most famous of London dining clubs—The Club. Among the original members, besides Reynolds and Johnson, were Edmund Burke, Topharn Beauclerk, Bennet Langton, and Oliver Goldsmith, and nine years later it was one of the proudest moments of Boswell's life when he was admitted to membership.

Edition of Shakespeare. Meanwhile, Johnson had been working on his edition of Shakespeare. Adumbrated in 1745 and formally announced in 1756, it had occupied him for much more than the two years in which he had hoped to complete it, and the long delay had provoked some typical satire from the pen of the scurrilous poet Charles Churchill:

> He for subscribers baits his hook,
> And takes their cash—but where's the book?

The book, in eight volumes, appeared at length in 1765. Johnson had been a student of Shakespeare all his life. He was no idolater, and his basic criticism is that of the moralist:

> He is so much more careful to please than to instruct, that he seems to write without any moral purpose. . . . He carries his persons indifferently through right and wrong, and at the close dismisses them without further care, and leaves their examples to operate by chance. This fault the barbarity of his age cannot extenuate; for it is always a writer's duty to make the world better.

As an editor Johnson set out, first, to correct textual corruptions; second, to elucidate obscurities of language;

and, last, in treating of Shakespeare's sources, to examine the very books that Shakespeare consulted. Of the critics' complaint of Shakespeare's neglect of the unities of time and place he made short work. The demand for these unities arose from "the supposed necessity of making the drama credible," but in Johnson's view no drama was either credible or credited. "The truth is that the spectators are always in their senses, and know, from the first act to the last, that the stage is only a stage, and that the players are only players." Johnson recognized that Shakespeare wrote his plays not for the reader at his desk but for an audience in the theatre. But Johnson himself could never appreciate the contribution made by the actor to dramatic interpretation. For him the actor was a reciter who said his piece "with just gesture and elegant modulation," and it was not in the theatre but in his study that he was most deeply moved by "the perpetual tumult of indignation, pity and hope." "He that peruses Shakespeare," he wrote about *Macbeth*, "looks round alarmed and starts to find himself alone," and he was so deeply shocked by his first reading of Cordelia's death in *King Lear* that he could not bear to read it again until he revised the play as an editor. The romantic critics of the 19th century belittled Johnson's work on Shakespeare, but the prophecy by the English critic Sir Walter Raleigh (1861–1922) that Johnson would receive more respect in the 20th century has been abundantly fulfilled by modern editors. It was in the year of the publication of his Shakespeare that Johnson received the degree of LL.D. from Trinity College, Dublin; his own university did him a similar honour ten years later.

Johnson's household. In whatever literary work Johnson was engaged his greatest terror was solitude, and the composition of his household reflects his efforts to avoid it. He had many habitations in London of which the most famous is the house in Gough Square, just north of Fleet Street, in which he lived from 1749 to 1759. When Boswell first called upon him in 1763 he was at 1, Inner Temple Lane. From there he moved back to Johnson's Court (7, Fleet Street) in 1765 and in 1776 to Bolt Court, which was his home until his death. The house in Gough Square has been well preserved and is now a Johnson museum. There the *Dictionary* was compiled; there the *Rambler*, the *Idler*, possibly *Rasselas* and much else was written; there his wife died in 1752. Even before her death and long before the grant of his pension, he had begun to make his house a refuge for the poor and unfortunate.

The house in Gough Square

Anna Williams, daughter of Zachariah Williams, who had received Johnson's help in writing his *Longitude at Sea*, came to London in the hope of being cured of a cataract but later became totally blind. She was a constant visitor in Mrs. Johnson's lifetime and, after her death, came to live in the house. For the rest of her life she had either a room in Johnson's house or lodgings near at hand. Blindness made her peevish in manner, but Johnson's regard and affection for her remained constant. When she died in 1783, he mourned a companion to whom he had had recourse for domestic amusement for 30 years. Shortly after Mrs. Johnson's death, Johnson's beloved friend Richard Bathurst presented to him Francis Barber, a Negro slave who had been freed by Bathurst's father. Johnson made Francis his friend as well as his servant, sent him to school at Bishop's Stortford, and provided for him handsomely in his will. Another humble friend was Robert Levett, "an obscure practiser in physick amongst the lower people," who was given a room at the top of the house in Johnson's Court. Johnson insisted that Levett was indebted to him for nothing more than houseroom, a share of a penny loaf at breakfast, and an occasional Sunday dinner. His death provoked one of Johnson's most moving poems:

> Well try'd through many a varying year,
> See Levett to the grave descend;
> Officious, innocent, sincere,
> Of every friendless name the friend.

Yet another beneficiary was Mrs. Desmoulins, daughter of Dr. Swynfen, Johnson's godfather, who joined the household in the early London years. After Mrs. Johnson's death she appears to have been in charge of the cooking. But she quarrelled constantly with Anna Williams, and anarchy reigned in the kitchen.

Friendship with the Thrales

It was from this domestic background that in 1765 Johnson was introduced to a family that was to provide one of the most comforting friendships of his life. Henry Thrale, owner of a Southwark brewery, had been married to Hester Lynch Salusbury in 1763. Two years later he was elected member of Parliament for Southwark, and his wife became a famous hostess and, in particular, "the provider and conductress of Dr. Johnson." For the first time in his life Johnson, who was invited not only to dine but to spend weeks, or even months, at the Thrales's country house at Streatham, was free to enjoy the luxury of solid comfort. No man, as he said, is a hypocrite in his pleasures, and at Streatham he could enjoy them all—a good library, intelligent conversation, pretty women, late hours, tasteful cookery. Furthermore, he became the confidant and counsellor of both parents and children; a room was set apart for him at the Southwark house as well as at Streatham; he was one of the family.

Political pamphlets. Already Johnson's pension had made two important differences in his way of life: he was no longer obliged to write for a living, and he could afford the time and money for holidays. Not that he was entirely idle as a writer: Shakespeare and the *Dictionary* called for revision; friends called for dedications and prologues (he wrote one for Goldsmith's *Good-Natur'd Marl)*; and he was also moved to write a series of political pamphlets. Of these tracts, the general view was, for long, that they exhibited the sad spectacle of a great man giving vent, unworthily, to the abusive expression of his conservative prejudice against the rights of the people. Today, more trouble is taken to understand Johnson's point of view. The pamphlets of the 1770s are seen not as a sudden whim of his later years but as part of the expression of a lifelong concern with political morality, which in the 1730s and 1740s had produced pamphlets vigorously denouncing the alleged tyranny and corruption of the Walpole government and, in the 1750s, a large amount of acute journalistic analysis (and condemnation) of the commercial and imperialist motives underlying the Seven Years' War.

The False Alarm (1770) was written not so much to defend the right of the House of Commons to refuse readmittance to a member already expelled as to protest against the absurdity of raising the controversy to the level of a constitutional crisis. What the country was arguing about, he said, was whether Middlesex should be represented, or not, by a criminal from jail. To raise the cry that liberty was in danger was to raise a false alarm.

Thoughts on the Late Transactions respecting Falkland's Islands (1771) was a reply to the anonymous political controversialist Junius and others, who had been urging the British government to resist the Spanish claim to those islands by aggressive action, and who now reproached it for having avoided war by a diplomatic compromise that left Britain in possession of the islands, though without fully establishing its title to them as against Spain's. Johnson gives a thorough history of the rival claims and of the diplomatic manoeuvring and then, in vigorous rhetoric, denounces the jingoism of his opponents, charging them with warmongering for commercial gain.

> These are the men who, without virtue, labour, or hazard . . . rejoice when obstinacy or ambition adds another year to slaughter and devastation; and laugh from their desks at bravery and science, while they are adding figure to figure, and cipher to cipher, hoping for a new contract from a new armament, and computing the profits of a siege or tempest.

The Patriot (1774), a short piece written just before the election of that year, defines the qualities of the true patriot and contrasts them with the behaviour of those who had unjustifiably arrogated the title to themselves. As for the supposed defenders of liberty of conscience, had they not opposed the act of Parliament that gave the Catholics of Quebec the right to practice their own religion?

Taxation No Tyranny (1775) is *a* reply to the resolutions passed by the American Continental Congress of 1774, which assert a position similar to that contained in the

Declaration of Independence in the year 1776. Here it is important to remember that throughout his life Johnson had shown little sympathy toward European colonization and colonists in general and had published bitter denunciations of the exploitation (and worse) of native populations by the Portuguese in Africa, the Spanish in Central and South America, and the "English barbarians that cultivate the southern islands of America'— the West Indies. In an "Idler" essay (81; 1759) Johnson puts into the mouth of an Indian chief a fierce tirade against the American colonists' oppression of the Indians and Negroes. By comparison, the colonists' complaints of "oppression" by the British government (in the form of taxation) seem to Johnson trivial. For mercenary reasons, they had settled on the other side of the Atlantic under English protection and with the authority of English charters. That they had no representation at Westminster was their own fault. They had the protection of English arms. Why should they not pay for it? And, if there were 3,000,000 Whigs in America "fierce for liberty," why were the loudest yelps for liberty heard among the drivers of Negroes?

Johnson's pragmatic philosophy of politics

For Johnson, these tracts were an opportunity to expound his essentially pragmatic philosophy of politics. He ascribed divine right neither to kings nor to people. Talk about liberty in the abstract or about "natural" rights he dismissed as cant. What was essential for a civilized community was a stable government and respect for its laws. But he clearly recognized that there was an ultimate safeguard. If government abused its power, mankind would not bear it; against a tyrant the people would rise and cut off his head.

The politics of the 1770s did not deter Johnson from the enjoyment of his new-found liberties. With the Thrales he spent weeks at Brighton as well as at Streatham, and his "annual midland ramble" included long visits to Oxford, to Lichfield, and to his old friend John Taylor at Ashbourne.

Journey to the western islands of Scotland. In 1773 Boswell persuaded Johnson to accompany him on a more exciting expedition. Johnson had wanted to visit the Hebrides, the western islands of Scotland, for longer than he could remember and now felt that Boswell's "gaiety of conversation and civility of manners were sufficient to counteract the inconveniences of travel." The travellers left Edinburgh on August 18 and followed the coast road. At St. Andrews Johnson's sadness at the sight of archiepiscopal ruins was mitigated by the kindness of the professors; at Aberdeen he was made a freeman of the city; dining with the governor of Ft. George, he talked learnedly about gunpowder. After Inverness, horses were substituted for the post chaise, and by the shore of Loch Ness Johnson encountered an old woman, who could speak very little English, boiling goat's flesh in a kettle. They crossed to Skye from Glenelg. There they had rain for the first time and were warned that a succession of three dry days was not to be expected for many months. In the island of Raasay they found "nothing but civility, elegance and plenty," and Johnson was delighted with the patriarchal life that he had come to see. Similarly, at Dunvegan he "tasted lotus." On the voyage to the island of Mull in the Inner Hebrides they were driven by a stormy sea into Coll, where, to Boswell's delight, Johnson strutted about one night with a broadsword and target (shield). But they were weather-bound for a week, and Johnson began to long for the mainland—"to go on with existence." It was Boswell, not Johnson, who insisted on visiting Iona, a small island southwest of Mull, but Johnson's comment is one of the most famous paragraphs in all his writings: "That man is little to be envied, whose patriotism would not gain force upon the plain of Marathon, or whose piety would not grow warmer among the ruins on Iona." On the way back to Edinburgh a visit was paid to Boswell's father at Auchinleck, where, to Boswell's distress, Johnson did not succeed in avoiding controversial topics. After a fortnight's stay in Edinburgh, Johnson returned to London and declared that his tour was the most pleasant journey he had ever made.

From many points of view the tour was a triumph. Johnson had his 64th birthday at Dunvegan, and his endurance of the physical strain of riding his pony up the rough tracks of Mam Rattachan, of sleeping in Highland huts, and of being tossed in small boats on stormy seas was in itself remarkable. Socially, with rare exceptions, the tour was a complete success, both from Johnson's point of view and from that of his hosts. For Boswell, of course, it was a glorious opportunity for him to write what was, in effect, the first installment of his *Life* of Johnson. Johnson himself was moved, for once, to write a book for its own sake. It was at Anoch in Glenmoriston, on August 31, 1773, that the travellers "entered a narrow valley not very flowery, but sufficiently verdant." While the horses grazed, Johnson sat down on a bank "such as a writer of Romance might have delighted to feign" and conceived the thought not of a romance but of an account of his journey. Of course, he did not attempt to compete with Boswell, whose diary he read, and approved, at intervals. For an intimate personal record of the "minute particulars" of Johnson's behaviour and conversation, the reader naturally turns to Boswell's ***Journal of a Tour to the Hebrides,*** published after Johnson's death. Johnson, in his own ***Journey to the Western Islands of Scotland*** **(1775),** was concerned to describe the customs, religion, education, trade, and agriculture of a society that was new to him. His narrative is far from being impersonal, but it contains no gossip. It was in his many letters to Mrs. Thrale that he wrote more freely. In one letter he offered her a definition of travel highly characteristic of his general approach to life: "The use of travelling is to regulate imagination by reality, and instead of thinking how things may be, to see them as they are."

Boswell's diary

Not long after his return, Johnson was saddened by the death of Oliver Goldsmith. They had met perhaps as early as 1760, and Johnson had been one of the first to recognize the quality of Goldsmith's writings. He had contributed a famous couplet to ***The Traveller,*** and ***She Stoops to Conquer*** had been dedicated to him. He frequently crushed Goldsmith in argument at The Club and elsewhere, but his final judgment was sincere: "Let not his frailties be remembered," he wrote, "he was a very great man."

Tours of Wales and France with the Thrales

Johnson's intimacy with the Thrales at this time is clearly illustrated by his joining them in two long tours—to North Wales in 1774 and to France in the following year. On the way to Wales he was able to introduce them to his Lichfield friends, and some time was spent in the Vale of Clwyd, where Mrs. Thrale's ancestors had lived. On the return journey Johnson was interested to see Matthew Boulton's "enginery" at Birmingham, where Boulton built steam engines in partnership with James Watt, and at Oxford the party was entertained in the hall of University College. In France he was interested more in people than in places. He admired some of the cathedrals, but his most significant comments were on social conditions, particularly on the gulf between rich and poor. There was no provision for the maintenance of the poor, and there was no comfortable middle class.

"The Lives of the English Poets." In 1777 three booksellers waited upon Johnson and asked him to write a series of ***Lives*** for an "elegant and accurate" edition of the English poets that they and others had in preparation. Johnson agreed. The choice of the poets had already been made, and he strongly objected to the volume being described as "Johnson's Poets." Only five names (and some of them rather odd names) were added at his suggestion. Though he said that he wrote his pieces "dilatorily and hastily, unwilling to work, and working with vigour and haste," it was in this work that he came nearest to actual enjoyment of writing.

The ***Lives*** are not a series but a miscellany. They follow no plan; they have no uniformity of design. On certain major poets (Milton, Dryden, Pope, and others) Johnson wrote long essays that remain as part of the stock-in-trade of English criticism. He was led on, he said, by "the honest desire of giving useful pleasure." Thus, in writing of Abraham Cowley, he seized the opportunity of a lengthy examination of the characteristics of the Metaphysical Poets, those poets who were "more desirous of being admired than understood." On the other hand, in a laconic piece of **300** words on Richard Duke, he concluded that his poems "were not below mediocrity." When he came to James Thomson (one of his own additions) he praised his ***Seasons*** highly, but of the poem ***Liberty*** he wrote: "When it first appeared, I tried to read and soon desisted. I have never tried again . . . ."

Criticism of Milton and Gray

The biographical part of literature was what Johnson loved best, and he made no attempt to separate the poetry from the man who had written it. If he disapproved of the man, he found it difficult to be a detached critic of his work. The ***Lives*** of Milton and Gray might be cited as two instances. Milton's religion and politics are scornfully exposed—he was not a member of any church and politically he was an acrimonious and surly republican. And some of this prejudice comes through in the discussion of Milton's poetry. To be sure, Johnson had high praise for ***Comus*** and "L'Allegro" and "Il Penseroso," and thought Paradise Lost one of the great "productions of the human mind." But he did not wish it longer, and insisted on pointing out major faults. In his criticism of "Lycidas" (a judgment that was to irritate the Romantics, though even this has been persuasively defended by modern critics), Johnson made a firm break with the older pastoral tradition, which he found stale and appropriate only to descriptions of country life.

Personal dislike is also evident in the ***Life*** of Gray. To the ***Elegy*** Johnson paid a famous tribute of genuine admiration, but he approached his other poems with "the neutrality of a stranger and the coldness of a critick." In fact, he was far from neutral in his harsh unfavourable criticisms of Gray's linguistic inversions and antiquarian epithets. It was Gray's attitude to life and literature that irritated Johnson. One who had spent 30 years in Grub Street could have little patience with a fastidious scholar who asked for leisure to be good and wrote only when he was in the humour.

LAST YEARS

Publication of the ***Lives*** was completed in **1781,** and a storm of criticism broke out. But Johnson was too old a campaigner to be disturbed by criticism, and, in fact, the ***Lives*** remain the best loved, or the least neglected, of Johnson's works. What distressed him far more that year than unfavourable reviews was the death of Henry Thrale. No loss since that of his wife had so much oppressed him; he felt like a man beginning a new course of life. To some extent he was saved from brooding over his loss by his attention to his duties as executor. Always interested in trade and commerce, he had enjoyed discussing brewery policy with Thrale in his lifetime, and now he was dealing not with parcels of boilers and vats but with "the potentiality of growing rich beyond the dreams of avarice."

Death of Henry Thrale

To his disappointment, but to Mrs. Thrale's delight, the brewery was sold to "a knot of rich Quakers" for £135,000. For a time Johnson continued to regard Streatham as his home, but in August **1782** Mrs. Thrale told him of her decision to sell the house. On his last visit in October he humbly thanked God for the comforts he had enjoyed there. A room was allotted to him in the house that Mrs. Thrale took in Argyll Street, but the intimacy of the Streatham days was weakening. For long a victim of asthma and dropsy, Johnson had a paralytic stroke in June **1783,** from which he did not properly recover until the spring of the following year. Mrs. Thrale wrote to him with "the attention and tenderness of ancient time," but she dared not tell him that she had for long been trying to decide whether or not to marry Gabriel Piozzi, an Italian musician. When, at the end of June **1784,** Johnson heard of her decision to marry, he wrote a letter of angry condemnation; but a week later he recognized that Mrs. Thrale's marriage (which was, in fact, a happy one) was her own affair and thanked her for the kindness that had soothed **20** years of a life radically wretched.

Johnson was now bereft of many old friends. He was a sick man, but to the end he fought untiringly against his worst enemy—solitude—and the melancholy that it induced. In December **1783** he had formed yet another club in Essex Street, which was conveniently near his house in Bolt Court, and in the following July he set out upon his last Midland ramble. At Lichfield everyone was pleased to see him; he stayed for two months at Ashbourne, where he was comfortable but "hungry for conversation," for John Taylor, his friend from school days, went to bed at nine; at Birmingham he talked over old times with Edmund Hector; at Oxford he was welcomed to his old college by the master. On November 16 he returned to London. Though his mind remained alert, his bodily state grew worse, and he died on December **13, 1784. A** week later he was buried in Westminster Abbey. One journalist noted that there was only one man of hereditary title among the mourners; but he added, rightly, that he who was followed by Reynolds and Burke did not go unhonoured to the grave.

REPUTATION AND CHARACTER

The history of Johnson's fame is curious. In his own day he was acknowledged as an outstanding writer and thinker, and his fame still persists. Yet today his reputation rests on a dual tradition. There is the "folk image," based largely on Boswell's accounts, and perpetuated by Macaulay, which stresses Johnson the witty talker, who often takes the wrong side of an argument. This is the colourful eccentric, the bear with the heart of gold, whom everybody quotes but nobody reads, remembered chiefly as a character in a great book. Then there is Johnson the man of letters, so admired in his own day and now, after a long eclipse, once again dominating the scene. This second tradition does not rule out Johnson the talker, so admirably described by Boswell, but it concentrates on the Great Cham's own writings.

Johnson the moralist

In scores of recent books the emphasis is not on Johnson's amusing foibles and forceful remarks but rather on his vigorous reasoning intelligence, his keen understanding of human frailty, his detestation of hypocrisy, and his practice of applying ethical standards to nations as well as to individuals. Many of Johnson's so-called obtuse prejudices, when properly understood, derive from his insistence on making strict moral judgments. Johnson was not a sentimental Tory reactionary. He violently opposed war and pleaded for more humane treatment of prisoners of war. He attacked censorship in the **1737** licensing act; he argued for mitigation of laws against prostitutes and debtors and the death penalty for forgery. He was the consistent champion of the poor and oppressed and argued for the rights of blacks and other indigenous peoples. Yet at the same time he saw the necessity of a strong central government in order to withstand the pressures of wealthy mercantile interests.

Moreover, it is now evident that Johnson was not, as was formerly assumed, a strictly neoclassical literary critic; rather, his approach tended to be empirical. He rejected any slavish following of rules and insisted on judging each work separately, making his own decision as to its merits and defects. Johnson is now seen to be the father of 20th-century New Criticism.

Although he scorned those optimists who had a naïve faith in the perfectibility of human institutions, Johnson himself was not a complete pessimist. Along with his skeptical doubts, he had an enormous zest for living. Participation and struggle were always necessary, and while there were some areas where improvement might be possible, everything had to have a strong ethical base. Johnson was not only a moralist but a Christian moralist. The sinfulness of man and his need of redemption by the Passion of Jesus Christ were the basis of his personal faith. His constitutional melancholy deprived him of a feeling of joy in his religion, and the sin of which he was most deeply conscious was idleness. In his ***Prayers and Meditations*** **(1785),** published after his death, his repentance and his good resolutions, constantly repeated but seldom kept, show the sincerity of his heart-searching and his humility. But his faith prevailed, and in his last days he refused to take opiates because he had prayed that he might render his soul to God unclouded.

Born in a bookshop, Johnson had the qualities, and the conscience, of a scholar. By the writing of books he strove to earn his daily bread; by the reading of books he sought to enlarge the range of his ideas and of his scholarship. And what, he asked in later years, should books teach but the art of living? Few men have left finer examples of the art of living than Samuel Johnson.

MAJOR WORKS

VERSE: *London: A Poem* (1738); *The Vanity of Human Wishes* (1749); *Irerze: a Tragedy* (1749).

PROSE: *Marmor Norfolciense* (1739), satirical essay; *A Compleat Virtdication of the Licensers of the Stage* (1739), ironical defense of the suppression of Henry Brooke's play *Gustavus Vasa; An Account of the Life of Mr Richard Savage* (1744); *Miscellaneous Observations on the Tragedy of Macbeth* (1745); *The Plan of a Dictionary of the English Language* (1747); *The Rambler* (1750–52); *A Dictionary of the English Language*, 2 vol. (1755); *The Prince of Abissinia*, 2 vol. (1759), better known as *Rasselas*; *The Idler* (1758–60); *The Plays of William Shakespeare*, 8 vol. (1765); *The False Alarm* (1770); *Thoughts on the Late Transactions respecting Falkland's Islands* (1771); *The Patriot* (1774); and *Taxation No Tyranrzy* (1775), four political tracts; *A Journey to the Western Islands of Scotland* (1775); *Prefaces, Biographical and Critical, to the Works of the English Poets,* 10 vol. (1779–81, rev. as *The Lives of the most eminent English Poets, 4* vol., 1781); *Prayers and Meditations, composed by Samuel Johnson, LL.D.* (1785), published posthumously by George Strahan.

BIBLIOGRAPHY. The standard bibliography is that by W.P. COURTNEY and D. NICHOL SMITH (1915), reissued in 1925 and supplemented by R.W. CHAPMAN and A.T. HAZEN in the *Proceedings of the Oxford Bibliographial Society* (1939). *The R.B. Adam Library Relating to Dr. Samuel Johnson and His Era,* 4 vol. (1929–30), contains facsimiles of many manuscripts and letters. The most extensive Johnson collection is now in the Hyde Library, near Somerville, New Jersey, described in *Four Oaks Library,* 2 vol. (1967). J.L. CLIFFORD and D.J. GREENE in *Samuel Johnson: A Survey and Bibliography of Critical Studies* (1970), list the most important works on Johnson from his day through 1968.

Collected editions: The Oxford edition, 9 vol. (1825), with two supplementary volumes of *Debates,* has long been the most cited, but its text is deficient by modern standards. The new Yale edition under the editorship of A.T. HAZEN and J.H. MIDDENDORF (1958–) now undertakes to supply a more satisfactory text. *Diaries, Prayers, and Annals,* ed. by E.L. MCADAM, JR., with DONALD and MARY HYDE •(1958); *The Idler and The Adventurer*, ed. by W.J. BATE, J.M. BULLITT, and L.F. POWELL (1963); *Poems*, ed. by E.L. MCADAM, JR., with GFORGE MILNE (1964); *Johnson on Shakespeare*, ed. by ARTHUR SHERBO, 2 vol. (1968); *The Rambler,* ed. by W.J. BATE and A.B. STRAUSS, 3 vol. (1969); *A Journey to the Western Islands*, ed. by MARY LASCELLES (1971). Until superseded, the standard edition of *The Lives of the English Poets* is that by G.B. HILL, 3 vol. (1905, reprinted 1967); and of the correspondence, *The Letters of Samuel Johnson,* ed. by R.W. CHAPMAN, 3 vol. (1952). The most extensive anthology is *Johnson, Prose and Poetry,* selected by MONA WILSON (1950).

Biography and criticism: The number of biographical and critical studies of Johnson is enormous. Apart from Boswell's *Journal of a Tour to the Hebrides* (1785) and *Life of Johrzson* (1791), there are a number of contemporary authorities, including the official *Life* by SIR JOHN HAWKINS (1787; abridged ed. by B.H. DAVIS, 1961); HESTER LYNCH PIOZZI, *Anecdotes* (1786; ed. by S.C. ROBERTS, 1925); *Lives* by ARTHUR MURPHY (1792) and by ROBERT ANDERSON (1795; 3rd expanded ed., 1815). Portions of these and many other anecdotes are included in G.B. HILL (ed.), *Johnsonian Miscellanies,* 2 vol. (1897, reprinted 1966). Two of the most famous essays on Johnson were those of MACAULAY: one a review of J.W. Croker's edition of Boswell *(Edinburgh Review,* 1831), in which the depreciation of Johnson as a writer set the fashion for many years; the other, a more balanced account, in the *Encyclopædia Britannica,* 8th ed. (1856). Among shorter Victorian studies there are *Lives* by LESLIE STEPHEN (1878) and FRANCIS R. GRANT (1887).

In the 20th century much new biographical evidence has been assembled by A.L. READE in *Johnsonian Gleanings,* 11 vol. (1909–52); A.T. HAZEN, *Samuel Johnson's Prefaces and Dedications* (1937); E.L. MCADAM, JR., *Dr. Johnson and the English Law* (1951); B.B. HOOVER, *Samuel Johnson's Parliamentary Reporting* (1953); and in the various volumes of the *Boswell Papers* (ed. by F.A. POTTLE *et al.),* particularly *The Correspondence and Other Papers of James Boswell Relating to the Making of the Life of Johnson,* ed. by MARSHALL WAINGROW (1969). Much of this new evidence is used in J.W. KRUTCH, *Samuel Johnson* (1944); J.L. CLIFFORD, *Young Sam Johnson* (1955); PAUL FUSSELL, *Samuel Johnson and the Life of Writing* (1971); CHRISTOPHER HIBBERT, *The Personal History of Samuel Johnson* (1971); and other critical works.

Some special studies are: WALTER RALEIGH, *Six Essays on Johnson* (1910, reprinted 1965), an early revaluation correcting Macaulay's view; B.H. BRONSON, *Johnson Agonistes, and Other Essays* (1946); W.K. WIMSATT, JR., *The Prose Style of Samuel Johnson* (1941); J.H. HAGSTRUM, *Samuel Johnson's Literary Criticism* (1952); J.H. SLEDD and G.J. KOLB, *Dr. Johnson's Dictionary* (1955); ARTHUR SHERBO, *Samuel Johnson, Editor of Shakespeare* (1956); EDWARD A. BLOOM, *Samuel Johnson in Grub Street* (1957); D.J. GREENE, *The Politics of Samuel Johnson* (1960); M.J. QUINLAN, *Samuel Johnson: A Layman's Religion* (1964); C.F. CHAPIN, *The Religious Thought of Samuel Johnson* (1968); R.B. SCHWARTZ, *Samuel Johnson and the New Science* (1971). There have also been various collections of shorter articles: F.W. HILLES (ed.), *New Light on Dr. Johnson* (1959); *Johnson, Boswell and Their Circle: Essays Presented to L.F. Powell* (1965); D.J. GREENE (ed.), *Samuel Johnson: A Collection of Critical Essays* (1965); and J.L. CLIFFORD (ed.), *Twentieth Century Interpretations of Boswell's Life of Johnson* (1970).

General evaluations that stress Johnson the moralist, thinker, and critic are: W.J. BATE, *The Achievement of Samuel Johnson* (1955); R.B. VOITLE, JR., *Samuel Johnson the Moralist* (1961); P.K. ALKON, *Samuel Johnson and Moral Discipline* (1967); ARIEH SACHS, *Passionate Intelligence: Imagination attd Reason in the Works of Samuel Johnson* (1967). D.J. GREENE, *Samuel Johnson* (1970), is an excellent introduction and analysis embodying the new approach.

(S.Rs./J.L.Cl.)

Joint

In anatomy, a joint is a structure that separates two or more adjacent elements of the skeleton. Depending on the type of joint, such separated elements may or may not move on one another. Though this article is intended to deal with human joints, its content is applicable as well to those of vertebrates in general and mammals in particular.

The Latin stem word for joint is art— (as in articular, articulation); the Greek stem word is *arthro— (as* in arthritis, arthrology). The international vocabulary of arthrology (the science of joints) uses Latin and Greek terms, but the more important of these have well-known English equivalents that will be used here when possible.

Classification of joints

Joints can be classified in two ways: temporally and structurally. Each classification is associated with joint function.

Temporal classification. Joints are either transient or permanent. The bones of a transient joint fuse together sooner or later, but always after birth. All the joints of the skull, for example, are transient, excepting those of the middle ear and those between the lower jaw and the braincase. The bones of a permanent joint do not fuse except as the result of disease or surgery. Such pathological or surgical fusion is called arthrodesis.

All permanent and some transient joints permit movement. Movement of the latter may be temporary, as with the roofbones of an infant's skull during birth, or long-term, as with the joints of the base of the skull during postnatal development.

Structural classification. There are two basic structural types of joint: diarthrosis, in which fluid is present, and synarthrosis, in which there is no fluid. All the diarthroses (commonly called synovial joints) are permanent. Some of the synarthroses are transient; others are permanent.

Before proceeding further with joint classification, it is necessary to describe the types of motion made possible by mobile joints. These motions include spinning, swinging, gliding, rolling, and approximation; they can be briefly described as follows: Spin is a movement of a bone around its own long axis denoted by the term rotation in anatomy. Swing, or angular movement, brings about a change in the angle between the long axis of the moving bone and some reference line in the fixed bone. Approximation denotes the movement caused by pressing or pulling one bone directly toward another—*i.e.*, by a "translation" in the physical sense. The reverse of approximation is separation. Gliding and rolling movements occur only within synovial joints and cause a moving bone to swing.

Flexion (bending) and extension (straightening) of the elbow are examples of swing. A swing (to the right or left) of one bone away from another is called abduction; the reverse, adduction. An important example of spin is provided by the radius ("outer" bone) of the forearm; this bone can spin upon the lower end of the humerus (upper arm) in all positions of the elbow. When an individual presses the back of his hand against his mouth, his forearm is pronated, or twisted; when the palm of his hand is pressed against his mouth, his forearm is supinated, or untwisted. Pronation is caused by medial (inward) rotation of the radius, and supination by lateral (outward) rotation. Approximation and separation normally occur only at the symphyses and gomphoses, joints that will be discussed later.

All of the most common movements of bones involve some degree of spin. Pure swings, which are not accompanied by spin, are most often seen in military drill and gymnastic exercises. This topic is dealt with relative to conjunct rotation at synovial joints.

SYNARTHROSES

Classes of synarthroses

Synarthroses are divided into three classes: fibrous, cartilaginous, and fibrocartilaginous; the last class is called symphysis. The tissues that make up these joints are described in the article CONNECTIVE TISSUE, HUMAN; and the composition of bone, in the article BONE.

Fibrous joints. In fibrous joints the articulating parts are separated by white connective tissue (collagen) fibres, which pass from one part to the other. There are two types of fibrous joints: suture and gomphosis.

A suture is formed by the fibrous covering, or periosteum, of two bones passing between them. In the adult, sutures are found only in the roof and sides of the braincase and in the upper part of the face. In the infant, however, the two halves of the frontal bone are separated by a suture (the metopic suture), as are the two halves of the mandible at the chin. Excepting those of the fetus and newborn infant, all sutures are narrow. In the late fetus and newborn child, the sagittal suture, which separates the right and left halves of the roof of the skull, is quite wide and markedly so at its anterior and posterior ends. This enables one of the halves to glide over the other during the passage of the child through the mother's pelvis during birth, thus reducing the width of its skull, a process called molding. (The effects of molding usually disappear quickly.) After birth all sutures become immobile joints. The expanded anterior and posterior ends of the sagittal suture are called fontanels; they lie immediately above a large blood channel (superior sagittal sinus). The anterior fontanel provides an easy access for obtaining a blood sample from the very young infant.

From Cunningham's Textbook of Anatomy edited by G.J. Romanes and published by Oxford University Press, as an Oxford Medical Publication

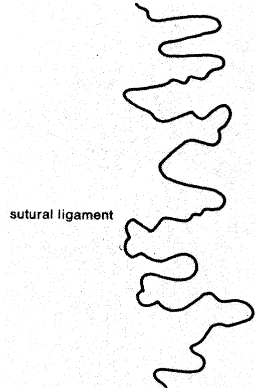


Figure 1: Surface view of portion of sagittal suture.

Sutures are transient, for they are unossified parts of the skeleton that become fused at various dates from childhood to old age. The fusion is effected by direct conversion of the sutures into bone. Until maturity the sutures are active sites of growth of the bones they separate.

A gomphosis is a fibrous, mobile peg-and-socket joint. The roots of the teeth (the pegs) fit into their sockets in the mandible and maxilla and are the only examples of this type of joint. Bundles of collagen fibres pass from the wall of the socket to the root; they are part of the circumdental, or periodontal, membrane. There is just enough space between the root and its socket to permit the root to be pressed a little further into the socket during biting or chewing. Comphoses are permanent joints, in the sense that they last as long as do the roots of the teeth involved in them, unless, of course, they are obliterated by disease.

The movement of the root within a gomphosis has a threefold effect. It lessens some of the impact between the upper and lower teeth in biting; it pumps blood and lymph from the periodontal membrane into the dental veins and lymph channels; and it stimulates sensory nerve terminals in that membrane to send signals to the brain centres controlling the muscles of mastication.

Fibrocartilaginous joints; symphysis. A symphysis is a joint in which the body (physis) of one bone meets the body of another. All but two of the symphyses lie in the vertebral (spinal) column, and all but one contain fibrocartilage as a constituent tissue. The short-lived suture between the two halves of the mandible is called the symphysis menti (mentum, meaning chin) and is the only symphysis devoid of fibrocartilage. All the other symphyses are permanent.

The symphysis pubis joins the bodies of the two pubic bones of the pelvis. The adjacent sides of these bodies are covered by cartilage through which collagen fibres run from one pubis to the other. On their way they traverse a plate of cartilage, which, in some instances (especially in the female), may contain a small cavity filled with fluid. Surrounding the joint and attached to the bones is a coat of fibrous tissue, particularly thick below (the subpubic ligament). The joint is flexible enough to act as a hinge that allows each of the two hipbones to swing a little upward and outward, as the ribs do during inspiration of air. This slight movement is increased in a woman during childbirth because of the infiltration of the joint and its fibrous coat by fluid toward the end of pregnancy, the fluid making the joint even more flexible. In both sexes the joint acts as a buffer against the shocks transmitted to the pelvic bones from the legs in running and jumping.

Intervertebral disks

The symphysis between the bodies of two adjacent vertebrae is called an intervertebral disk. It is composed of two parts: a soft centre (nucleus pulposus) and a tough, flexible ring (anulus fibrosus) around it. The centre is a jellylike (mucoid) material containing a few cells derived from the precursor of the spine (notochord) of the embryo. The ring consists of collagen fibres arranged in concentric layers like those of an onion bulb. These fibres reach the adjacent parts of the vertebral bodies and are attached firmly to them. The mass of fibres is embedded in chondrousia—the mixture of ground substances characteristic of cartilage—but yet the ring, although called fibrocartilaginous, is predominantly fibrous, and this fact determines its mechanical properties.

There are 23 intervertebral disks, one found between each pair of vertebrae below the first cervical, or neck (atlas), and above the second sacral (just above the tailbone). The lumbar (lower back) disks are thickest, the thoracic (chest or upper back) thinnest, and the cervical are of intermediate size. These differences are associated with the function of the disks.

In general these disks have two functions: to allow movement between pairs of vertebrae and to act as buffers against the shocks caused by running, jumping, and other stresses applied to the spine. Each of these functions merits separate consideration.

If an intervertebral disk were the only joint between a pair of vertebrae, then one of these could move on the other in any direction; but each pair of vertebrae with an intervertebral disk has also a pair of synovial joints, one on each side of the vertebral (neural) arch. These joints limit the kinds of independent movement possible so that the thoracic vertebrae move in only two directions, and the lumbar in only three; only the cervical vertebrae below the atlas have full freedom of movement.

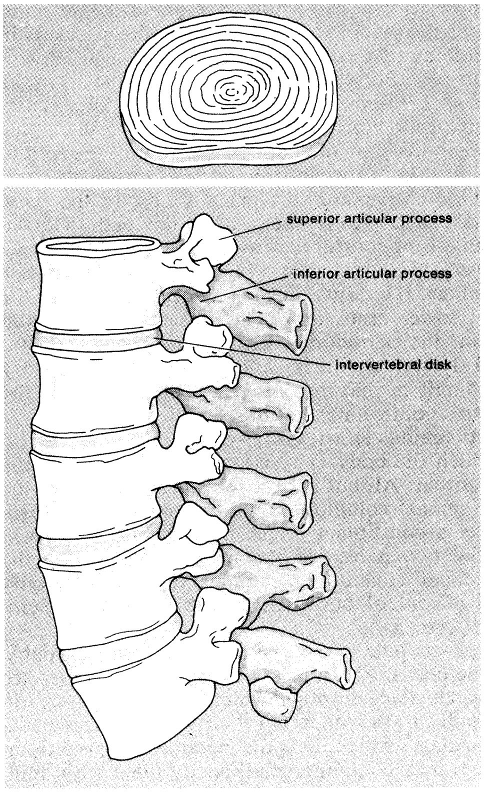


Figure 2: (Top) Cross section of an intervertebral disk. (Bottom) The lumbar vertebrae (left-side view).
Adapted, with permission, from *The CIBA Collection of Medical Illustrations* by Frank H. Netter, M.D.

All intervertebral disks allow approximation and separation of their adjacent vertebrae. This is caused partly by movement brought about by muscle action and partly by the weight of the head and the trunk transmitted to the pelvis when a person is upright. The effect of weight is of special importance. The mucoid substance in the centre of the disk behaves like a fluid. It is acted upon by the person's weight and any other pressure forces transmitted along the spine. Hence the disk flattens from above downward and expands in all other directions. After arising in the morning, and as the day progresses, a person decreases in height because of this compression of the disks. An average decrease of one millimetre in the height of each disk would mean an overall shortening of 2.3 centimetres, or about an inch. The spine lengthens again, of course, during sleep.

In the infant the greater part of the disk consists of the soft centre. Later the fibrous ring becomes relatively thicker in such a way that the soft part is nearer to the back of the disk. As middle age approaches, there is an increase in the fibrous element, the soft centre is reduced in size, and the amount of cartilage is increased. There is a tendency for the posterior part of the fibrous ring to degenerate, so that a sudden violent pressure may rupture the disk and allow the central part to protrude backward against the spinal cord. This condition is commonly referred to as slipped disk.

Cartilaginous joints. These joints, also called synchondroses, are really the unossified masses found between bones or parts of bones that pass through a cartilaginous stage before ossification. Examples are the synchondroses between the occipital and sphenoid bones and the sphenoid and ethmoid bones of the floor of the skull. As already stated, these permit growth of the adjacent bones and act as virtual hinges at which the ethmoid and occipital bones swing upward upon the sphenoid; this allows backward growth of the nose and jaws during postnatal life. The juxta-epiphyseal plates separating the ossifying parts of a bone are also an example. Growth of the whole bone takes place at these plates when they appear, usually after birth. All synchondroses are transient; and all normally have vanished by the age of 25.

DIARTHROSES: SYNOVIAL JOINTS

Structure and elements of synovial joints. A synovial joint is nothing but a bursa (a closed sac lined with a membrane and containing fluid; see BURSA) at which movement of one or more bones occurs on one or more others. Two or more parts of the bursal wall become cartilage (chondrify) during prenatal life. These are the parts of the bursa that are attached to the articulating bones, and they constitute the articular cartilage of the bones.

The bursal fluid is called synovia, whence the common name for this class of joints. Bursas are found between structures that glide upon each other, and all motion at diarthroses entails some gliding, the amount varying from one joint to another.

General structure of synovial joints

A synovial joint consists of a wall enclosing a joint cavity that is wholly filled with synovial fluid. The wall consists of two layers: an outer, complete, fibrous layer and an inner, incomplete, synovial layer (Figure 3). Parts of the outer layer are either chondrified as articular cartilages or partly ossified as sesamoid bones (small, flat bones developed in tendons that move over bony surfaces). Parts of the synovial layer project into the cavity to form fatty pads. In a few diarthroses the fibrous layer also projects inward to become intra-articular disks, or menisci. These various structures will be discussed in connection with the layer to which they belong.

The bones appertaining to a synovial joint are, strictly speaking, outside the joint because it is a complete bursa. Nevertheless, those parts of them that can be seen only after the bursal wall has been opened are usually said to be inside the joint. This custom will be followed here whenever it makes for clarity.

From *Cunningham's Textbook of Anatomy* edited by G.J. Romanes and published by Oxford University Press, as an Oxford Medical Publication

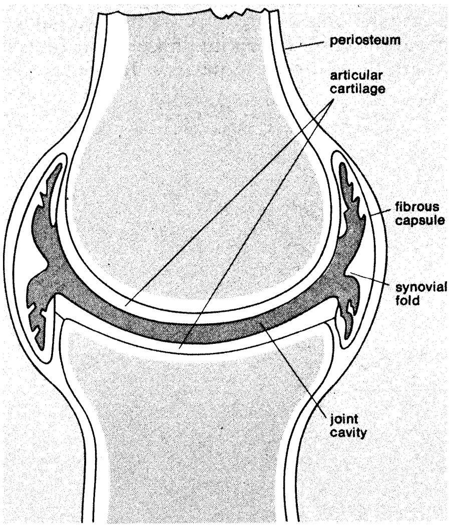


Figure 3: A synovial joint.

The fibrous layer. The fibrous layer is comprised of two continuous sections—one manifest and one hidden. The manifest part is that which is visible in an unopened joint cavity and is referred to as the investing ligament. At the point where it reaches the articulating bones, it covers their articulating ends as a complete periosteum. This is the hidden part. It is made of white fibrous tissue, composed of collagen fibres along with the cells that produce and maintain them.

Articular cartilage. Where the periosteal part of the fibrous layer covers the articulating part of a bone it becomes permeated with cartilaginous material, and the collagen fibres form a spongework in the meshes of which lie cartilage cells, thus forming articular cartilage. The continuity of the periosteal collagen with that of the articular cartilage can be seen by means of polarized light. Although the articular cartilages of a joint do complete the wall of the articular bursa, they are mechanically and visually distinguishable and merit the separate descriptions they usually receive.

Physical characteristics of articular cartilage

Articular cartilage is of the type called hyaline (glass-like) because thin sections of it are translucent, even transparent, to light. Unlike bone, it is easily cut by a sharp knife. It is deformable but elastic, so that it recovers its shape quickly when the deforming stress is removed. These properties are important for its function.

The surface of articular cartilage is smooth to the finger, like that of a billiard ball, and the cartilage's coefficient of friction is low indeed—0.02 or less. The electron scanning microscope has recently shown, however, that the surface is actually irregular, more like that of a golf ball. The part of the cartilage nearest to the bone is impregnated with calcium salts. This calcified layer appears to be a barrier to the passage of oxygen and nutrients to the cartilage from the bone, so that the cartilage is largely dependent upon the synovial fluid for its nourishment.

Every articular cartilage is divisible into two parts: a central, articulating part and a marginal, nonarticulating part. The marginal part is much smaller than the central and is covered by a synovial membrane. It will be described later in connection with that membrane.

The central part is either single, if only two bones are included in the joint, or divided into clearly distinct portions by sharp ridges, if more than two bones are included. Thus the upper articular surface of the arm bone (humerus) is single, for only this bone and the shoulder blade (scapula) are included in the shoulder joint. The lower articular surface of the humerus is subdivided into two parts, one for articulation with the radius and one for articulation with the ulna, both being included in the elbow joint. There is a functional reason for the subdivision, or partition, of articular cartilage when it does occur.

Within a diarthrosis joint bones articulate in pairs, each pair being distinguished by its own pair of conarticular surfaces. Conarticular surfaces constitute "mating pairs." Each mating pair consists of a "male" surface and a "female" surface, terms explained below. As has just been pointed out, there is only one such pair of bones within the shoulder joint; hence there is only one pair of conarticular surfaces. There are two such pairs within the elbow joint—the humeroradial and humero-ulnar. The radius moves on one of the two subdivisions of the lower humeral articular cartilage, the ulna moves on the other subdivision. There are, then, two pairs of conarticular surfaces within the elbow joint even though there are only three bones in it.

Ovoid and sellar classes of articular surfaces

Articular surfaces are divisible into two primary classes: ovoid and sellar. An ovoid surface is either convex in all directions or concave in all directions; in this respect it is like one or other of the two sides of a piece of eggshell, hence the name (ovum, egg). A sellar surface is convex in one direction and concave in the direction at right angles to the first; in this respect it is like the whole or part of a horse saddle (sella, saddle). There are no flat articular surfaces, although custom permits slightly curved ovoid or sellar surfaces to be called flat. Following an engineering convention, an ovoid surface is called male if it is convex, female if it is concave. In any diarthrosis having ovoid conarticular surfaces, the male surface is always of larger area than the female. For this reason the larger of two sellar conarticular surfaces is called male, the smaller is called female. The larger the difference in size between conarticular surfaces the greater is the possible amount of motion at the joint.

In all positions of a diarthrosis, except one, the conarticular surfaces fit imperfectly. This incongruence may not be great and may be lessened by mutual deformation of the opposed parts of the surfaces, a consequence of the deformability of articular cartilage. The exceptional position is called the close-packed position; in it the whole of the articulating portion of the female surface is in complete contact with the apposed part of the male surface, and the joint functionally is no longer a diarthrosis but synchondrosis. Every joint has its close-packed position brought about by the action of the main ligaments of the joint. A good example is that of the wrist when the hand is fully bent backward (dorsiflexed) on the forearm; another is that of the knee when the thigh and leg are in the military position of attention. It is not assumed constantly because it requires special muscular effort. It is also a dangerous position, for in it two bones in series are converted temporarily into a functionally single, but longer, unit, more likely to be injured by sudden, torsional stresses. Thus a sprained, or even fractured, ankle usually occurs when that joint, when close packed, is suddenly and violently bent.

No articular surface is of uniform curvature; neither is it a "surface of revolution" such as a cylinder is. That part of a male conarticular surface that comes into contact with the female in close pack is both wider and of lesser curvature than is the remainder. Inspection of two articulating bones is enough to establish their position of close pack, flexion, extension, or whatever it may be.

Intra-articular fibrocartilages. Intra-articular fibrocartilages are complete or incomplete plates of fibrocartilage that are attached to the manifest part of the fibrous layer (the investing ligament) and that stretch across the joint cavity between a pair of conarticular surfaces. When complete they are called disks; when incomplete they are called menisci. Disks are found in the temporomandibular joint of the lower jaw, sternoclavicular (breastbone and collarbone), and ulnocarpal (inner forearm bone and wrist) joints. A pair of menisci is found in each knee joint, one between each femoral condyle and its female, tibial counterpart. A small meniscus is found in the upper part of the acromioclavicular joint. These fibrocartilages are really parts of the fibrous layer of the diarthrosis in which they occur and effect a complete or partial division of the articular bursa into two parts, depending upon whether they are disks or menisci, respectively. When the division is complete, there are really two synovial joints—*e.g.*, the sternodiskal and discoclavicular.

A disk or meniscus is a chondrified ingrowth of the fibrous layer, the chondrification being slight and the fibrous element predominating, especially in the part nearest to the investing ligament. Both animal experiment and surgical experience have shown that a meniscus of the knee can be grown again if removed. The growth takes place from the fibrous layer and is wholly fibrous in structure. The function of these intra-articular plates is to assist the gliding movements of the bones at the joints that contain them.

The synovial layer. The inner layer of the articular bursal wall is called the synovial layer (stratum synoviale) because it is in contact with the synovial fluid. Unlike the fibrous layer, it is incomplete and does not extend over the articulating parts of the articular cartilages and the central parts of articular disks and menisci. One can, therefore, designate a "synovial periosteum," covering such parts of the bones as are found within the joint, and a "synovial perichondrium," covering the nonarticulating parts of the articular cartilages.

Synovial membrane cells

The layer, commonly called the synovial membrane, is itself divisible into two strata, the intima and the subintima. The intima is smooth, moist, and glistening on its free (synovial) surface. It could be described as an elastic plastic in which cells are embedded. Its elasticity allows it to stretch when one of the articulating bones either spins or swings to the opposite side and to return to its original size when the movement of the bone is reversed. It contains many elastic fibres. These are extremely fine but are easily seen when the membrane is stained by methylene blue. The cells of a synovial membrane can be divided into two classes: synovioblasts and protective cells. The synovioblasts are responsible for the generation and maintenance of the matrix. Their form depends upon their location. They are flattened and rounded at or near the internal surface of the membrane, more elongated and spindle shaped elsewhere. They appear to be quite mobile, able to make their way to the free surface of the membrane and even to enter the synovial fluid. Excepting the regions in which the synovial membrane passes from the investing ligament (fibrous capsule) to the synovial periostea, these cells are scattered and do not form a continuous surface layer as do, for example, the cells lining the inner surface of the gut or of a blood vessel. In this respect they resemble the cells of other connective tissues such as bone and cartilage. Apart from the generation and maintenance of the matrix of the membrane, their function is obscure. They seem to be the only cells capable of secreting hyaluronic acid, the characteristic component of synovial fluid, but direct evidence that they do so is still lacking.

The protective cells are scattered through the membrane. They are of two kinds, mast cells and phagocytes. The mast cells secrete heparin and play the same part in synovial membrane as they do elsewhere, for example in the skin and the gums. The phagocytes ("swallowing cells") ingest unwanted particles, even such large ones as those of injected India ink; they are, in short, scavengers here as elsewhere. The subintima is the connective tissue base on which the intima lies and may be fibrous, fatty, or areolar (loose). In it are found the blood vessels and nerves that have penetrated the fibrous layer. Both the blood vessels and the nerves form plexuses, to be described later. The areolar subintima forms folds (synovial fringes) or minute, fingerlike projections (villi) that project into the synovial fluid. The villi become more abundant in middle and old age. The fatty parts of the subintima may be quite thin; but in all joints there are places where they project into the bursal cavity as fatty pads (plicae adiposae); these are wedge shaped in section, like a meniscus, the base of the wedge being against the fibrous capsule. The fatty pads are large in the elbow, knee, and ankle joints. In the hip joint a fatty pad is found in the deepest part of the female surface (acetabulum). It is usually, but not always, connected to the male surface (head of femur) and is then called the ligamentum teres.

Functions of the fatty pads

The function of fatty pads depends upon the fact that fat is liquid in a living body and that, therefore, a mass of fat cells is easily dcformable. When a joint is moved, the synovial fluid is thrown into motion because it is adhesive to the articular cartilages, the motion of the fluid being in the direction of motion of the moving part. The fatty pads project into those parts of the synovial space in which there would be a likelihood of an eddying (vortical) motion of the fluid if those parts were filled with fluid. In short, the pads contribute to the "internal streamlining" of the joint cavity. Their deformability enables them to do this effectively. Of equal importance is the fact that the fatty pads by their very presence keep the synovial fluid between the immediately neighbouring parts of the male and female surfaces sufficiently thin, with proper elasticity as well as viscosity, to lubricate the joint.

Fatty pads are well provided with elastic fibres that bring about recovery from the deformation caused by pressure across a moving joint and prevent the pads themselves from being squeezed between two conarticular surfaces at rest. Such squeezing can happen, however, as the result of an accident and is very painful because of the large number of pain nerve fibres in these pads.

Three characteristics of the synovial fluid

The synovial fluid. The main features of synovial fluid are: (1) Chemically, it is dialyzate of blood plasma—that is, the portion of the plasma that has "filtered" through a membrane—but contains a larger amount of hyaluronic acid than other plasma dialyzates. (2) Physically, it is a markedly thixotropic fluid; that is, one which is both viscous and elastic. Its viscosity decreases with an increase of the speed of the fluid when it is in motion. Its elasticity, on the other hand, increases with increase of speed of the fluid. Its thixotropy is due to the hyaluronic acid in it. (3) Functionally, it has two parts to play: nutrition and lubrication. It has now been established that it alone, by virtue of its being a blood-plasma dialyzate, can nourish the articulating parts of the articular cartilages. Its thixotropic properties make it suitable for forming what are called elastohydrodynamic lubricant films between the moving and the fixed conarticular surfaces of any mating pair. The motion of the synovial fluid, referred to earlier in connection with the fatty pads, assists its nutritional function by distributing it over the articular surfaces, from which it slowly passes into the interior of the cartilage. The bulk of the fluid comes from the synovial blood capillaries. The source of the hyaluronic acid is not known at present.

Types of synovial joints. Recognition of the bursal nature of synovial joints makes it possible to describe them simply in terms of the bursal wall and to group together a number of types of structures. There are seven types of synovial joints: plane, hinge, pivot, sellar, ellipsoid, spheroidal, and bicondylar (two articulating surfaces). This classification is based on the anatomical form of the articular surfaces.

Plane type or arthrodial. The plane-type or arthrodial joint has mating surfaces slightly curved and may be either ovoid or sellar. Only a small amount of gliding movement is found. Examples are the joints between the metacarpal bones of the hand and those between the cuneiform bones of the foot.

Hinge type or ginglymus. The hinge-type or ginglymus joint is a modified sellar type with each mating surface ovoid on its right and left sides. This modification reduces movement to a backward–forward swing like that allowed by the hinge of a box or a door. The swing of the joint, however, differs from that of a hinge in that it is accompanied by a slight spin of the moving bone around its long axis. This brings the joint either into or out of its close-packed position, which is always that of extension.

The joints between the bones of the fingers (phalanges) and that between the ulna ("inner" bone of the forearm) and the humerus at the elbow are classical examples. When a finger is completely straightened, each of its interphalangeal joints is in close pack, as is the ulnohumeral joint when the arm and forearm are in one line.

Pivot type or trochoid. The pivot-type or trochoid joints are of two forms: in one a pivot rotates (spins) within a ring; in the other a ring moves around a pivot. In each case the ring is composed of fibrous tissue, part of which is converted into cartilage to form a female surface; the remainder may be ossified. Similarly, only part of the pivot is covered by a male articular cartilage. Pivot joints are always of the ovoid class; from a functional aspect, they are the ovoid counterparts of hinge joints.

The atlantoaxial (first and second cervical vertebrae) and the two radio-ulnar joints are examples of pivot joints. In the former, a thick, bony spike projects upward from the axis vertebra to form the pivot. The ring is formed in front by the anterior part of the (ring-shaped) atlas vertebra and completed behind by a transverse band of fibrous tissue that divides the opening of the atlas into two parts; the articulating surfaces are in front. In this joint the ring (atlas) moves around the pivot. It is the chief, though not the only, joint involved when the head is turned from side to side. In the upper radio-ulnar joint, the pivot is formed by the side of the upper end of the radius, and the ring is formed by a female surface on the ulna and a band of fibrous tissue that starts from the ulna behind and returns to it in front. This band lies in the capsule of the elbow joint and is called the radial annular ligament. In the lower radio-ulnar joint, the pivot is formed by the lower end of the ulna, and the ring by anterior and posterior bands of fibrous tissue and a female surface on the ulnar side of the radius. Thus, when the forearm is twisted (pronated) or untwisted (supinated), the limb being free, the pivot moves on the ring in the upper joint, but the ring moves on the pivot at the lower joint.

Movements of sellar joints

Sellar type. The sellar type has already been described in the section dealing with articular surfaces in general. It has two types of movement—both swings—flexion–extension and abduction–adduction. In addition to these it allows movements combining these two; that is, swings accompanied by rotation (spin) of the moving bone.

There are many examples of sellar joints. The best known is the carpometacarpal joint of the thumb, in which the free, first metacarpal moves upon the trapezium. (The carpal bones are those of the wrist, the metacarpals of the hand, and the phalanges of the fingers. The trapezium is a carpal bone.) The articular surface of the latter bone is convex from behind forward and concave from side to side. The thumb can be swung from side to side or from behind forward. But the most frequent movement is that which swings it so that it comes "face to face" with one or another of the fingers, as in

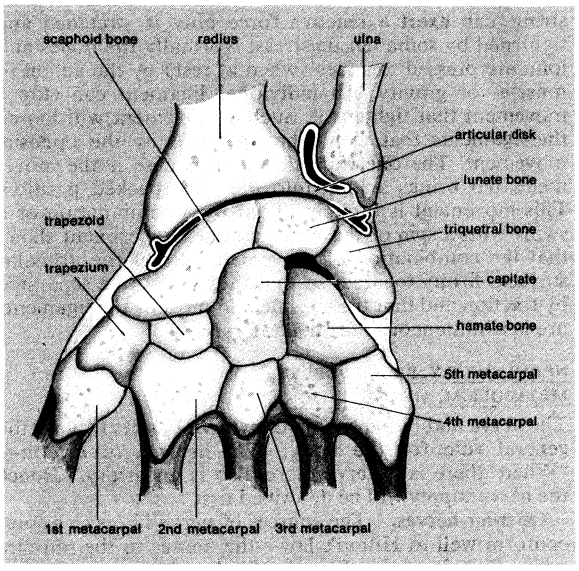


Figure 4: Coronal section through the radiocarpal, carpal, carpometacarpal, and intermetacarpal joints of the wrist.
From *Cunningham's* Textbook of *Anatomy* edited by G.J. Romanes and published by Oxford University Press as an Oxford Medical Publication

grasping a needle or a ball. This movement is called opposition (*i.e.*, of thumb to fingers). The reverse movement is called reposition. During opposition the thumb is rotated around its long axis; it has been said that our civilization depends upon the opposition of the thumb.

Sellar joints are found wherever only two types of movement are required, and a movement of the opposition kind is the habitual one.

Ellipsoid type. The ellipsoid is also a joint having two types of movement but allowing opposition movement only to a small degree. Its surfaces are ovoid and vary in both length and curvature as they are traced from front to back or from side to side, just as the diameter and curvature of an ellipse vary in directions at right angles to each other (hence the name). The joint between the second metacarpal and the first phalanx of the second finger is a good example. The male surface is on the head (finger end) of the metacarpal, at its extremity, and extends, broadening from side to side, on to its palmar aspect. It allows the finger to flex and extend, to swing toward or away from its neighbouring finger, and to swing forward with a slight amount of rotation. Its close-packed position is full flexion. The remaining joints between metacarpals and fingers are of the same type. So also are those between the lower end of the radius and the lunate and scaphoid bones of the wrist (radiocarpal) joint. In this case, however, the close-packed position occurs when the hand is swung backward (dorsiflexed) completely. If a person falls violently on his dorsiflexed hand, fracture of the scaphoid and the lower end of the radius is likely to occur.

Adapted from *Cunningham's* Textbook of *Anatomy* edited by G J. Romanes and published by Oxford University Press as an Oxford Medical Publication

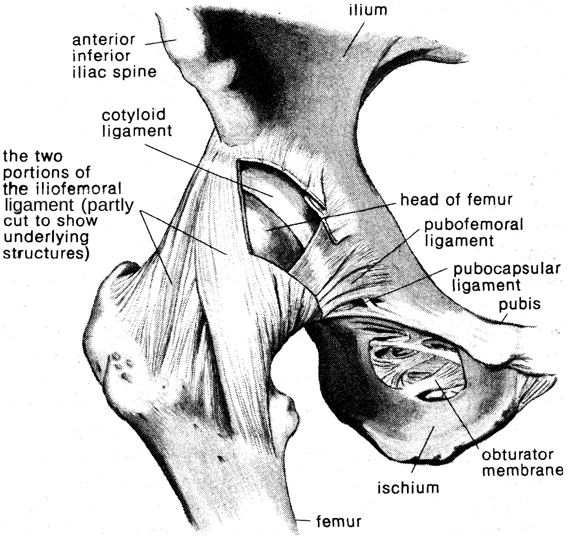

ilium
anterior inferior iliac spine
cotyloid ligament
the two portions of the iliofemoral ligament (partly cut to show underlying structures)
head of femur
pubofemoral ligament
pubocapsular ligament
pubis
obturator membrane
ischium
femur

Figure 5: Anterior view of the hip joint, showing attachment of ligaments to femur, ilium, ischium, and pubis.

Spheroidal type. The spheroidal type, also known as a ball-and-socket joint (see Figure 5), is the only one with three types of movement. It is an ovoid joint the male element of which could be described as a portion of a slightly deformed sphere. It is this fact that allows one of the component bones to swing in various directions and also to spin upon the other bone. The largest spheroidal joints are those at the hip and shoulder. At the shoulder the male surface is on the humerus and faces inward toward the body.

The female surface (glenoid cavity) is at the upper front angle of the scapula and is quite shallow. This allows freedom of movement of the arm upon the body in all directions except backward (extension).

Nevertheless, the utmost amount of swing forward (flexion) or outward (abduction) is 90"; the hand can then be swung further upward by movements of the scapula on the trunk. In accordance with this freedom of movement, the capsule of the joint is long and loose in most positions, and the joint itself is extremely susceptible to dislocation if the surrounding muscles be "off guard." It is in close pack when the arm is fully abducted and outwardly rotated.

The hip joint

At the hip joint the male surface is on the femur and also faces inward; it is larger than that on the humerus. The female surface is on the side of the hipbone, It is contained within a deep, cuplike socket (acetabulum), which permits flexion and extension of about the same amount as at the shoulder but restricts abduction-adduction and rotation of the femur around its long axis. The fibrous capsule is not as loose as that at the shoulder. The hip joint is close packed in the position of fullest extension.

Bicondylar type. Condylar type is better called the bicondylar type, for in it two distinct surfaces on one bone articulate with corresponding distinct surfaces on another bone. The two male surfaces are on one and the same bone and are of the same type (ovoid or sellar). The male surfaces are called male condyles; the female surfaces are female condyles. These joints have two types of movement: one is always a swing; the other is either another swing or a spin.

Bicondylar joints are quite common. The largest is the tibiofemoral joint in which both pairs of mating surfaces are within a single joint. At this joint, flexion and extension are the main movements; but active rotation of the leg on the femur is possible in most people when the leg and thigh are at right angles to each other.

Every vertebra of the cervical, thoracic, and lumbar series is connected to (or separated from) the one below it by a pair of synovial joints as well as by an intervertebral disk. This pair of joints constitutes a bicondylar joint, the shape of whose articular surfaces determines the amount of movement permitted between the vertebra (Figure 2). The atlanto-occipital joint between the skull and the vertebral column is also a bicondylar joint. Finally, the right and left temporomandibular joints, between the lower jaw and the skull, are really two parts of a bicondylar joint, not only by definition—if the base of the skull be considered as a single bone—but also functionally, for one mandibular condyle cannot move without the other moving also.

JOINT LIGAMENTS

Any set of collagen fibres joining one bone of an articulating pair to the other is called a ligament. Thus all the manifest part of the articular bursal wall is a ligament, called either the fibrous capsule or the investing ligament. The former name will be used in the text that follows.

In descriptive medical anatomy, the term ligament is used for special sets of collagen fibres associated with joints. There are two types of these sets: capsular and noncapsular.

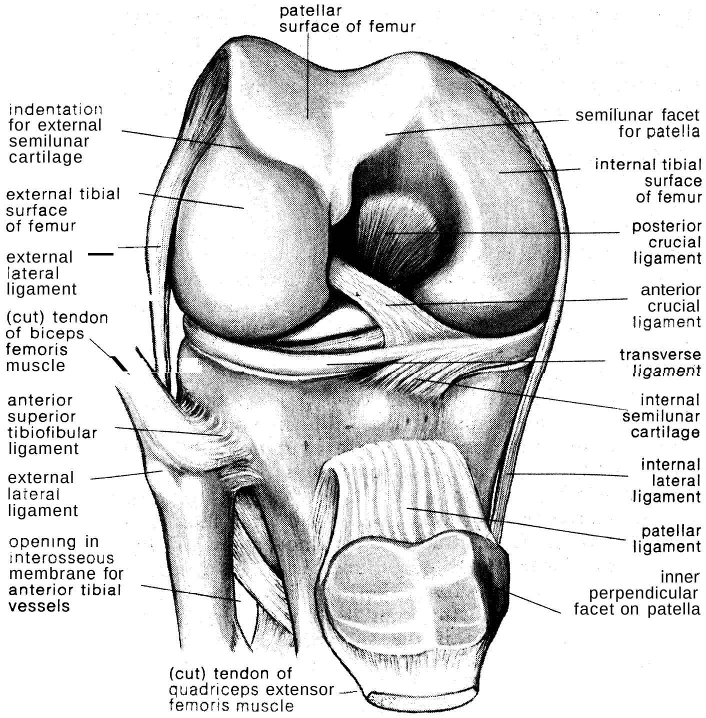

patellar surface of femur
indentation for external semilunar cartilage
external tibial surface of femur
external lateral ligament
(cut) tendon of biceps femoris muscle
anterior superior tibiofibular ligament
external lateral ligament
opening in interosseous membrane for anterior tibial vessels
(cut) tendon of quadriceps extensor femoris muscle
semilunar facet for patella
internal tibial surface of femur
posterior crucial ligament
anterior crucial ligament
transverse ligament
internal semilunar cartilage
internal lateral ligament
patellar ligament
inner perpendicular facet on patella

Figure 6: Dissection of knee joint from the front. The patella is thrown down to show the internal semilunar cartilage and transverse ligament.
Adapted from *Cunningham's Textbook of Anatomy* edited by G. J. Romanes and published by Oxford University Press as an Oxford Medical Publication

Capsular ligaments

Capsular ligaments are simply thickenings of the fibrous capsule itself, taking the form of either elongated bands or triangles the fibres of which radiate from a small area of one articulating bone to a line upon its mating fellow. The iliofemoral ligament of the hip joint is an example of a triangular ligament. Capsular ligaments are found on the outer surface of the capsule. There is one exception to this rule: ligaments of the shoulder joint (glenohumeral ligaments) are found on the inner surface.

Noncapsular ligaments are free from the capsule and are of two kinds, internal and external.

The internal type is found in the knee, wrist, and foot. In the knee there are two, both arising from the upper surface of the tibia; each passes to one of the two femoral condyles and lies within the joint cavity surrounded by synovial membrane. They are called cruciate ligaments because they cross each other X-wise. At the wrist most of the articulations of the carpal bones share a common joint cavity, and neighbouring bones are connected sideways by short internal ligaments. The same is true of the tarsal bones that lie in front of the talus and calcaneus.

The external, noncapsular ligaments are of two kinds, proximate and remote. The proximate ligaments pass over at least two joints and are near the capsules of these joints. They are found only on the outer side of the lower limb. Examples are the outer (fibular) ligament of the knee, which passes from the femur to the upper part of the fibula over both the knee and tibiofibular joints, and the middle part of the outer ligament of the ankle joint, which passes from the lowest part of the fibula to the heel bone. These two ligaments, particularly that passing over the ankle, are especially liable to damage (sprain).

The remote ligaments are so called because they are far from, rather than close to, the joint capsule. A notable example is that of the ligaments that pass between the back parts (spines and laminae) of neighbouring vertebrae in the cervical, thoracic, and lumbar parts of the spinal column. These are the chief ligaments of the pairs of synovial joints between the vertebrae of these regions. Unlike most ligaments they contain a high proportion of elastic fibres that assist the spinal column to return to its normal shape after it has been bent forward or sideways.

Functions of the ligaments

Contrary to the opinion of earlier anatomists, ligaments are not normally responsible for holding joint surfaces together. This is because a set of collagen fibres, like a string, can exert a reactive force only if stretched and tightened by some tensile stress. Normally the bones at a joint are pressed together (when at rest) by the action of muscles or gravity. An individual ligament can stop a movement that tightens it. Such a movement will loosen the ligaments that would be tightened by the opposite movement. The one exception to this case is the movement that brings a joint into the close-packed position. This movement is brought about by a combination of a swing with a spin of the moving bone. Experiment shows that the combination of movement screws the articular surfaces firmly together so that they cannot be separated by traction and that the capsule and most of the ligaments are in simultaneous maximal tautness.

NERVE SUPPLY, BLOOD SUPPLY. METABOLISM, AND NUTRITION

The nerve and blood supply of synovial joints follows the general rule for the body: ***Ubi nervus ibi arteria***—"Where there is a nerve there also is an artery." Hence the nerve supply will be described first.

Articular nerves. The sources of nerve fibres to a joint conform well to Hilton's law—the nerves to the muscles acting on a joint give branches to that joint as well as to the skin over the area of action of these muscles. Thus the knee joint is supplied by branches from the femoral, sciatic, and obturator nerves that among them supply the various muscles moving the joint. Some of these nerves go to the fibrous capsule and ligaments; others pierce this capsule and reach the synovial membrane. Some of these nerves are sensory; others give both motor and sensory fibres to the arteries that accompany them.

Types of sensory fibres to the fibrous capsule

The sensory fibres to the fibrous capsule are of two kinds: (1) algesic, responsible for painful sensation, particularly when the capsule or other ligaments are overstretched or torn; and (2) proprioceptive, that terminate in various forms of specialized structures and convey information to all parts of the central nervous system, including the cerebellum and the cerebrum. It has been established that this information includes the posture of a resting joint and both the rate and extent of motion at a moving joint. The latter is supplemented by impulses conveyed by the nerves from the muscles acting and the skin affected by the movement.

The sensory fibres to the synovial membrane reach it by piercing the fibrous capsule at various points and form wide-meshed networks in the subsynovial layer. They are mainly algesic in function, stimulation of them giving rise to diffused rather than localized pain (unlike the corresponding fibres to the fibrous capsule). They are found wherever the synovial membrane is, being especially abundant in the fatty pads, and are also present over the peripheral (nonarticulating) parts of the articular cartilage, disks, and menisci. This fact accounts for the great, sickening pain that accompanies injury of these latter structures. The articulating part of the articular cartilage has no nerve supply.

Articular blood and lymph vessels. The arteries in the vicinity of a synovial joint give off subdivisions that join (anastomose) freely on its outer surface. From the network of vessels so formed, branches lead to the fibrous capsule and ligaments and to the synovial membrane. Blood vessels to the synovial membrane are accompanied by nerves, and, when these vessels reach the subsynovial membrane, they proliferate to form another anastomotic network from which capillaries go to all parts of the membrane. These subsynovial arteries also ramify to the fatty pads and the nonarticulating parts of the articular cartilage, disks, and menisci. None, however, go to the articulating part of an articular cartilage, which, therefore, depends for its nourishment upon the synovial fluid. The vessels to the nonarticulating part of the cartilage pass into cartilage canals in its interior. (These canals have a structure resembling that of the haversian canals of bone except that their walls are not ossified; see BONE.) In one form of connective-tissue disease (osteoarthritis), the peripheral parts of the cartilage become ossified and give rise to bony projections. The blood vessels encircling the joint are numerous.

Veins, of course, are found with the arteries. In addition, a joint has a well-developed set of lymphatic vessels the ultimate channels of which join those of the neighbouring parts of the limb or body wall.

Joint metabolism and nutrition. The metabolism and nutrition of the fibrous capsule and ligaments are similar to that of fibrous tissues elsewhere. Their blood supply is relatively small, pointing to a low rate of metabolic changes. Unlike skin, for example, they heal slowly if injured.

The metabolism of articular cartilage, however, is of interest here. It is primarily dependent upon that of its cells (chondrocytes). Carbohydrate metabolism in these cells is similar to that of cells elsewhere and is unaffected by age. The oxygen consumption of the chondrocytes, on the other hand, decreases with age once the cells have matured. All the evidence suggests that the intracellular combustion is of glucose and protein, in that order of preference, rather than of fat. Sulfur passes from the blood to the synovial fluid and from there to the chondrocytes. From these it is transferred to the matrix to help to form chondroitin sulfate and keratosulfate molecules, the main constituents of the cartilaginous material. Chondroitin sulfate could be described as a sulfonated form of hyaluronic acid, the characteristic constituent of synovial fluid. Its presence in the matrix of the cartilage but not in the synovial fluid shows that the chondrocytes are necessary for its formation. After the second decade of life, the proportion of chondroitin sulfate falls and that of keratosulfate rises, as would be expected in view of the corresponding diminution of metabolic activity of the cells.

Formation of chondroitin sulfate

Excepting the articulating parts of the articular cartilages, disks and menisci, all the tissues of synovial joints are nourished directly by the blood vessels. The excepted parts are nourished indirectly by the synovial fluid. This is distributed over the surface of the articulating cartilage by the movements of the joint. The need for keeping joints healthy by frequent exercise of all of them is thus apparent.

BIBLIOGRAPHY. C.H. BARNETT *et al.*, *Synovial Joints: Their Structure and Mechanics* (1961), the first text on the general arthrology of synovial joints published in English since vol. 4 of the 11th ed. of *Quain's* Anatomy—contains all the advances made in the subject for about 50 years, with an extensive bibliography; M.A. MACCONAILL and J.V. BASMAJIAN, *Muscles and Movements* (1969), a text embodying both the rigorous modern theory of kinematics of muscles and joints and the information about how movements are brought about at individual joints or joint systems as revealed by electromyography; G. and C. OLIVIER, *Mdcanique articulaire* (1963), a simple presentation of the movements of bones and joints in man and their correlation with articular structure; A. STEINDLER, *Kinesiology of the Human Body Under Normal and Pathological Conditions* (1955), a study of the kinematics and kinetics of bones and joints, on classical lines of approach; and the *Journal of Bone and Joint Surgery* (8/year), current research in both basic and applied arthrology from all parts of the English-speaking world.

(M.A.MacC.)

Joint Diseases and Injuries

Diseases of the joints, although rarely life threatening, probably cause more invalidism in Western countries than any other conditions except heart and mental disease. It has been estimated from the National Health Survey that more than 1,000,000 persons in the United States are unemployable because of them. The most sorely afflicted are the elderly, but no age is immune. The disorders may be variously short-lived or exceedingly chronic; agonizingly painful or merely nagging and uncomfortable; they may be confined to one joint or may affect many parts of the skeleton. Many different disorders are involved, and several classifications of them have been proposed by professional organizations. For the purposes of this article two principal categories are distinguished: joint diseases in which inflammation (heat, swelling, pain, redness) is the principal set of signs or symptoms; and diseases, called in this article "noninflammatory," in which inflammation may be present to some degree, as after an injury, but is not the essential feature.

INFLAMMATORY JOINT DISEASES

"Arthritis" is a generic term for inflammatory joint disease. Inflammation restricted to the lining of a joint (the synovial membrane) is referred to as "synovitis." "Arthralgias" simply are pains in the joints; as ordinarily used, the word implies that there are no other accompanying evidences of arthritis. "Rheumatism," which is not synonymous with these, does not necessarily imply an inflammatory state but refers to all manners of discomfort of the articular apparatus including the joints and also the bursas, ligaments, tendons, and tendon sheaths. Inflammation of the spine and its joints is called spondylitis. Regardless of the cause, the inflammation may cause pain, stiffness, swelling, and some redness of the skin about the joint. Effusion of fluid into the joint cavity is common, and examination of this fluid is often a valuable procedure for determining the nature of the disease. The inflammation may be of such a nature and of such severity as to destroy the joint cartilage and underlying bone and bring irreparable deformities. Adhesions between the articulating members are frequent in such cases, and the resulting fusion with loss of mobility is called ankylosis. When apposed joint surfaces are displaced from each other they are said to be "dislocated" or "luxated." Incomplete dislocation is called "subluxation."

Types and routes of infection

Infectious **arthritis.** Joints may be infected by many types of micro-organisms (bacteria, fungi, viruses) and occasionally by animal parasites. There are three routes of infection: by direct contamination, by way of the bloodstream, and by extension from adjacent bony infections (osteomyelitis). Direct contamination usually arises from penetrating wounds but may also come during surgery on joints. Blood-borne infections may enter the joints through the synovial blood vessels. Commonly, however, they first set up foci of osteomyelitis in the long bone near the end of its shaft or next to the joint. The infection then extends into the joint through natural openings or pathological breaches in the outside layer or cortex of the bone. Characteristically hematogenous (blood-borne) infectious arthritis affects one joint (monarthritis) or a very few (oligoarthritis) rather than many of them (polyarthritis), and usually large joints (knee and hip) rather than small ones.

Inflammatory reactions in the joints are like inflammation in other tissues. Joint cartilage may be damaged rapidly by formation of pus in infections with such bacteria as staphylococci, hemolytic streptococci, and pneumococci. Tuberculosis of the joint, by contrast, can result in extensive destruction of the adjacent bone and open pathways to the skin without destruction of the cartilage. Tuberculous spondylitis, Pott's disease, is the most common form of this infection. It occurs mostly in young children. Before treatment with the antibiotic streptomycin and the antituberculous drugs isoniazid and para-aminosalicylic acid was available, hunchback deformities were frequent and the mortality was about 30 percent. At one time bovine tubercle bacilli caused much of human tuberculosis of the bones and joints, but pasteurization of milk and control of dairy herds has virtually eliminated this source of infection. The most frequent American fungus infection is caused by *Coccidioides immitis,* an organism indigenous to the arid regions of the Southwestern United States. As in the case of tuberculosis, seeding from the lung to the bone usually precedes involvement of a joint. Aside from this destructive form of coccidioidomycosis, transient pains in the joints are frequent during the period of dissemination of the disease through the bloodstream. Brucellosis, like tuberculosis, has a particular affinity for the spine. *Brucella suis* is the most likely brucellar organism to cause skeletal disease.

These infections are most directly identified by finding the causative agent in the synovial fluid or in specimens of synovial tissue. At times it may be preferable to investigate the regional lymph nodes. Infections of the joints, like infections elsewhere in the body, often cause fever and other systemic indications of inflammation. Appropriate antibiotics or chemicals are used to eliminate the infection, and early drainage of the exudate is desirable if there is pus.

Among the better recognized virus infections that can cause joint discomforts are rubella (German measles) and serum hepatitis, both of which usually are of short duration and have no permanent effect. Several tropical forms of synovitis have also been reported to be viral. Destructive arthritis sometimes complicates smallpox, but it probably results from supra-imposition of a bacterial infection rather than from invasion of the joint by the pox virus proper. Parasitic infestations are unusual, but there have been documented cases of joint involvement by the Guinea worm, *Dracunculus* medinensis, a nematode that affects more than 50,000,000 persons in tropical countries. Ordinarily it penetrates only into the soft tissues of the foot, but occasionally it spreads into the knee joint.

Complicating factors in venereal diseases

Infectious arthritis complicates several venereal diseases, including gonorrhea. The gonococcal organism is highly sensitive to penicillin, and early treatment with penicillin usually effects a prompt cure and prevents the marked destruction of the joint that may otherwise ensue. The arthritis of Reiter's disease, which is occasionally confused with gonorrhea and apparently is spread by sexual contact, does not respond to antibiotics but usually improves spontaneously over the course of several months. Characteristically Reiter's disease involves inflammation of the joints, the urethra, and the conjunctiva. Syphilis appears not to infect the joints directly except in the most advanced stage of the disease and in congenital syphilis. The latter frequently causes destructive inflammation in the growing cartilaginous ends of the bones of newborn infants. Untreated it leads to deformity and restriction of growth of the involved part (Parrot's pseudoparalysis), but early treatment with penicillin results in complete recovery. Clutton's joint is another type of congenital syphilitic lesion. It is a true inflammation of the synovial membrane that occurs in youngsters between 6 and 16 years; although it causes swelling of the knees, it is a relatively benign condition. Lesions characteristic of tertiary syphilis sometimes occur in the joints of children who have congenital syphilis. Yaws, a nonvenereal infection by an organism closely related to that causing syphilis, or identical with it, leads to similar skeletal lesions. The condition has largely been eradicated but only a few years ago affected many millions of persons in tropical areas and must once have been the principal cause of infectious disease of the bones and joints.

Deformities and destructive changes in the joints in leprosy (Hansen's disease) arise from infection of the nerves by the leprosy bacillus or from infection by some other bacteria.

Infectious arthritis occurs in many animal species and is of considerable economic importance in livestock commerce. A chronic "rheumatoid-like" arthritis in swine has for many years been the first or second cause for the condemnation of pork in the United States and other countries. One body of opinion attributes it to infection by *Erysipelothrix* insidiosa (rhusiopathiae), another to Mycoplasma hyorhinis. Mycoplasma organisms are minute filterable bacteria that lack cell walls. They cause arthritis in many avian and mammalian species. Certain types of these organisms cause pneumonia, but there is no proof that they are responsible for rheumatoid or other sorts of arthritis in man. The situation with another group of filterable agents, the Bedsonia (Miyagawanella), is analogous to this except that rare instances of arthritis occur in the venereal disease lymphogranuloma venereum, which is caused by this organism. Streptobacillus *moniliformis,* an organism commonly found in rats, causes arthritis in other species as well and in man has been responsible for Haverhill fever, characterized by rash, especially on the arms and legs, and by inflammation of several joints. It usually does not last more than about two weeks, and is effectively treated with penicillin.

Rheumatoid arthritis and allied disorders. In several types of arthritis that resemble infectious joint disease no causative agent has been isolated. Principal among these is rheumatoid arthritis. This disorder may appear at any age but is most usual in the fourth and fifth decades. A type that affects children is called Still's disease. Rheumatoid arthritis typically affects the same joints on the two sides of the body. Almost any movable joint can be involved, but the fingers, wrists, and knees are particularly susceptible. The joints are especially stiff when the affected person awakes in the morning. Rheumatoid arthritis is not only a disease of the joints. Fatigue and anemia indicate that there is a more generalized systemic involvement. A slight fever may sometimes be present. Lesions also occur in sites outside the joints. Involvement of bursas, tendons, and tendon sheaths is an integral part of the disease. Approximately one of five affected persons has nodules in the subcutaneous tissue at the point of the elbow or elsewhere. Inflammatory changes also are found sometimes in small arteries and the pericardium—the membrane enclosing the heart—less often in other structures. The eyes occasionally are involved but not as frequently as in Still's disease or in ankylosing spondylitis, a spinal arthritis that causes vertebral joints to fuse together (ankylose). The natural course of rheumatoid arthritis is variable. It may be episodic or sustained. Many affected persons are benefitted over the course of several months by rest, by analgesics such as aspirin, and by therapeutic exercises. In approximately one-third of the instances of the disease, it progresses and causes serious incapacity. In the absence of proper physical therapy, the joints may become greatly deformed and ankylosed.

Signs and course of rheumatoid arthritis

The cause of rheumatoid arthritis is unknown. The two principal hypotheses at this time are that rheumatoid arthritis may be an infection whose causative agent cannot be recovered by presently available techniques; or that it may be a disorder of hypersensitivity, conceivably to an infectious agent. In approximately 85 percent of cases, the blood serum contains abnormal, large antibody-like globulins called "rheumatoid factors." Rheumatoid factors are characterized by their property of combining selectively with serum gamma globulin, the blood protein that contains antibodies. They are thus directed against a normal body constituent (the gamma globulins) and may be classified as autoantibodies. Autoantibodies are involved in causing several other disorders, but a pathogenic role for rheumatoid factors has not been proven in rheumatoid arthritis. The rheumatoid factors form the basis for useful, but not absolutely specific, diagnostic tests for rheumatoid arthritis.

Although there is no cure for rheumatoid arthritis, much can be done to make the affected person comfortable and self-sufficient. Basic to the treatment are the conservative methods already noted. Adrenal corticosteroid hormones often induce initial dramatic improvements but do not in any way cure rheumatoid arthritis. Their effectiveness generally diminishes with time, and there are definite disadvantages in their use, such as a greater susceptibility to infection and to peptic ulcers. Several chemicals that have been tried—phenylbutazone, chloroquines, and indomethacin—may be of benefit but are all somewhat poisonous. Intramuscular injections of gold compounds have proven effective in a certain proportion of persons. Surgery is often of value in correcting established deformities. Much interest in early surgical excision of the inflamed synovial tissue existed in the early 1970s. Though thousands of victims of the disease have been treated in this way, insufficient time has elapsed to permit the profession to evaluate its long-term benefits or deleterious consequences. The long continuation of the disease causes much demoralization of the afflicted and makes them prey for innumerable nostrums and frauds. There is no place for vaccines, cobra or bee venoms, or sulfur compounds in the management of rheumatoid arthritis, and antibiotics also are valueless. A mild, dry climate seems to be beneficial in occasional cases, but the improvement is generally not sufficient to justify a move that would disrupt the total pattern of the affected person's life.

There is at times a close association between rheumatoid arthritis and seemingly unrelated disorders. In about one-third of the cases of Sjögren's syndrome, characterized by greatly diminished flow of tears and of saliva with resultant dry eyes and dry mouth, there is also rheumatoid arthritis, and high levels of rheumatoid factors are usually present in the bloodstream. In Felty's syndrome, rheumatoid arthritis coexists with enlargement of the spleen and diminution in the number of circulating blood cells, particularly the white blood cells. Removal of the spleen restores the blood to normal but has no effect on the arthritis.

Diseases resembling rheumatoid arthritis

Several other types of polyarthritis resemble rheumatoid arthritis but characteristically lack the rheumatoid factors in the bloodstream. One is associated with the skin disease psoriasis. The frequency of the association is greater than chance would dictate; roughly 5 percent of patients with psoriasis also have arthritis. Psoriatic arthritis also differs from rheumatoid arthritis insofar as it has a predilection for the outer rather than the inner joints of the fingers and toes; furthermore, it results in more destruction of bone. Another type of arthritis is associated with chronic intestinal diseases—ulcerative colitis, regional enteritis, and Whipple's disease. Ankylosing spondylitis, also known as Marie-Strümpell, or Bechterew's, disease, affects some of the peripheral joints, such as the hip; but its principal location is in the spine and sacro-iliac joints. In the spine, the small synovial joints and the margins of the intervertebral disks are both involved. These structures become bridged by bone, and the spine is accordingly rigid. Ankylosing spondylitis is a disease of men, the sex ratio being approximately 8:1 over women. The age of onset is lower than that in rheumatoid arthritis. The general management of the two disorders is much the same, but phenylbutazone is more effective in ankylosing spondylitis than in rheumatoid arthritis.

"Collagen diseases." The collagen diseases are so called because in all of them abnormalities develop in the collagen-containing connective tissue. These diseases are primarily generalized or systemic, frequently accompanied by joint complaints. The latter may be trivial in comparison to the systemic problems or may be severe enough to simulate rheumatoid arthritis. One of these diseases, systemic lupus erythematosus (SLE) may affect any structure or organ in the body. It usually causes characteristic eruptions on the face and inflammation and pain in the joints; it may inflame the eyes, cause kidney disease, or affect the central nervous system. The disease is characterized by serum antibodies that react with one or another component of nucleoproteins. This gives rise to the "L.E. cell phenomenon," a useful diagnostic test. The cause of SLE is not known. Some sort of overlap with rheumatoid arthritis is suggested by the fact that one-quarter of those with SLE have positive serological tests for rheumatoid factor; and perhaps as many patients with rheumatoid arthritis have positive L.E. cell tests. In another collagen disease, generalized scleroderma, the skin becomes thickened and tight. Similar changes occur in other organs, particularly the gastrointestinal tract. Restriction of motion of the joints of the fingers is in large part accounted for by the condition of the skin, but true synovitis also is frequent.

Rheumatic fever

Rheumatic fever often is classified with the collagen diseases. It has certain similarities to rheumatoid arthritis, as the name suggests, but the differences are more impressive. In both conditions, arthritis and subcutaneous nodules occur, and inflammation of the pericardium is frequent. Nevertheless the joint manifestations of rheumatic fever typically are evanescent, while those of rheumatoid arthritis are more persistent. The reverse is true of cardiac involvement in the two disorders. There is no compelling evidence that streptococcal infection is an important causative factor in rheumatoid arthritis, but it appears well established in rheumatic fever.

Arthritis more or less resembling rheumatoid arthritis occurs in roughly one-fourth of youngsters who lack gamma globulins in the blood. In this circumstance there is a deficit in the body's mechanisms for forming antibodies. This has sometimes been cited as an argument against the hypothesis holding that rheumatoid arthritis results from hypersensitivity.

Miscellaneous arthritides. Several types of arthritis appear to be related to an altered state of hypersensitivity. Erythema nodosum is a skin disease characterized by the formation of reddened nodules on the anterior surface of the lower extremities. In the majority of cases pain may arise in various joints and sometimes swelling appears. Lymph nodes at the hilus of the lung (the site of entrance of bronchus, blood vessels, and nerves) are enlarged. The synovitis disappears in the course of several weeks or months. Although this is not always so, many cases are associated with drug hypersensitivity, with evolving infections such as tuberculosis, coccidioidomycosis, and leprosy, and with Boeck's sarcoidosis, a systemic disease in which nodules form in the lymph nodes and in other organs and structures of the body. Synovitis of this sort occurs in 10 to 15 percent of patients with sarcoidosis. Occasionally, sarcoid joint disease becomes chronic and may simulate rheumatoid arthritis.

Palindromic (recurring) rheumatism is an arthritis of unknown cause. There is no fever. Each attack lasts but a day or two and leaves no permanent effects. Nevertheless palindromic rheumatism rarely remits completely and in perhaps one-third of cases eventuates in rheumatoid arthritis. Polymyalgia rheumatica, a relatively frequent although only recently recognized condition occurring in older people, is characterized by aching and stiffness in the muscles in the region of the hips and shoulders, but the joints proper seem not to be involved. There does seem to be some relationship to one type of arterial inflammation, temporal or giant cell arteritis. Polymyalgia rheumatica is not usually accompanied by serious systemic abnormalities and is treated with small doses of corticosteroids.

NONINFLAMMATORY JOINT DISEASES

Traumatic joint diseases. Blunt injuries to joints vary in severity from mild sprains to overt fractures and dislocations. A sprain is a damage to ligaments, tendons, and muscles that follows a sudden wrench and momentary subluxation of a joint. There is some slight hemorrhage into these tissues and healing usually takes place in several days. More violent stresses may cause tears in ligaments and tendons. Because the ligaments and tendons are so strong they frequently are torn from their bony attachments rather than ripped into segments. Ligamentous, tendinous, and capsular tears are able to heal by fibrous union provided that the edges are not totally separated from each other. Internal derangements of the knee most often arise from tears in the semilunar cartilages (menisci). Usually it is the medial meniscus that is disrupted. These tears are particularly frequent in athletes and develop as the knee is twisted while the foot remains fixed on the ground. Locking of the knee is a characteristic symptom. Because the semilunar cartilages have little capacity for repair, they must be removed surgically. Bleeding into the joint, called hemarthrosis, may also result from injuries. Other things being equal, it is desirable to remove the blood that has leaked from the blood vessels.

Sprains, tears, and dislocations

Most traumatic dislocations are treated by prolonged immobilization to permit the capsular and other tears to heal over. In some instances surgical repairs are required. Fractures of bone in the vicinity of joints may or may not extend into the joint space. Whether they do or not, the normal contour of the joint must be restored or arthritic complications are likely to develop. Other sorts of complication are also seen at times. In Sudeck's atrophy the nearby but unfractured bone becomes porous and is painful. The syndrome is believed to result from neurovascular reflexes originating in the injured tissue. Treatment is therefore directed at interrupting the sympathetic nerve pathways. In Volkmann's contracture, the muscles about the injured joint (most often the elbow) become converted into a shrunken fibrous mass, and the extremity becomes deformed and useless.

Degenerative joint disease. Osteoarthritis is a ubiquitous disorder affecting all adults to a greater or lesser degree by the time they have reached middle age. It is not restricted to man but occurs in many species regardless of their position in the taxonomic scale. The name "osteoarthritis" is a misnomer insofar as its suffix implies that the condition has an inherently inflammatory nature. For this reason it frequently is called degenerative joint disease or, on the European continent, osteoarthrosis or arthrosis deformans. When the spine is involved, the corresponding term is spondylosis. Unlike rheumatoid arthritis, osteoarthritis rarely causes crippling deformities. Indeed, in the majority of instances, the milder anatomical changes are not accompanied by appreciable symptoms. There are no systemic complaints directly attributable to the joint disease. The changes are characterized by abrasive wearing away of the articular cartilage concurrent with a reshaping of adjacent ends of the bones. As a result, masses of newly proliferated bone (osteophytes) protrude from the margins of the joints. The polished ends of the bone resemble ivory.

Osteoarthritis of the hip

The clinical manifestations of osteoarthritis vary with the location and severity of the lesions. The most disabling form occurs in the hip joint, where it is known as malum coxae senilis. Osteoarthritis of the hip, like that of other joints, is classified as primary and secondary. In secondary osteoarthritis, the changes come about as a consequence of some antecedent structural or postural abnormality of the joint. In perhaps half or so of cases, however, even rigorous examination fails to disclose such an underlying abnormality; in these instances, the osteoarthritis is called primary.

Probably the most frequent cause of osteoarthritis of the hip is congenital dysplasia (dislocation or subluxation of the hip). This term refers to a poor fit of the head of the femur, the long bone of the thigh, with its socket in the pelvis, the acetabulum. There is impressive evidence that many cases arise in infancy as a consequence of swaddling infants or carrying them in headboards, procedures that keep the thighs in an extended position. Before the child is ready for walking, the hip joint has frequently not yet fully developed and the head of the femur is forced out of its normal position by this extension. There thus arises a congenital dislocation or subluxation of the hip.

Osteoarthritis of the hip occurring in relatively young persons—in their 30s or 40s—frequently follows a progressive course and requires surgical treatment. Two rather different strategies of surgery have been employed: one, an osteotomy, involves reshaping the upper end of the femur so that the load borne by the joint is distributed more efficiently; the other requires removal of the diseased tissue and replacement by an artificial joint.

Aside from the rapidly developing forms, osteoarthritis of the hips also appears frequently in aged persons. It has been estimated that 175,000 Americans 65 years and older now require assistance for even their elementary hygiene because of degenerative hip disease. Chronologic aging is an important factor in the development of other forms of degenerative joint disease as well, since the lesions increase in frequency and severity as the decades pass by.

Causes of osteoarthritis

Considerations like these have led to the view that the principal causative factors in degenerative arthritis are faulty mechanical loading and senescent deterioration of joint tissue. Single injuries, unless they leave a joint permanently deformed, rarely if ever result in osteoarthritis. Recurrent small athletic and occupational injuries, such as those arising from heavy pneumatic drill vibrations, apparently are more likely to do so. Lifting heavy weights has been implicated in some studies of spinal involvement. Compensation boards wrestle every day with the problem that injuries may aggravate or direct attention to the pathologic process and not necessarily be its cause. Obesity is often considered mechanically harmful to joints, but there is increasing evidence that it is only a minor part of the problem.

Systemic factors play a less well-defined role. Dietary factors neither cause nor cure degenerative arthritis. Endocrine influences are suggested by the fact that Heberden nodes, the osteoarthritic swellings so common in the outermost joints of the fingers, occur far more frequently in women than in men and often appear at the menopause. Complaints about joint pain are particularly frequent at this age, but it is the consensus that menopausal arthritis reflects primarily a lowered threshold to discomfort attending this emotionally stressful period of life. Therapy with natural or synthetic hormones thus has no effect on the joint disease. There is no indication for treatment with thyroid hormone. Aside from surgery of the sort noted in the hip and sometimes the knee, treatment is directed toward an adjusted program of rest and proper exercise, avoidance of injury, the use of analgesics to relieve pain, and several types of physical therapy.

Chondromalacia patellae is a common and distinctive softening of the articular cartilage of the kneecap in young persons, particularly those with athletic inclinations. It results in "catching" and discomfort in the region of the patella, or kneecap, as the knee is bent and straightened out. Pathologically the changes are indistinguishable from changes that occur early in osteoarthritis. The condition often responds poorly to conservative measures and is treated surgically, sometimes by shaving off the degenerated cartilage. The wisdom of total excision of the patella is debated by orthopedists.

Degeneration of the invertebral disk is a frequent and in some ways analogous disorder. Often this occurs acutely in young and middle-aged adults, and the pulpy centre of the disk pushes out through tears in the fibrous outer ring (so-called slipped disk). When this takes place in the lumbosacral region, the displaced centre (the nucleus pulposus) impinges on the adjacent nerve roots and causes shooting pains in the distribution of the sciatic nerve—hence the name "sciatica." Pain in the small of the back is loosely referred to as lumbago, or lumbosacral sprain. It may be associated not only with degeneration of the intervertebral disk and spondylosis but also with structural anomalies of the region. Principal among these is spondylolisthesis, in which there is an anterior displacement of one lumbosacral vertebral body on another. The episodes respond to bed rest and mechanical support from wearing an abdominal corset or brace. Muscle relaxants may be of value. Recurrences are prevented by avoidance of back strains. Muscle-strengthening exercises are frequently helpful. The protruding tissue is removed by surgery only in cases in which pain and neurological defects are severe and fail to improve after less drastic measures. Surgery is not uniformly successful.

CONGENITAL AND HEREDITARY ABNORMALITIES

Clubfoot and other deformities

Congenital abnormalities are not necessarily transmitted from generation to generation but can be acquired during fetal life or soon after delivery. The latter abnormalities usually are structural; the inherited defects may be structural or appear later in life as the consequence of a systemic metabolic defect present from conception. Mention has already been made of the most important of these: congenital dysplasia of the hip. The joint proper may be normal initially in this and in several other congenital disorders; only after other supporting tissues have altered the proper relationships does the contour of the bone and joint become distorted. In arthrogryposis multiplex congenita ("multiple congenital crooked joints"), many joints are deformed at birth, particularly the hip. The deformities are the consequence of weakness of muscles that, in turn, sometimes results from spinal cord disease. Clubfoot (talipes equinovarus) is a congenital deformity of the foot, which is twisted downward and inward because the ligaments and tendons are too short. Only infrequently are the muscles at fault as they are in arthrogryposis. Idiopathic scoliosis (lateral curvature of the spine) usually makes its appearance during early adolescence. Its ultimate cause is unknown. There is considerable plasticity of the tissues with latitude for correction of these various deformities and for preventing their progression as the child grows. For this reason, the application of splints and other mechanical supports as soon as the condition is recognized is the major part of their management. Surgery is resorted to in certain cases when other measures have proved unsuccessful.

Structural variations in the lumbosacral spine are common and often are harmless. Incompletely ossified interarticular portions of the neural arches of a vertebra constitute a congenital anomaly referred to as spondylolysis; it predisposes to forward slipping of the vertebra later in life and so to the congenital type of spondylolisthesis described above. By contrast, when the failure of bony fusion exists between the right and left halves of the neural arch, the condition is called spina bifida occulta. This is a very common incidental X-ray finding and does not favour development of condition., that cause backache.

Gout, pseudogout, and ochronotic arthropathy

Several genetically influenced metabolic diseases have important articular manifestations. Gout is the most frequent of these. It occurs in approximately three of every thousand persons in the United States and western Europe. Chalklike masses of sodium acid urate crystals are deposited in and about the joints. Acute episodes of gouty arthritis are among the most painful conditions known. There is a tendency toward involvement of the great toe, a condition known as podagra (see also METABOLISM, DISEASES OF).

Ochronotic arthropathy results from another but rarer inborn error of metabolism. It is characterized pathologically by pigmentation and degeneration of hyaline cartilage, which covers the articular surfaces of bones, and biochemically by defective breakdown of the amino acids tyrosine and phenylalanine. The normal breakdown of these products is arrested at an intermediate level, and large amounts of homogentisic acid accumulate in body fluids and the urine. The urine turns black when exposed to air, a phenomenon called alkaptonuria. After many years the cartilages also become blackened, and severe degenerative changes occur in the peripheral joints and in the spine.

In yet another metabolic disease, chondrocalcinosis, or pseudogout, crystals of calcium pyrophosphate are deposited in joint cartilages. There are several forms of the disease. Sometimes there are no symptoms. In other cases, symptoms are sufficiently severe to cause confusion with rheumatoid arthritis. Some cases run in families. Unlike the situation in gout, specific methods for treating ochronosis and chondrocalcinosis have not yet been devised.

Other inherited disorders

Joints also are affected by several relatively rare heritable defects in the metabolism of connective tissue matrices. Depending on the specific biochemical defect, these disorders may be widespread throughout the body or limited to a few structures. There also is much variation in their inheritance. In Hurler's syndrome, for example, the most important manifestations are mental retardation and heart failure, though skeletal growth also is abnormal. Most affected persons do not survive adolescence. Morquio-Brailsford disease, by contrast, is a recessively inherited form of severe dwarfism that is not associated with mental deficiency or cardiac insufficiency. X-ray films of the spine reveal a characteristic misshapen flattened appearance of the vertebral bodies. Premature and severe degenerative changes in the peripheral and spinal joints are usual. Polyepiphyseal dysplasias (abnormal development of a number of epiphyses—the ends or outlying portions of bone, in childhood separated from the main body of the bone by cartilage) are a vaguely similar, though much milder, group of conditions in which precocious osteoarthritis and spondylosis are the first abnormalities to appear. Pre-existing changes in the skeleton, resembling a milder form of Morquio-Brailsford disease, may then be discovered by X-ray examination. The hip joint is most severely affected. In some cases the inheritance is dominant, in others recessive. Abnormalities in the fibrous components of connective tissue matrices are exemplified by Marfan's syndrome. Many organs are affected in this condition, and the articular manifestations are relatively unimportant among these. The joints are, however, excessively loose, and painful complications develop in about half of these patients. Hypermobility of joints also occurs in many other circumstances.

SECONDARY JOINT DISEASES

Nutritional diseases. The principal nutritional deficiencies that affect joints are scurvy and rickets. In both conditions, the principal location of the skeletal abnormality is at some distance from the joint proper, but the latter is sometimes misshapen as well. In addition, hemorrhage into the joint space is an intrinsic part of joint defects associated with scurvy.

Hemorrhagic **joint** diseases. Hemarthrosis (bleeding into the joints) is a major complication of hemorrhagic disorders. It constitutes, aside from the life-threatening episodes of bleeding, the principal disability arising from the hemophilias.

Most persons with these clotting defects are affected, and usually within the first years of life. Bleeding into the joints is usually precipitated by relatively minor injury but may leave several residual deformities and loss of mobility of the part. Hemarthrosis sometimes is a feature of the bleeding tendency associated with acute injury from radiation but is not ordinarily present in thrombocytopenic purpura; that is, bleeding associated with a shortage of blood platelets. Recurrent hemorrhage into an isolated joint, in the absence of a systemic tendency to bleed, is a characteristic of pigmented villonodular synovitis, a disease characterized by abnormal thickening and coloration of the synovial membrane. The precise source of the bleeding here is not known, and one theory is that a hidden benign tumour or malformation of the blood vessels in the joint is at fault. This is not a primary inflammatory disease of joints, despite the name, the thickening of the synovial tissue being a response to the hemorrhagic exudation. Large joints, usually of the lower extremity, are affected. This condition occurs in adults. It usually responds well to excision of part of the synovial membrane.

Joint involvement in aseptic necrosis

Aseptic necrosis. Because joint cartilages are without blood vessels, they are more immune to failures in the blood supply than are the underlying bones. Nevertheless. several clinically important joint diseases arise in association with aseptic necrosis—tissue death not caused by infection—of bone next to the joints. The precise nature of the failure of the blood supply is not always known. Fractures are one obvious cause. In caisson disease the obstructive elements are minute gas bubbles formed in the circulating blood from excessively rapid decompression. Decompression syndromes occur principally in divers and "sandhogs" (tunnel workers). Acute cases take the form of "bends" and frequently are fatal. However, in a large proportion of workers in these occupations, even if they have not experienced bends, extensive infarcts (areas of dead tissue) of bones and secondary osteoarthritis develop after many years. Analogous changes in sickle-cell anemias presumably result from blood clotting related to the abnormality of the red blood cells. There is no entirely persuasive explanation for other types of aseptic necrosis that occur in adults. In each instance, the hip is the joint most affected. Osteochondritis dissecans is a basically similar disorder in which a piece of joint cartilage and of underlying bone breaks off and lodges in the joint cavity. Usually the person affected can remember having injured the joint. The knee is the most frequent site. The condition usually occurs during the second and third decades of life. The displaced fragment causes a creaking sound when the joint is moved and is therefore called a "joint mouse." It must be removed by surgery.

Two different patterns of aseptic necrosis with joint involvement occur in growing children. One type (so-called slipped epiphysis) is characterized by partial or complete tearing away of an epiphysis, usually as the result of injury. The epiphysis at the upper end of the thighbone is particularly susceptible. Osgood-Schlatter disease is an analogous lesion, but it affects a growth centre (anterior tibial tubercle) at a slight distance from the joint rather than in its immediate vicinity. In the second type of aseptic necrosis in children, the necrosis is not the consequence of mechanical tearing away of the part. The most frequent site is in the head of the thighbone; necrosis at this site is known as Legg-Calvè-Perthes disease. It occurs in youngsters between the ages of 3 and 13 and is much more frequent in boys than in girls. Persistent pain is the most prominent symptom. Uncorrected severe lesions lead to arrest of growth, deformity, and arthritic changes in the hip joint. Treatment is directed first toward protection of the part from injury, together with rest and immobilization. In severe cases, detached fragments from the joint are removed by surgery. Other epiphyses, including those of the spine, may be affected.

Endocrine factors. The only joint lesion clearly related to a malfunctioning of the ductless (endocrine) glands is acromegaly. This disease results from excessive secretion of growth hormone by a tumour of the anterior pituitary gland. The hormone stimulates the proliferation of several skeletal soft tissues including the joint cartilage of the peripheral joints. It is this that causes the enlargement of the fingers that is characteristic of the disease. The enlarged joints are particularly prone to undergo osteoarthritic degeneration. Other hormonally influenced joint diseases are less clearly established. Cretinism, which is related to hypothyroidism, causes dwarfism and abnormally developed bony epiphyses but apparently does not lead to joint disease. Severe diabetes mellitus, however, may result in joint disease arising from the effect of diabetes on the nervous system.

Enlarged joints in acromegaly

Neurogenic arthropathy. A Charcot joint is a severe degenerative disease that develops when the sensory mechanisms of joints are impaired. The current view is that these joints become excessively strained because the ability to receive stimuli from bodily structures and organs necessary for normal limitation of motion is lacking. As a result the supporting tissues are torn and extreme disintegration of the structure results. Neurogenic arthropathy (joint disease related to nerve lesions) is most often associated with tabes dorsalis, a late form of syphilis affecting the posterior columns of the spinal cord. Other neurological lesions, including those associated with diabetes mellitus, syringomyelia (a disease in which cavities develop in the gray substance of the spinal cord), pernicious anemia, and leprosy, also are sometimes responsible. The disease usually is localized to one joint or one group of joints, depending on the location of the nerve defect. In tabes the knee joint is the most frequent site, in diabetes the foot, in syringomyelia the upper extremity. Pain is frequently mild considering the massive distortion of the joint. Treatment is difficult and is based primarily on immobilization and restriction of weight-bearing. Surgical fusion of the joints succeeds in only a small number of persons.

Hypertrophic osteoarthropathy. In approximately 5 to 10 percent of persons who have primary tumours within the chest, the ends of the bones near the joints become enlarged and painful. New bone is formed in the periosteum (the thick membrane covering the bone except within joints) and only occasionally do abnormalities develop within the joints themselves. Just how the chest abnormality leads to hypertrophic osteoarthropathy (disease of bones and joints with abnormal growth of bone) is somewhat of a mystery, but there is reason to believe that the vagus nerve is involved, since the condition is usually relieved promptly by cutting the vagus. It is also relieved by removal of the tumour. In this disorder the tips of the fingers become club-shaped, a painless lesion that was known to Hippocrates and that occurs in many other circumstances, as well.

Shoulder–hand syndrome. The shoulder–hand syndrome, so called because pain in the shoulder is associated with pain, swelling, and stiffness of the hand, only rarely develops in the wake of external injury. Most often it follows a heart attack (myocardial infarction) or is associated with disease in the neck vertebrae; frequently there is no apparent explanation. Most often the syndrome begins with pain and stiffness of a shoulder, followed shortly by pain and swelling of the hand, with vascular (blood vessel) changes in the skin of the hand. Over the course of several months the swelling and vascular changes subside but the skin and soft tissues become tightened. These changes sometimes disappear completely, but in other cases they leave permanent contractures; *i.e.*, flexion and loss of mobility due to the tightening of the fingers. Loss of mineral occurs in the bones of the shoulder, wrist, and fingers. Blocking (interruption of functioning) of sympathetic nerves serving the area, administration of adrenal hormones (corticosteroids), and therapeutic exercises are used in the management of the condition.

Tumours of joints. Tumours of joints are uncommon. In synovial chondromatosis, a benign condition, numerous cartilaginous nodules form in the soft tissues of the joint. The lesion is usually confined to one joint, particularly the knee, and occurs in young or middle-aged adults. It may or may not cause pain or swelling and usually is cured by excision of a portion of the synovial membrane. The tumour rarely becomes malignant. The cartilaginous nodules sometimes also contain islands of bone; in this circumstance the lesion is called synovial osteochondromatosis. Like synovial chondromatosis, synovial osteochondromatosis is often a spontaneous or primary disorder of unknown cause. In many cases, however, it is a development secondary to other diseases of the synovium, such as rheumatoid arthritis and even osteoarthritis.

Synoviomas, malignant tumours of the joints, arise in the tissues around the joints—the capsule, the tendon sheaths, the bursas, the fasciae, and the intermuscular septa or divisions—and only rarely within the joint proper. Although they may occur at any age, they are most frequent in adolescents and young adults. The legs are more often involved than the arms. The tumours spread locally and also to regional lymph nodes and lungs. Synoviomas are not particularly sensitive to X-rays, and treatment with drugs has been ineffective. If distant spread has not occurred at the time the condition is identified, radical excision, usually with amputation of the part, is the recommended treatment.

SPECIAL DIAGNOSTIC PROCEDURES

Four diagnostic procedures are widely employed in the diagnosis of joint diseases and injuries: X-ray examination, serologic tests, analysis of synovial fluid, and biopsy of synovial tissue.

Four diagnostic procedures

In X-ray films the soft tissues of joints are not ordinarily visible, although the bones are. Disappearance of articular cartilage is therefore manifested by a narrowing of the "roentgenographic joint space"; *i.e.*, the gap between the apposed subchondral bone surfaces, the bone surfaces beneath the cartilage. Only infrequently is the true joint cavity made visible by the injection of air, a procedure called pneumoarthrography. In myelography a contrast medium is injected into the spinal canal to disclose a protruding intervertebral disk or other soft tissue mass. The radiopacity of bone is reduced when bone is involved in fractures or in inflammatory processes. Osteophytes (bony outgrowths) protrude from the margins of the joints in osteoarthritis and spondylosis; they are likely to involve the periosteum in inflammatory diseases. Eburnation, degeneration of bone into a hard, ivory-like substance, appears as a zone of increased density at the joint surface. Calcification is often described about joints in X-ray reports, but with few exceptions the mineral is actually part of new bone formation rather than deposits of calcium salt in other tissues.

Analysis of synovial fluid is a safe and frequently helpful diagnostic procedure. The fluid is drawn from the joint easily with a needle after local anesthesia. Its viscosity typically is reduced in inflammatory diseases, and the number of white blood cells is increased. In infectious and acute inflammations of the joints, polymorphonuclear leukocytes (white blood cells with nuclei of varying shapes) predominate; in chronic rheumatoid arthritis, lymphocytes (white blood cells formed in lymph nodes and other lymph tissue) and mononuclear cells (large white blood cells with one nucleus each). The so-called ragocyte or rheumatoid arthritis cell now is regarded as having no diagnostic specificity. Microscopic examination of the fluid with plane-polarized light is of great value in finding crystals of gout and chondrocalcinosis.

Synovial biopsy is useful occasionally in differentiating rheumatoid from gouty or infectious arthritis. Micro-organisms are sometimes seen in the tissues, but they are more readily found if the specimen is cultured. In general, infectious agents disappear from synovial fluid before they do from the synovial tissue. For this reason, it is preferable to culture the latter rather than the former.

The serologic tests for rheumatoid factors are described above.

HISTORICAL AND GENERAL CONSIDERATIONS

Joint disease in antiquity

Most of what is known about disease in prehistoric periods necessarily concerns the skeletal system because the latter, unlike the soft tissues, is preserved through fossilization. Degenerative joint disease has been recorded in fossil reptiles as far back as the Mesozoic Era, and its frequency can perhaps be gauged by the fact that vertebral ankylosis is present in 6 of 11 extant specimens of the giant sauropod dinosaur ***Dipodocus longus.*** Spondylosis has been found in several neanderthaloids and probably accounts for previously held views that primitive man had a semi-erect posture. Changes consistent with rheumatoid arthritis have been found in Egyptian mummies and of ankylosing spondylitis in Pre-Columbian Indians.

Descriptions of dislocations of joints and directions for their treatment were included in the Hippocratic writings. All forms of chronic arthritis were regarded as manifestations of gout from that era until relatively recent times. Drinking of natural mineral waters was prescribed and carefully regulated by the ancient Greeks, and the Romans built baths for these disorders throughout their empire. Although there were some earlier attempts to discriminate between gout and other arthritic conditions, the recognition of rheumatism by Guillaume de Baillou of Paris in 1591 is an historic landmark. Arthritis as a sequel to gonorrhea was described by William Musgrave of Exeter in 1703 and confirmed by John Hunter 83 years later. Musgrave also associated "white swelling" of the joints with scrofula long before the tubercle bacillus was discovered (1882). The turn of the 19th century witnessed a great clarification of these disorders as the classical theory of humoral pathology went into oblivion. Rheumatic fever emerged as a systemic disease affecting the heart as well as the joints with the publication of a book by J.-B. Bouillaud in 1836. The definitive separation of gout from other rheumatic diseases came with the identification of its relation to disordered urate metabolism by Sir Alfred Garrod in 1859. Surprisingly, rheumatoid arthritis and osteoarthritis were confused with each other by the father of cellular pathology, Rudolf Virchow (1869), and a clear delineation of the two was made only in the early years of the 20th century.

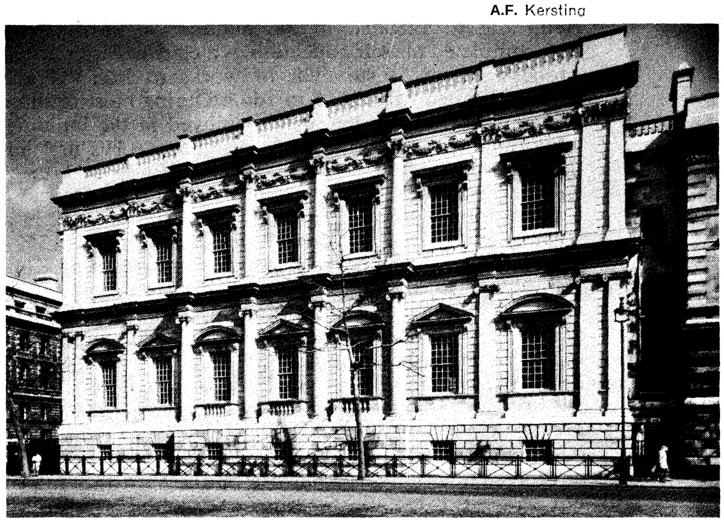
Diseases of the joints fall into two major provinces of medicine: orthopedics and rheumatology. There is some overlap between the two, but by and large they correspond to surgery and internal medicine, respectively. Accordingly, injuries to joints and surgical management of arthritic diseases are tended to by orthopedists, while other aspects of joint disease are handled by rheumatologists. There are professional and governmental agencies for these problems in most developed countries. The International League Against Rheumatism was founded in 1928. The World Health Organization has an expert committee on rheumatic diseases. In the United States, unlike certain other countries, rheumatology is not a certified specialty; but the American Rheumatism Association has some 1,800 physician members who devote their practices exclusively or primarily to these disorders. The Public Health Service conducts and supports research through its National Institute of Arthritis and Metabolic Diseases. At present there are more than a dozen medical journals in the world dealing only with arthritis and rheumatism, but they are able to publish only a fraction of the extensive relevant literature (see also JOINT).

BIBLIOGRAPHY. W.S.C. COPEMAN, *A Short History of the Gout and the Rheumatic Diseases* (1964), a scholarly history of arthritis and rheumatism, and (ed.), *Textbook of the Rheumatic Diseases,* 4th ed. (1969); and J.L. HOLLANDER (ed.), *Arthritis and Allied Conditions,* 7th ed. (1966), two authoritative texts; H.L. JAFFE, *Tumors and Tumorous Conditions of the Bones and Joints* (1958), on the pathology of neoplastic skeletal diseases, including a chapter on the joints; J.A. KEY and H.E. CONWELL, *The Management of Fractures, Dislocations and Sprains,* 6th ed. (1956), a textbook of orthopedic traumatology; V.A. MCKUSICK, *Heritable Disorders of Connective Tissues,* 3rd ed. (1966), a comprehensive monograph by a pioneer in the field; L. SOKOLOFF, *The Biology of Degenerative Joint Disease* (1969), a systematic consideration of the pathogenesis of osteoarthritis; J.H. TALBOTT, *Gout,* 3rd ed. (1967); R. WATSON-JONES, *Fractures and Joint Injuries* 5th ed. (in press), a highly respected clinical text.

(L.S.)

Jones, Inigo

Inigo Jones, surveyor of works to James I and Charles I of England, was the designer of court masques and architect of royal buildings during their reigns. Founder of the English classical school of architecture, he exerted a wide influence in his own time and found even greater acceptance in the 18th century. He set his mark on London by designing the first of its civic squares.


A.F. Kersting

The Banqueting House, Whitehall, London, by Inigo Jones, 1619–22.

Baptized on July 19, 1573, in the church of St. Bartholomew the Less, Smithfield, London, Jones was the son of a clothworker also called Inigo. Of the architect's early life little is recorded, but he was probably apprenticed to a joiner. By 1603 he had visited Italy long enough to acquire skill in painting and design and to attract the patronage of Christian IV of Denmark, at whose court he was employed for a time before returning to England. There, he is next heard of as a "picture maker" (easel painter). Christian IV's sister, Anne, was the queen of James I of England, a fact that may have led to Jones's employment by her in 1605 to design the scenes and costumes of a masque, the first of a long series he designed for her and later for the king. These masques had their origin in the intermezzi, scenographic and musical episodes, which had developed at the Medici Court in Florence and with which Jones was evidently familiar. The words were often supplied by Ben Jonson; the scenes, costumes, and effects nearly always by Jones. Over 450 drawings by him, representing work on 25 masques, a pastoral, and two plays ranging in date between 1605 and 1641, survive at Chatsworth House, Derbyshire.

Works for the Earl of Salisbury

From 1605 till 1610 Jones probably regarded himself as primarily under the Queen's protection, but he was patronized also by the Earl of Salisbury, for whom he produced his earliest known architectural work, a design for the New Exchange in the Strand (c. 1608; demolished in the 18th century). Though a somewhat immature design, the work was more sophisticated than anything being done in England at the time. Some designs (later superseded) for the restoration and improvement of old St. Paul's Cathedral also date from this period, and in 1610 Jones was given an appointment that confirmed the direction of his future career. He became surveyor of works to the heir to the throne, Henry, prince of Wales.

This appointment, with all its promise, was short-lived, and Jones did little or nothing for the Prince before the latter's death in 1612. In 1613, however, he was compensated by the guarantee of still higher office on the death of the King's surveyor of works, Simon Basil. To this office Jones succeeded in 1615, in the meantime having taken the opportunity offered him by Thomas Howard, 2nd earl of Arundel, to revisit Italy. Arundel and his party, including Jones, left England in April 1613 to escort Elizabeth, the daughter of James I, and her newly married husband, the elector palatine Frederick V, to their home at Heidelberg. They then proceeded to Italy, spending the winter of 1613–14 in Rome. In the course of the visit Jones had ample opportunity to study the works by modern masters as well as antique ruins. Of the masters, the one to whom he attached the greatest importance was Andrea Palladio (1508–80), the Italian architect who had gained wide influence through his Quattro *libri* dell' architettura, which Jones took with him on his tour. Returning to England in the autumn of 1614, Jones had completed his self-education as a classical architect.

Jones's career as surveyor of works to James I and Charles I lasted from 1615 to 1642. During most of those 29 years he was continuously employed in the building, rebuilding, or improvement of royal houses. His first important undertaking was the Queen's House at Greenwich, near London, based to some extent on the Poggio a Caiano, the Medici villa near Florence, but detailed in a style closer to Palladio or Vincenzo Scamozzi (1552–1616). Work there was suspended on the death of King James's queen in 1619 and only completed in 1635 for Charles's queen, Henrietta Maria. The building, considerably altered, now houses the National Maritime Museum. In 1619 the Banqueting House at Whitehall was destroyed by fire; and between that year and 1622, Jones replaced it with what has always been regarded as his greatest achievement. The Banqueting House consists of one great chamber, raised on a vaulted basement. It is conceived internally as a basilica on the Vitruvian model but without aisles, the superimposed columns being set against the walls, which support a flat, beamed ceiling. For the main panels of this ceiling allegorical paintings by Rubens were commissioned by Charles I and set in place in 1635. The exterior echoes the arrangement of the interior, with pilasters and regular columns set against rusticated walling.

Banqueting House at Whitehall

The Banqueting House has only two complete facades. The ends were never completed and this has given rise to the supposition that the building was intended to form part of a larger whole. This may have been so, and it is certain that Charles I, nearly 20 years after the Banqueting House was built, instructed Jones to prepare designs for rebuilding the whole of Whitehall Palace. These designs exist (at Worcester College, Oxford, and at Chatsworth House) and are among Jones's most interesting creations. They owe something to the Escorial Palace near Madrid, Spain, but are worked out in terms deriving partly from Palladio and Scamozzi and partly from Jones's own studies of the antique.

Apart from the Queen's House at Greenwich and the Banqueting House at Whitehall, the only one of Jones's royal buildings to survive is the Queen's Chapel at St. James's Palace (1623–27), a single-cell building with a magnificent coffered ceiling. Of Jones's many decorative works and minor additions to the royal palaces nothing remains.

Jones's work was not confined to royal palaces. He was much involved in the regulation of new buildings in London and out of this work emerged the project that he planned in 1630 for the 4th Earl of Bedford on his land at Covent Garden. This comprised a large open space bounded on the north and east by arcaded houses, on the south by the Earl of Bedford's garden wall, and on the west by a church with flanking gateways connecting to two single houses. The design probably derives partly from the piazza at Leghorn, Italy, and partly from the Place Royale (now des Vosges) at Paris. None of the original houses survive, but the church of St. Paul still stands, though much altered. Its portico is an instance, unique in Europe at its date, of the use of the primitive 'Tuscan order.

Covent Garden

With Covent Garden, Jones introduced formal town planning to London—it is the first London "square." He was probably instrumental, from 1638, in creating another square by planning the layout of the houses in Lincoln's Inn Fields, one of the houses (Lindsey House, still existing at No. 59 and 60, Lincoln's Inn Fields) being attributed to him.

The most important undertaking of Jones's later years in office was the restoration of St. Paul's Cathedral in 1634–42. This included not only the repair of the 14th-century choir but the entire recasing, in rusticated masonry, of the Romanesque nave and transepts and the building of a new west front with a portico (56 feet [17 metres] high) of ten columns. This portico, among Jones's most ambitious and subtly calculated works, tragically vanished with the rebuilding of the cathedral after the great fire of 1666. Jones's work at St. Paul's, however, considerably influenced Sir Christopher Wren and is reflected in some of his city churches as well as in his early designs for rebuilding the cathedral.

At the outbreak of the Civil War in 1642, Inigo Jones was compelled to relinquish his office as surveyor of works and left London. He was captured at the siege of Basing House in 1645. His estate was temporarily confiscated, and he was heavily fined. In the following year, however, his pardon was confirmed by the House of Lords and his estate restored. In the year of Charles I's execution, 1649, he was doing work at Wilton for the Earl of Pembroke, but the great double-cube room there is probably mostly due to his pupil John Webb, whom he had trained as his successor and who survived to re-establish something of the Jones tradition after the restoration of 1660. Jones died in London on June 21, 1652. He was buried with his parents in the Church of St. Benet, Paul's Wharf, where a monument, for which he had provided in his will, was erected, only to be destroyed in the great fire of 1666.

Of Jones's personal life little is known beyond what is told in the adulatory writings of John Webb and the slanderous satires of Ben Jonson, with whom he quarrelled. He was certainly an able administrator; he was also vain and dictatorial and wholly identified with the absolutism of his royal masters. His architecture stands out in high relief against the traditional English building work of his time and bears comparison with contemporary achievements in France and Italy, from which it is in some important respects markedly independent. His legacy to his successors was immense, eventually providing the main incentive for the "Palladian" movement that, beginning around 1715, dominated English architecture for 40 years.

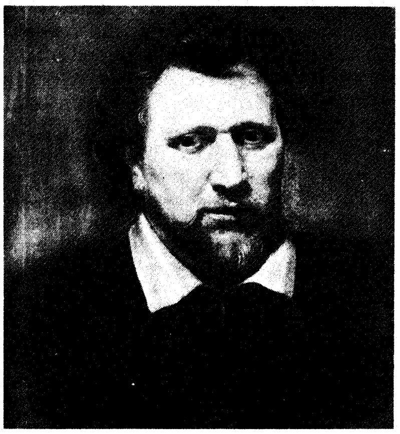
BIBLIOGRAPHY. J. WEBB, ***The Most Notable Antiquity of Great Britain, Vulgarly Called Stone-Heng . . .*** (1655), based on notes by INIGO JONES, includes almost the only early biographical material available. Further biographical notes are contained in Webb's ***Vindication of Stone-Heng Restored . . .*** (1665; 2nd ed., 1725); P. SIMPSON and C.F. BELL, ***Designs by Inigo Jones in Masques and Plays at Court*** (1924, reprinted 1966), an illustrated catalog; J.A. GOTCH, ***Inigo Jones*** (1928, reprinted 1968), an adequate biographical summary but critically inadequate; and J.N. SUMMERSON, ***Inigo Jones*** (1964).

(J.Sum.)

Jonson, Ben

Eight years the junior of William Shakespeare, Ben Jonson occupies by common, though not universal, consent the second place among English dramatists of the reigns of Elizabeth I and James I. He was a man of contraries. For "twelve years a papist" (as he admitted to the Scottish poet William Drummond of Hawthornden), he was also—in fact though not in title—Protestant England's first poet laureate. His major comedies express a strong distaste for the world in which he lived and a delight in exposing its follies and vices; yet, in his masques written as diversions for the court, he could compliment King James as the ideal king and present his realms as constituting the perfect human society. A gifted lyric poet, he wrote two of his most successful plays (Epicoene and *Bartholomew* Fair) entirely in prose, an unusual mode of composition in his time. He could rebuke Shakespeare for his extravagance yet also write verses acclaiming his great rival as the equal of the dramatists of classical antiquity. Often an angry and stubborn man, no one had more disciples than he. Much the most learned dramatist of his time, he attended no university, and—though his learning permeates his writings—he was master of theatrical plot, language, and characterization. It is a measure of his reputation that his dramatic works were the first to be published in folio (the term refers to the size of page but effectively means the "collected works") and that his plays held their place on the stage until the period of the Restoration, whose dramatists he influenced. Later his plays fell into neglect, though The *Alchemist* was revived during the 18th century, and in the last 40 years or so several have come back into favour: *Volpone, The Alchemist*, and *Bartholomew Fair* especially have been staged with striking success. Jonson, indeed, belongs, securely and brilliantly, to this age as much as to his own.

BY courtesy of the National Portrait Gallery, London



Jonson, oil painting by an unknown artist.
In the National Portrait Gallery, London.

Theatrical career. Jonson was born, probably on June 11, 1572, in or near London, two months after his father had died. His stepfather was a bricklayer, but by good fortune the boy was able to attend Westminster School (and later acknowledged a debt to his honoured master there, the antiquary William Camden, by dedicating two plays to him). His formal education, however, ended early, and he at first followed his stepfather's trade, then fought with some success in the Netherlands. On returning to England he became an actor and playwright, experiencing the life of a strolling player. There is reason for believing that at this time he acted the leading role of Hieronimo in Thomas Kyd's *The Spanish Tragedie;* in 1602 he was paid for making "additions" to it, and yet later he made fun of this early revenge drama. In a duel he killed a fellow actor, and, though he escaped capital punishment by pleading "benefit of clergy" (the ability to read from the Latin Bible), he could not escape branding. This, however, possibly through the help of friends, was done with a cold iron. At this time he was writing plays for Philip Henslowe, the leading impresario for the public theatre. With one exception *(The Cave is Altered)*, these early plays are known, if at all, only by their titles. He apparently wrote tragedies as well as comedies in these years, but his extant writings include only two tragedies, the brilliant *Sejanus* (1603) and the more stolid *Catiline* (1611).

Early
plays for
the public
theatre

The year 1598 marked an abrupt change in Jonson's status, when *Every Man In his Humour* was successfully presented by the Lord Chamberlain's theatrical company (a legend has it that Shakespeare himself recommended it to them), and his reputation was established. In this play Jonson tried to bring the spirit and manner of Latin comedy to the English popular stage by presenting the story of a young man with an eye for a girl, who has difficulty with a phlegmatic father, is dependent on a clever servant, and who is ultimately successful—in fact, the standard plot of the Latin dramatist Plautus. But at the same time he sought to embody in four of the main characters the four "humours" of medieval and Renaissance medicine, which were thought to determine the physical and mental makeup of human beings. These "humours" were choler, melancholy, phlegm, and blood. Following this success, the same company acted Jonson's *Every Man Out of his Humour* a year later. This was even more ambitious than its predecessor: it was the longest play ever written for the Elizabethan public theatre; it strove to provide an equivalent of the Greek comedy of Aristophanes, although most of the audience knew little of that playwright; and "induction," or "prelude," and regular between-act comment made plain the author's views on what the drama should be. (It was also a seminal play for Jonson and contained in germ most of the comic ideas that were later to be fully realized in his masterpieces.)

Plays for
the
"private"
theatres

The play, however, proved a disaster, and Jonson had to look elsewhere for a theatre to present his work. The obvious place was the "private" theatres, in which only young boys acted. The high price of admission they charged meant a select audience, and they were willing to try strong satire and formal experiment; for them Jonson wrote *Cynthia's Revels (c.* 1600) and *Poetaster* (1601). Even in these, however, there is the paradox of contempt for human behaviour hand in hand with a longing for human order. The great majority of the characters are presented for the audience's disapproval, but the goddess Cynthia finally rebukes them and indicates that men must do as she requires. This links up with Jonson's praise of the monarch's position in his masque. *Cynthia's Revels* ends when the goddess rebukes those characters who have exhibited divergent and eccentric "humours" throughout the play, demanding their reformation and their contribution to a celestial harmony. Yet it is an ending that neither Jonson nor his spokesman in the play, Crites, could properly believe in.

Unlike Shakespeare, Jonson was not a member of an acting company during his years as an important dramatist, and the fluctuations of adherence already noted continued throughout his career. Of his major plays, *Sejanus, Volpone,* and *The Alchemist* were acted by the King's Men (as the Lord Chamberlain's Company had become known on James I's accession). *Eastward Hoe,* written in collaboration with George Chapman and John Marston, and *Epicoene* were acted by the children and *Bartholomew Fair* by the Lady Elizabeth's Company. The plays of the last years—which contained much shrewd satire and authoritative writing but won little applause—were uniformly done by the King's Men, who were by that time increasingly associated with the Blackfriars Theatre, a private playhouse taken over from the children.

His masques at court. It appears that Jonson had won royal attention by his *Entertainment at Althorpe,* given before James I's queen as she journeyed down from Scotland in 1603, and in 1605 *The Masque of Blackness* was presented at court. The "masque" was a quasi-dramatic entertainment, an elaborate game of pretense in which everyone present was imagined as taking part. With its roots in medieval pageantry, especially the tournament, it slowly became recognizable during the 16th century as an independent form. The locale was a royal court or noble house. A monarch or some other emblem of human authority presided over an entertainment given for his guests and attendants; a "presenter" then entered and announced the arrival of a group of strangers who begged admission; assent being given, the strangers entered, disguised, and offered their homage; they performed their "masquing dance" in honour of the presiding figure; finally came the "revels," when the masquera danced with members of the assembled company.

This elementary pattern was much elaborated during the reign of James I, when the architect Inigo Jones provided increasingly magnificent costumes and scenic effects for masques at court. The few spoken words that the masque had demanded in Elizabethan days expanded into a "text" of a few hundred lines and a number of set songs. Thus the author became important as well as the designer: he was to provide not only the necessary words but also a special "allegorical" meaning underlying the whole entertainment. It was Jonson, in collaboration with Jones, who gave the Jacobean masque its characteristic shape and style.

Jonson's
innovations
in *The
Masque of
Blackness*

He did this primarily by introducing the suggestion of a "dramatic" action. In *The Masque of Blackness* the "presenters," Oceanus and Niger, announce the arrival of certain daughters of Niger (James's queen herself played one of them), whose beauty had been turned black by the fierceness of the sun in their own country, Ethiopia. Wishing for a lighter complexion, they have been told to travel to a land whose name ends with "-tania." They anxiously ask the name of the land they have now reached and are at first told it is "Albion." This, however, is soon corrected to "Britannia" (a tribute to James's sovereignty over both England and Scotland), and they are informed that here they will receive their hearts' desire. So there is occasion for song and dance. The degree of elaborate presentation, and also the length of time taken up by the entertainment, was far greater than a perusal of the text might suggest. The concluding "revels" could last throughout the night, and, indeed, a final song often suggests that morning is at hand. Thus, although the poet could be regarded as a mere auxiliary, it was he who provided the informing idea and dictated the fashion of the whole night's assembly. *The Masque of Blackness* was clearly a success, for during the following years Jonson was repeatedly called upon to function as poet at court. In his plays he had never been content to repeat himself, and, even if two plays share a governing idea (as in *Every Man In his Humour* and *Every Man Out of his Humour; Volpone* and *The Alchemist),* the mode of presentation is widely different. Similarly, in his masques he was fertile in inventing new motives for the arrival of the strangers. But this was not enough: he also invented the "antimasque" (sometimes "antemasque"), which preceded the masque proper and which figured grotesques or comics who, while they might dance and sing, were primarily actors (the "masquers" were played by ladies and gentlemen of the household; accomplished dancers and musicians, they could not necessarily act). A structural contrast developed between antimasque and masque. A hint of this appears as early as *Hymenaei* (1606), performed to celebrate the marriage of the Earl of Essex and Lady Frances Howard; the four "humours" appear in an attempt to disrupt the ceremony but are expelled by "reason" so that the masquers (in this case, gods favourable to marriage, with attendants) could enter for the celebration. But here the device is embryonic, and the antimasquers do not speak: that was soon to come, and from one antimasque there came to be several—though Jonson, though he must have loved the opportunity it gave him for expressing the duality of viewpoint that was so strong within him, generally managed to keep the device in its place.

Rivalry with Inigo Jones

Important though Jonson was at the court in Whitehall, it was undoubtedly Inigo Jones's contributions that caused most stir. That tension should arise between the two men was inevitable. As early as 1614 Jonson was surely making fun of Jones in *Bartholomew Fair* through his characterization of the puppet master, Lantern Leatherhead, although some scholars are doubtful of this interpretation of the role. Eventually friction led to a complete break: Jonson wrote the Twelfth Night masque for the court in 1625 but then had to wait five years before the court again asked for his services; in 1631 he wrote the poem "An Expostulation with Inigo Jones" and was ready to declare his contempt for the view that "Painting and carpentry are the soul of the masque." But the new king (Charles I) and his court preferred the display to the text. As an additional grievance, Jones was a permanent court employee, whereas Jonson was merely called in as an auxiliary. Even so, he was eventually recalled in 1631, and in noble houses away from Whitehall his services still proved welcome, even where the King himself was a guest.

His prime and later life. Troubles, meantime, had been numerous. After his tragedy *Sejanus*, he was in 1603 called before the privy council—doubtless because the fall of the Roman statesman Sejanus brought the downfall of the Earl of Essex to men's minds. His fun at the Scots' expense in *Eastward Hoe* in 1605 also led to trouble, but even so—although a "papist"—he was confidentially employed by the privy council in an attempt to make contact with a Catholic priest who might, they hoped, provide information about the attempted regicide (the Gunpowder Plot) of that year. In 1606 Jonson and his wife (whom he had married in 1594) were brought before the consistory court in London to explain their absence from church. He denied that his wife was guilty but admitted that his own opinions held him aloof from attendance. The matter was patched up through his agreement to have conference with learned men, who might persuade him if they could. Apparently it took six years for him to decide to conform. For some time before this he and his wife had lived apart, Jonson taking refuge in turn with his patrons Sir Robert Townshend and Esmé Stuart, Lord Aubigny.

During this period, nevertheless, he made a mark second only to Shakespeare's in the public theatre. *Volpone* and *The Alchemist* were among the most popular and esteemed plays of the time. Each exhibited man's folly in his pursuit of gold. Set respectively in Italy and London, they demonstrate Jonson's enthusiasm both for the typical Renaissance setting and for his own town on Europe's fringe. Both plays are eloquent and compact, sharp-tongued and controlled.

Other episodes reveal man's intransigence and vulnerability. In 1612 Sir Walter Raleigh, a prisoner in the Tower of London (but frequently visited by famous people), made Jonson his son's tutor for a visit to the Continent. The young man was difficult, and, soon after their return to England, Jonson may have mirrored their relationship in his depiction of Humphrey Wasp and Bartholomew Cokes in *Bartholomew Fair.* During this visit Jonson was present at a disputation between a Catholic and a Protestant in Paris, and he met Cardinal Duperron, assuring him that his translations from Virgil were "naught." The poet William Drummond, quoting Jonson's words on this matter in a record of the conversations he had with him in 1619, remarked that Jonson knew no French.

Conversations with Drummond

The occasion of Jonson's conversations with Drummond was a walking tour that Jonson made in 1618–19, which took him to Edinburgh and places nearby. From Drummond's report comes Jonson's view "that Shakespeare wanted art" and that "Donne was the first poet in the World, for some things" but " . . . for not keeping of accent, deserved hanging" and "for not being understood, would perish." During the visit the city of Edinburgh made him an honorary burgess and guild brother, giving a special banquet in his honour. On his return to England he received an honorary master of arts degree from Oxford University, a most signal honour in his time. He had already told Drummond that "He was Master of Arts in both the Universities by their favour not his study"; so by then he must have known of his proposed induction to the Oxford degree. No record survives of Cambridge's recognition. A letter of 1621 suggests that he narrowly escaped being knighted by James I, an honour that would have proved costly.

In 1623 his personal library was destroyed by fire (remembered in the poem "An Execration upon Vulcan"). As already noted, his services were seldom called on for the entertainment of Charles I's court, and his last plays failed to please. In 1628 he suffered what was apparently a stroke and, as a result, was confined to his room and chair, ultimately to his bed. His indignation at the failure of a play, *The Newe Inne*, in 1629 led him to write "Ode to Himself," beginning "Come, leave the loathed stage, And the more loathsome age." When he published the play in 1631, he described it as "never acted, but most negligently play'd, by some, the King's Servants. And more squeamishly beheld and censured by others, the King's subjects." In 1632 John Pory wrote to Sir Thomas Puckering that he "thought Ben Jonson had been dead."

He lived still. In 1628 he had been made city chronologer (thus theoretically responsible for the city's pageants, especially the lord mayor's show) in succession to Thomas Middleton, who had died in 1627. Middleton, a distinguished dramatist, had done the city proud: there was restiveness because Jonson did nothing to earn his salary. As so often with Jonson, the quarrel was patched up: in 1634 his salary was declared to be a pension.

Meanwhile he had other gratification, other anguish. He told Drummond that his wife "was a shrew yet honest." His admission on this occasion that he was "given to venery" suggests a reason for his long separation from her. They lived apart for five years, but several children came of the marriage, and there are poignant epitaphs from their father on the baby Mary and the seven-year-old Benjamin, who died of the plague. Thus his family responsibilities should be borne in mind when considering his generally successful career: he complained to

Drummond that poetry "had beggared him, where he might have been a rich lawyer, physician or merchant." That he always needed money may partly—but probably only partly—xplain his willingness to flatter James at court. When writing "The Prologue to the King's Majesty," which preceded *Bartholomew Fair,* he could address James as an equal, as a man who—like the playwrights and actors whose theatres were constantly under attack—had had much to put up with from the Puritans.

Patronage of James I

Certainly James was ready to help him: in 1621 Jonson had received a royal pension of 100 marks and soon afterward was given the reversion to the mastership of the revels. A further sign of honour, in 1624, was his inclusion in a list of 84 "able and famous laymen" proposed as members of the British Royal Academy. Perhaps the most interesting of all these honours is indicated in a legal document of 1623: there Jonson declared that his current residence was Gresham College. That was a town house of Sir Thomas Gresham, the famous founder of the Royal Exchange, where learning of a high order was invited. Possibly Jonson was acting as deputy to Henry Croke, who had been appointed as professor of rhetoric four years earlier. It was in this year, as already noted, that Jonson's library was destroyed; so he may have merely found a temporary refuge as a lodger in Gresham College.

His life was a life of talk as well as of writing. He engaged in "wit-combats" with Shakespeare, and, first at the Mermaid Tavern and then at the Devil Tavern, Jonson reigned supreme. It was a young man's ultimate honour to be regarded as a "son of Ben." The most prominent of his "sons" among the Caroline dramatists was Richard Brome, whose spirited comedies show how much he learned from his master: he had apparently been some sort of a "servant" to Jonson, perhaps as secretary, perhaps more menially. Others, more genteel, also flocked around him.

Jonson died on August 6, 1637. He was buried in Westminster Abbey, and on his gravestone there is an inscription "O rare Ben Jonson." It has been suggested that the first two words should be joined to make the Latin word *Orare*, the whole reading "Pray for Ben Jonson"—perhaps as a surreptitious tribute to Jonson's former Catholicism. The following year appeared *Jonsonus Virbius,* a collection of elegies written in his honour.

The first folio edition of his works had appeared in 1616; posthumously, in a second Jonson folio (1640), appeared *Timber: or, Discoveries,* a series of observations on life and letters. It was also called *Sylva* before its epigraph and *Explorata* in the title before the text. Here Jonson held forth on the nature of poetry and drama and noted his various debts to reading; here, too, he paid what was probably his final tribute to Shakespeare, in which, despite acknowledging a belief that his great contemporary was, on occasion, "full of wind"—*sufflaminandus erat*—he declared that "I loved the man, and do honour his memory, on this side idolatry, as much as any." Another prose work that also first appeared in this folio was *The English Grammar.* If it is true that for a time Jonson acted as professor of rhetoric at Gresham College, both this and the *Discoveries* may have been written with lectures in mind.

Importance in* English *drama. Jonson exerted a great influence on the playwrights who immediately followed him. In the late Jacobean and Caroline years, it was he, Shakespeare, and Beaumont and Fletcher who provided all the models. But it was he, and he alone, who gave the essential impulse to dramatic characterization in comedy of the Restoration and also in the 18th and 19th centuries. The Restoration dramatists' use of type names for their characters (Cockwood, Witwoud, Petulant, Pinchwife, Wishfort, Marwood, and the like) was a harking back to Jonson. So, too, was it in the 18th century, with such characters as Macheath, Peachum, Lumpkin, Candour, Languish. And although, as the 18th century proceeded, comic dramatists increasingly used names quite arbitrarily, the idea of the Jonsonian "type" or "humour" was always at the root of their imagining. In 19th-century comedy and melodrama type names were still employed, especially for minor characters; again the idea of the "type" was basic. In the mid-19th century, Dion Boucicault's brilliant pastiches of Restoration comedy (tempered to please the popular stage of his own day) continued the line from Jonson. In today's drama, though the type name is not so popular, the type itself remains.

Use of comic types

But Jonson did not deal exclusively in "humours." In some of the plays (notably *Every Man In his Humour)* the stock types of Latin comedy contributed as much as the humours theory did. What the theory provided for him and for his contemporaries was a convenient mode of distinguishing among human beings when such an elementary distinction was of use, as today the terms "introvert" and "extrovert" or "schizoid" and "cyclothymic" or "cerebrotonic" and "viscerotonic" may be conveniently employed. Jonson and his major contemporaries had primarily and ultimately to do with the diversity of human nature, its general surprisingness and shockingness as well as its recurrent shapes. He could also recognize the possibility of simple goodness on occasion, or at least wish for it (as in the contribution of Celia and Bonario to *Volpone),* but he found it atypical. For the rest, it was convenient to have a starting point—in the "humours," in Latin comic types, or, as in *Volpone,* in the assimilation of human beings to different members of the animal kingdom. Volpone, Mosca, Sir Epicure Mammon, Face, Subtle, Dol Common, Overdo, Ursula are not simply "humours": they are glorious type figures, so vitally rendered as to take on a being that transcends the type. Because his method was one of simplification, of typification (yet simultaneously of vitalization), his characters cannot be imagined as existing in any circumstances other than those of the play. This is not always true of Shakespeare's characters, but even in his case the exceptions are rare. The theatre encloses for the most part. In Jonson's plays it does so totally.

Jonson's influence has not been great outside the realm of comedy. His two extant tragedies, *Sejanus* and *Catiline,* belong with many other attempts in his time to bring Elizabethan and Jacobean theatre closer to that of the ancients. *Sejanus* is magnificent, but neither it nor the later tragedy was acceptable to his age. His masques belonged to and largely constituted a form that was essentially of the Jacobean and Caroline years, and attempts to revive it after the Restoration were tired things: even Jonson must sometimes have found the exercise strenuous, much as he surely wanted to celebrate the established order. In later days the atmosphere was too much filled with an awareness that real power lay elsewhere. Modern revivals of the masques have only an "archaeological" interest; modern imitations can be only a sad pastiche. As a lyric poet, however, Jonson has provided a model by no means forgotten by poets. His directness became less influential than the involuted manner of 17th-century Metaphysical poets such as John Donne, but poets and critics through the ages have honoured him.

Continental reputation

On the continent of Europe his position was for long obscure. Stefan Zweig adapted *Volpone* in 1927, and this version (in many ways a softened one) was translated into French and in that language made into a film. In general, as with so many of Shakespeare's contemporaries and successors, the European continent—even Germany, which is accustomed to speak of *unser* ("our") Shakespeare—has been slow to grapple with him. So, in some ways, has North America. Increasingly university and professional theatres have realized that his plays and those of other outstanding dramatists of his time are stage-worthy; but the theatre seems at the moment in advance of the classroom, which will surely come to confirm him as a living dramatist.

MAJOR WORKS

Every Man In his Humour (1598); *Every Man Out of his Humour* (1599); *Sejanus* (1603); *The Masque of Blackness* (1605); *Eastward Hoe* (1605); *Volpone* (1606); *Epicoene* (1609); *The Alchemist* (1610); *Bartholomew Fair* (1614); *The Metamorphos'd Gypsies* (1621); *The Newe Inne, or The light Heart* (1629); *Timber; or, Discoveries* (1640).

BIBLIOGRAPHY. For bibliographies of Jonson, see S.A. TANNENBAUM, *Ben Jonson: A Concise Bibliography* (1938; *Supplement by* S.A. and D.R. TANNENBAUM, 1947); and G.E. BENTLEY, *The Jacobean and Caroline Stage*, vol. 4, pp. 604–606 (1956). Extant contemporary manuscripts (not holograph) are listed in ALFRED HARBAGE, *Annals of English Drama*, 975–1700, rev. by SAMUEL SCHOENBAUM, pp. 311–312 (1964). STEVEN L. BATES and SIDNEY D. ORR (comps.), *A Concordance to the Poems of Ben Jonson* (1978), is keyed to Herford and Simpson's standard edition.

Editions: The major collected edition of his works is C.H. HERFORD and PERCY and EVELYN SIMPSON (eds.), *Ben Jonson*, 11 vol. (1925–52). E.A. HORSMAN (ed.), *Bartholomew Fair* (1960); and F.H. MARES (ed.), *The Alchemist* (1967), have been published in the Revels Plays. The "Yale Ben Jonson" edition includes: *Volpone*, ed. by ALVIN B. KERNAN (1962); *Bartholomew Fair*, ed. by EUGENE M. WAITH (1963); *Sejanus*, ed. by JONAS A BARISH (1965); *The Complete Masques*, ed. by STEPHEN OROEL (1969); *Every Man in His Humor*, ed. by GABRIELE BERNHARD JACKSON (1969); *Epicoene*, ed. by EDWARD PARTRIDGE (1971); and *The Alchemist*, ed. by ALVIN B. KERNAN (1974).

Biography: Biography has been best handled in Herford and Simpson's edition, vol. 1 (1925), vol. 11 (1952); and in MARCHETTE CHUTE, *Ben Jonson of Westminster* (1953). A useful brief account may be found in F.H. MARE'S edition of *The Alchemist* (cited above). G.E. BENTLEY, *Shakespeare and Jonson: Their Reputations in the Seventeenth Century Compared* (1945), is indispensable for Jonson's reputation. PERCY SIMPSON and C.F. BELL (eds.), *Designs by Inigo Jones for Masques and Plays at Court* (1924), fully indicates Jones's contribution to the performance of Jonson's masques at Whitehall.

Criticism: Critical writings of major importance include L.C. KNIGHTS, *Drama and Society in the Age of Jonson* (1937); E.B. PARTRIDGE, *The Broken Compass: A Study of the Major Comedies of Ben Jonson* (1958, reprinted 1976); and JONAS A. BARISH, *Ben Jonson and the Language of Prose Comedy* (1960). On the masques, see especially STEPHEN ORGEL, *The Jonsonian Masque* (1965); and JOHN C. MEAGHER, *Method and Meaning in Jonson's Masques* (1966). For the stage history of the plays, see R.G. NOYES, *Ben Jonson on the English Stage*, 1660–1776 (1935). *A Celebration of Ben Jonson*, ed. by WILLIAM BLISSETT, JULIAN PATRICK, and R.W. VAN FOSSEN (1973), is a collection of papers given at a conference commemorating the 400th anniversary of Jonson's birth.

(Cl.L.)

Jordan

The Hashemite Kingdom of Jordan (al-Mamlakah al-Urdunniyah al-Hāshimīyah), an Arab state of Southwest Asia, is a young nation that occupies an ancient land associated with the civilizations of antiquity. Situated on both banks of the Jordan River, it is bounded to the north by Syria, to the east by Iraq, to the southeast and south by Saudi Arabia, and to the west by Israel. The western border divides Jerusalem. Jordan has 12 miles (19 kilometres) of coastline on the Gulf of Aqaba in the southwest, where al-'Aqabah, its only port, is located. The total area is 36,833 square miles (95,396 square kilometres), and the population in the 1979 census was 3,044,000. Jordan's capital is Amman.

Since the Arab–Israeli War of June 1967, Israel has occupied the territory west of the Jordan River, known as the West Bank, including the Jordanian sector of Jerusalem. The occupied area is about 2,270 square miles, 6 percent of Jordan's territory and one-half of its agricultural land. About one-third of Jordan's de jure population lives in the West Bank.

Formation of the modern nation

As an international entity, Jordan came into being after World War I, gaining its independence from the United Kingdom as a hereditary constitutional monarchy in 1946. King Hussein ibn Talal ascended the throne in 1953; he has attempted to maintain Jordan's traditional policy of friendship with the West despite strong local and international pressures. Although Jordan has meagre natural resources, its most pressing problems are political. The intellectual and ideological divisions and frustrations that pervade the Arab world are reflected within Jordan—all the more so because the country has provided a haven for thousands of Arab refugees who fled from their homes as a result of the Arab–Israeli wars.

An economically developing country, Jordan throughout its existence has had to depend upon outside aid. This came first from the United Kingdom, later from the United States, and then—since 1967—from other Western and Arab countries, including Saudi Arabia, Kuwait, and the United Arab Emirates. Between 1954 and 1967, despite population pressures, a paucity of natural resources, and an influx of refugees, Jordan's gross natural product nearly quadrupled. To sustain this progress, Jordan adopted a seven-year plan in 1964, envisaging a self-sufficient economy by 1972. The 1967 Six-Day War with Israel, however, and the resulting influx of still more refugees, temporarily halted some of the measures included in the plan, overstrained the economy, which was already in difficulty, and increased Jordan's dependence on outside aid. Since then Jordan has carried out a three-year plan (1973–75) and a five-year plan (1976–80) and in 1981 introduced its second five-year plan. The object of these plans has been the revitalization of Jordan's economy, lessening its need for outside aid, and promoting general economic and social development. (For coverage of associated physical features, see ARABIAN DESERT; DEAD SEA; GALILEE, SEA OF; and JORDAN RIVER; for historical aspects, see JERICHO; NEAR EAST, ANCIENT; and SYRIA AND PALESTINE, HISTORY OF.)

The* landscape.** ***Relief and drainage. There are four major physiographic regions in Jordan: the Jordan desert, the East Bank uplands, the West Bank uplands, and the Jordan rift valley (a branch of the great African rift-valley system).

The desert is located in the eastern part of the country, occupying about four-fifths of the territory of the East Bank. Its northern part is composed of volcanic lava and basalt, and its southern part of outcrops of sandstone and granite. It is much eroded, primarily by wind. The East Bank uplands, an escarpment overlooking the rift valley, has an average altitude of between 2,000 and 3,000 feet (600 and 900 metres); the elevation increases to about 5,750 feet in the south. There are outcrops of sandstone, chalk, limestone, and flint extending to the extreme south, where igneous rocks solidified from the molten state predominate. In the northern uplands several valleys and perennial streams run west; around al-Karak they run west, east, and north; south of al-Karak nonperennial valley streams run east toward the Jafr Depression.

The Jordan Valley

The Jordan Valley, approximately 1,315 feet below sea level at the Dead Sea, contains the lowest point on the Earth's surface. Meandering south, the Jordan River drains the waters of the Sea of Galilee (Lake Tiberias), the Yarmūk, and the valley streams of both plateaus into the Dead Sea. The soil of its lower reaches is strongly saline, with the shores of the Dead Sea consisting of salt marshes that do not support vegetation. The Dead Sea occupies the central area of the valley. To its south, Wadi al-'Arabah, a completely desolate region, is thought to contain mineral resources.

The West Bank uplands, historically known as the Samarian and Judean mountains, are a complex anticlinal structure (*i.e.*, declining in opposite directions from a central axis) with an average height of about 3,000 feet above sea level. This plateau is cut into by some large valleys draining toward the Mediterranean; the valleys draining east are shorter and are precipitous, often consisting of deep gorges. Though the soils are generally thin and often poor, there are some rich alluvial plains among the hills and in certain basins. On these plains limestone, sandstone, and chalky rocks prevail.

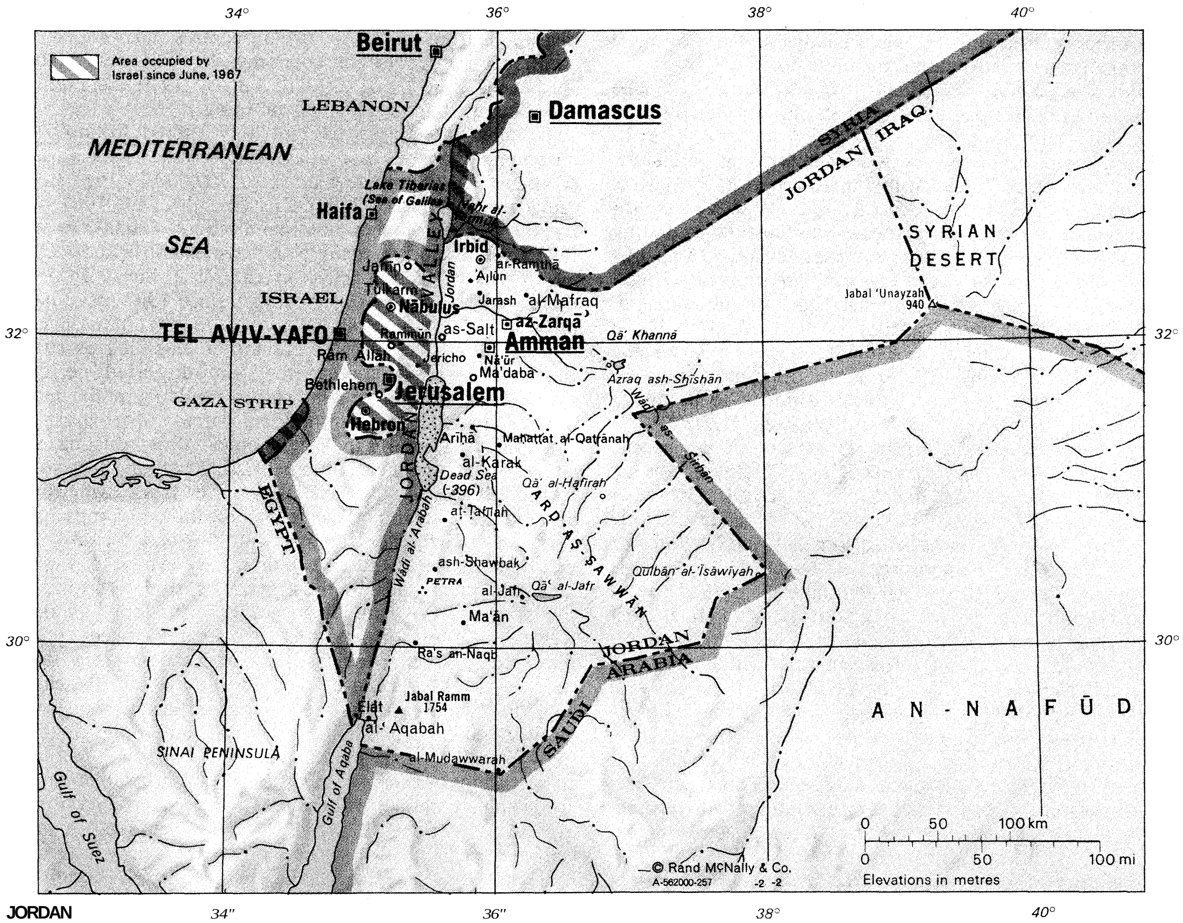
Climate. The climate varies from the Mediterranean type in the west to the desert type in the east, but the land is generally arid. The proximity of the Mediterranean Sea is the major climatic influence, although this influence is modified by continental air masses and by altitude. Mean monthly temperatures for Jericho (825 feet below sea level) range between 61° and 90° F (16° and 32° C), while at Amman (2,362 feet above sea level) they range from 44° to 87° F (7" to 31° C). The prevailing wind throughout the country is westerly to southwesterly. In the north the wind is generally more humid. Rainfall occurs in the short, cool winters, decreasing from 16 inches (400 millimetres) in the north to four inches in the south. The average rainfall in the east and west uplands totals about 16 inches annually. The valley itself has a yearly average of eight inches. The desert regions receive less than two inches. Occasional snow and frost occur in the uplands but are rare in the rift valley.

Vegetation and animal life. The plant and animal life of Jordan falls into three distinct types: that associated with the Mediterranean, with the steppe (treeless plains), and with the desert. In the uplands the Mediterranean type predominates. In the drier steppe region sagebrush predominates. Grassland is prevalent there, and some trees and shrubs, such as lotus fruit and the Mt. Atlas pistachio, also occur. In the desert scant vegetation grows in depressions and on the sides and floors of the valleys.

There is a great variety of animal life, including wild boars, as well as the ibex, a species of wild goat found in the gorges and in the 'Ayn al-Azraq oasis. Hares, jackals, foxes, wildcats, hyenas, wolves, gazelles, mole rats, mongooses, and a few panthers also occur. Among the domesticated animals, horses, mules, donkeys, camels, cattle, sheep, and goats are most common. Centipedes, scorpions, and various types of lizards are also found. Birds include the golden eagle and the vulture, while wild fowl include the pigeon and the partridge.

The nomads

The landscape under human settlement. The landscape falls into two regions—the desert zone and the cultivated zone—each of which is associated with its own mode of living. The nomads (Bedouins, or Badu) generally inhabit the desert and some areas of the steppe and the uplands. The number of nomads—only about 20,000 by the early 1980s—has decreased dramatically because of successful government efforts at resettlement. The eastern Bedouins are principally camel breeders and herders, while the western Bedouins are sheep and goat herders, There are some seminomads, in whose existence the modes of life of the desert and the cultivated zones merge. These people adopt a nomadic existence, living in tents only in the winter months after they have planted their lands, upon which some have also built modest homes; they return to their homes again in the spring or at harvest time. The two largest nomadic tribes of Jordan are Banū Sakhr and al-Ḥuwayṭāt. The grazing grounds of both are entirely within Jordan, as is the case with the smaller tribe of as-Sirhān. Other, lesser tribes, include Banū Hasan, al-Banū Khalid, al-Ajarmeh, al-Adwan, Banū Attiyeh, al-Hajayah, and as-Sleet, as well as the smaller tribes of al-Hawazim, as-Sulaylat, and ash-Sherarat who are obliged to pay protection money to larger tribes. The Rwalah tribe,



JORDAN

MAP INDEX

Cities and towns

Ajlūn	32 20n	35 45e
al- Aqabah	29 31n	35 00e
al-Jafr	30 18n	36 13e
al Karak	31 I I n	35.42e
al-Khalīl see Hebron		
al-Mafraq.	32 21n	36.12e
al-Mudawwarah .	29 19n	36.01e
al-Quds, see Jerusalem		
Amman.	31.57n	35 56e
Ariha	31.25n	35.47e
Ariḥa, see Jericho		
ar-Ramthā	32.34n	36.00e
ash-Shawbak. . . .	30.32n	35.34e
as-Salt.	32.02n	35.44e
aṭ-Ṭafīlah	30.50n	35.36e
Azraq ash-Shīshān (oasis):, . . .	31.50n	36.49e
az-Zarqa . .	32.05n	36.06e
Bethlehem (Bayt Laḥm)	31 43n	35.12e
Hebron (al-Khalil). .	31.32n	35.06e
Irbid	32 33n	35 51e
Janin.	37 28n	35.18e
Jarash	32 17n	35 54e
Jericho (Arīhā) .	31 52n	35.27e
Jerusalem (al-Quds).	31.46n	35.14e
Ma'ān	30.12n	35.44e
Madaba	31.43n	35.48e
Maḥaṭṭat al-Qaṭrānah	31.15n	36.03e
Nabulus	32.13n	35.16e
Nā'ūr	31.53n	35.50e
Ram Allah	31.54n	35.12e
Rammūn	31.56n	35.18e
Ra s an Naqb . .	30 00n	35.29e
Tūlkarm	32.19n	35.02e

Physical features and points of interest

Aqaba, Gulf of .	29.00n	34 40e
Arabah, Wādī al-, watercourse	30.10n	35.10e
Ard-as-Sawwan, plain	30 45n	37.15e
Dead Sea . .	31.30n	35.30e
Ḥafīrah, Qā'al- *depression* . .	31 06n	36 14e
'Īsawiyah, Qulbān al-, *wells*	30.38n	37.53e
Jafr Qā'al- *depression* . .	30 17n	36.20e
Jordan (Nahr al-Urdunn), *river*	31 46n	35.33e
Jordan Valley	32.00n	35.35e
Khannā, Qa, flats.	32.04n	36.26e
Ramm, Jabal, mountain.	29.35n	35.24e
Sham. Bādiyat ash- see Syrian Desert		
Sirhan, Wadi as- vailey	30.30n	38 00e
Syrian Desert (Bādiyat ash-Sham)	32.00n	40.00e
'Unayzah Jabal, mountain.	32.12n	39.18e
Urdunn, Nahr al-, see Jordan		
Yarmūk, Nahr al-, *river*	32.38n	35.34e

which is not indigenous, passes through Jordan in its yearly wandering from Syria to Saudi Arabia.

There are some 1,000 villages on the East Bank and 800 on the West Bank. Including nomads, rural residents represent about 40 percent of the population. The average village is a cluster of houses and other buildings, including an elementary school and a mosque, with pasturage on the outskirts. A medical dispensary and a post office may be found in the larger villages, together with a general store and a small cafe, whose owners are usually part-time farmers. Bedouin influence is strongest in the East Bank villages because of the interaction between villages and tribes characteristic of this region. Kinship relationships are patriarchal, while extended-family ties govern social relationships and tribal organization. Increases in the literacy rate and the influence of the mass media, in addition to extensive migration from rural to urban areas, have had a marked influence in recent years.

Of the total population of 3,044,000, about 60 percent live in the dozen or so major cities and towns. Amman alone has a population of about 650,000, but the smaller towns have only a few thousand inhabitants. Most towns have hospitals, banks, government and private schools, mosques, churches, libraries, and entertainment facilities, and some have institutions of higher learning and newspapers. Amman and Jerusalem, and to some extent Irbid, have urban characteristics, while smaller towns are more reluctant to accept modernizing influences.

Languages and dialects

People and population. *Ethnic and religious groups.* The Jordanian Constitution states that all Jordanians are equal. The majority of the people are Arabic-speaking, In addition to the difference between the written, or classic, Arabic and the colloquial form, there are various dialects with local inflections and accents. The Qaysi–Yemeni dichotomy—a pre-Islamic split that was introduced to the area with the Arab conquests and that cut across religious and ecological lines—was at one time an important broad social division. The Arabs, whether Muslim or Christian, used to trace their ancestry from the north Arabian Qaysi (Ma'di, Nizārī, Adnani, or Ismāʿīlī) tribes or from the south Arabian Yemeni (Banū Kalb or Qahtani) tribes. Only a few tribes and towns have continued to be aware of the split, but Bethlehem considers itself Yemeni.

According to the 1979 census 93 percent of the people are Sunni Muslim, and 5 percent are Christian. Among Christians adherents of the Rum, or Greek Orthodox Church, form a majority. Other Christian groups include the Rum, or Greek, Catholics, also called the Melkites or Eastern Catholics, who recognize the supremacy of the pope; the Roman Catholic community, headed by a patriarch appointed by the pope; the Assyrian (East Syrian), or Nestorian, Church, whose members number about 1,000 and live mostly in Jerusalem and Bethlehem; and the small Syrian Orthodox, or Jacobite, Church, whose members, like the Nestorians, use Syriac in their liturgy. Most non-Arab Christians are Armenians; the majority belong to the Gregorian, or Armenian, Orthodox Church, the rest to the Armenian Catholic Church. There are several Protestant denominations representing relatively recently formed communities whose converts came almost entirely from other Christian sects.

As-Samarah, the Samaritans, are remnants of an ancient Jewish sect; they number about 200 in Jordan and reside in Nābulus, at the foot of Mt. Shekim. The Druze, an offshoot of the Ismāʿīlī Shi'ah sect, also number about 200 and reside in and around Amman. The Baha'i—who in the 19th century also split off from Shi'ah Islām and who number about 1,000—live in al-Adasiyah in the Jordan Valley. The Armenians, Nestorians, Samaritans, Druze, and Baha'i are at once religious and ethnic communities. The Shishan (Chechen) are a Cherkess (Circassian) Sunni Muslim group, numbering about 1,000, who are descended from 19th-century immigrants. With the Cherkess, also Sunni, they make up the most important non-Arab minority. Another small non-Arab group consists of some Turkmen, who inhabit the village of Rammūn near Jarash.

Demography. Within the cultivated area of Jordan, the population density is 670 per square mile. In 1979, more than half of the population in the East Bank was under the age of 15; life expectancy at birth was 58 years. The birth and death rates were 46 and 12 per 1,000 respectively, giving a natural rate of increase of 3.4 percent. Internal migration from rural to urban centres has added an additional burden to the economy. In 1961 only 60,000 Jordanians were living abroad, but by 1979 the number was estimated to be 500,000 or more.

Jordan, Area and Population

	area*		population	
	sq mi	sq km	1971 estimate	1979 census†
Governorates				
East Bank			1,668,000	2,152,000
al-Aṣimah	962	2,491	943,000	1,188,000
al-Balqāʿ	431	1,116	102,000	151,000
al-Karak	1,872	4,849	88,000	126,000
Irbid	1,500	3,885	476,000	612,000
Maʿān	1,555	4,028	58,000	75,000
West Bank			680.000	. . .
al-Khalīl	418	1,082	. . .	. . .
al-Quds	795	2,059	. . .	. . .
Nābulus	969	2,509	. . .	. . .
Total Jordan	36,617 36,833	94,839 95,396	2,348,000	3,044,000‡

*Area figures are for settled areas only. Of the two total area figures, the first is the land area, the second the total area.
†Preliminary. $Including Palestinian refugees numbering 682,000 in the East Bank and 310,000 in the West Bank.
Source: Official government figures.

The refugees

The influx of refugees has not only altered Jordan's demographic map but also affected its political, social, and economic life. Numbering close to 700,000 in 1966, their ranks were swelled by 350,000 as a result of the 1967 war. By the 1980s refugees numbered about 680,000 in the East Bank, including many who have established Jordanian citizenship, and 310,000 in the West Bank. In 1955 there were 24 large refugee camps; since 1967 more have been added. In some of the camps the inhabitants live in huts or tents. The refugees cannot be classified as either rural, urban, or nomadic: they are displaced persons, whose situation and present status result from the interplay of regional and international forces. Most insist on their right to return to their homes, thus resisting any efforts at permanent resettlement. The condition of some of the refugees is wretched; they suffer from overcrowding and poor sanitary facilities. Some of the camps have neither electricity nor running water, which t causes extreme hardship to the inhabitants, especially in winter. Unemployment adds to the frustration and bitterness of the refugees. They are supported by the United Nations Relief and Works Agency for Palestine Refugees in the Near East, which provides dole at a subsistence level. UNRWA is also primarily responsible for their health, education, and welfare. All of the refugees have been granted full Jordanian citizenship and have refugee status only for relief purposes.

The national economy. Despite its basic problems, the Jordanian economy before 1967 showed resilience and growth. The Seven-Year Plan of 1964 looked toward economic self-sufficiency. Economic growth, which was halted by war in the second half of 1967, continued thereafter at a slower pace but was revitalized by the 1972–75 three-year plan and the 1976–80 five-year plan; real growth in the gross national product (GNP) was 7 and 10 percent, respectively, during these two periods. The GNP increased 20-fold from the mid-1950s to 1980. The West Bank's contribution to domestic income in 1965 was 38 percent of the total, but its occupation by Israel in 1967 required that the government apply its social and economic plans to the East Bank only.

Sources of revenue

By 1980 the major sources of revenue were tourism, mining and industry, trade, communications and transportation, agriculture, and construction. Income from tourism, which has grown dramatically, is mostly in foreign reserves, and tourism has become a major factor in Jordan's efforts to reduce its balance of payments deficit.

Exports, though growing. never cover the value of imports; the deficit is financed by foreign grants, loans, and other forms of capital transfers. Although Jordan's trade deficit has grown, it is offset somewhat by earnings from tourism, remittances sent by Jordanians working abroad, earnings from foreign investments made by the Jordan Central Bank, and subsidies from Arab and other governments.

In the past there has been a great outflow of skilled labour from Jordan to neighbouring countries, but the problem eased somewhat in the early 1980s. This reversal was a result both of better employment opportunities within Jordan itself and of a curb on foreign labour by the neighbouring Persian Gulf states.

Biological resources. There are only about 1,200,000 *dunums* (one *dunum* equals one-fourth of an acre) of forest, most of which are on the rocky highlands of the East Bank. Although unprotected, these forests have survived the depredations of villagers and nomads alike, as well as constant overgrazing. The Jordanian government embarked on a reforestation program in 1948. In the higher regions of the uplands, the predominant types of trees are the Aleppo oak, the Kermes oak, the Palestinian pistachio, the Aleppo pine, and the Oriental strawberry tree. Wild olives are also found there, and the Phoenicia Juniper occurs in the regions with lower rainfall.

Pastureland is so degraded that it can hardly support Jordan's livestock; it has, moreover, been reduced on the East Bank by the extension of land devoted to olive and fruit trees. Artesian wells have been dug to increase the pasturage area. Sheep and goats are by far the most important livestock, but there are some cattle, camels, horses, donkeys, and mules. Livestock decreases when droughts occur. There is fishing in the Gulf of Aqaba.

Mineral resources. Mineral resources include large deposits of phosphates, potash, and cement, as well as iron, phosphorus, manganese, and copper. In addition, newly discovered minerals include barite (the principal ore of the metallic element barium), quartzite, gypsum (used as a fertilizer), and feldspar. Marbles of various colours and limestone also are quarried.

All power in Jordan is generated by fossil fuel, mostly oil. There are several generating plants; the two major power stations, at Amman and az-Zarqāʾ, are linked by a transmission system. By the early 1980s the government had nearly completed a program to link the major cities by a country-wide grid, which later is to include rural areas.

Management of the economy. The economy is primarily based on private enterprise. To reduce its dependence on outside assistance, raise the standard of living, and reduce the imbalance between imports and exports, the government, in 1964, began the process of planning the economy. All of the government plans have given a high priority to agriculture, mining and industry, tourism, and the service industries, in addition to the strengthening and widening of the social, health, welfare, and education infrastructure. Unemployment, estimated at between 12 and 14 percent of the employable male population in 1961, had been virtually eliminated by 1980. Per capita income increased sixfold between 1963 and 1980.

Aside from a licensing system, a moderate taxation on luxury items, and the establishment of standards and health measures, both internal and international trade are virtually free of restraint. Governmental support takes the form of ensuring the basic framework of internal security, guaranteeing a stable currency, and maintaining the transport and communication routes and other facilities required for economic progress. The government also has participated with private enterprise in establishing the largest mining, industrial, and tourist firms in the country. By 1980 the government's share of the 32 largest companies was about 30 percent.

Taxation

Fiscal policy has aimed at increasing revenue by raising various tax rates and by reforming the tax system. Measures applied since 1964 include increases in customs and excise duties and an increase in income taxes, Although the government has placed great effort on reforming the income tax, both to increase revenue and to redistribute income, revenue from indirect taxes continues to exceed that from direct taxes. Tax measures have been adopted to increase the rate of savings necessary for financing investments. Exemptions on foreign investments and transfers of foreign profits and capital have been continued.

The existence of labour unions and employer organizations is recognized by law. The weakness of the trade-union movement is partly compensated for by the government, which has special procedures for settling labour disputes; if governmental efforts fail, the union or employers may resort to the judicial process.

Economic growth depends to a large extent on the return of the West Bank, where most of the sites attractive to the tourist trade are located and where a large proportion of the economically active population lives. The small size of the Jordanian market, the fluctuations in agricultural production because of irregular rainfall, lack of capital, political instability, and the presence of refugees all combine to make the continuation of outside help a necessity. Because of its basic weaknesses, the economy is highly sensitive both to domestic and to international policy.

The road system

Transportation. By 1980 the East Bank had a main, secondary, and rural road network about 4,300 miles long; of this about 3,100 miles are hard-surfaced. This roadway system, maintained by the Ministry of Public Works, links the major cities and towns and also links the kingdom with neighbouring countries. Within cities, towns, and villages, however, the local authorities are responsible for road upkeep, One of the main traffic arteries is the Amman–Jarash–ar-RamthB highway, which links Jordan with Syria to the north. The route from Amman via Maʿān to the port of al-'Aqabah is the principal route to the sea. From Maʿān the Desert Highway passes through al-Mudawwarah, linking Jordan with Saudi Arabia. The Amman–Jerusalem highway, passing through Nāʿūr, is a major tourist artery. The Hejaz Railway is government operated and extends from Darʿā in the north via Amman to Maʿān in the south. By the early 1980s part of an extension of this line, to Medina in Saudi Arabia, was being completed. Air transport developed into a major activity in the late 1960s. The Royal Jordanian Airline, Alia, operates flights linking Jordan to Arab, African, Asian, American, and European countries. Jordan's major airports, at Amman, al-'Aqabah, and Jerusalem—the latter Israeli-occupied—also receive international carriers. Queen Alia international airport near al-Jizah, south of Amman, was completed in 1982, eventually to replace the older airport at Amman.

Before 1948 Jordan's Mediterranean Sea trade was through Haifa. The Arab–Israeli conflicts severed that link, and Jordan's outlet since has been Beirut, in Lebanon. The expansion of the Jordanian economy, especially the export of phosphate, led to the development of the port of al-'Aqabah, which was handling about 6,200,000 tons of goods by 1980.

Administration and social conditions. The 1952 constitution is the most recent of a series of legislative instruments that, both before and after independence, moved toward increased executive responsibility. The constitution declares Jordan to be a constitutional hereditary monarchy with representative government; Islam is the religion of the state, and Jordan is declared to be part of the Arab *ummah* ("nation"). The king wields wide powers over the executive, legislative, and judicial branches. Jordan's central government is headed by a prime minister appointed by the king; the prime minister then chooses his cabinet. According to the constitution, the appointments of both prime minister and cabinet are subject to parliamentary approval. The cabinet coordinates the work of the different departments and establishes general policy.

Under the constitution the membership of the upper house of the bicameral legislature, composed of *al-A'yan* ("notables"), is appointed by the king for four years. Elections for *nuwwāb* ("deputies") of the lower house are to be held at least every four years. Since 1969, however, no parliamentary elections have been held; the ninth parliament, elected in 1965. was prorogued several times before being replaced in 1978 by the National Consultative Council, an appointed body with reduced power that debates government programs and activities.

Persons 18 years of age and over may vote provided they meet the legal requirements and are not members of the royal family. Voting participation has varied and has run as high as 70 percent, as it did in the 1962 elections. Political parties have been banned since 1957, however. Before that date several parties—Communist, Arab Ba'ath Socialist, National Socialist, Muslim Brotherhood, Liberation Movement, Arab Constitutional, Nahda (Renaissance), and Ummah (Community)—ran candidates. None of them, however, had a mass following. The Muslim Brotherhood was the only party exempted from the 1957 ban, but the group is kept under close surveillance.

Jordan is divided into eight administrative *muḥāfaẓāt* (governorates), which in turn are divided into districts and subdistricts, each of which is headed by an official appointed by the minister of the interior. Cities and towns have mayors and elected councils. In the West Bank the town and village municipal councils are mostly still in operation, but are under the careful scrutiny of the Israeli occupation authorities.

Three categories of courts

The judiciary. The judiciary is constitutionally independent, though judges are appointed and dismissed by royal *irddah* ("decree") following a decision of the Justices Council. There are three categories of courts. The first category consists of regular courts, including magistrates courts, courts of first instance, and courts of appeals and cassation in Amman, which hear appeals passed on from lower appeals courts. The constitution also provides for the Diwān Khāṣṣ (Special Council), which interprets the laws and passes on their constitutionality. The second category consists of Shari'ah Muslim courts and other religious courts for non-Muslims; these exercise jurisdiction over matters of personal status. The third category consists of special courts, such as land, government, property, municipal, tax, and customs courts.

The armed forces. The Jordan Army is a trained force of about 75,000. The small air force, equipped with modern jet aircraft, developed from the Arab Legion, which was originally commanded by British officers. Jordan also has a small navy. The king is commander in chief of the armed forces.

Education. There are three types of schools in Jordan—government schools, private and missionary schools, and the UNRWA schools for refugee children. Schooling consists of six years of elementary, three years of preparatory, and three years of secondary education. The Ministry of Education supervises all schools and establishes the curricula, teachers' qualifications, and state examinations; it also distributes free books to students in government schools and enforces compulsory education to the age of 14. Almost 70 percent of the students attend government schools. In addition to Khadduri Agricultural Training Institute, there are agricultural secondary schools, as well as a number of vocational, labour, and social affairs institutes, a Shari'ah (Qur'ānic) seminary, and nursing, military, and teachers' colleges. The State University of Jordan, established in Amman in 1962, had approximately 12,000 students by 1980; Yarmūk University, established in 1976, had 6,000 students. A third university, Mu'tah, opened in 1981.

Health and welfare services. Infectious diseases, except for dysentery and eye infections, have been brought under control. The number of doctors has grown rapidly, and by 1980 there was one doctor per 1,000 people. Comprehensive health facilities are operated by the government. A national health insurance program covers medical, dental, and eye care at a modest cost; service is provided free to the indigent. There are hospitals at Amman, as-Zarqā', Irbid, al-Karak, as-Salf, aṭ-Ṭafīlah, Janin, Jerusalem, and al-'Aqabah.

Traditionally, welfare services were private until the Ministry of Social Affairs was established in 1951. Besides supervising and coordinating some 400 social and charitable organizations, the ministry administers welfare programs.

Housing. The housing situation has remained critical despite the continuing house-building boom. A survey of Amman conducted in 1960 indicated that half of the city's families were living in one-room dwellings. The 1961 survey of the East Jordan Valley showed that three-quarters of the households consisted of single-room dwellings. The 1979 census estimated that 16,000 units were needed annually. The Housing Corporation and the Jordan Valley Authority build units for low-income families. Urban renewal projects in Amman and az-Zarqā' have provided new units and renovated others. The Housing Bank issues home building loans.

Wages and cost of living. Wages are higher in industry than in agriculture and in the cities than elsewhere. There are minimum wages for both skilled and unskilled labourers. The development of local services and resources, the availability of hard currencies from foreign aid, and a competitive import policy have given Jordan a degree of price stability not often encountered among developing nations.

Cultural life and institutions. Culturally, Jordan is an integral part of the Arab world and thus cannot be said to have a separate and distinct culture of its own. As in the rest of the Arab world, the highest form of artistic expression remains oral. Jordan's most famous poet was Muṣṭafā Wahbah aṭ-Ṭāl, whose style and content rank him among the major Arab poets of the 20th century. After World War II a number of important poets and prose writers emerged, though few have achieved an international reputation.

Since 1966 both private and governmental efforts have been made to foster the arts, and an art gallery has been opened. Modernity has lessened the influence of the traditional Islamic injunction against the portrayal of animate objects. Thus, in addition to the traditional decorative design, architecture, and the various handicrafts, it is possible to find sophisticated forms of painting and sculpture.

Folk and popular arts

Folk art survives in tapestry work and in the making of leather, pottery, and ceramics, as well as in the manufacturing of wool and goat-hair rugs with varicoloured stripes. Popular culture is manifested in such oral arts as songs, ballads, and storytelling. The villagers have special songs for births, circumcisions, weddings, funerals, planting, plowing, and harvesting. Several types of *debkah* (dances characterized by the pounding of feet on the floor to mark the rhythm) are danced on festive occasions, while the *sahjeh* is a well-known Bedouin dance. The Circassian minority have a sword dance, as well as several other Cossack dances. Government interest in preserving folk arts has resulted in the formation of a national troupe; it is regularly featured on state radio and television programs.

The press is relatively free, and the government only intervenes if flagrant abuse occurs. There are several literary magazines and scientific and topical periodicals. Among major newspapers the most prominent are *ar-Ra'i* and *al-Dustūr*. Most professional groups and government departments issue their own periodicals. Radio and television stations, which are government owned, feature programs from both Arab and foreign, mostly Western, countries. Most major towns have movie theatres that offer both Arab and foreign films. There is no legitimate theatre in Jordan, but amateur groups perform in institutions of learning, on radio and television, and in the various foreign cultural centres in Amman, Irbid, and Jerusalem.

BIBLIOGRAPHY. ABDULLAH, KING OF JORDAN, *My Memoirs Completed* (Eng. trans. from the Arabic, 1954); MUHAMMAD ABŪ ḤASSĀN, *Turāth al-badw al-qaḍā'ī* (1974), a study in Arabic of the life-styles, customs, and heritage of the Bedouins; NASEER H. ARURI, *Jordan: A Study in Political Development* (1972), a discussion of the political development of Jordan; C.S. COON, *Caravan: The Story of the Middle East*, rev. ed. (1958), an anthropological study of the Middle Eastern peoples and their environment; C.D. CREMEANS, *The Arabs and the World* (1963), a discussion of Arab politics, with a section on each country; H.M. DAVIS (comp.), *Constitutions, Electoral Laws, Treaties of States in the Near and Middle East*, 2nd ed. rev. (1953); FAO MEDITERRANEAN DEVELOPMENT PROJECT, *Jordan* (1967), a study of Jordan's economy, with emphasis on social and agricultural aspects; W.B. FISHER, *The Middle East: A Physical, Social, and Regional Geography*, 6th ed. (1970); SIR JOHN BAGOT GLUBB, *A Soldier with the Arabs* (1959), the author's personal insights; L.G. HARDING, *The Antiquities of Jordan*, rev. ed. (1967), an excellent resume; G.L. HARRIS,

Jordan: Its People, Its Society, Its Culture (1958), a general reference book; RICHARD F. NYROP (ed.), ***Jordan: A Country Study,*** 3rd ed. (1980), a general work; A.H. HOURANI, ***Minorities in the Arab World*** (1947), a scholarly account of the various minority groups and their backgrounds; J.C. HUREWITZ, ***Middle East Politics: The Military Dimension*** (1969), a voluminous work with emphasis on the role of the military; HUSSEIN, KING OF JORDAN, ***Uneasy Lies the Head*** (1962), an account that traces his early life and experiences in Jordan, and ***Mihnati Ka-Malik*** (1978), a personal expose of Jordanian politics; INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT, ***The Economic Development of Jordan*** (1957); A.M. LABADI, ***Land and Water Use in the Hashemite Kingdom of Jordan*** (1969); G. MOUNTFORT, ***Portrait of a Desert*** (1965), an illustrated description of life in the East desert of Jordan; NATIONAL PLANNING COUNCIL, ***The Jordan Five Year Socio-Economic Development Plan, 1981–1985*** (1981), an outline of the social and economic plans for the five-year period; R. PATAI, ***The Kingdom of Jordan*** (1958), a comprehensive general work; G. SPARROW, ***Modern Jordan*** (1961), a personal account of the author's travels; CHRISTINE OSBORNE, ***An Insight and Guide to Jordan*** (1981), an overview of the landscape, society, and culture; Y.T. TONI, ***Jordan: A Geographical Introduction*** (1968); BAHA-UDDIN TOUKAN, ***A Short History of Trans-Jordan*** (1945); P.J. VATIKIOTIS, ***Politics and the Military in Jordan*** (1967), a historical sketch outlining the influence of the military on Jordanian politics. SHELAGH WEIR, ***The Bedouin*** (1976), an illustrated study of the arts and crafts of the Bedouins of Jordan.

(K.S.A.J.)

Jordan River

The Jordan River (Nahr al-Urdunn), which flows southward from Syria across Israel and into Jordan, is the lowest river in the world. It rises on the southeastern slopes of Mt. Hermon in Syria (other headstreams rise in Lebanon), flows through a deep trench flanked by high plateaus, and—having crossed the Ḫula Basin and traversed the Sea of Galilee—drains into the Dead Sea at a depth of 1,315 feet (400 metres) below sea level. The Jordan is more than 223 miles (360 kilometres) in length, although, because its course is meandering, the actual distance between its source and the Dead Sea is less than 124 miles.

The river's role as a boundary

The Jordan has often served as an international boundary and, since 1948, has marked the frontier between Israel to the west and Jordan to the east from a few miles south of the Sea of Galilee until the Ytibis River flows into it from the left bank. Since 1967, however, when Israeli forces occupied the Jordanian territory on the west bank of the river south of its confluence with the Yābis, the Jordan has served as the cease-fire line as far as the Dead Sea.

The river was called the Aulon by the Greeks and ha-Yarden by the Hebrews, and it is sometimes called ash-Shari'ah (Watering Place) by the Arabs. Christians, Jews, and Muslims alike revere the Jordan; it was in its waters that Jesus was baptized by St. John the Baptist (for associated physical features, see DEAD SEA; GALILEE, SEA OF).

The natural environment. ***Physiography of the river valley.*** What is generally called the Jordan Valley is, in effect, a rift valley running north and south, forming part of the gigantic rift-valley system that extends from southern Turkey southward via the Red Sea and into East Africa. The valley itself is a long and narrow trough averaging about six miles in width but becoming narrower in some places—for example, at each end of the Sea of Galilee—and wider in others, such as the Bet She'an region, where the Ḫarod-Qishon depression gives access to the plains of Haifa to the west; the greatest width occurs in the Jericho region, where its plain is about 15 miles wide.

Throughout its course the valley lies much lower than the surrounding landscape, above all toward the south, where the Judaean plateaus rise to heights of about 3,000 feet, and those of ʿAjlūn rise to about 3,600 feet. The valley walls are steep, sheer, and bare, and they are broken only by the gorges of tributary wadis (seasonal watercourses).

Headstreams

The river course. The Jordan has three principal sources, all of which rise at the foot of Mt. Hermon. The longest of these is the Ḥāṣbānī, which rises in Lebanon, near Ḥāṣbayyā, at a height of 1,800 feet. From the east, in Syria, flows the Nahr Bāniyās; between the two is the Dan, the waters of which are particularly fresh. These three rivers join together in the Ḫula Basin, into which flow other streams, notably the 'Enot ʿEnan. The plain was formerly occupied by a lake and by marshes filled with a dense vegetation composed of papyrus, water lilies, and yellow pond lilies; in the 1950s, however, 15,000 acres (6,000 hectares) were drained to form excellent agricultural land, while 800 acres were preserved for their vegetation and animal life. At the southern end of the basin, the Jordan has cut a gorge through a basaltic barrier.

The Zor floodplain

The river then drops sharply down to the Sea of Galilee, which is itself 686 feet below sea level; this lake also serves to govern the river's rate of flow. After crossing the lake, the river valley narrows once more, and the Jordan receives its principal tributary, the Yarmūk, which marks part of the frontier between Syria and Jordan. It is then joined by two more tributaries, the Ḫarod on the right bank and the Yābis on the left. The river's plain then spreads out to a width of about 15 miles and becomes very regular; this is the Ghawr (Ghor) Plain, which is formed of marls (friable earthy deposits consisting of clay and calcium carbonate) and gypsum. Its flat and arid terraces are cut here and there by wadis or rivers (the Fāriʿah on the right bank, the Zarqā' and the Nusayrat on the left) into rocky towers, pinnacles, and badlands, a maze of ravines and sharp crests that resemble a lunar landscape and create obstacles to travel. There are three principal terraces on the plain, of which the highest is about 650 feet above the river. The Jordan has cut a valley into the plain of between about 1,300 and 10,000 feet in width and from between about 50 and 200 feet in depth. Along this stretch, the Jordan's floodplain is known as the Zor; it describes so many meanders that, although it runs for 135 miles, the actual distance it covers between the Sea of Galilee and the Dead Sea is only 65 miles. The Zor, which floods frequently, was formerly covered with thickets of reeds, tamarisk, willows, and white poplars sheltering an abundant wildlife, including boars. Since dams were built to control the river's flow, however, this land has been covered with irrigated fields. Finally, the Jordan drains into the Dead Sea through a vast, gently sloping delta.

Climate and wildlife. Although the bordering plateaus receive a relatively abundant rainfall, the depression itself is not well watered. The Ḫula Basin receives 22 inches (550 millimetres) a year, whereas only about three inches fall north of the Dead Sea. The winter is very mild, above all in the south, where the temperature in January is 52° F (11" C) in the Ḫula Basin, while at Jericho it is about 57° F (14° C). The summers, however, are torrid, with August temperatures of 81° F (27° C) in the Hula Basin and 90° F (32° C) at Jericho.

The wildlife of the Ḫula Swamp Nature Reserve includes otters, jungle cats, wild boars, and nesting populations of the common tern, the stilt (a long-legged, three-toed bird), and two types of herons. The river's fishes include some of those found in the Sea of Galilee; migrating southward, they die upon reaching the saline waters of the Dead Sea.

Volume of flow

Hydrology. The Jordan is fed by the rains falling on the neighbouring plateaus, the waters then flowing downward through rivers or wadis. The river itself is shallow. Upstream from the Sea of Galilee, the river's annual average flow is about 21,200,000,000 cubic feet (600,000,000 cubic metres); its minimum, reached in 1959 to 1960, was about 8,800,000,000 cubic feet, and its maximum, in 1968 to 1969, was more than 35,000,000,000 cubic feet. The Jordan's high-water period lasts from January to March, reaching a monthly maximum of about 3,200,000,000 cubic feet in February. Its low-water period occurs at the end of the summer and the beginning of autumn; the August flow averages about 780,000,000 cubic feet.

The waters of the Yarmūk, which average about 16,000,000,000 cubic feet, almost double the Jordan's flow; these are, however, most irregular, and the floodwaters in February are very sudden, so that the rate of flow is capable of rising from 2,000 to 60,000 cubic feet per second from one day to the next. The Zarqā' contributes an average of 1,600,000,000 cubic feet per year.

As one proceeds downstream, the rate of flow diminishes as a result of evaporation losses and the seeping away of water. As the average degree of the slopes over which the river flows is comparatively steep, the speed of the current is relatively fast, amounting to about four and a half feet per second at low water at the Allenby Bridge, a short distance north of the Dead Sea.

As the Ghawr region consists of marl, the Jordan transports a considerable load of silt, amounting to an average of more than 2 percent of its total flow and to more than 5 percent during the flood period.

The existence of thermal springs, notably in the Tiberiade region on the western side of the Sea of Galilee, as well as the frequency of gypsum, results in the Jordan's waters having a relatively high degree of salinity—averaging 0.77 percent—thus creating problems when the water is used for irrigation.

The imprint of man. *Settlements and river crossings.* Where irrigation permits, the Jordan Valley has been settled by Arab and Jewish agricultural communities. Notable settled regions are the Ḫula Valley in the north; the string of agricultural communities south of the Sea of Galilee on the West Bank, including Deganya—the oldest kibbutz (collective agricultural settlement) in Israel, founded in 1909—Afriqim, Ashdot Ya'aqov, and Ḫawwat Shemu'el; the area along the Ghawr Canal on the East Bank; and the area of the Wādī Fāriʿah on the West Bank.

Navigation is impossible because of the river's precipitous upper course, its seasonal flow, and its shallow, twisting lower course. The principal river crossings are those at the Benot Ya'aqov Bridge, on the road from Tiberias to Damascus; the Dāmiyā Bridge, on the road from as-Salf to Nābulus; the King Hussein Bridge, on the road from Irbid to ʿAfula; the Allenby Bridge, on the highway from Amman to Jerusalem; and the pipeline crossing at al-Manshiyah.

Irrigation and other projects

The river's resources. The Jordan waters are of especial importance for irrigation. For a long time the water was not used, except for several oases located in the bordering foothills—for example, at Jericho—which used the waters of springs that fed the river. The Ghawr region was formerly barren, desolate, and uninhabited except in winter by nomadic herdsmen and their livestock, who went up to the plateaus in summer. The Ghawr irrigation canal—43 miles long—was, however, completed in 1967 on the East Bank and has permitted the cultivation of oranges, bananas, early vegetables, and sugar beet in the region. In Israel, apart from the draining of the Ḫula Basin and the construction of a canal from the Sea of Galilee to Bet She'an, a water-supply grid has been constructed that permits 11,300,000,000 cubic feet of the Jordan's waters to be pumped each year to the centre and south of Israel.

Prospects for the future. Several projects have been drawn up that envisage the use of the waters of the entire basin. In addition, it has been suggested that the level of the Dead Sea could be maintained by channelling in salt water from the Mediterranean, at the same time using the flow to generate hydroelectricity. Because of Arab–Israeli tensions, however, only a partial use of the Jordan's waters has been made. Syria and Jordan envisage the construction of an irrigation and hydroelectric dam on the Yarmiik.

BIBLIOGRAPHY. *Handbook of the Geology of Jordan,* issued by the Government of the Hashemite Kingdom of Jordan, Amman (1959), deals with the eastern part of the Ghor region; L. PICARD, *Structure and Evolution of Palestine, with Comparative Notes on Neighboring Countries* (1943), a basic geological study; N. SCHULMAN, "The Geology of the Central Jordan Valley," *Bull. Res. Council Israel,* 8:36–90 (1959), a further geological study; I. SCHATTNER, *The Lower Jordan Valley: A Study in the Fluviomorphology of an Arid Region* (1962), a study of erosion in the Ghor region; THE HASHEMITE KINGDOM OF JORDAN, CENTRAL WATER AUTHORITY, HYDROLOGY DIVISION, *Technical Papers,* an annual publication giving information on rainfall and hydrology; ISRAEL HYDROLOGICAL SERVICE, *Hydrological Papers,* includes detailed hydrological statistics; M. LOUVISH *et al.* (trans.), *The Huleh and the Upper Jordan Region* (1954), on the problems of development of the Upper Jordan Valley; ANDREW M. WATSON, *The East Jordan Valley: A Social and Economic Survey* (1966), deals with the Ghawr Canal; K.B. DOHERTY, *Jordan Water Conflicts* (1965), on the political problems posed by the utilization of the Jordan's waters.

(P.Sa.)

Joseph II, Emperor

Joseph II, Holy Roman emperor from 1765 to 1790, was one of his century's "enlightened despots." He introduced numerous reforms in a vain attempt to modernize and unify the Habsburg domains. Born in Vienna on March 13, 1741, the eldest son of Maria Theresa of Austria and Francis Stephen of Lorraine (the future emperor Francis I), Joseph was strictly and thoroughly educated.

Early years

When Maria Theresa appointed him to the Council of State, he exhibited unusual intelligence and a passionate interest in politics. Joseph's first marriage in 1760 to the Bourbon princess Isabella of Parma, whom he loved passionately, ended in tragedy when she died of smallpox three years later. In 1765 he married Maria Josepha of Bavaria, who also died of smallpox in 1767. His inability to make decisions necessarily limited his ambition. After his father died, in 1765, he became emperor, but Maria Theresa still reserved all important decisions for herself.

By courtesy of the Kunsthistorisches Museum, Vienna

Joseph II, with Peter Leopold of Tuscany (left), painting by Pompeo Batoni, 1769. In the Kunsthistorisches Museum. Vienna.

After her death in 1780, Joseph tried to finish her work of reform. The educational system had been consolidated throughout the monarchy. For the University of Vienna, no longer under the influence of the church, Joseph tried to find the best scholars and scientists. The judiciary and the executive had already been separated at the top; Joseph extended this process to the lower administrative levels. In 1786 the Universal Code of Civil Law was issued. Under Maria Theresa the physician Gerard van Swieten had organized a public-health service, and in Joseph's time the General Hospital in Vienna was considered one of the best equipped in Europe. The monarchy's finances were balanced. The reorganization of the army secured Joseph's position in Europe. He ordered the abolition of serfdom; by the Edict of Toleration he established religious equality before the law, and he granted freedom of the press. The emancipation of the Jews within a short time endowed cultural life with new vitality. The artistic life of Vienna rose to new heights when the Burgtheater became the German National Theater. By transferring the management of the theatres to the actors, Joseph introduced an artistically fruitful concept.

Domestic reforms

Joseph's conflict with the Roman Catholic Church, however, posed more difficult problems. He established national training colleges for priests, deprived the bishops of their authority, and limited their communications with the Pope. The power of the church was even more affected by the dissolution of more than 700 monasteries not engaged in such useful activities as teaching or hospital work. The 36,000 monks forced to leave their orders were given an annuity or money to return home; those so returning could continue as secular priests. Some measures intended to forestall a relapse into monasticism, such as the foundation of new parishes, bore good results. The Emperor's impatience in turning the monks out of the monasteries, however, caused many works of art to be destroyed. At the climax of the crisis, Pope Pius VI visited the Emperor in Vienna, but the visit changed nothing; nor did a later journey by Joseph to Rome. Joseph's passionate zeal to change everything and to force a new form of life on his subjects met with embittered resistance, chiefly in such strongly traditional countries as the Austrian Netherlands and Hungary.

Failure in foreign affairs

In foreign policy, Joseph had obtained some success even as co-regent with his mother. When a civil war occurred in Poland under King Stanistaw I1 Poniatowski, the lover of Catherine II the Great of Russia, who was completely dependent on Russia, Joseph met with Poland's third neighbour, King Frederick the Great of Prussia, to plan the partition of Poland, with each neighbour taking a part of the country and the remaining part to be given a last chance at independence. Frederick took what was later West Prussia, Austrian Galicia, and Catherine took as much border territory as she thought necessary. In a later treaty with Turkey, Joseph annexed Bukovina to his country.

To obtain a personal view of the situation in eastern and western Europe, Joseph visited France, where he was enthusiastically received by the intellectual elite, and then also visited Catherine of Russia. The banquets given in his honour in Paris could not conceal the truth from him: France was headed for catastrophe. His Russian visit gave him the impression of a state retarded in its development compared with the West, but the loyalty of its enormous population to Catherine and her nearly unlimited power seemed to make her the best ally for political manoeuvres in Europe.

After his mother's death, Joseph had involved himself fruitlessly in 1784 in an attempt to force the Dutch to lift their blockade to secure a passage to the sea for the Austrian Netherlands. The Emperor hoped for more success with his unusual plan of exchanging the Austrian Netherlands for Bavaria. The Wittelsbach dynasty had been extinguished in Bavaria, and the heir, the count palatine Charles Joseph, was in favour of moving from Munich to Brussels. But Joseph left Prussia out of his calculations. Frederick protested, and his troops marched into Bohemia. The threat of war ended without a battle being fought, for in 1785 Frederick had formed the Fürstenbund (Princes' League) against Joseph to prevent the exchange. Deeply disappointed, Joseph now saw his only hope in Catherine. Though he was in bad health, he decided to visit her again; the Austrian Netherlanders and Hungarians, enraged at his reforms, resisted the move. Both publicly and secretly Catherine proposed a complete sharing of power in the east and southeast. Joseph signed an alliance giving her a free hand for her far-reaching plans, the conquest of Constantinople and the Dardanelles, and assuring Austria of substantial territorial gains. When Catherine declared war on Turkey sooner than expected, Joseph raised an army of 250,000 men. Yet, despite careful preparations, the organization of this large army was weak. Revolutionary unrest in the Austrian Netherlands and Hungary grew in the belief that preoccupation with the war would prevent the Emperor from taking on the revolutionaries as well. Joseph spent several months with his army; but both his illness and the domestic crisis made progress dangerous, and he had to return to Vienna before a victory could be won. There the Emperor attempted to establish peace in the Austrian Netherlands by delaying negotiations, but he failed in this as he did in Hungary, where his refusal to be crowned had deprived him of a legal foundation for his reign.

In Hungary topographical surveys and the replacing of Latin by German as the official legal language drove the Hungarian gentry into opposition, and in the Austrian Netherlands immigrants who had fled from Holland opened hostilities against the occupation forces, and finally the country declared its independence.

When Joseph died in Vienna on February 20, 1790, deserted by all his friends. his efforts to achieve freedom and general welfare for his subjects seemed to have been in vain. Today, however, Joseph is generally judged differently. He is considered the benefactor of his subjects, the people's emperor, who by his many voyages and inspections devoted himself to all matters of greater or lesser importance. In his century he had predicted there would be a time when general welfare took precedence over individual prosperity. But he was born one century too soon.

BIBLIOGRAPHY. PAVEL MITROFANOV, *Joseph II: Seine politische und kulturelle Tatigkeit, 2* vol. (1910; orig. pub. in Russian, 1907), long considered the most basic work about the life and times of Joseph II; SEBASTIAN BRUNNER, *Correspondances intimes de l'Empereur Joseph II avec . . . le comte de Cobenzl et . . . le Prince de Kaunitz* (1871), contains anecdotes and personal details that give a good picture of Joseph's life and personality: FRANCOIS FEJTO, *Un Habsbourg révolutionnaire, Joseph II* (1953; German trans., *Joseph II., Kaiser und Revolutionar: Ein Lebensbild,* 1956), the most valuable of recent biographies; SAUL K. PADOVER, *The Revolutionary Etnperor: Joseph the Second, 1741–1790* (1934), shows the significance of Joseph's reign in the context of German and European history; HAROLD W.V. TEMPERLEY, *Frederic the Great and Kaiser Joseph,* 2nd ed. (1968), a very good analysis of the contrast in character, temper, and significance between the two great opponents.

(F.Sc.)

Josephus, Flavius

Jewish priest and scholar, historian of the Jews, military commander against the Romans during the Jewish revolt of AD 66–70, Flavius Josephus was born Joseph ben Matthias in AD 37 or 38 of an aristocratic priestly family in Jerusalem. According to his own account, he was a precocious youth who by the age of 14 was consulted by high priests in matters of Jewish law. At age 16 he undertook a three-year sojourn in the wilderness with the hermit Bannus, a member of one of the ascetic Jewish sects that flourished in Judaea around the time of Christ. Returning to Jerusalem, he joined the Pharisees—a fact of crucial importance in understanding his later collaboration with the Romans. The Pharisees, despite the unflattering portrayal of them in the New Testament, were for the most part intensely religious Jews and adhered to a strict though nonliteral observance of the Torah. Politically, however, the Pharisees had no sympathy with the intense Jewish nationalism of such sects as the military patriotic Zealots and were willing to submit to Roman rule if only the Jews could maintain their religious independence.

Bv courtesv of the Hessische Landesbiblioihek. Fulda. West Germanv

IOSEPHVS
VESPASIANVS

Josephus before Vespasian, miniature portrait from a Josephus manuscript, 14th century. In the Hessische Landesbibliothek, Fulda, West Germany.

In AD 64 Josephus was sent on an embassy to Rome to secure the release of a number of Jewish priests of his acquaintance who were held prisoners in the capital. There, he was introduced to Poppaea Sabina, Emperor Nero's second wife, whose generous favour enabled him to complete his mission successfully. During his visit, Josephus was deeply impressed with Rome's culture and sophistication—and especially its military might.

Military commander in Galilee

He returned to Jerusalem on the eve of a general revolt against Roman rule. In AD 66 the Jews of Judaea, urged on by the fanatical Zealots, ousted the Roman procurator and set up a revolutionary government in Jerusalem. Along with many others of the priestly class, Josephus counselled compromise but was drawn reluctantly into the rebellion. Despite his moderate stance, he was appointed military commander of Galilee, where (if his own untrustworthy account may be believed) he was obstructed in his efforts at conciliation by the enmity of the local partisans led by John of Giscala. Though realizing the futility of armed resistance, he nevertheless set about fortifying the towns of the north against the forthcoming Roman juggernaut. The Romans, under the command of the future emperor Vespasian, arrived in Galilee in the spring of AD 67 and quickly broke the Jewish resistance in the north. Josephus managed to hold the fortress of Jotapata for 47 days, but after the fall of the city he took refuge with 40 diehards in a nearby cave. There, to Josephus' consternation, the beleaguered party voted to perish rather than surrender. Josephus, arguing the immorality of suicide, proposed that each man, in turn, should dispatch his neighbour. They were to cast lots, the first man killing the second, and so on. Curiously, Josephus contrived to draw the last lot, and, as one of the two surviving men in the cave, he prevailed upon his intended victim to surrender to the Romans.

Led in chains before Vespasian, Josephus assumed the role of a prophet and foretold that Vespasian would soon be emperor—a prediction that gained in credibility after the death of Nero in AD 68. The stratagem saved his life, and for the next two years he remained a prisoner in the Roman camp. Late in AD 69 Vespasian was proclaimed emperor, by his troops: Josephus' prophecy had come true, and the agreeable Jewish prisoner was given his freedom. From that time on, Josephus attached himself to the Roman cause. He adopted the name Flavius (Vespasian's family name), accompanied his patron to Alexandria, and there married his third wife. (His first wife had been lost at the siege of Jotapata, and his second had deserted him in Judaea.) Josephus later joined the Roman forces under the command of Vespasian's son and later successor, Titus, at the siege of Jerusalem in AD 70. He attempted to act as mediator between the Romans and the rebels, but, hated by the Jews for his apostacy and distrusted by the Romans as a Jew, he was able to accomplish little. Following the fall of Jerusalem and the destruction of the Temple, Josephus took up residence in Rome, where he devoted the remainder of his life to literary pursuits under imperial patronage.

Literary career

Josephus' first work, *Bellum Judaicum (History of the Jewish War)*, was written in seven books between AD 75 and 79, toward the end of Vespasian's reign. The original Aramaic has been lost, but the extant Greek version was prepared under Josephus' personal direction. After briefly sketching Jewish history from the mid-2nd century BC, Josephus presents a detailed account of the great revolt of AD 66–70. He stressed the invincibility of the Roman legions, and apparently one of his purposes in the works was to convince the Diasporan Jews in Mesopotamia, who may have been contemplating revolt, that resistance to Roman arms was pure folly. The work has much narrative brilliance, particularly the description of the siege of Jerusalem; its fluent Greek contrasts sharply with the clumsier idiom of Josephus' later works and attests the influence of his Greek assistants. In this work, Josephus is extremely hostile to the Jewish patriots and remarkably callous to their fate. The *Jewish War* is not only the principal source for the Jewish revolt but is especially valuable also for its description of Roman military tactics and strategy.

In Rome, Josephus had been granted citizenship and a pension. He was a favourite at the courts of Vespasian, Titus, and Domitian, and he enjoyed the income from a tax-free estate in Judaea. He had divorced his third wife, married an aristocratic heiress from Crete, and had given Roman names to his children. He had written an official history of the revolt and was loathed by the Jews as a turncoat and traitor. Yet despite all of this, Josephus had by no means abandoned his Judaism. His greatest work, *Antiquitates Judaicae (The Antiquities of the Jews)*, completed in *20* books in AD 93, traces the history of the Jews from Creation to just before the outbreak of the revolt of AD 66–70. It was an attempt to present Judaism to the Hellenistic world in a favourable light. By virtually ignoring the Prophets, by embellishing biblical narratives, and by stressing the rationality of Judaic laws and institutions, he stripped Judaism of its fanaticism and made it appealing to the cultivated and reasonable man. Historically, the coverage is patchy and shows the fatigue of the author, then in his middle 50s. But throughout, sources are preserved that otherwise would have been lost, and, for Jewish history during the Second Commonwealth, the work is invaluable.

Book XVIII of the *Antiquities* contains a celebrated reference to Christ. But the implication in the text of Christ's divinity could not have come from Josephus, and the passage undoubtedly represents the tampering (if not outright invention) of a later Christian copyist.

Appended to the *Antiquities* was a *Vita (Life)*, which is less an autobiography than an apology for his conduct in Galilee during the revolt. It was written to defend himself against the charges of his enemy Justus of Tiberias, who claimed that Josephus was responsible for the revolt. In his defense, he contradicted the account given in his more trustworthy *Jewish War*, presenting himself as a consistent partisan of Rome and thus a traitor to the rebellion from the start. Josephus appears in a much better light in a work generally known as *Contra Apionem* (*Against Apiort*, though the earlier titles *Concerning the Antiquity of the Jews* or *Against the Greeks* are more apposite). Of its two books, the first answers various anti-Semitic charges levelled at the Jews by Hellenistic writers, while the second provides an argument for the ethical superiority of Judaism over Hellenism and shows Josephus' commitment to his religion and his culture.

Since *Against Apion* mentions the death of Agrippa II, it is probable that Josephus lived into the 2nd century; but Agrippa's death date is uncertain, and it is possible that Josephus died earlier, in the reign of Domitian, sometime after AD 93.

Assessment

As a historian, Josephus shares the faults of most ancient writers: his analyses are superficial, his chronology faulty, his facts exaggerated, his speeches contrived. He is especially tendentious when his own reputation is at stake. His Greek style, when it is truly his, does not earn for him the epithet "the Greek Livy" that often is attached to his name. Yet he unites in his person the traditions of Judaism and Hellenism, provides a connecting link between the secular world of Rome and the religious heritage of the Bible, and offers many insights into the mentality of subject peoples under the Roman Empire.

Personally, Josephus was vain, callous, and self-seeking. There was not a shred of heroism in his character, and for his toadyism he well deserved the scorn heaped upon him by his countrymen. But it may be said in his defense that he remained true to his Pharisee beliefs and, being no martyr, did what he could for his people.

BIBLIOGRAPHY. The best scholarly text is that of B. NIESE in the "Teubner Series," 7 vol. (1887–95); a sound text with English translation may be found in the Loeb Classical Library edition, translated by H. ST. JOHN THACKERAY, R MARCUS and L.N. FELDMAN (1926–65). For Josephus' life and times, see H. ST. JOHN THACKERAY, *Josephus, the Man and the Historian* (1929); NORMAN BENTWICH, *Josephus* (1914, reprinted 1945); WILLIAM R. FARMER, *Maccabees, Zealots, and Josephus* (1956); R.J.H. SHUTT, *Studies in Josephus* (1961); and GEOFFREY A. WILLIAMSON, *The World of Josephus* (1964).

(Ga.P.)

Joyce, James

James Joyce's abilities as a novelist and his subtle yet frank portrayal of human nature coupled with his mastery of language and brilliant development of new literary methods have made him one of the most commanding influences on the writers of our time. Although originally banned in most countries, his novel *Ulysses*, describing the events of a single day (June 16, 1904) in Dublin, has come to be accepted as a major masterpiece, two of its characters, Leopold Bloom and his wife, Molly, being portrayed with a fullness and warmth of humanity unsurpassed in fiction. His semi-autobiographical novel *A* Portrait of the Artist as a Young Man is also remarkable for the intimacy of the reader's contact with the central figure and contains some astonishingly vivid passages. *Dubliners* is a collection of 15 short stories of Dublin life, mainly focussed upon its sordidness, but the last, "The Dead," is one of the world's great short stories. Critical opinion is divided over Joyce's last work, Finrzegans Wake, a universal dream about an Irish family, composed in a multilingual style on many levels and aiming at a multiplicity of meanings; but although seeming unintelligible at a first reading, the book is full of poetry and wit, containing passages of great beauty. Joyce's other works—some verse (Chamber Music, 1907; Pornes Penyeach, 1927; Collected Poems, 1936) and a play, Exiles (1918)—though competently written, add little to his international stature.

Gisele Freund

Joyce, photograph by Gisèle Freund, 1939.

Nearly all Joyce's works are imaginative reconstructions of his own life and early environment and, although too close an identification of the author with his fictional counterpart can mislead, a knowledge of his life history helps toward understanding his work.

John Stanislaus Joyce

Early life and work. James Augustine Joyce (he rarely used his middle name after about 1907) was born at Rathgar, Dublin, on February 2, 1882. At that time the burning question was that of Irish independence (Home Rule), and the Irish leader was Charles Stewart Parnell. John Stanislaus Joyce, James's father, was an ardent follower of Parnell, for whom he worked as election agent, an occupation for which his sociable temperament and ready tongue made him very suitable. In 1880 he had succeeded in getting two Parnellites returned as members of Parliament for Dublin and, partly as a reward, had been appointed collector of taxes for the city at the then considerable salary of £500 a year. Soon after, he married Mary Jane Murray, from Longford. With his salary, an income of £315 a year from inherited property, and what was left of his grandfather's 21st birthday present of £1,000, the young pair began married life in comfortable circumstances. James was their eldest son, and when he was six and a half (September 1888), he was sent to Clongowes Wood College, a Jesuit boarding school that has been described as "the Eton of Ireland." Such evidence as survives suggests that he was happy there. But his father was not the man to stay affluent for long; he drank, neglected his affairs, and borrowed money from his office. The dividing line came in 1890–91 with Pamell's fall following the scandal over the O'Shea divorce, and in John Joyce's mind the two events were always connected. He believed that, with his leader, he had been betrayed and suspected that the Catholic clergy were somehow responsible. The facts suggest a different interpretation: he had been absent from his office without permission while campaigning for the Parnellite candidates in Cork, and there were deficiencies in his accounts, which he eventually met by mortgaging his property. The new Irish Local Government Act had abolished his position, and he was not appointed to the equivalent one that replaced it, although he was allowed a pension. His son James shared his views and wrote, at the age of nine, a poem attacking T.M. Healy, who, with Michael Davitt, led the Irish opposition to Parnell. The poem was printed at his father's expense, but no copy is known to survive.

Education

During the years that followed, the Joyce family sank deeper and deeper into poverty. Ten children survived infancy, and they became accustomed to conditions of increasing soididness, subject to visits from debt collectors, having household goods frequently in pawn, and often moving to another house, leaving the rent and tradesmen's bills unpaid. James did not return to Clongowes after the summer vacation of 1891 but—apart from some months at a Christian Brothers' school, which he never afterward mentioned—stayed at home for the next two years and, according to his brother Stanislaus' autobiography, My Brother's Keeper (1958), tried to educate himself, asking his mother to check his work. In April 1893 both brothers were admitted, without fees, to Belvedere College, a Jesuit grammar school in Dublin. James did well there academically and was twice elected president of the Marian Society, a position virtually that of head boy. He left, however, under a cloud, as it was thought (correctly) that he had lost his Catholic faith.

He entered University College, Dublin, then staffed by Jesuit priests, although their control was limited by the Royal University Act of 1879. There he was taught neither theology nor philosophy; he studied languages, reserving his energies for extracurricular activities, reading widely—particularly in books not recommended by the Jesuits—and taking an active part in the college's Literary and Historical Society. Greatly admiring Henrik Ibsen, he learned Dano-Norwegian to read the original aad had an article, "Ibsen's New Drama"—a review of *When* We Dead Awaken—published in the London Fortnightly Review in 1899 just after his 18th birthday. This early success confirmed Joyce in his resolution to become a writer and persuaded his family. friends, and teachers that the resolution was justified. In October 1901 he published an essay, "The Day of the Rabblement," attacking the Irish Literary Theatre (later the Dublin Abbey Theatre) for catering to popular taste. Joyce had previously supported the theatre and had refused to join a students' protest against the "heresy" of William Butler Yeats's Countess Cathleen. His next publication was in the unofficial college magazine, Saint *Stephen's*, in May 1902. An essay, "James Clarence Mangan," on an Irish poet who, Joyce claimed, was unjustly neglected, it was written in an over-elaborate prose and based on a highly praised lecture he had given to the students' society.

The "epiphanies"

He was leading a dissolute life at this time but worked sufficiently hard to pass his final examinations, matriculating with "second class honours in Latin" and obtaining the degree of B.A. on October 31, 1902. Never did he relax his efforts to master the art of writing. He wrote verses and experimented with short prose passages that he called "epiphanies." The word means the manifestations, by the gods, of their divinities to mortal eyes; but Joyce used it to describe his accounts of moments when the real truth about some person or object was revealed. His experiments were useful in helping him to develop a concise style while recording accurate observation. His lifelong care in preserving his work for future use is first shown in the use he made of his "epiphanies" in his next two books.

To support himself while writing, he decided to become a doctor, but after attending a few lectures in Dublin, he borrowed what money he could and went to Paris. After a fortnight there, during which he found that fees were payable in advance and his Dublin qualifications insufficient, he returned home for a month's holiday on December 23, 1902, and became friendly with Oliver St. John Gogarty, whom he later pilloried as Mulligan in Ulysses. On returning to Paris he abandoned the idea of medical studies, wrote some book reviews, and, on the proceeds of these and of a few English lessons, with small remittances from his mother, he studied in the Sainte-Geneviève Library and compiled notes on a theory of aesthetics he was evolving from Aristotle, Aquinas, and Flaubert. The notes and the book reviews were published in Critical *Writings* of James Joyce (ed. by E. Mason and R. Ellmann, 19.59).

Recalled home in April 1903 because his mother was dying, he tried various occupations, including teaching, and lived at various addresses, including (from September 9–19, 1904) the Martello Tower at Sandycove, now Ireland's Joyce Museum. He had begun writing a lengthy naturalistic novel, Stephen *Hero*, based on the events of his own life, when in 1904 George Russell (Æ) offered $1 each for some simple short stories with an Irish background to appear in a farmers' magazine, *The* Irish Homestead. In response Joyce began writing the stories published as Dubliners (1914). Three stories, "The Sisters," "Eveline," and "After the Race," had appeared under the pseudonym Stephen Daedalus before the editor decided that Joyce's work was not suitable for his readers. Meanwhile Joyce had met, on June 10, a girl named Nora Barnacle. He met her next on June 16, the day that, mainly in celebration of their meeting, he chose as what is known as "Bloomsday" (the day of his novel Ulysses). On June 16 he fell in love with her, and eventually he persuaded her to leave Ireland with him, although he refused, on principle, to go through a ceremony of marriage.

Dubliners

Trieste—1905. Joyce and Nora left Dublin together in October 1904. Joyce obtained a position in the Berlitz School, Pola, working in his spare time at his novel and short stories. In 1905 they moved to Trieste, where James's brother Stanislaus joined them and where their children, George and Lucia, were born. In 1907, while giving English lessons to a Triestine businessman, Ettore Schmitz, he learned that Schmitz had written under the name Italo Svevo but had become discouraged. Joyce admired his work and exerted influence to obtain recognition for it. In 1906–07, for eight months, he worked at a bank in Rome, disliking almost everything he saw. Ireland seemed pleasant by contrast; he wrote to Stanislaus that he had not given credit in his stories to the Irish virtue of hospitality and began to plan a new story, "The Dead." The early stories were meant, he said, to show the paralysis from which Dublin suffered, but they are written with a vividness that arises from his success in making every word and every detail significant. His studies in European literature had interested him in both the Symbolists and the Realists; his work began to show a synthesis of these two rival movements. He decided that Stephen Hero lacked artistic control and form and rewrote it as "a work in five chapters" under a title—A Portrait of *the* Artist as a *Young* Man—intended to direct attention to its focus upon the central figure. In Trieste, too, Joyce wrote, but never tried to publish, a poetical short story, Giaconzo Joyce, describing his feelings toward one of his pupils, Amalia Popper, whose father was named Leopoldo and whose Jewish charms contributed to the character of Molly Bloom. The manuscript, found among Stanislaus' papers, was edited and published with a foreword by Richard Ellmann in 1968.

A Portrait
of the
Artist

In 1909 he visited Ireland twice to try to publish *Dubliners* and set up a chain of Irish cinemas. Neither effort succeeded and he was distressed when a former friend told him that he had shared Nora's affections in the summer of 1904. Another old friend proved this to be a lie. Joyce's reactions can still be read in a series of letters now in the Cornell University Library. Only the milder passages could be published in the "collected" Letters of *James* Joyce; the series as a whole includes some of the most astonishing examples of erotica ever written. Joyce always felt that he had been betrayed, and the theme of betrayal runs through much of his later work. So, almost equally, does that of tenderness. "The Dead" mentions a boy named Michael Furey who dies for love of the heroine, Gretta. He is based on a boy called Michael Bodkin, who had once courted Nora. On his next and final visit to Ireland in 1912, Joyce visited "the lonely churchyard on the hill where Michael Furey lay buried," and he brought its "crooked crosses" into the closing cadences of his story. But printers' objections to the themes, to the frequent mention of real places and people, and to occasional use of the word bloody prevented the publication of Dubliners till 1914, by which time A Portrait of the Artist was being serialized in The Egoist (a review financed by Harriet Shaw Weaver). Both publications were favourably reviewed.

Ziirich—1915. When Italy declared war in 1915 Stanislaus was interned, but James and his family were allowed to go to Zürich. At first, while he gave private lessons in English and worked on the early chapters of Ulysses—which he had first thought of as another short story about a "Mr. Hunter^{n}—his financial difficulties were great. He was helped first by a grant of E75 from the Royal Literary Fund; then by a large grant from Mrs. Edith Rockefeller McCormick; and finally by a series of grants from Miss Weaver, which by 1930 had amounted to more than $23,000. Her generosity resulted partly from her admiration for his work and partly from her sympathy with his difficulties, for, as well as poverty, he had to contend with eye diseases that never really left him. From February 1917 until 1930 he endured a series of 25 operations for iritis, glaucoma, and cataracts, sometimes being for short intervals totally blind. Despite this he kept up his spirits and continued working, some of his gayest passages being composed when his health was at its worst. Unable to find an English printer willing to set up A Portrait of *the* Artist for book publication, Miss Weaver published it herself, having the sheets printed in the United States, where it was also published, on December 29, 1916, by B.W. Huebsch, in advance of the English Egoist Press edition. Encouraged by the acclaim given to this, in March 1918, the American Little Review began to publish episodes from Ulysses, continuing until the work was banned in December 1920.

Ulysses

Paris—1920. After World War I Joyce returned for a few months to Trieste, then—at the invitation of Ezra Pound—in July 1920 he went to Paris. Ulysses was published there on February 2, 1922, by Sylvia Beach, proprietor of a bookshop called "Shakespeare & Co." The book, already well known because of the censorship troubles, became immediately famous. Joyce had prepared for its critical reception by having a lecture given by Valery Larbaud who pointed out the Homeric correspondences in it and that "each episode deals with a particular art or science, contains a particular symbol, represents a special organ of the human body, has its particular colour . . . proper technique, and takes place at a particular time." Joyce never published this scheme (see Table); indeed, he even deleted the chapter titles in the book as printed. It may be that this scheme was more useful to Joyce when he was writing than it is to the reader. Sometimes the technical devices become too prominent, particularly in the much-praised "Oxen of the Sun" chapter (II, 11) where the language goes through every stage in the development of English prose from Anglo-Saxon to the present day to symbolize the growth of a fetus in the womb. The execution is brilliant, but the process itself seems ill-advised. More often the effect is to add intensity and depth, as, for example, in the "Aeolus" chapter (II, 4) set in a newspaper office, with rhetoric as the "art." Joyce inserted into it hundreds of rhetorical figures and many references to winds—something "blows up" instead of happening, people "raise the wind" when they are getting money—

Science of Symbolism in Chapters *of Ulysses*

title	time	scene	bodily organ	art
I. Telemachia				
1 Telemachus	8 AM	Tower	—	Theology
2 Nestor	10 AM	School	—	History
3 Proteus	11 AM	Shore	—	Philology
II. Ulysses				
1 Calypso	8 AM	House	Kidney	Economics
2 Lotus Eaters	10 AM	Bath	Genitals	Botany, chemistry
3 Hades	11 AM	Graveyard	Heart	Religion
4 Aeolus	Noon	Newspaper office	Lungs	Rhetoric
5 Lestrygonians	1 PM	Lunch	Esophagus	Architecture
6 Scvlla & Charybdis	2 PM	Library	Brain	Literature
7 Wandering Rocks	3 PM	Streets	Blood	Mechanics
8 Sirens	4 PM	Concert room	Ear	Music
9 Cyclops	5 PM	Tavern	Muscle	Politics
10 Nausicaa	8 PM	Rocks	Eye, nose	Painting
11 Oxen of the Sun	10 PM	Hospital	Womb	Medicine
12 Circe	Midnight	Brothel	Locomotor apparatus	Magic
III. Nostos (Homecoming)				
1 Eumeus	I AM	Shelter	Nerves	Navigation
2 Ithaca	2 AM	House	Skeleton	Science
3 Penelope	—	Bed	Flesh	—

and the reader becomes aware of an unusual liveliness in the very texture of the prose. The famous last chapter, in which we follow the stream of consciousness of Molly Bloom as she lies in bed, gains much of its effect from being written in eight huge unpunctuated paragraphs. Nevertheless, the main strength of the book lies in its depth of character portrayal and its breadth of humour. Joyce claimed to have taken his "stream-of-consciousness" technique from a forgotten French writer, Édouard Dujardin (1861–1949), who had used the *monologue intérieure* in his novel *Les Lauriers sont coupés* (1888), but many critics have pointed out that it is at least as old as the novel, although no one before Joyce had used it so continuously.

Finnegans Wake

In Paris Joyce worked on *Finnegans Wake*, the title of which was kept secret, the novel being known simply as "Work in Progress" until published in its entirety in May 1939. In addition to his chronic eye troubles Joyce suffered great and prolonged anxiety over his daughter's mental health. What had seemed slight eccentricity grew into unmistakable and sometimes violent mental disorder that Joyce tried by every possible means to cure, but it became necessary to place her in a mental hospital near Paris. In 1931 he and Nora visited London, where they were married, his scruples having yielded to his daughter's complaints.

Meanwhile he wrote and rewrote sections of his new book; often a passage was revised more than 14 times before he was satisfied. Every word, every letter was scrutinized and pondered over. He usually began with a simple narrative. Basically the book is, in one sense, the story of a publican in Chapelizod, near Dublin, his wife, and their three children; but Mr. Humphrey Chimpden Earwicker, Mrs. Anna Livia Plurabelle, Shem, Shaun, and Isabel are every family of mankind, the archetypal family about whom all mankind is dreaming. The 18th-century Italian Giambattista Vico provides the basic theory that history is cyclic; to demonstrate this the book begins with the end of a sentence left unfinished on the last page. Ideally it should be bound in a circle. It is thousands of dreams in one. Languages merge: Anna Livia has "vlossyhair"—*włosy* being Polish for "hair"; "a bad of wind" blows; *bâd* being Turkish for "wind." Characters from literature and history appear and merge and disappear as "the intermisunderstanding minds of the anticollaborators" dream on. On another level, the protagonists are the city of Dublin and the River Liffey which flows enchantingly through the pages, "leaning with the sloothering slide of her, giddygaddy, grannyma, gossipaceous Anna Livia." And throughout the book James Joyce himself is present, joking, mocking his critics, defending his theories, remembering his father, enjoying himself.

Despite much scholarly study the book remains imperfectly understood; Joyce said he expected his readers to spend their lives on his book. It will remain a book for the minority, but it will always be loved by that minority. Since its publication it has had a great effect on many serious writers, as well as providing a new technique of word distortion and word creation for writers of advertisements.

After the fall of France in World War II (1940), Joyce took his family back to Zürich, where he died on January 13, 1941, still disappointed with the reception given to his last book. He would be pleased to know that, of the two periodicals now dealing with his work, one is entirely devoted to *Finnegans Wake*.

MAJOR WORKS

NOVELS AND STORIES: *Dubliners* (1914), 15 short stories: "The Sisters," "An Encounter," "Araby," "Eveline," "After the Race," "Two Gallants," "The Boarding House," "A Little Cloud," "Counterparts," "Clay," "A Painful Case," "Ivy Day in the Committee Room," "A Mother," "Grace," and "The Dead"; *A Portrait of the Artist as a Young Man* (1916); *Ulysses* (1922); *Finnegans Wake* (1939), sections published as parts of *Work in Progress* from 1928 to 1937.

VERSE: *Chamber Music* (1907); *Pomes Penyeach* (1927); *Collected Poems* (1936).

PLAY: *Exiles* (1918).

BIBLIOGRAPHY. The definitive biography, RICHARD ELLMANN, *James Joyce* (1959), is reliable and exhaustive. Critical studies of the works are very numerous; a list to December 1961 is given in R.H. DEMING, *A Bibliography of James Joyce Studies* (1964). The most concise account is A. WALTON LITZ, *James Joyce* (1966), which contains a well-selected bibliography with helpful comments. H. BLAMIRES, *The Rloomsday Book* (1966), is the best single guide for anyone reading *Ulysses* for the first time. FRANK BUDGEN, *James Joyce and the Making of Ulysses*, new ed. (1960), gives an intimate account of Joyce at work. For the earlier works MARVIN MAGALANER, *Time of Apprenticeship* (1959); and a collection of essays, *James Joyce's Dubliners* (1969), contain useful material; and R. SCHOLES and R.M. KAIN, *The Workshop of Daedalus* (1965), reprints Joyce's *Epiphanies* and the Pola and Trieste notebooks. No fully satisfactory account of *Finnegans Wake* has yet been written. W.Y. TINDALL, *A Reader's Guide to Finnegans Wake* (1969), is the most up-to-date comprehensive study but is not always reliable; ADALINE GLASHEEN, *A Second Census of Finnegans Wake* (1963), provides many necessary identifications. The *James Joyce Quarterly* (1963–), gives regular surveys of current critical work; *A Wake Newslitter* (1962–), now published by the English Department of Dundee University, Scotland, continues explicating *Finnegans Wake*.

(J.S.A.)

Juárez, Benito

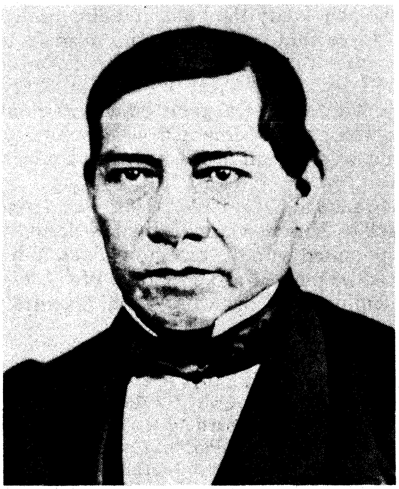
Benito Pablo Juárez, Mexico's first president of Indian descent, led his country's struggle to destroy the remnants of feudalism and to create a democratic, federal republic. To prevent foreign domination and conquest, Juárez successfully rallied the resistance of his countrymen. It may, in fact, be said that it was he who, after 50 years of nominal nationhood, brought the Mexican nation into being. To this day his countrymen revere him as their national hero.

Juárez was born in San Pablo Guelatao in the state of Oaxaca on March 21, 1806, of Indian parents, both of whom died when he was three years old. When he was 12 he left the uncle who was caring for him and joined his sister in the city of Oaxaca, where he began his formal education.

Early career

He originally studied for the priesthood, but in 1829 he entered the Oaxaca Institute of Arts and Sciences to study law and science. In 1831 he received his law degree and also won his first public office, a seat on the municipal council. Impeccably honest, he never used public office for personal gain, and his modest way of life reflected his simple tastes even after his marriage in 1843 to Margarita Maza, a Oaxaca girl 17 years younger than he. Politics soon became his life's work: he was a member of both the state and national legislatures, he became a judge in 1841, and he served as governor of his state, a post that brought him into national prominence.

During these years in politics Juárez began to formulate



Juárez.
By courtesy of the Library of Congress. Washington. D.C

solutions for his country's many problems. The road to economic health, he concluded, lay in substituting capitalism for the stifling economic monopoly held by the church and the landed aristocracy. He also believed that political stability could only be achieved through the adoption of a constitutional form of government based on a federal system.

But the return of the conservatives to power in the elections of 1853 doomed imminent reform. Many prominent liberals were exiled, including Jubrez. From December 1853 until June 1855 he lived in New Orleans in the United States in semipoverty, occupying himself by exchanging ideas with other Mexicans and laying plans to return home. The opportunity to put his ideas into action finally came in 1855 when the liberals took control of the national government, and Jubrez left the United States to join the new administration as minister of justice and public instruction.

The liberals carried out three major reforms, all supported by Jubrez. As minister of justice he was responsible for the law bearing his name that abolished special courts for the clergy and military, for he felt that juridical equality would help promote social equality. In June 1856 the government published the Ley Lerdo (called after the minister of finance), which, although it forced the church to sell its property, contained no threat of confiscation. By breaking up large landed estates, the government hoped that many Mexicans would be able to acquire property and thus create the middle class that it believed was essential for a strong and stable Mexico. The climax of the reform was the liberal constitution promulgated in February 1857.

In the same year, Ignacio Comonfort was elected president, and the new Congress chose Jubrez to preside over the Supreme Court and therefore, according to the constitution, also to serve as the effective vice president of Mexico. The court position was critical in determining his future career, for when the conservatives revolted and ousted Comonfort in January 1858, Juárez had a legal claim to the presidency. Lacking troops to control the area around Mexico City, however, he retired to the eastern port city of Veracruz.

At Veracruz Jubrez faced serious difficulties, for he had to create a government and hold it together through quarrels, betrayals, and defeat; to enforce and implement the constitution; to maintain armies in the field and defeat the conservative forces. But he was an extraordinarily tenacious and self-sufficient man, able to concentrate his energy and interest, and he proved himself the master of his government.

Because the clergy was supporting the conservatives against the legal government, Juárez enacted several laws to curb ecclesiastical power. He nationalized all church property, exempting only those buildings actually used for worship and instruction. To weaken clerical influence still further, he also nationalized the cemeteries and put birth registrations and marriages under the civil authority. Finally, the government separated church and state and guaranteed religious liberty to all citizens.

Presidency

By late 1860 the conservatives were faltering, and in January 1861 Jubrez was able to return to Mexico City and was constitutionally elected president. He was, however, faced with many serious problems: the opposition's forces still remained intact, the new Congress distrusted its president, and the treasury was virtually empty. As a solution to this latter problem Judrez decided in July 1861 to suspend payment on all foreign debts for two years. To safeguard their investments, England, Spain, and France decided to intervene, and by January 1862 the three countries had landed troops at Veracruz. But when Britain and Spain realized that Napoleon III intended to conquer Mexico and control it through a puppet, Archduke Maximilian of Austria, they withdrew their forces. Despite a major defeat at Puebla on May 5, 1862, reinforcements enabled the French to occupy Mexico City in June 1863, and Maximilian soon followed to take control of the government.

Forced to leave the capital again, Jubrez kept himself and his government alive by a long series of retreats that ended only at Ciudad Jubrez at the Mexican–American border. Early in 1867, as a result of continued Mexican resistance, increased United States pressure, and criticism at home, Napoleon decided to withdraw his troops. Soon afterward, Mexican forces captured Maximilian and executed him.

Jubrez then made the greatest mistake of his political career. In August 1867, shortly after his return to Mexico City, he issued a call for national elections and for a referendum on whether Congress should make five amendments to the constitution. Public opinion did not object to the President's running for re-election, but the constitutional changes aroused immediate and violent reaction in many quarters, including those sympathetic to Jubrez. Both procedure and content of his proposed changes came under fire because amendments enacted by Congress alone were unconstitutional and the changes would strengthen the executive power. Jubrez was re-elected, but the controversy had created such a crisis of confidence that the administration did not even bother to count the votes on the amendments.

Despite illness and personal loss—in October 1870 Juárez suffered a stroke and three months later his wife died—he decided to run again in 1871. After a bitter campaign he was re-elected, but many of his countrymen, refusing to accept the result as final, took up arms against him. Jubrez spent the last few months of his life trying to restore peace. On July 18, 1872, he died of a heart attack.

Assessment

Jubrez' political rise was a continual struggle to transform his liberal ideas into a permanent political reality and to overcome the prevalent social attitudes toward his Indian background. Only in the 20th century did the Mexicans come to admire and respect their Indian heritage; the prejudices of the 19th century serve to emphasize and enhance Juárez' extraordinary qualities and achievements. His domestic reforms, which freed Mexico from the most flagrant remnants of neocolonialism, and his leadership against the French earned Jubrez his place in the national pantheon.

BIBLIOGRAPHY. The best general work on Juárez is RALPH ROEDER, *Juarez and his Mexico*, 2 vol. (1947, reprinted 1968). The best biography, although uncritical, is CHARLES A. SMART, *Viva Juarez* (1963). with Juárez' important autobiographical essay, "Notes for My Children." WALTER V. SCHALES, *Mexican Politics During the Juarez Regime, 1855–1872* (1969), is a study of the domestic aspects of the Juárez government.

(W.V.S.)

Judah ha-Levi

Judah ha-Levi was a Hebrew poet and philosopher of the 11th and 12th centuries whose poetic works were the culmination of the development of Hebrew poetry within the sphere of Arabian culture. He was strongly influenced by Arabian literature, elements of which he ingeniously assimilated. His great collection of poems entitled *Dīwān* includes secular and religious poetry, both of which express passionate attachment to Zion (the land of Israel). For the poet, the Holy Land was not only a site where the Jewish people would one day gather after their deliverance from exile; immigration and settlement in Palestine would also hasten the coming of the Messiah. He celebrated Jerusalem in song as had none of his medieval predecessors. He also expounded his views on the nature of Judaism in an Arabic prose work consisting of dialogues between a learned Jew and the Khazar king who was converted to Judaism in the 8th century. It was widely circulated in Hebrew translation under the title *Sefer ha-Kuzari* ("Book of the Khazar").

Early life

Judah ha-Levi (in Hebrew, Yehuda ben Shemuel ha-Levi; Arabic surname, Abū al-Ḥasan) was born, at the latest, around 1075 in the town of Tudela in northern Spain and died in 1141 in Egypt. At the time of his birth, most of Spain, including his native town, was still under Muslim rule, but the Reconquista, the Christian sovereigns' struggle to regain the territories lost to the Muslims, was already under way. In 1085 King Alfonso VI of Castile conquered Toledo and made it his capital, and the exploits of the Cid, the celebrated national hero of Spain, also fall into the same period. Judah ha-Levi, whose poetic gifts manifested themselves unusually early, spent his childhood in the Christian part of the country. but even as a boy he felt himself drawn to Muslim Spain, then one of the principal cultural centres of Europe. He went to Andalusia in southern Spain some time before 1090, where he established contact with local Hebrew poets and intellectuals, and justly attracted considerable attention by his impressive talent. The most famous Hebrew poet of the time, Moses ibn Ezra from Granada, invited Judah ha-Levi to visit him and the two sealed a bond of lifelong friendship. His stay in Granada, enjoyed in the company of Ibn Ezra, was a period of success and happiness. He expressed his good spirits in several poems. This pleasant period ended in 1090 when Granada was stormed by the Almoravids, North African Berber disciples of a zealous Muslim movement, who now established an orthodox and intolerant regime in Andalusia. It is not known with any certainty whether Judah ha-Levi witnessed the Almoravid invasion in Granada or elsewhere, but the event greatly influenced the remainder of his life and his world view.

In his youth he also spent time in other Jewish centres of Andalusia, for example, in Lucena, a town of predominantly Jewish population in which a noted yeshiva, or academy, for Jewish theological studies was located. He composed a poetic epitaph when Isaac Alfasi, the head of the institution, died in 1103 and maintained very friendly relations with his successor, Joseph ibn Migash, for whom he even wrote letters. Judah ha-Levi also spent a certain amount of time in Seville, where he was poorly received by some wealthy Jews, on whom he revenged himself by denouncing their greed and ignorance in biting satirical verses. There are intimations in his poems that he must once have known material distress and depended on the good will of generous patrons.

Move to Christian Castile

He finally made his way, however, and became independent. Disappointed with the Almoravid regime, he turned toward Christian Castile and settled in its capital city of Toledo. There he worked untiringly as a physician, one of the professions open to Jews in Christian surroundings, a profession which in fact brought them into close contact with those surroundings.

As a resident of Toledo he celebrated prominent Castilian Jews in his verses, particularly the successful courtier Joseph ibn Ferruziel, better known by his Hispano-Arabic sobriquet Cidellus, who distinguished himself as a physician and adviser to King Alfonso VI. Judah ha-Levi for a while believed that the fortunes of his sorely tried people would flourish in Castile, but his hopes were destroyed by successive disappointments. Solomon ibn Ferruziel, a nephew of Cidellus who was also actively in the service of the Castilian state, was to return to Toledo from an important mission in Aragon. Along the way he was assassinated by Christian Spaniards on May 3, 1108. Judah ha-Levi had already composed a very elaborate poem to celebrate the reception of the Jewish statesman, which he had to set aside. He composed a long official elegy for the murdered man, ending it with a curse against the "Daughter of Edom," sinful Christianity. Additional acts of violence were committed against Jews in Castile, and still worse, it was often they who suffered in the clashes between the Almoravid realm and the Christian kingdoms in Spain. Distrusted, plundered, and slain by both sides, it was as though they were between hammer and anvil. Judah ha-Levi recognized the complete hopelessness of their situation and portrayed it in his poems.

Medieval Jews tried again and again to decipher the mysterious dates of their deliverance cited in the Book of Daniel and sought to apply them to their own time. Judah ha-Levi's works contain a reference to Daniel in a prophetic poem, in which the poet said that he had learned in a dream of the impending collapse of the Muslim empire in 1130. In the last years of his life he apparently returned in resignation to Muslim Spain and lived in Cordova, which remained an important centre of Jewish culture even in the period of decline. Judah ha-Levi had a very wide circle of acquaintances and maintained relationships with many famous contemporaries in Spain as well as abroad. He managed to gain a certain prosperity and lived in his house surrounded by a loving family and a few disciples. Yet he was thoroughly dissatisfied with his life. As old age approached he felt an increasing need to travel to Jerusalem, writing about it at length in verse and prose. The epilogue of the *Kuzari* explains his attachment to Zion and sounds like a farewell to Spain. Among his many poems celebrating the Holy Land is "Zionide" ("Ode to Zion"), his most famous work and the most widely translated Hebrew poem of the Middle Ages. He also carried on a heated controversy in verse with the opponents of his Zionist ideas.

Departure from Spain

He thought about and prepared for his journey to the Holy Land for many years. He was aided by a good friend, Halfon ha-Levi-Aldamyati, a very rich and cultivated Egyptian Jew whose trade relations extended as far as Yemen and India and who also frequently visited Spain. Judah ha-Levi left Spain in 1140. According to his carefully laid plans, he was first to embark for Egypt and then to proceed from there via the land route to Palestine. Aboard ship he composed a whole series of sea songs, which in both theme and mood represented a considerable innovation in Hebrew literature. His ship entered Alexandria Harbour on May 3, 1140, where he, along with a large Jewish party, was splendidly received. He was lodged in the magnificent home of Aaron ibn al-'ammānī, a noted Jewish physician and judge, and stayed in Egypt for several months. Many prominent Jews of the country came to admire him and to make his acquaintance, and he acquired many friends. From Alexandria he went to Cairo, or Fustat, the city where lived Samuel ben Hananiah, the Nagid, or head, of all Egyptian Jews, and there he was further acclaimed. Judah ha-Levi felt deep awe and humility in the land in which some of the biblical miracles had occurred and at the same time a kind of delight in all the beauties that revealed themselves to him. It seemed to him that his youth was restored; creative forces stirred within him and he wrote prolifically and easily. But he certainly always bore in mind his sacred destination and was often disturbed by the thought that death might yet intervene.

Death in Egypt

He did not in fact go beyond Egypt, although it is not known what detained him there. He died in July 1141 and was deeply mourned in Egypt. His death was romantically embellished in a legend that arose much later, according to which he was slain by a hostile Muslim just as he had arrived in Zion and was reciting his famous "Zionide." The legend found wide circulation and was repeated in detail by two well-known 19th-century poets, in German by Heinrich Heine in the *Romanzero* of 1851 and in Hebrew by Micah-Joseph Lebensohn in *Rabbi Yehudah ha-Levi* in 1869 (see also JEWISH PHILOSOPHY).

BIBLIOGRAPHY. J.H. SCHIRMANN'S biography of Judah ha-Levi (in Hebrew) in *Tarbiz*, 9:35–54, 219–240, 284–305, with corrections in vol. 10 and 11 (1937–40), is the only available, comprehensive study. Previous biographies are in-

complete and partly obsolete. SHELOMO D. GOITEIN discovered autographs of Judah ha-Levi and important documents concerning him in the manuscripts from the Cairo Geniza. He summarized his discoveries in "The Biography of R. Judah ha-Levi in the Light of the Cairo Geniza Documents," in *Proc. Am. Acad. Jewish Res.,* 28:41–56 (1959). For characteristics of Judah ha-Levi's personality, see J. JACOBS, "Judah Halevi: Poet and Pilgrim," in *Jewish Ideals and Other Essays* (1896); I. HEINEMANN, "Introduction," to the English translation of Judah ha-Levi's *Kuzari* (1947); S.W. BARON, "Yehudah Halevi: An Answer to an Historic Challenge," *Jewish Social Studies,* 3:243–272 (1941); J.M. MILLAS Y VALLICROSA, *Yehudd ha-Levi como poeta y apologistn* (1947); and Y.F. BAER, A *History of the Jews in Christian Spain,* vol. 1, pp. 59–77, 389–393 (1961, Eng. trans. from the 2nd Hebrew edition of 1959). The best existing edition of Judah ha-Levi's collected Hebrew poems (his *Dīwān*) has been published by H. BRODY, 4 vol. (1894–1930). The English translations by N. SALAMAN, *Selected Poems of Jehudah Halevi* (1924), with the Hebrew originals; by D. GOLDSTEIN, *Hebrew Poems from Spain,* pp. 89–120, 162–169 (1965); and the notes of FRANZ ROSENZWEIG to his German translation, 92 *Hymnen und Gedichte* (1926), are of interest. (J.Sc.)

Judaism

Judaism, the religion of the Jews, is the complex expression of a religious and ethnic community, a way of life as well as a set of basic beliefs and values, which is discerned in patterns of action, social order, and culture as well as in religious statements and concepts.

This article is divided into the following sections:

- Nature and characteristics
  - The religion of history
  - The "narrative theology" and later extrapolations
  - Basic affirmations, dispositions, and aims
- Basic beliefs and doctrines
  - God
  - Torah
  - Israel (the Jewish people)
  - Man
  - Ethics and society
  - The universe
  - Eschatology
- Basic practices and institutions
  - The hallowing of everyday existence
  - The traditional pattern of individual and familial practices
  - The traditional pattern of synagogue practices
  - Ceremonies marking the individual life cycles
  - The cycle of the religious year
  - Holy places: the land of Israel and Jerusalem
  - The sacred language: Hebrew and the vernacular tongues
  - The rabbi: legal, judicial, and congregational roles
  - Chief rabbinates
  - General councils or conferences
  - Modern variations
- Art and iconography
- Relation with non-Judaic religions
  - Exclusivist and universalist emphases
  - Relation to Christianity
  - Relation to Islām
  - Relations with other religions
- The role of Judaism in Western culture and civilization
  - Its historic role
  - Its present role
  - Future prospects
- Present-day forms of Judaism

NATURE AND CHARACTERISTICS

God's presence in history

The religion of history. It is history that provides the clue to an understanding of Judaism, for its primal affirmations appear in early historical narratives. Many contemporary scholars agree that although the biblical (Old Testament) tales report contemporary events and activities, they do so for essentially theological reasons. Such a distinction, however, would have been unacceptable to the authors, for their understanding of events was not superadded to but was contemporaneous with their experience or report of them. For them, it was primarily within history that the divine presence was encountered. God's presence was also experienced within the natural realm, but the more immediate or intimate disclosure occurred in human actions. Although other ancient communities saw a divine presence in history, this was taken up in its most consequent fashion within the ancient Israelite. community and has remained, through many developments, the focus of its descendants' religious affirmations. It is this particular claim—to have experienced God's presence in human events—and its subsequent development that is the differentiating factor in Jewish thought. As ancient Israel believed itself through its history to be standing in a unique relationship to the divine, this basic belief affected and fashioned its life-style and mode of existence in a way markedly different from groups starting with a somewhat similar insight. The response of the people Israel to the divine presence in history was seen as crucial not only for itself but for all mankind. Further, God had—as person—in a particular encounter revealed the pattern and structure of communal and individual life to this people. Claiming sovereignty over the people because of his continuing action in history on its behalf, he had established a *berit* ("covenant") with it and had required from it obedience to his Torah (teaching). This obedience was a further means by which the divine presence was made manifest–expressed in concrete human existence. The corporate life of the chosen community was thus a summons to the rest of mankind to recognize God's presence, sovereignty, and purpose—the establishment of peace and well-being in the universe and in mankind.

History, moreover, disclosed not only God's purpose but also manifested man's inability to live in accord with it. Even the chosen community failed in its obligation and had, time and again, to be summoned back to its responsibility by divinely called spokesmen—the prophets—who warned of retribution within history and argued and reargued the case of affirmative human response. Israel's role in the divine economy and thus Israel's particular culpability were dominant themes sounded against the motif of fulfillment, the ultimate triumph of the divine purpose, and the establishment of divine sovereignty over all mankind.

The "narrative theology" and later extrapolations. When the biblical text is thus viewed, the community's basic affirmations come into relief. A paradigmatic statement is made in the narrative that begins with Genesis and concludes with Joshua. In the early chapters of Genesis the divine is described as Creator of the natural order, including mankind. In the Eden, Flood, and Tower of Babel stories, man is recognized as rebellious and disobedient. In the patriarchal stories (about Abraham, Isaac, Jacob, and Joseph) a particular family is called out of humankind to restore the thwarted relationship through personal and communal responsiveness. The subsequent history of the community thus formed is recounted so that the divinely sought restoration may be recognized and the nature of the obedient community may be observed: the Egyptian servitude, the going out from Egypt, the revelation of the "teaching," the wandering years, and finally fulfillment through entrance into the "land" (Canaan). The prophetic books (in the Hebrew Bible these include the historical narratives up to the Babylonian Exile—*i.e.*, Joshua, Judges, Samuel, and Kings) continue to deal with the rebellion–obedience tension, interpreting it within the changing historical context and adding new levels of meaning to the fulfillment–redemption motif.

Philosophical, mystical, and exegetical additions

It is from this "narrative theology," as it has been recited throughout the centuries, that new formulations of the primal affirmations have been drawn. These have been clothed in a number of vocabularies: philosophical, mystical, ethnic, political, and others. The emphases have been various, the disagreements often profound. No single exposition has exhausted the possibilities of the affirmations or of the relationship between them. Philosophers have expounded them on the highest level of abstraction, using the language of the available philosophic systems. Mystics have enveloped them in the extravagant prose of speculative systems and in simple folktales. Attempts have been made to encompass them in theoretical ethical statements and express them through practical ethical behaviour. Yet, in each instance, the proposed interpretations have had to come to terms with the biblical affirmations and with the particular mode of understanding them required by the spiritual and intellectual demands arising out of the community's experience. The biblical texts, themselves the products of a long period of transmission and embodying more than a single outlook, were subjected to extensive study and interpretation over many centuries and, when required, were translated into other languages. This activity gave rise to the Midrashim (expository and homiletic commentaries; singular Midrash), the Targumim (expository translations into Aramaic; singular Targum), and to Greek translations that also involved interpretation and explanation. Side by side with the biblical text there also existed oral tradition, concerned particularly with the legal aspect of the community's life but also containing other matters. This oral material—brought together at the end of the 2nd century CE (Common Era = AD) in an authorized collection, the Mishna—was in subsequent centuries studied and commented upon, both in Palestine and in Babylon. The results of these two processes were embodied in the Palestinian and Babylonian Talmuds (commentaries on the Mishna). The whole literature continued to provide the basis of further developments, so that any attempt to formulate a statement of the affirmations of Judaism must, however contemporary it seeks to be, give heed to the scope and variety of speculation and formulation in the past (see TALMUD AND MIDRASH).

Basic affirmations, dispositions, and aims. Judaism is not and cannot be viewed as an abstract intellectual system, although some of its affirmations may be couched in such terms. It affirms divine sovereignty disclosed in creation (nature) and in history, without necessarily insisting upon—but at the same time not rejecting—metaphysical speculation about the divine. It insists that the community has been confronted by the divine not as abstraction but as person, with whom the community and its members enter into relationship. It is—as the concept Torah (connoting "guidance" or "direction") indicates—a program of human action, rooted in this personal confrontation. Further, the response of this particular people to its encounter with God is viewed as significant for all mankind. The community is called upon to express its loyalty to God and the Covenant by exhibiting solidarity within its corporate life on every level—including every aspect of human behaviour, from the most public to the most private. Thus, even Jewish worship is communal celebration of the meetings with God in history and in nature. Yet the particular existence of the Covenant people is not thought of as contradicting but rather as enhancing human solidarity. This people, together with all men, is called upon to create political, economic, and social forms that will affirm divine sovereignty—embody it in communal existence. This task is carried out in the belief not that man will succeed solely by his own efforts in these endeavours but that these sought-after human relationships have both their source and their goal in God—who assures their actualization. Within the sphere of his existence in the community, each Jew is called upon to realize the Covenant in his personal intention and behaviour.

The Judaic and the human

In considering the basic affirmations of Judaism from this point of view it is best to allow indigenous formulations rather than systematic statements borrowed from other traditions to govern the presentation.

BASIC BELIEFS AND DOCTRINES

God. An early statement of basic beliefs and doctrines emerged in the liturgy of the synagogue some time during the last pre-Christian and first Christian centuries, although there is evidence that such formulations were not absent from the Temple cult that came to an end in the year 70 CE. A section of the Siddur (order of worship, or prayerbook) that has as its focus the recitation of a series of biblical passages (Deut. 6:4–9; Deut. 11:13–21; Num. 15:37–41) takes its name from the first of these, Shema ("Hear"): "Hear, O Israel! the Lord is our God, the Lord alone" (or ". . . the Lord our God, the Lord is one"). In the Shema—often regarded as the Jewish confession of faith, or creed—the biblical material and accompanying benedictions are arranged to provide a unified statement about God and his relationship to the world and Israel, as well as Israel's obligations toward and response to God. In this statement, God, who is the Creator of the universe and who has chosen Israel in love ("Blessed art thou, O Lord, who has chosen thy people Israel in love"), expressed by the giving of Torah, is declared to be "one"; his love is to be reciprocated by men who lovingly obey Torah and whose obedience is rewarded and rebellion punished. The goal of this obedience is God's "redemption" of Israel, a role foreshadowed by his action in bringing Israel out of Egypt.

Unity and uniqueness. At the centre of this liturgical formulation of belief is the concept of the divine unity. In its original setting, it may have served as the theological statement of the reform under Josiah, king of Judah, in the 7th century BCE (Before the Common Era = BC) when worship was centred exclusively in Jerusalem, and all other cultic centres were rejected, so that the existence of one shrine only was understood as affirming one deity. The idea, however, acquired further meaning. It was understood toward the end of the pre-christian era to proclaim—over against dualistic religious formulations in the Greco-Roman world—the unity of divine love and divine justice, as expressed in the divine names Yahweh and Elohim, respectively. A further expansion of this affirmation is found in the first two benedictions of this liturgical section, which together proclaim that the God who is the Creator of the universe and the God who is Israel's ruler and lawgiver are one and the same—as over against religious positions that insisted that the Creator God and the lawgiver God were separate and even inimical. Subsequently, this affirmation was developed in philosophical and mystical terms by both medieval and modern thinkers.

Divine love and justice

Creativity. As has been noted, this "creed," or "confession of faith," underscores in the first benediction the relation of God to the world as that of Creator to creation. "Blessed art thou, O Lord our God, King of the Universe, who forms light and creates darkness, who makes peace and creates all things." It adds the assertion that his activity is not in the past but is ongoing and continuous, for "he makes new continually, each day, the work of creation"; thus, unlike the deity of the Stoic world view, he remains actively present in nature. This "creed" is concerned as well to come to terms with the ever-present problem of evil. Paraphrasing Isa. 45:7, "I form the light and create darkness; I make peace, and create evil," it changes the last word to "all" (or "all things") rather than "evil." The change was clearly made to avoid the implication that God is the source of moral evil. Judaism, however, did not flinch from confronting the problem of pain and suffering in the world and affirming the paradox of suffering and divine sovereignty, of pain and divine providence, refusing to accept the concept of a partial God—a God that is Lord over only the harmonious and pleasant aspects of reality.

Divine Providence

Activity in the world. The second and the third benedictions deal with divine activity within the realm of history and human life. God is teacher of men through the giving of instruction (Torah); he acts in the life of mankind in historical events; he has chosen a particular people—Israel—in love to witness to his presence and his desire for a perfected society; he will, as redeemer, enable man to experience that perfection. These activities, together with creation itself, are understood to express divine compassion and kindness as well as justice (judgment), recognizing the sometimes paradoxical relation between them. Taken together, they disclose Divine Providence—God's continual activity in the world. The constant renewal of creation (nature) is itself an act of compassion overriding strict justice and affording mankind further opportunity to fulfill the divinely appointed obligation. Yet the basically moral nature of God is asserted in the second of the biblical passages that form the core of this liturgical statement (Deut. 11:13-21). Here, in the language of its agricultural setting, the community is promised reward for obedience and punishment for disobedience. The intention of the passage is clear: obedience is rewarded by the preservation of order, so that *the* community and its members find wholeness in life; while disobedience—rebellion against divine sovereignty—shatters older, so that the community is overwhelmed by adversity. The passage of time has made the original language unsatisfactory (promising rain, crops, and fat cattle), but the basic principle remains, affirming that, however difficult it is to recognize the fact, there is a divine law and judge. Support for this affirmation is drawn from the third biblical passage (Num. 15:37–41), which explains that the fringes the Israelites are commanded to wear on the corners of their garments are reminders to observe the commandments of God, who brought forth Israel from Egyptian bondage. The theme of divine redemption is elaborated in the concluding benediction to point toward a future in which the as yet fragmentary rule of God will be brought to completion: "Blessed is his name whose glorious kingdom is for ever and ever."

Otherness and nearness. Within this complex of ideas, other themes are interwoven. In the concept of the divine Creator there is a somewhat impersonal or remote quality—of a power above and apart from the world—which is underscored by such expressions as the trifold declaration of God's holiness, or divine otherness, in Isaiah 6:3: "Holy, holy, holy is the Lord of hosts. . . ." The development of surrogate divine names for biblical usage, as well as the substitution of Adonai ("my Lord") for the tetragrammaton (YHWH) in the reading of the Bible itself, suggests an acute awareness of the otheiness of God. Yet this has as its countertheme the affirmation of divine nearness. In the biblical narrative it is God himself who is the directly active participant in events, an idea that is emphasized in the liturgical narrative (Haggada) recited during the Passover meal (seder): "and the Lord brought us forth out of Egypt—not by an angel, and not by a seraph, and not by a messenger. . . ." The surrogate divine name *Shekhina*, the Present One, is derived from a Hebrew root meaning "to dwell," again calling attention to divine nearness ("presence"). The relationship between these two affirmations, otherness and nearness, is expressed in a Midrashic statement, "in every place that divine awesome majesty is mentioned in Scripture, divine abasement is spoken of, too."

The divine "thou"

Closely connected with these ideas is that of divine personhood, most particularly disclosed in the use of the pronoun "thou" in direct address to God. The community and the individual, confronted by the Creator, teacher, redeemer, addresses the divine as living person, not as theological abstraction. The basic liturgical form, the *berakha* ("blessing"), is usually couched in the second person singular: "Blessed art thou. . . ." This relationship, through which remoteness is overcome and presentness is established, illuminates creation, Torah, and redemption, for it reveals the meaning of love. From it flow the various possibilities of expressing the divine–human relationship in personal, intimate language. Sometimes, especially in mystical thought, such language becomes extravagant, foreshadowed by such vivid biblical metaphors as the husband–wife relation in Hosea; the "adoption" motif in Ezek. 16; and the firstborn-son relation (Ex. 4:22). Nonetheless, although terms of personal intimacy are used widely to express Israel's and man's relationship with God, such usage is restrained by the accompanying sense of divine otherness. This is to be seen in the liturgical "blessings," where, following the direct address to God, in which the second person singular pronoun is used, the verbs, with great regularity, are in the third person singular, thus providing the requisite tension between nearness and otherness, between the impersonal and the personal.

Modern views of God. The Judaic affirmations about God have not always been given the same emphasis nor have they been understood in the same way. This was true in the Middle Ages, among both philosophers and mystics, and in modern times down to the 1970s. In the 19th century, western European Jewish thinkers attempted to express and transform these affirmations in terms of German Idealist philosophy: more recently, philosophical Naturalism was offered as the suitable content of Judaism, while still retaining the traditional God language. The meaningfulness of the whole body of such affirmations, moreover, has been called into question by the philosophical schools of Logical Positivism and Linguistic Analysis. Most recently, the destruction of 6,000,000 Jews during the Nazi period has raised the issue of the validity of such concepts as God's presence in history, divine redemption, the covenant, and the chosen people. In every case, however, it is with the structure of ideas here noted that these challenges must deal.

The divine teaching or guidance

Torah. The concept "Giver of Torah" played a central role in the understanding of God, for it is Torah, or "teaching," that confirms the events recognized by the community as the act of God. Thus, for example, in the synagogue's reckoning of the Ten Commandments (more accurately, the Ten Utterances), a paradigmatic expression of Torah, the first is the historically oriented statement: "I am the Lord your God, who brought you out of the land of Egypt, out of the house of bondage" (Ex. 20:2). This provides the basis for the binding character of the provisions that follow, to which Israel is now to respond in obedience.

The meanings of Torah. The word Torah (plural *torot*) comes from a Hebrew root meaning "to point the way, give direction." While this sense is clearly suggested by the history recounted in the Bible, the term must be understood as involving more than guidance through the tangle of events. It implies the pattern of behaviour called for on every level of communal and individual living because of the ultimate normative guidance that has been given. Further, the source of the pattern is acknowledged to be divine, an affirmation conceptualized in the phrase *Torah min ha-shamayim,* "Torah from Heaven." When these two aspects of the term are recognized, it becomes evident that to translate Torah by the word law (as is commonly done) is, while not entirely erroneous, a grave distortion of its full meaning. Viewed from this larger perspective, Torah is the entire content of Judaism: its sacred Scriptures; its oral traditions; its theological affirmations; its historical recollections; its ethical obligations; its ritual and ceremonial observances; and its interpretations of its authoritative texts. It also has more restrictive or specific references. As the New Jewish Version of *The Torah: The Five Books of Moses* (1962) indicates, it may mean ritual prescription in a very narrow sense—"This is the ritual [*torah*] of the burnt offering," (Lev. 6:2)—or legal prescription—"These are the laws, rules, and directions [*torot*] . . ." (Lev. 26:46).

Written and oral Torah. A wider and more general usage is as the name for the Pentateuch (the five books of Moses) in the threefold division of Scriptures: Torah, Nevi'im *(Prophets),* Ketuvim (Writings). This reflects a narrowing of the concept "Torah from Heaven" to "Torah (given) to Moses at Sinai," although even the latter has been understood to include more than the five books of Moses. Nonetheless, traditionally the concept was understood to affirm not only the divine source but also the Mosaic authorship of Genesis, Exodus, Leviticus, Numbers, and Deuteronomy.

The believed Sinaitic origin of oral Torah

As indicated, while the phrase "Torah (given) to Moses at Sinai" may be understood in a restricted sense, the Pharisaic–rabbinic tradition (originated by the Pharisees and continued by the Talmudic rabbis) viewed it as referring to a wide body of teaching. According to this position, which dominated Jewish thought until the modern era and still commands the allegiance of traditionalists, the encounter between God and Israel at Sinai deposited with Israel not only a written Torah *(Torah she-bikhtav)* but also an oral Torah *(Torah she-beʿal* pe) that was transmitted from generation to generation.

Oral Torah is not merely the tradition brought together in the Mishna. It also comprises the interpretations of the written Torah, collected in the Midrashic literature, and the commentaries (Gemara) on the Mishna, collected in the Talmud. Over the centuries much of this oral tradition, eventually set down in writing, was commented upon so that a vast literature grew up embodying many levels of interpretation and reflecting the varying experiences and situations of the Jewish community.

Central to this vast structure was, of course, the Jewish community's concern to live in accordance with the divine will embodied and expressed in Torah in the widest sense. Scripture, Halakhic and Haggadic Midrash, Mishna, and Gemara were the sources from which the leaders of the communities drew in order to provide both stability and flexibility. The dispersion of the Jews outside Palestine confronted communities and individuals with novel and unexpected situations that had to be dealt with in such a way as to provide continuity while at the same time making it possible to exist with the unprecedented.

Prophecy and religious experience. Torah in the broad sense includes the whole Hebrew Bible, including the prophetic books. In biblical prophecy, God is seen as continuing to be disclosed in the nexus of historical events and as making ethical demands upon the community. According to rabbinic Judaism, this source of Torah—the charismatic person—dried up in the period of Ezra (*i.e.*, about the time of the return from the Babylonian Exile in the 5th century BCE). This opinion may have been a defensive reaction to the luxuriant growth of apocalyptic speculation about the end of the world and the kingdoms of this world, a development that was considered dangerous and unsettling in the period after the Bar Kokhba revolt (132–135 CE). Indeed, there appears to have developed an ongoing suspicion that unrestrained individual experience as the source of Torah was inimical to the welfare of the community. Such an attitude was by no means new. Deuteronomy (13:2–19) had already warned against such "misleaders." The culmination of this attitude is to be found in a Talmudic narrative in which even the bat qol, the divine "echo" that announces God's will, is ignored on a particular occasion. Related to this is the reluctance on the part of teachers in the early Christian centuries to point to wonders and miracles in their own time. Far from expressing an ossification of religious experience—the development of the Siddur and the Talmudic reports on the devotional life of the rabbis contradict such an interpretation—the attitude seems to be a response to the development of such religious enthusiasm as was exhibited, for example, in the behaviour of the Christian Church in Corinth (I Cor.) and among Gnostic sects and sectarians. Thus, even among the speculative mystics of the Middle Ages, where allegorization of Scripture abounds, the structure of the community and the obligations of the individual are not displaced by the deepening of personal religious life through mystical experience. The decisive instance of this is Joseph Karo (16th century), who was thought to be in touch with a supernal guide, but who was, at the same time, the author of an important codification of Jewish law, the Shulhan *ʿarukh.*

Anticharismatic trends in postbiblical Judaism

Admittedly, there have been occasions when Torah, even in the wide sense, has been rigidly viewed and applied. In certain historical situations, the dynamic process of rabbinic Judaism has been treated as a static structure. What is of greater significance, however, is the way in which this tendency toward inflexibility has been checked and reversed by the inherent dynamism of the rabbinic tradition.

Modern views of Torah. In modern times—since the end of the 18th century—the traditional position has been challenged both in detail and in principle. The rise of biblical criticism has raised a host of questions about the origins and development of Scripture and thus about the very concept of Torah, in the senses in which it has functioned in Judaism. Naturalistic views of God have required a reinterpretation of Torah in sociological terms as the ideals and sancta (holy things) of the Jewish people. Other and varying positions of many sorts have been and undoubtedly will be forthcoming. What is crucial, however, is the concern of all these positions to retain—with whatever modifications are required—the concept of Torah as one of the central and continuing affirmations of Judaism.

Israel (the Jewish people). *Choice and covenant.* The concluding phrase of the second benediction of the liturgical section referred to above reads: "who has chosen thy people Israel in love." Here the basis of the relationship between God and Israel set forth in the biblical narrative is clearly and succinctly stated: the choice of this people was determined by no other factor than divine love. The patriarchal narratives, beginning with the 12th chapter of Genesis, presuppose the choice, which is set forth explicitly in Deut. 7:6–8 in the New Jewish Version:

> For you are a people consecrated to the Lord your God: of all the peoples on earth the Lord your God chose you to be His treasured people. It is not because you are the most numerous of peoples that the Lord set His heart on you and chose you—indeed you are the smallest of peoples; but it was because the Lord loved you and kept the oath He made with your fathers that the Lord freed you with a mighty hand and rescued you from the house of bondage, from the power of Pharaoh king of Egypt.

Later rabbinic traditions on occasion sought to base the choice upon some special merit of Israel, and the medieval poet and theologian Judah ha-Levi suggested that the openness to divine influence originally present in Adam continued only within the people of Israel.

However understood, the background of this choice is the recurring disobedience of mankind narrated in Genesis 2–11. Abraham and his descendants are singled out not merely as the object of the divine blessing but also as its channel to all mankind. The choice, however, requires a reciprocal response from Abraham and his lineage. That response is obedience, as exemplified in the first instance by Abraham's readiness to leave his "native land" and "father's house" (Gen. 12:1). This twofold relationship was formalized in a mutually binding agreement, a covenant between the two parties. The covenant, thought by some modern biblical scholars to reflect the form of ancient suzerainty treaties, indicates (as in the Ten Utterances) the source of Israel's obligation—the acts of God in history—and the specific requirements those acts impose. The formalization of this relationship was accomplished by certain cultic acts that may, according to some contemporary scholars, have been re-enacted on a regular basis at various sacred sites in the land, eventually being centralized in Jerusalem. The content of the covenantal obligations thus formalized was Torah. Israel was bound in obedience, and Israel's failure to obey provided the occasions for the prophetic messages. The prophets, as spokesmen for God, called the community to renewed obedience, threatened and promised disaster if such was not forthcoming, and—recalling the source of the choice in divine love—sought to explain its persistence even when, strictly understood, the covenant should have been repudiated by God.

The binding agreement between God and Israel

The choice of Israel has its concrete expression in the requirements of the precepts (mitzwot, singular mitzwa) that are part of Torah. The blessing recited before the public reading of the pentateuchal portions on Sabbath, festivals, holy days, fasts, and certain weekdays refers to God as "He who chose us from among all the peoples and gave us His Torah," thus emphasizing the intimate relationship between the elective and revelatory aspects of God.

Priest and witness to the nations

Israel's role was not defined solely in terms of its own obedience to the commandments. As noted, Abraham and his descendants were seen as the means by which the estrangement of disobedient mankind from God was to be overcome. Torah was the formative principle underlying the community's fulfillment of this obligation. Israel was to be "a kingdom of priests and a holy nation" (Ex. 19:6) functioning within mankind and for its sake. This task is enunciated with particular earnestness in the writings of the prophets. In Isa. 43–44, Israel is declared to be God's witness and servant who is to bring the knowledge of God to the nations. In chapter 42 of the same book Israel is declared to be a "covenant of the people, a light to the nations, to open the blind eyes, to bring out the prisoners from the prisons, and them that sit in darkness out of the prison house" (42:6–7). This double motif of a chosen people and witness to the nations, joined to that of the righteous king, developed in the biblical and postbiblical periods into messianism in its several varieties (see below Eschatology).

The intimate relation between choice, covenant, and Torah determined the modality of Israel's existence. Religious faith, far from being restricted to or encapsulated in the cult, found its expression in the totality of communal and individual life. The obligation of the people was to be the true community, in which the relationship between its members was open, in which social distance was repudiated, and in which response to the divine will expressed in Torah was called for equally from all. One of the recurring themes of the prophetic movement was the rejection of any tendency to limit divine sovereignty to the partial area of "religion," understood as the realm of the priesthood and cult. Subsequent developments continued this theme, albeit in other forms. Pharisaic Judaism and its continuation, rabbinic Judaism, down to modern times resolutely held to the idea of the all-pervasive functioning of Torah, so that however the various Jewish communities over the centuries may have failed to fulfill the ideal, the self-image of the people was that of "holy community."

Israel and the nations. The double motif of "treasured people" and "witness" was not without its tensions as it functioned in ongoing history. Tensions are especially visible in the period following the return from the Babylonian Exile at the end of the 6th and beginning of the 5th centuries BCE. It is, however, doubtful whether the use of such terms as nationalism, particularism, or exclusivism (as opposed to universalism) are of help in understanding the situation. Emphasis has, for example, been laid upon Ezra 9:2 and 10:2, in which the re-established community is commanded to give up wives taken from "the peoples of the land." This is taken as indication of the narrow, exclusivistic, nationalistic nature of Judaism, without reference to the situation in which a harassed contingent of returned exiles sought to maintain itself in a territory surrounded by politically unfriendly if not hostile neighbours. Nor does this represent racialism, since foreigners were admitted to the Jewish community, and in the following centuries some groups engaged in extensive missionary activities, appealing to the nations to join themselves to the God of Israel, who was the one true God, the Creator of the heavens and earth.

The welding of exclusivism and universalism

A more balanced view recognizes that within the Jewish community religious universalism was affirmed at the same time and by the same people who understood the nature of Jewish existence in politically particularistic (*i.e.*, nationalistic) terms. To neglect either side is to distort the picture. In no case was the universalism disengaged from the reality of the existing community, even when it was expressed in terms of the ultimate fulfillment of the divine purpose, the restoration of the true covenantal relationship between God and all mankind. Nor was political particularism, even under circumstances of great provocation and resentment, misanthropic. The most satisfactory figure in describing the situation of the restored community, and one that continues to be useful in dealing with later episodes, is that of the human heartbeat, made up of two functions, the systole, or contraction, and the diastole, or expansion. There have been periods of contraction and of expansion in the history of Judaism. The emphasis within the abiding tension has been determined by the historical situation in which the community found itself. To generalize in one direction or the other is fatal to an understanding of the history and faith of the "holy community."

The people and the land. Closely related to the concept of Israel as the chosen, or Covenant, people is the role of the land of Israel. In the patriarchal stories, settlement in Canaan is an integral part of the fulfillment, from the divine side, of the Covenant. The goal of the Israelites escaping from Egypt is the same land, and entry into it is understood in the same fashion. The return from the Babylonian Exile, too, is seen in the same light. As there was the choice of a people, so was there the choice of a land—and for much the same reason. It was to provide the setting in which the community could come into being as it carried out the divine commandments. This choice of the land contrasts with the predominant ideas of other peoples in the ancient world, in which the deity or divinities were usually bound to a particular parcel of ground outside of which they lost their effectiveness or reality. Though some such concepts may have crept into Israelite thought during the period of the kings (from Saul to Jehoiachin), the crisis of the Babylonian Exile was met by a renewal of the affirmation that the God of Israel was, as Lord of all the earth, free from territorial restraint, although He had chosen a particular territory for this chosen people. Here again the twofold nature of Jewish thought is apparent, and both sides are to be affirmed or the view is distorted. Following the two revolts against Rome (66–73 CE and 132–135 CE), the Jews of the ever-widening dispersion continued, as they had before these disasters, to cherish the land. Once again it became the symbol of fulfillment, so that return to it was looked upon as an integral part of messianic restoration. The liturgical patterns of the community, insofar as they were concerned with natural phenomena (*e.g.*, planting, rainfall, harvest, and the annual cycle) rather than historical events, were based on geography, topography, and agricultural practices of the land, viewed as paradigmatic. Although Jews continued to live in the land, yet for most in the distant dispersion it was idealized and viewed primarily in eschatological terms—at the end of days, in the world to come. The 11th-century poet Judah ha-Levi not only longed for it in verse but also gave it a significant role in his theological interpretation of Judaism and eventually sought to return to it from his native Spain. It was not, however, until the 19th century that the land began to play a role other than the goal of pilgrimage or of occasional settlement by pietists and mystics. At the end of that century the power of the utopian concept was released in eastern Europe in a cultural renaissance that focussed, in part, on a return to the land and, in western and central Europe, in a political movement coloured by nationalist motifs in European thought. The coming together of these two gave rise to Zionism. The political movement reflected a dissatisfaction with the view of the Jews as merely a body or organization of religious believers—like the Christian churches—an interpretation that had become dominant following the political emancipation of the Jews in the period after Napoleon. The political emphasis of Zionism aroused considerable opposition from those Jews who were convinced of the necessity of a churchly definition of Judaism parallel to the Roman Catholic and Protestant communions. While this conflict erupted in bitter debate during the first half of the 20th century, the events of the Nazi period in Europe brought it to a close, except for some sporadic renewals on the part of numerically insignificant groups. For the most part, although there are few satisfactory formulations—theological or secular—there is a working consensus that acknowledges a significant role to the land and recognizes that a churchly definition of the Jewish community, while strategically acceptable in some situations, does not do justice to history and is not theologically sound if it suggests that Judaism merely consists of abstract doctrines and dogmas. Some Jews, however, argue that whatever the past has been, the future of the Jewish community is with those movements in the modern world that cut across or transcend the particularity Zionism represents.

The land as ideal and reality

Modern reinterpretations or rejections of chosenness

Modern views of the people Israel. The nature of the people Israel and of the land of Israel has been variously interpreted in the history of Jewish thought. In modern times, some interpretations have been deeply influenced by contemporary political and social discussions in the general community. Thus, for example, Zionist theoreticians were influenced by concepts of political nationalism on the one hand and by socialist ideas on the other. Further, the challenge to traditional theological concepts in the 19th century raised issues about the meaning of the choice of Israel, and Jewish thinkers borrowed from romantic nationalism such ideas as the "genius" of the people. Most recently, attempts have been made to approach the question sociologically, dismissing the theological mode as unhelpful. The concept of the chosen people is then understood to indicate a specific role deliberately undertaken by the Jewish people and similar to that espoused by other groups (*e.g.*, "Manifest Destiny" by the

American people). The establishment of the State of Israel has motivated some thinkers to call for a repudiation of the idea, in keeping with the position that normal existence for the Jews requires the dismissal of such concepts. Although only a small minority of Jewish thinkers espouse this position, the doctrine of the choice is not without its theological difficulties even for those who continue to affirm it.

Man. *The image of God.* In Gen. 1:26, 27; 5:1; and 9:6 two terms occur, "image" and "likeness," that seem to indicate clearly the biblical understanding of man's essential nature: he is created in the image and likeness of God. Yet the texts in which they are used are not entirely unambiguous; the idea they point to does not appear elsewhere in Scriptures; and the concept is skirted cautiously in the rabbinic interpretations. What the image and likeness of God or the divine image refer to in the biblical text is not made explicit, and, in the light of the psychosomatic unity of man that dominates the biblical concepts, it is not possible to escape entirely from the implication of "bodily" similarity. What the terms meant in their context at the time and whether they reflect mythological usages taken over from other Near Eastern thought is a question that is by no means answered. Evidence of the problematic nature of the concept is found in rabbinic Judaism. Akiba (2nd century CE) ignored the usages in Gen. 1 and 5 and emphasized 9:6, understanding it to mean, contrary to the usual interpretation, "after an image, God made man," that is, in the Platonic sense of a heavenly archetype. He did not wish to allow any resemblance between God and any created being. Other interpretations sought to avoid the difficulty by rendering ***elohim*** (a plural form) not as "God" but as "divine beings" (*i.e.*, angels: "God created man after the image of divine beings ***[elohim]").***

"Soul," "spirit," or "life"

The earthly-spiritual creature. In those parts of the Jewish community of antiquity that were deeply influenced by Greek philosophical ideas, a dualistic interpretation of man was offered. Here the divine likeness suggested is that of the immortal, intellectual soul as contrasted to the body. Still other thinkers, both ancient and modern, have understood that likeness to be ethical, with particular emphasis placed on freedom of the will. What is evident is that no doctrine of man can be erected on the basis of these several verses alone, but that a broader view must be taken, in which they are assimilated. A careful examination of the biblical material, particularly the words ***nefesh, neshama,*** and ***ruah,*** which are often too broadly translated as "soul" and "spirit," indicates that these must not be understood as referring to the psychical side of a psycho-physical pair. A man did not possess a ***nefesh*** but rather was a ***nefesh,*** as Gen. 2:7 says: ***"wa-yehi ha-adam le-nefesh hayya"*** (". . . and the man became a living being"). Man was, for most of the biblical writers, what has been called "a unit of vital power," not a dual creature separable into two distinct parts of unequal importance and value. While this understanding of the nature of man dominated biblical thought, in apocalyptic literature (2nd century BCE–2nd century CE) the term ***nefesh*** began to ***be*** viewed as a separable psychical entity with existence apart from body. Although this was not entirely divorced from the unitary biblical view, nonetheless a functional body–soul dualism was present in such literature. In the Alexandrian version of Hellenistic Judaism the orientation toward Greek philosophy, particularly the Platonic view of the soul imprisoned in the flesh, led to a clear-cut dualism with a negative attitude toward the body. Rabbinic thought remained closer to the biblical position, at least in its understanding of man as a psychosomatic unit, although the temporary separation of the components after death was an accepted position.

Sheol, resurrection, and immortality

The biblical view of man as an inseparable psychosomatic unit meant that death was understood to be his dissolution. Yet, although man ceased to be, this dissolution was not utter extinction. Some of the power that functioned in the unit may have continued to exist, but it was not to be understood any longer as life. The existence of the dead in ***sheol,*** the netherworld, was not living but the shadow or echo of living. For most of the biblical writers this existence was without experience, either of God or of anything else; it was unrelated to events. To call it immortality is to empty that term of any vital significance. However, this concept of ***sheol,*** along with belief in the possibility of occasional miraculous restorations of dead individuals to life, and perhaps even the idea of the revival of the people of Israel from the "death" of exile, provided a foothold for the development of belief in the resurrection of the dead body at some time in the future. The stimulus for this may have come from ancient Iranian religion, in which the dualistic cosmic struggle is eventually won by life through the resurrection of the dead. This idea began to appear in sketchy form in postexilic writings (Isa. 26:19; Dan. 12:2). In this view there is life only in the psychosomatic unit now restored. This restoration was bound up with the eschatological hope of Israel (see ***Eschatology,*** below) and was limited to the righteous. In subsequent apocalyptic literature a sharper distinction between body and soul was entertained, and the latter was conceived of as existing separately in a disembodied state after death. Although at this point the doctrine of the resurrection of the body was not put aside, nonetheless, the direction of thinking changed. The shades of *sheol* were now thought of as souls, and real personal survival—with continuity between life on earth and in *sheol*—was posited. Now Greek ideas, with their individualistic bent, began to have influence, so that the idea of resurrection that was in some way related to a final historical consummation, began to recede. True life after death was now seen as release from the bondage of the body, so that in place of, or alongside of, the afterlife of physical resurrection was set the afterlife of the immortal soul.

It was not the status of the soul, however, that concerned either the biblical or the rabbinic thinkers. What emerged from the latter's discussions of the biblical themes was an emphasis on the ethical import of man's composite makeup. Man is in a state of tension or equilibrium between the two foci of creation, the "heavenly" and the "earthly." He necessarily participates in both, and, as such, is the one responsible creature who can truly serve his Creator, for he alone, having both sides of creation in him, may choose between them. It is the ability to make an ethical choice that is the distinguishing mark of man. This ability is not derived from the "heavenly" side but resides in the double basis of man's existence. It is important to recognize this as something other than a body-soul dualism in which the soul is the source of good and the body the basis of evil. Such an attitude, however, did appear in some rabbinic material and was often affirmed in medieval philosophical and mystical speculations and by some of the later moralists. These are genuine variations and developments of the biblical-rabbinic ideas and may not be dismissed as aberrations. They represent authentic attempts to come to terms with other currents of thought and with the problems and uncertainties inherent within the earlier materials themselves.

The ethically ***bound creature.*** Mankind is then viewed, however this position is arrived at, as ethically involved. The first 11 chapters of Genesis are posited upon this responsibility, for the implicit assumption of the prepatriarchal stories is man's ability to choose between obedience and disobedience. Rabbinic Judaism, taking up the covenant-making episode between God and Noah (Gen. 9:8–17), developed it as the basis of mankind's ethical obligation. All men, not merely Israel, were engaged in a covenant relationship with God, which was spelled out in explicit precepts—variously enumerated as six, seven, or even 10 and occasionally as many as 30—that reflect general humanitarian behaviour and are intended to assure the maintenance of the natural order by the establishment of a proper human society. The Covenant with Israel was meant to bring into being a community that would advance the development of this society through its own obedience and witness.

Man's nature, viewed ethically, was explained in rabbinic Judaism not only as a tension between the "heavenly" and "earthly" components but also as a tension between two "impulses." Here again, fragmentary and allusive

The good and evil impulses

biblical materials were developed into more comprehensive statements. The biblical word *yetzer* means "plan," that which is formed in man's mind. In the two occurrences of the word in Genesis (6:5; 8:21), the plan or formation of man's mind is described as *ra,* perhaps "evil" in the moral sense or maybe no more than "disorderly," "confused," "undisciplined." The other biblical occurrences do not have this modifier. Nonetheless, the Aramaic translations (Targumim) invariably denominated it as *bisha* ("wicked") wherever it occurred. Rabbinic literature created a technical term *yetzer ha-raʿ* ("the evil impulse") to denote the source within man of his disobedience, and, subsequently, the counterterm *yetzer ha-ṭov* ("the good impulse") to indicate man's obedience. These more clearly suggest the ethical quality of man's duality, while their opposition and conflict point to man's freedom and the ethical choices he makes. Indeed, it is primarily within the realm of the ethical that Judaism posits freedom, recognizing the bound, or determined, quality of much of his existence (*e.g.*, his natural environment or physiological makeup).

It is this ethically free creature who stands within the covenant relationship and who may choose to be obedient or disobedient. Sin, then, is ultimately deliberate disobedience or rebellion against the divine sovereign. This is more easily observed in relation to Israel, for it is here that the central concern of Judaism is most evident and the subject discussed in greatest detail. It should be noted, however, that since, according to Judaic tradition, all mankind stands within a covenant relation to God and is commanded to be moral and just, essentially the same choice is made universally. In technical language, the acceptance of divine sovereignty by the people of Israel and by the individual within that community is called "receiving the yoke of the kingship." This involves intellectual commitment to a basic belief, as expressed by the Deuteronomic proclamation: "Hear, O Israel, the Lord, our God, the Lord is one!" At the same time it imposes obligations in terms of communal and individual behaviour. These two responses are understood to be inextricably bound together, so that rejection of the divine sovereign is manifest as denial of God both intellectually and practically. It amounts to "breaking the yoke of the kingship." In more specific terms, sin is sometimes summed up under three major, interrelated headings: idolatry, murder, and illicit sexual behaviour, each and all of which explicitly and implicitly involve rebellion, for they involve activities that deny—if not God's existence—his commanding relationship and the requirement of man's response. Such behaviour destroys the community and sets individual against individual, thus thwarting the ultimate purpose of God, the perfected human society.

Teshuva ("turning")

If, however, man is free to choose rebellion and to suffer its consequences, he is also able to turn back to God and to become reconciled with him. The Bible—most particularly the prophetic writings—is filled with this idea, although the term *teshuva* ("turning") came into use only in rabbinic sources. Basically it grows out of the covenant and God's unwillingness—despite man's failures—to break off his relationship. In rabbinic thought it is apparently assumed that even the direst warnings of utter disaster and rejection imply the possibility of turning back to God, motivated by remorse and the desire for restoration. Divine readiness and human openness are the two sides of the process of reconciliation. What was expressed in prophetic literature in the immediate historico-political situation was stated in the synagogal liturgy in connection with pentateuchal and prophetic lessons and the homilies developed from them. Thus, the divine invitation was constantly being offered. Man was called upon to atone for his rebellion by positive action that repudiated his failure. He was summoned to reconstitute wholeness in his individual life and community in society.

Historically viewed, Jewish existence, as it developed under rabbinic leadership, following after the two disastrous rebellions against Rome, was an attempt to reconstitute a community of faith expressed in worship and in an ordered society that would enable the individual to live a hallowed life of response to the divine will. Although this plan was not spelled out in detail, it was probably understood to be the paradigm for the eventual reconstruction of humanity.

Medieval and modern views of man. The Jewish view of man is certainly less clearly articulated than its affirmations concerning God. Nonetheless, it is evident that its central concern was ethical. The question of how man as individual and community was to behave was the focus of interest. Yet it is also clear that metaphysical concerns, however rudimentary in the beginning, were included in the developing discussion. Medieval philosophers sought an accommodation between the doctrine of the resurrection of the body and the concept of the immortality of the soul. The greatest of them, Moses Maimonides (1135–1204), propounded an extremely subtle position that equated immortality with the cleaving of the human intellect to the active intellect of the universe, thus limiting it to philosophic adepts. In the modern period, the impact of various philosophical and psychological schools has further fragmented the situation so that little or no consensus is evident, although resurrection or immortality language is still used even when its content is uncertain. But alongside this lack of agreement, the view that man is to be understood, however else, as a creature who makes free ethical choices for which he is responsible remains—although variously articulated—the basic affirmation of Judaism about man.

The imitation of God

Ethics and society. *The ethical emphasis of Judaism.* Jewish affirmations about God and man intersect in the concept of Torah as the ordering of human existence in the direction of the divine. Man, however else understood, is an ethically responsible creature responsive to the presence of God in nature and in history. Although that responsiveness is expressed on many levels, it is within the horizontal relationship of man to man that it is most explicitly called for. The pentateuchal legislation sets down, albeit within the limitations of the structures of the ancient Near East, the patterns of interpersonal relations. The prophetic messages are deeply concerned with these demands and see the disregard of them as the source of social and individual disorder. No segment of society, even the most exalted, is free of ethical obligation. Indeed, the transformation of prophetism from its earlier form as ecstaticism and soothsaying is seen in the ethical confrontation of David by Nathan ("Thou art the man") for seducing Bathsheba and arranging to have her husband killed (II Sam. 12). What is particularly striking is the affirmation that God is not only the source of ethical obligation but is himself the paradigm of it. In the so-called Code of Holiness (Lev. 19), it is imitation of divine holiness that is offered as the basis of human behaviour in the ethical sphere as well as the cultic-ceremonial. Concern for the economically vulnerable members of the community; obligations toward neighbours, hired labourers, and the physically handicapped; interfamilial relationships; and attitudes toward strangers (*i.e.*, non-Israelites) were all motivated by the basic injunction, "You shall be holy, for I, the Lord your God, am Holy." Acceptable human behaviour is, therefore, "walking in all His ways" (Deut. 11:22). The dialectic relation between God and man in the literary prophets also exhibits divine righteousness and divine compassion as patterns to be emulated in the life of the community.

This theme, *imitatio Dei* ("imitation of God"), as developed in rabbinic Judaism, is expressed succinctly in a comment on the verse from Deuteronomy quoted above. In response to the question of how it is possible to walk "in all His ways," the reply is made (*Sifre* Deut. 85a): "As He is merciful and gracious, so be you merciful and gracious. As He is righteous so be you righteous. As He is holy, strive to be holy." Indeed even more daringly, God is described as clothing the naked, nursing the sick, comforting the mourners, burying the dead, so that man may recognize his own obligations.

Interpenetration of communal and individual ethics. What stands out in the entire development of Jewish ethical formulations is the constant interpenetration of communal and individual obligations and concerns. Al-

The just man in a just society

though in the Book of Ezekiel (see especially chapter 18) emphasis is laid on individual responsibility, "the person who sins shall die," in contrast to the more widespread statement of communal involvement, "visiting the sins of the fathers upon the children" (*e.g.*, Ex. 20:5), these two aspects of ethical conduct are never entirely distinguished in Judaism. The just society requires the just man, and the just man functions within the just society. The concrete expression of ethical requirements in legal precepts took place with both ends in view, so that the process of beginning the holy community and the formation of the *ḥasid* ("pious"), the man of steadfast devotion to God, were concomitant processes. The relationship between the two is, of course, often mediated by the historical situation, so that in some periods one or the other moves to the centre of practical interest. In particular, the end of the Judaean state (70–135 CE) truncated the communal aspect of ethical obligations, often limiting discussion to apolitical responsibilities rather than to the full range of social involvements. The re-establishment of the State of Israel in the 20th century has, therefore, reopened for discussion areas that have for millennia been either ignored or relegated to the realm of abstraction. What this implies is that the full ethical responsibility of the Jew cannot be carried out solely within the realm of individual relationships but must include involvement in the life of a fully articulated community.

This double involvement is most vividly apparent in the biblical period: when both were equally present as divine command and demand. In the rabbinic period, because of the new political context, the communal aspect receded, so that discussion was mainly oriented toward the relationships between the members of the Jewish community or between individuals as such, and away from political responsibilities in the larger society. Nonetheless, the virtues that were understood to govern these relationships were, in their biblical setting, communal as well. Righteousness and compassion had been obligations of the state, governing the relationship between political units, as the first two chapters of Amos make evident. At the same time, as Micah 6:8 shows, doing justly, loving mercy, and walking humbly with God made up the pattern of the individual's obligations as well. Given the situation of the dispersion of the Jews following the revolts against Rome in the 1st and 2nd centuries CE, the individual pattern became the object of primary considerations. It is important to recognize that while theoretical ethical systems were not developed until the Middle Ages under the influence of philosophical concerns, nonetheless, even in the early period it was understood that behind the practical system of Halakha, the enumeration of legal precepts, there stood the dynamic of ethical theory. An attempt was made to reduce the hundreds of precepts to a small number expressing the ethical essence of Torah.

Protection of the weak

The key *moral* virtues. In keeping with the rabbinic understanding of Torah, study also was viewed as an ethical virtue. A passage in the traditional Prayer Book enumerates a series of virtuous acts—honouring parents, deeds of steadfast love, attendance twice daily at worship. hospitality to wayfarers, visiting the sick, dowering brides, accompanying the dead to the grave, devotion in prayer, peacemaking in the community and in family-life—and concludes by setting study of Torah as the premier virtue. Here is exhibited the complex variety of ethical behaviour called for within the Jewish tradition. To parental respect and family tranquillity are added, in other contexts, the responsibility of parents for children, the duties of husband and wife in the establishment and maintenance of a family, and ethical obligations that extend from the conjugal rights of each to the protection of the wife if the marriage is dissolved. The biblical description of God as upholding the cause of the fatherless and the widow and befriending the stranger, providing him with food and clothing (Deut. 10:18), remained a motivating factor in the structure of the community. Ethical requirements in economic life are expressed concretely in such a passage as Lev. 19:35–36: "You shall do no wrong in judgment, in measures of length or weight or quantity. You shall have just balances, just weights, a just ephah, and a just hin"; and in Amos' bitter condemnation of those who "sell the righteous for silver, and the needy for a pair of shoes" (Amos 2:6). Such injunctions, together with many other specific precepts and expressions of moral requirements, established the basis for a wide-ranging program that sought to govern, both in detail and in general, the economic life of the individual and the community. Not only are relations within the human sphere the object of ethical concern but nature also is so regarded. The animal world, in the biblical view, requires merciful consideration, so that not only man is commanded to rest on the sabbath but his domestic animals are to share the rest with him (Ex. 20:10; 23:12). Mistreatment of beasts of burden is prohibited (Deut. 22:4); and wanton destruction of animal life falls under the ban (*ibidem:* 6–7). In the rabbinic attitude toward brute creation, even inanimate nature is the object of human solicitude. Thus, for example, the food-yielding trees of a city under siege may not be destroyed, according to Deuteronomic legislation (Deut. 20:14–20). The enlargement of this and other biblical precepts resulted in the generalized rabbinic prohibition "You shall not destroy" that governs man's use of his environment.

The relation to non-Jewish communities and *cultures*. As noted above, the end of the Jewish state reduced the scope of ethical judgments in the political sphere; nonetheless, relations between the Jewish community and other societies—particularly political units: the Roman and Christian empires, the Islāmic states, and other regimes—provided opportunities for the exploration of the ethical implications of such encounters. Since most of these were victor-victim, superior-inferior, power-powerless situations, with the Jews the weaker party, prudential considerations were dominant. Despite this, Jewish authorities sought to bring to bear upon these external arrangements the ethical standards that governed the internal structures.

Isolation and assimilation

The whole problem of the relationship between the Jewish community, in whatever form it has existed and does exist, and other social units has been vastly complicated. Ideally, the relation is that of witness to the divine intent in the world. Practically, it has swung between the extremes of isolation and assimilation, in which the ideal has, on occasion, been lost sight of. Culturally, from its earliest beginnings, the people Israel has met and engaged the ideas, forms, behaviour and attitudes of its neighbours in constructive development. It borrowed as it contributed and reformulated what it received in terms of its own commitments and affirmations. On more than a few occasions, as in the period of settlement in Canaan, it rejected the religio-cultural ideas and forms of the native population. On others, it actively sought out—as in the Islāmic period in Spain (8th to 15th centuries)—ideas and cultural patterns of its neighbours, viewing them from its own perspective and embracing them when they were found to be of value. Indeed, the whole history of Israel's relationship with the world may be comprehended in the metaphor, used previously, of the heartbeat with its systole and diastole. No period of its existence discloses either total rejection of or abject surrender to other cultural and political structures but rather a tension, with the focal point always in motion at varying rates. Being more than a "confession" in the Christian sense, Judaism's adjustment to and relation with other sociopolitical units involved larger aspects of communal and individual life than merely the religious. Whether or not, under such circumstances, it is helpful to describe Judaism as a civilization, it is important to recognize that, viewed functionally, much more must be included than is usually subsumed under the common usage of "religion."

The *formulation* of Jewish ethical doctrines. The ethical concerns of Judaism have found frequent literary expression. Not only were rabbinic writings constantly directed toward the establishment of legal patterns that embody such concerns but in the medieval period the issues were dealt with in treatises on morals; in ethical wills, in which a father instructed his children about their obligations and behaviour; in sermons; and in

The Musar movement

other forms. In the 19th century the traditionalist Musar ("moral instructor") movement in eastern Europe and the philosophical discussions of the nascent Reform movement in the West focussed upon ethics. Indeed, since the political and social emancipation of the Jews, ethical and social rather than theological questions have tended to be given priority. Often the positions espoused have turned out to be, nonetheless, "judaized" versions of philosophic ethics or of political programs. In some instances, as in the case of the distinguished German-Jewish philosopher Hermann Cohen, the result has been a Jewishly compelling restatement of a secular philosophic ethics. In others, it has resulted in no more than a pastiche. More crucial, however, is the question of a unique Jewish ethics and of its authority. The re-establishment of a Jewish state renews the possibility that the full range of ethical decisions, including communal as well as individual responsibility, may be confronted. In such a situation, the ideal task of the people moves out of the realm of speculation to become actual again.

The universe. *Creation and Providence: God's world.* Although the first chapter of Genesis affirms divine creation, it does not offer an entirely unambiguous view of the origin of the universe, as the debate over the correct understanding of Gen. 1:1 in former as in modern times discloses. (Was there or was there not a pre-existing matter, void, or chaos?) Yet, basically, the interest of the author was not in the mode of creation, a later concern perhaps reflected in the various translations of the verse: "In the beginning God created," which could signify what medieval philosophers designated *creatio ex nihilo* ("creation out of nothing"); and "when God began to create," which could indicate some concept of prime matter. He was concerned rather to affirm that the totality of existence, inanimate (Gen. 1:3–19), living (20–25), and human (26–31), derived immediately from the same divine source; and, thus, that it is a universe. As divine creation, it is transparent to the presence of God, so that the Psalmist said: "The heavens declare the glory of God, and the expanse proclaims [that it is] the work of his hands" (19:1). Indeed, the repeated phrase: "And God saw how good it was" (Gen. 1:4, 10, 12, 18, 25, 31) may be understood as the ground of this affirmation, for the workmanship discloses the workman. The observed order of the universe is further understood by the biblical author to be the direct result of a covenantal relationship established between the world and God: "So long as the earth endures, seedtime and harvest, cold and heat, summer and winter, day and night, shall not cease." (Gen. 8:22). This doctrine of the providential ordering of the universe, reaffirmed in rabbinic Judaism, is not without its difficulties, as in the liturgical change made in Isa. 45:7 to avoid ascribing evil to God. Nonetheless, despite the problem of theodicy (the problem of evil in a world made and ordered by God), Judaism has not acquiesced to the mood reported in the Palestinian Targum to Gen. 4:8: "He did not create the world in mercy nor does he rule in mercy." Rather, it has affirmed a benevolent and compassionate God.

Creation and history

It is the physical world—divine creation—that provides the stage for history, which is the place of the divine-human encounter. An early Midrash, in response to the question as to why Scripture begins with the story of creation, points out that it was necessary in order to establish the identity of the Creator with the Giver of Torah, an argument basic to the liturgical structure of the Shema. This relationship is further emphasized in the *qiddush*, the prayer of sanctification recited at the beginning of the sabbath. That day is designated "a remembrance of creation" and "a recollection of the going-forth from Egypt." Thus, creation (nature) and history are understood to be inextricably bound up, for both derive from the same divine source. This being so, redemption —the reconciliation of God and man through and in history--does not ignore or exclude the natural world. Using the imagery of an extravagantly fecund world of nature, rabbinic thought expressed its view of the all-inclusive effects of the restored relationship.

Man's place in the universe. Man as creature is, of course, subject to the natural order. It is, indeed, in the world and through the world that man carries out his relationship to God. The commandments of Torah are obeyed not solely as observances between man and God but as actions between man and man, between man and the world. Although the creation story designates man as ruler over the earth and its inhabitants (Gen. 1:26–28; see also Ps. 8:5–9), nonetheless, far from being an arbitrary master, man's dominion is limited by Torah, for its regulations are concerned not only with transactions between man and man but also lay out his responsibilities to the land he cultivates, the produce of the soil, the animals he domesticates. Being bound in the network of existence he, as the moral creature, is responsible for it in all of its parts.

Man's responsibility to nature

Even the destruction of the Jewish commonwealth in the 1st and 2nd centuries CE did not alienate the Jew from these responsibilities, as the elaborate system of Mishna and Gemara gives evidence. The gradual but consistent exclusion of the Jewish community from immediate connection with large segments of the natural world, through legislation in Christendom and Islām, tended to dull the Jew's awareness of it; the recurring references to it in the religious calendar, however, and the observation of harvest festivals even by citydwellers continued to remind the community of its ties. Thus, at the end of the 19th century, the nascent Zionist movement recognized that the regeneration of the Jewish people involved, among other requirements, a responsible relation to the natural order expressed in its attitude toward and treatment of the land.

As indicated in other contexts, the particular emphasis placed on one or the other side of the frequent twofoldness of the Jewish view has depended upon the situation in which the community has found itself. If nature as the place of divine disclosure has, during long periods of Jewish existence, assumed a somewhat subordinate role, it has never been rejected or been seen to be irrelevant to the divine purpose. Indeed, in Jewish eschatology, its restoration is part of the goal of history.

Intermediary beings: angels and demons. The exact nature of the nonhuman beings mentioned in Scripture —angels or messengers—is not altogether clear and their roles seem ephemeral. In the postexilic period, perhaps under Iranian influence, and in the late biblical and post-biblical literature, these beings emerge as more complete and often as clearly identifiable individuals with their own personal names. The unfocussed biblical view gave way to an elaborate hierarchy of functionaries who acted, in some apocalyptic visions, as a veritable heavenly bureaucracy. Nevertheless, despite a consensus concerning their existence, there was little agreement as to their role or importance. In some Midrashim God takes counsel with them; in other sources the rabbis urge men not to involve them but to approach God directly. Actually, they belong to that marginal area between religion and folklore. Like their counterfigures, the demons, they have a residual existence rooted in various layers of the Jewish experience and interpretation of the universe. At some times they are highly individualized and sharply realized; at others, they flit in and out of the imagination like bats in the evening. The medieval philosophers Aristotelized or Platonized them; the early mystics Neoplatonized them; the Kabbalists continually invented new ones and fitted them into their complicated network of cosmic existence. Nonetheless, their role, even in periods of considerable emphasis, was peripheral. They were outside the great movements and meanings of Jewish thought.

Contemporary philosophical speculation about the nature of the universe has, of course, required a response from Jewish thinkers. But, given the particular temper of a period in which metaphysics has not been central to much of theological discussion, no major statement has yet developed that has taken hold of the dominant positions and attempted to view them from the Jewish creationist perspective. The attempt within Reconstructionism (see below) to provide a naturalistic framework for Judaism, while courageous, lacks the breadth and depth of the great philosophical approaches.

Eschatology. The future age of *mankind* and the world. The choice of Israel, according to the biblical writings, had occurred because of mankind's continual failure, by rebellion against its Creator, to fulfill its divine potential. The subsequent failure of Israel to become the holy community and thereby a witness to the nations gave rise to the prophetic movement that summoned the people to obedience. An integral part of prophetic summoning, side by side with threats of punishment and warnings of disaster, was the envisioning of a truly holy community, a society fully responding to the divine imperative. This kingdom of the future was conceived of as entirely natural, functioning as any normal sociopolitical unit and under the leadership of a human ruler, who would, however, carry out his tasks within the sphere of divine sovereignty, serving primarily to exhibit his own obedience and thus to stimulate the obedience of the entire people. This human monarch of the future was often, although not always, portrayed in terms of an idealized David, using such features of his life and reign as would underscore submission to God and emphasized social stability, economic satisfaction, and peace. During the period of the monarchy, the prophetic demand was directed toward each succeeding king, with the hope or even the expectation that he would be or become the new David, or the ideal ruler.

The Davidic model

The Babylonian Exile added a new measure of urgency to this expectation, although it was not expressed in any uniform fashion. The later chapters of the Book of Ezekiel provide in largely impersonal fashion the constitution for the new commonwealth but do not describe the peculiar characteristics of the ruler, while the later chapters of the Book of Isaiah focus on several figures—including Cyrus the Mede—who are seen as the divine instruments ushering in a new era. It is important to recognize that while such figures have extraordinary virtues ascribed to them, these virtues are neither superhuman nor suprahuman but such as are ultimately required of all Israel and of all men. The frustrations of the post-exilic period, when the several attempts to bring into being the holy community had no more than partial success and were thwarted by the imperial designs of the great powers—as they had been in the pre-exilic period as well—led to an emphasis upon the futuristic quality of the messianic hope. This was abetted undoubtedly by external influences, such as Iranian thought, in which the cosmic rather than the historic aspect of a future era dominated. Since ancient cosmic myths—in good measure demythologized—had been part of the Israelite intellectual inheritance, evidenced at least in literary usages throughout Scriptures, the impact of such neighbouring ideas was to reinvigorate the mythic elements. Thus, hopes for the future at the end of the Persian period and on through the Hellenistic developments after *c.* 330 BCE comprised both historical expectations focussed upon a sociopolitical community and cosmic-mythic visions that moved on a broader stage. The latter were, of course, never entirely absent from the historical expectations and situations, for a renewal of nature was viewed as integral to the functioning of true society. The obedient community required, and was to be granted, a natural world in which true human relations could exist. In its most vivid forms, apocalypses (*i.e.*, visionary disclosures of the future), the literature of the period affords a remarkable insight into the agonies and urgencies of the people. After the failures, saving events, and disappointments of the past are recounted, the present, in transparent disguise, is portrayed and the immediately hoped-for intervention of God is described in awesome detail as a means of affirming and confirming the faith of those who saw themselves as the remnant, or perhaps the promise, of the holy community.

The socio-political and cosmic-mythic aspects

The king-messiah and his reign. Put schematically, Israel's hope was for the restoration of divine sovereignty over all of creation. Concretely, that hope found a considerable variety of expressions. Of all such expressions, that which centred around the idealized king began to assume an ever more important (but never exclusive) role. Many of the writings that report the ideas and attitudes of the Jewish community in the period immediately preceding and following the rise of Christianity are either ignorant of or more probably indifferent to the personal element. God is envisioned as the protagonist of the end, actively intervening or sending his messengers (*i.e.*, angels), to perform specific acts in ending the old and inaugurating the new era. On the other hand, in some writings of the period the annointed king-messiah (Hebrew, *mashiah*, "anointed")—the title reflects the episode in I Sam. 16 in which David is thus singled out as the divinely chosen ruler—becomes more sharply defined as the central figure in the culminating events and, given the cosmic-mythic components, assumes suprahuman and in some instances, even quasidivine, aspects. It is clear, then, that the doctrine of last things in Judaism is not necessarily messianic, if that term is properly limited to an inauguration of a future era through the action of a human, suprahuman, or quasi-divine person. Nonetheless, it must be recognized that the messianic version of eschatology played a more compelling role in rabbinic Judaism than other modes. The same is true with regard to the locus of the "world (or age) to come." Given the ingredients noted above, it was possible to construct various eschatological landscapes ranging from the mundane to the celestial, from Jerusalem in the hills of Judah to a heavenly city. Indeed, confronted with an embarrassment of riches, the medieval theologians sought to combine them into an inclusive system that intricately involved as large a variety of the possibilities as could be brought together. In such patterns the messianic this-worldly emphasis was understood as a preliminary movement toward an ultimate resolution. The ideal ruler, the new David, would re-establish the kingdom in its own land (in "Zion," or Palestine) and would reign in righteousness, equity, justice, and truth, thus bringing into being the holy nation and summoning all mankind to dwell under divine sovereignty. As a component of this re-established kingdom, the righteous dead of Israel would be resurrected to enjoy life in the true community that did not exist in their days. This kingdom, however long it was destined to endure, was not permanent. It would come to an end either at a predetermined time or as victim of the unrepentant nations and cosmic foes, at which point the ultimate intervention by God would take place. All the wicked throughout history would be recalled to life, judged, and doomed; all the righteous would be transformed and transported into a new world; *i.e.*, creation would be totally restored. The particular emphases that one or the other of these ideas received, the ways in which they were interpreted—philosophically, mystically, or ethically—were determined most frequently by the situations and conditions in which the Jewish community found itself. With such a considerable body of ideas at its disposal and with the details of none of them ever receiving the kind of affirmation that statements about God, Torah, and Israel had, freedom of speculation in the realm of eschatology was little restricted. Thus, Joseph Albo (15th century) in his work on Jewish "dogmas" (*Ikkarim*) was not inhibited from denying that belief in the messiah was fundamental. The mystical movements of the Middle Ages found in eschatological hopes a crucial centre. The early Kabbala was little interested in messianism, for it interiorized the expectations in the direction of personal redemption. Following the disasters of the late 15th to 17th centuries (*e.g.*, the expulsion of the Jews from Spain and the Cossack massacre of the Jews in Poland) however, messianic speculation in all of its varieties underwent a luxuriant growth, finally running wild in the movements surrounding Shabbeiai Tzevi of Smyrna and later Jacob Frank of Offenbach. These tragedies for the Jewish communities once again resulted in a futurizing of the hopes or at least a limiting of their application (see also JEWISH MYSTICISM and MESSIAH AND MESSIANIC MOVEMENTS).

Resurrection of the dead and Last Judgment

Secularization of messianism. In the 19th century, with the political emancipation of the Jews in western Europe and the development of an optimistic evolutionism, messianism was transformed by many liberal thinkers into a version of the idea of progress whose goal was often thought of as immediately attainable through enlightened social and political action. When disillusionment with the emancipation set in, messianism was even more completely secularized in some segments of the community who saw its meaning and fulfillment in some form of socialism—again, rather close at hand. In others, it was absorbed into the emerging political nationalism—Zionism. Similar developments took place in eastern Europe, with parallel transformations. In more recent times, particularly since the events symbolized by the name Auschwitz (a Nazi death camp in Poland, where millions of Jews were exterminated), the earlier modern interpretations, particularly of messianism, but also of eschatology as a whole, have been considered inadequate. Although no compelling statement has been forthcoming, Jewish thinkers in the second half of the 20th century have been attempting once again to come to grips with eschatological concepts in all of their varieties and forms.

BASIC PRACTICES AND INSTITUTIONS

The centrality of Halakha

The hallowing of everyday existence. Systematic presentations of the affirmations of the Jewish community never served as the sole mode of expressing beliefs of the people. Side by side with speculation—Haggadic, philosophic, mystical, or ethical—there stood, not in a secondary role but as the other of the double focuses, Halakha ("practice," "rules of conduct"), the paradigmatic statement of the behaviour, individual and communal, that embodied concretely the beliefs conceptualized in speculation. Life in the holy community was understood to embrace every level of human existence, and attempts to confine it to a special religious realm met with opposition. The prophets vigorously resisted attempts to limit the sovereignty of the God of Israel to organized worship and ritual. The Pharisees, even while the Jerusalem Temple cult was still in existence, sought to reduce priestly exclusiveness by enlarging the scope of sacral rules to include, as far as possible, all of the people. Rabbinic Judaism, Pharisaism's surviving descendant, continued the process of democratization and sought, through its system of interpretation, to find in every occasion of life a means of affirming divine concern and presence. Viewed negatively by some Protestant theologians, this development has been judged to stifle spontaneity. Yet spontaneity is not necessarily lacking within a world governed by Halakha, although the danger of the stylized routine in religious and ethical life is apparent. Nonetheless, the intention of the Halakhic attitude is to remind the Jew constantly that each and every occasion of life is a locus of divine disclosure. This is most clearly seen in the *berakhot*, the "blessings," that are prescribed to accompany the performance of a broad spectrum of human actions, from the commonplace routines of daily life to the restricted gestures of the cultic-liturgical year. In these, God is addressed directly in the second person singular, his sovereignty is affirmed, and his activity as creator, Giver of Torah, or redeemer, expressed in a wide variety of eulogies, is proclaimed. There are no areas of human behaviour in which man cannot be met by God, and in terms of its intention, the Halakhic pattern is designed to make such possibilities experienced realities. Yet, again, it must be noted that the situation of the Jewish community determines in a very large way how the intention is actualized. On more than one occasion the Halakhic pattern has served as a defense against a hostile environment and has thus tended to become scrupulosity (an obsessive concern with minute details), but the dynamic of the intention itself has as often broken through to re-establish its integrity and hallow life in its wholeness.

The liturgy of daily life

The traditional pattern of individual and familial practices. Perspective on the traditional pattern of an individual's life is obtained by examining a passage from the Babylonian Talmud (tractate *Berakhot* 60b) that was subsequently reworked into a liturgical structure but which in its original form exhibits the intention discussed above. In this passage, the blessings accompanying a man's waking and returning to the routines of life are prescribed. There is a brief thanksgiving on awakening for being restored to conscious life; then the impingement of the external world is responded to in a benediction over the cock's crowing; following this, each ordinary act, opening one's eyes, stretching and sitting up, dressing, standing up, walking, tying one's shoes, fastening one's belt, covering the head, washing the hands and face, has its accompanying blessing, reminding a man that the world and the life to which he has returned exist in the presence of God. These are followed by a supplication in which the petitioner asks that his life during the day may be worthy in all of its relationships. Then, as the first order of daily business, Torah, both written (Bible) and oral (Mishna), is briefly studied, introduced by eulogies of God as Giver of Torah. Finally, there is a prayer for the establishment of the Kingdom of God, for each day contains within itself the possibility of ultimate fulfillment. As indicated, this was originally not a part of public worship (even today it is, strictly speaking, not part of the synagogue service, although it is most frequently recited there) but was personal preparation for a life to be lived in the presence of God.

From H. Hapgood, The *Spirit* of the Ghetto: @ 1965 by Funk & Wagnalls Co., Inc., reproduced by permission of Funk& Wagnalls Publishing Co.. Inc.

Boy putting on phylacteries in preparation for morning prayer, drawing by Jacob Epstein, c. 1902.

The central role of women in family religion

Such individual responsibility marks much of Jewish observance, so that the synagogue—far from being the focus of observance—shares with the home and the workaday world the opportunities for the divine-human encounter. The table blessings, *kiddush* (the "sanctification" of the sabbath and festivals), the erection of the booth (*sukka*) for Sukkot (the Feast of Tabernacles), the seder (the festive Passover meal) with its symbols and narration of the Exodus, the lighting of the lamps during the eight days of Ḥanukka (the Feast of Dedication), are all the obligation of the individual and the family and have their place in the home. It is here, too, that woman's role is defined and here, as contrasted with the synagogue, that she functions centrally. Given the traditional dietary regimen of the Jewish community—the exclusion of swine, carrion eaters, shellfish, and other creatures, the separation of meat and dairy products, the ritual slaughtering of animals, the required separation and burning of a small portion of dough (*ḥalla*) when baking, the supervision of the Passover food requirements, and many other stipulations—there exists a large and meticulously governed area within the home that is indeed the sphere of woman's religion. There seems not to have been a hierarchy of values in which the home-centred, as contrasted with the synagogue-oriented, practices were given an inferior status. In modern times,

Family from Yemen celebrating Passover in Israel.
Popperfoto

however, particularly in Western civilization where the pervasiveness of religious obligation has been replaced by ecclesiastical institutionalism, on the prevailing Christian model, this whole crucial area has lost much of its meaning as a place of divine-human meeting. Thus, for many it is only the synagogue that provides such an opportunity, and the individual act has been reduced on the scale of values. With this downgrading, woman's religion has lost its significance so that her status—when parallels are drawn to her role in the larger society—has been reduced to one of inferiority. However attenuated personal religious responsibility may have become in some environments or transformed into stylized cultural forms, the intention that informs the Halakhic structure, the hallowing of the individual's total existence, remains a potent force within the Jewish community.

The traditional pattern of synagogue practices. The other focus of observance is the synagogue. The origins of this institution are obscure and a number of hypotheses have been proposed to account for the appearance of this essentially lay-oriented form of worship. What seems certain is that during the period of the Second Temple—following the return from Babylon and continuing until the Temple destruction in 70 CE—there were, side by side with the official cult, other modes of worship more or less independent of the priesthood and nonsacrificial in form. The reports by the philosopher Philo and the historian Josephus in the 1st century, buttressed by the Qumrān document (Dead Sea Scrolls), provide some knowledge of the practices of the contemporary Essenes; rabbinic sources, including the earliest layers of the traditional order of worship, enable us to understand another, apparently Pharisaic, mode; the brief allusions to the practices of James and his Jewish Christian companions in the book of Acts suggest yet other varieties. In any case, the grouping that formed the cadre of what eventually became rabbinic Judaism observed some form of worship that, with the destruction of the Temple cult, was able to provide a new centre and even to absorb enough from the defunct priestly institution to suggest continuity and legitimacy. This was probably the basic pattern for synagogal liturgy in the millennia that followed.

Readings from the Torah and prophets

At the heart of synagogal worship is the public reading of Scriptures. This takes place at the morning service on sabbaths, holy days, and festivals, on Monday and Thursday mornings, and on sabbath afternoons. The readings from the Pentateuch are presently arranged in an annual cycle so that, beginning on the Sabbath following the autumnal festivals with Gen. 1:1, the entire five books are read through the rest of the year. The texts for festivals, holy days, and fasts reflect the particular significance of those occasions. In addition, a second portion from the prophetic writings (in the Jewish tradition these include Joshua, Judges, Samuel, and Kings, as well as the three major and 12 minor prophets, but not Daniel) is read on many of these occasions. All of this takes place within the structure of public worship and is provided with ceremonies during which the Sefer Torah ("Book of the Torah), the pentateuchal scroll, is removed from the ark (cabinet) at the front of the synagogue, and carried in procession to the reading desk; from it, the pertinent text is chanted by the reader. The text for the service is divided into subsections varying from seven on the sabbath to three at the weekday morning service, and individuals are called forward to recite the blessings eulogizing God as Giver of Torah before and after each of these. The order of worship is composed of the preparatory blessings and prayers noted above, to which are added passages recalling the Temple sacrificial cult (thus relating the present form of worship to the past); the recitation of a number of Psalms and biblical prayers; the Shema and its accompanying benedictions, introduced by a call to worship that marks the beginning of formal public worship; the prayer (tefilla) in the strict sense of petition; confession and supplication (*taḥanun*) on weekdays; the reading of Scripture; and concluding acts of worship. This general structure of the morning service varies somewhat, with additions and subtractions for the afternoon and evening services and for sabbath, holy days, and festivals.

The "Eighteen Benedictions"

The prayer (tefilla), just mentioned, is often called the shemone *ʿesre*, the "Eighteen Benedictions"—although it actually has 19—or the *ʿamida,* "standing," because it is recited in that position. It is made up of three introductory benedictions: praise of the God of the Fathers, of God the Redeemer who resurrects the dead, and of God the holy one who fills the earth with his glory, and of three concluding acts; a prayer for the acceptance of the service, a thanksgiving, and a prayer for peace—with a series of intermediate petitions for knowledge, well-being, acceptance of repentance, forgiveness of sin, and others. On the sabbath and festivals these are replaced by benedictions that mention the specific occasion but are not petitionary, it being considered inappropriate to attend to workday concerns at these times.

Radio Times Hulton Picture Library

Scroll of the Law being shown to the congregation on Simhat Torah in a synagogue on the Tunisian island of Jazirat Jarbah.

While the general outline of this order of service is found throughout the entire Jewish world, the details have varied, both in different periods and in geographic and cultural areas. The public service, requiring the presence of at least ten males, the minyan ("quorum"), is generally led by a synagogal official, the *ḥazzan,* or cantor, but any Jewish male with the requisite knowledge may act in this capacity since there is, quite strictly, no clerical class in the community to whom such leadership is limited (see The rabbi, below).

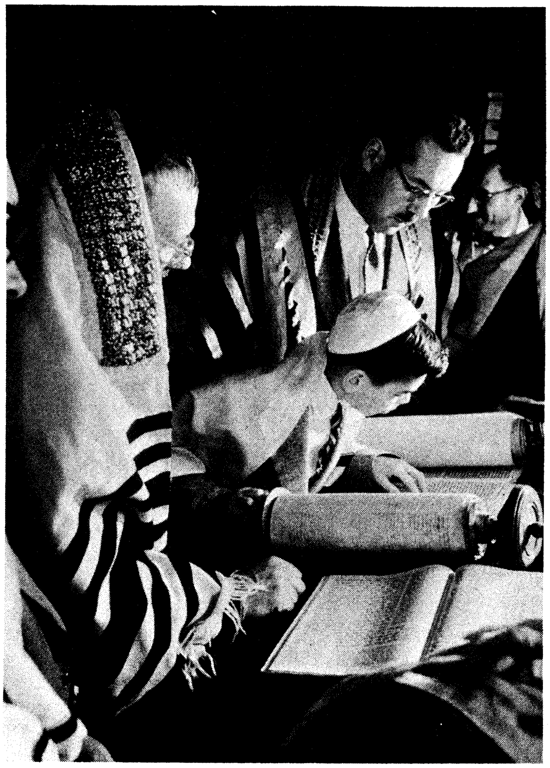
The synagogue room itself has a very simple basic form although, of course, it may be embellished considerably. The only requirements are a container for the Torah scroll(s), the *aron* ha-qodesh ("the holy ark")—a chest against the east wall, or a recessed closet with doors and a curtain; a prayer desk (*ʿamud*) facing the ark at which the reader stands when reciting the service; and the pulpit (*bima*)—according to some requirements in or close to the centre of the room—from which the Torah is read. In the Spanish-Portuguese tradition, only one desk (called teva) is used. The ark contains one or more scrolls, on which are written the five books of Moses. These are variously ornamented, depending upon the cultural region: European communities decking them in coverings of cloth; Oriental (North African and Near Eastern) placing them in wooden or metal containers. In addition, silver ornaments, in the form of towers or crowns, are often set on the tops of two rods on which the scroll is wound, and a breastplate and a pointer are suspended from them.

Accommodations for the worshippers vary according to the cultural milieu, from rugs and cushions in Oriental synagogues to pews and standing desks in European ones. Given this essential simplicity, the synagogue room itself may be used for other purposes than worship, *e.g.,* study and community assembly. Again, this varies with the cultural pattern.

Circumcision and naming

Ceremonies marking the individual life cycles. There are within Jewish life two cycles corresponding to the individual and the synagogal focuses, although they necessarily intermingle. The life of the individual is marked by observances that single out the notable events of personal existence. A male child is circumcised on the eighth day following birth, as a covenantal sign (Gen. 17); the rite of circumcision (berit *mila*) is accompanied by appropriate benedictions and ceremonies, including naming. Females are named in the synagogue, generally on the sabbath following birth, when the father is called to recite the benedictions over the reading of Torah. A firstborn son, if he does not belong to a priestly or a levitical family, is redeemed at one month (in accordance with Ex. 13:12–13 and Num. 18:14–16) by the payment of a stipulated sum to a *kohen* (a putative member of the priestly family). On arrival at the age of 13, a boy is called publicly to recite the Torah benedictions, thus signifying his religious coming-of-age; he is thenceforth obligated to observe the commandments as his own responsibility—he is now a Bar Mitzwa ("Son of the Commandment"). Marriage (*ḥatuna,* also *Qiddushin,* "sanctifications") involves a double ceremony, performed together in modern times but separated in ancient times by a year. First is the betrothal (erusin), which includes the reading of the marriage contract (ketubba) and the giving of the ring with a declaration, "Behold you are consecrated to me by this ring according to the law of Moses and Israel," accompanied by certain benedictions. This is followed by the marriage proper (*nissuʾin*), consisting of the reciting of the seven marriage benedictions. The ceremony is performed under a huppa, a canopy, that symbolizes the bridal bower.

The burial service is marked by simplicity. The body is prepared for the grave by the hevra' qaddisha' (the holy society), clad only in a simple shroud, and the interment takes place as soon after death as possible. In Israel no coffin is used. There are observances connected with death, many of which belong to the realm of folklore rather than Halakhic tradition. A mourning period of 30 days is observed, of which the first seven (*Shivʿa*) are the most rigorous. During the 11 months following a death, the bereaved recite a particular form of a synagogal doxology (Kaddish) during the public service as an act of memorial. The doxology itself, entirely devoid of any mention of death, is a praise of God and a prayer for the establishment of the coming Kingdom. It is also recited annually on the anniversary of the death (*yahrzeit*).



Boy reading the Torah at synagogue services, an important part of the Bar Mitzwa ceremony.
Cornell Capa—Magnum

The cycle of the religious year. The ceremonials related to the life of the individual have complementary observances in the synagogal cycle. Here the calendar rather than the stages of life provides the framework. The synagogal calendar itself reflects two cycles: one beginning in the spring; one in the autumn. The former celebrates those occasions that have an agricultural as well as an historical significance: Pesah (Passover), commemorating the Exodus from Egypt and reflecting as well the beginning of the agricultural year; Shavuot (the Feast of Weeks), marking the end of the grain harvest as well as, in rabbinic tradition, the giving of the Torah at Sinai; Sukkot (the Feast of Tabernacles), an autumnal harvest observance, interpreted in historical terms to commemorate divine guidance and protection during the wilderness wanderings. This cycle is brought to a close by Shemini Atzeret (the Eighth Day of Solemn Assembly), also serving as the conclusion and new beginning of the annual cycle of Torah readings.

The other cycle is less clearly articulated. It begins in the autumn with Rosh Hashana (New Year), a solemn festival introducing a ten-day period of penitence that concludes with the fast of Yom Kippur (Day of Atonement). At midwinter, Hanukka (Feast of Dedication) is observed, commemorating victory over the Hellenistic Syrian Empire in 160 BCE. In late winter, Purim, celebrating the deliverance of the Jews of the Persian Empire, is observed. Among the several fasts in addition to Yom Kippur, Tisha be-Av (the Ninth of Av), observed in late summer in remembrance of the destruction of the Jerusalem Temple by the Romans in 70 CE, is the most important. Running through both cycles and binding them together is the Sabbath.

All of these holy days, festivals, and fasts have both their communal and individual focuses. All, or nearly all, combine both synagogue and home observances and obligations. For a more detailed discussion see JEWISH RELIGIOUS YEAR.

The Jerusalem Temple

Holy places: the land of Israel and Jerusalem. The land of Israel, as is evident from the biblical narratives, played a significant role in the life and thought of the Israelites. It was the promised home, for the sake of which Abraham left his birthplace; the haven toward which moved the tribes who escaped from Egyptian servitude; the hope of the exiles in Babylon. In the long centuries following the destruction of the Judean state by the Romans, it remained inextricably bound up with messianic and eschatological expectations. During the early period of settlement, there seem to have been many sacred localities, with one or another functioning for a time as a central shrine for all of the tribes, without displacing the others. Even the establishment of Jerusalem as the political capital by David and the building of a royal chapel there by Solomon did not bring to an end local cult centres. It was not until the reign of Josiah of Judah (640–609) that a reform centralized the cult in Jerusalem and attempted—although not entirely successfully—to end worship at local shrines. However irregular was the effectiveness of this reform, the Babylonian Exile and the subsequent return saw Jerusalem and its Temple win out over its rivals and become—in law, in fact, and in sentiment—the centre of Jewish cultic life. As noted above, this did not inhibit the rise and development of other forms of worship and even—on a few occasions—other cult centres. Nonetheless, no matter how unpopular the priesthood of the Jerusalem Temple became with some segments of the population—the Qumran community seems to have denied its legality, and the Pharisees complained bitterly about its arrogance and exactions, attempting when politically feasible to impose and enforce Pharisaic regulations upon it—reverence for the Temple itself seems to have remained a widespread sentiment. With the destruction of the Temple by the Romans in 70 CE, such reverence was transformed both by messianic expectations and eschatological hopes into fervent devotion, which, over the following centuries, became idealized and even supernaturalized. The most ardently articulated statement of the crucial role of the land of Israel and the Jerusalem Temple is found in the Sefer *ha-Kuzari* of Judah ha-Levi in which the two are seen as absolutely indispensable for the proper relation between the people Israel and God. Symbolizing the significance of the land and of the city is the practice of facing in their direction during worship. The earliest architectural evidence derived from synagogue remains in Galilee indicates that the attempt was made to arrange the building in such a way that the worshippers faced directly toward Jerusalem. This practice may have continued even in the Diaspora, but at a later date the present practice of setting the holy ark in or before the east wall was established, so that "facing Jerusalem" is now more symbolic than actual.

Use of the vernacular in worship

The sacred language: Hebrew and the vernacular tongues. The transformation of Hebrew into a sacred language is, of course, bound up with the political fate of the people. In the period following the return from the Babylonian Exile, Aramaic, a cognate of Hebrew, functioned as the international or imperial language in official life and certainly gained a foothold as a vernacular. It did not, despite claims made by some scholars, displace the everyday Hebrew of the people. The language of the Mishna, far from being a scholar's dialect, seems to reflect—in the same way as the Koine (common) Greek of the New Testament—popular speech. Displacement of Hebrew—both in its literary form in Scriptures and in its popular usage—did take place in the Diaspora, however, as evidenced by the need to translate Scriptures into Greek in some communities and into Aramaic in others. As far as the emerging order of worship is concerned, there seems also to have been an inclination on the part of some authorities to permit even the recitation of the Shema complex in the vernacular. Struggles over these issues within the communities continued for a number of centuries in various places, but the development of formal literary Hebrew—a sacred tongue, to be used side by side with the Hebrew Scriptures in worship—brought them to an end. Although the communities of the Diaspora used the vernaculars of their environment in day-to-day living and even—as in the case of the communities of the Islāmic world—for philosophical, theological, and other scholarly writings, in worship, Hebrew remained the standard until modern times when some of the reform movements in western Europe sought partially—and a very small fraction even totally—to displace it.

The rabbi: legal, judicial, and congregational roles. The rabbinate, with its peculiar nature and functions, is the result of a series of developments going back to the period that followed the disastrous second revolt against Rome (132–135 CE). The term rabbi ("my teacher") was originally an honorific title for the graduates of the academy directed by the *nasi* or patriarch, the head of the Jewish community in Palestine in that era, who was also a Roman imperial official. The curriculum of the school was Torah, written and oral, according to the Pharisaic tradition and formulation. The *nasi* appointed rabbis to the law court (the Bet Din) and as legal officers of local communities: acting with the local elders, they supervised and controlled the life of the community and its members in all of its aspects. A similar situation obtained in Babylon under the Parthian and Sāsānian empires, where the *resh galuta* or exilarch ("head of the exile") appointed rabbinical officials to legal and administrative posts. In time the patriarchate and exilarchate disappeared, but the rabbinate, nourished by independent rabbinical academies, survived. An authorized scholar, when called to become the judicial officer of a community, would at the same time become the head of the local academy and would, after adequate preparation and examination, grant authorization to his pupils, who were then eligible to be called to rabbinical posts. There was, thus, a diffusion of authority, the communities calling, rather than a superior official appointing, their rabbis. What must be kept in mind is that these rabbis were not ecclesiastical personages but communal officials, responsible for the governance of the entire range of life of what was understood to be the *kehilla qedosha,* the "holy community."

The modern rabbi

In modern times and particularly in the Western world, the total change in Jewish communal existence required a transformation of this ancient structure. The rabbinate became, for the most part, an ecclesiastical rather than a communal agency, reflecting the requirements of civic life in modern national states. The education of rabbis who now function within this new situation is carried on in seminaries whose structure and curriculum have been influenced by European and American academic institutions. The majority of their graduates serve as congregational rabbis, in roles similar to those of ministers and priests in the Christian denominations, but with some other functions deriving from the particular situation and nature of the Jewish community.

Even in the State of Israel, where certain larger areas, such as that of family law, are still reserved to the rabbinate, it nonetheless functions more as a counterpart to other ecclesiastical organizations, Christian and Muslim, than as an overarching and all-inclusive communal agency that embodies, as in the past, involvement in every aspect of community and personal life.

Chief rabbinates. The existence of the offices of chief rabbi in the State of Israel derives from the situation in the Turkish Empire when the various religious communities functioned as quasi-political entities in that multiethnic conglomerate. Israel has two chief rabbis, one for the Ashkenazic (European) and one for the Sefardic (Oriental) communities—they no longer function, however, as the heads of whole communities but only of ecclesiastical organizations. The same is true in those countries outside Israel that have the office of chief rabbi; *e.g.,* Great Britain and France. Here they function vis à vis the governments like their ecclesiastic counterparts in the Christian churches. While they have certain kinds of limited authority because of their official position, their jurisdiction extends only over those members of the total Jewish community who are ready to accept it; others form their own ecclesiastical units and act without reference to the chief rabbinate. In some situations, particularly in the

United States where there is no similar structure, the title chief or grand rabbi has been assumed occasionally by individuals as the means of asserting superior dignity or even (fruitlessly) authority.

General councils or conferences. The precise nature of the Sanhedrin (Council Court) in the last years of the Jewish commonwealth is a much disputed matter. The several councils mentioned in Talmudic literature are equally difficult to define with exactitude. There are references scattered throughout medieval literature that suggest the existence of councils and synods but their composition and authority are also uncertain. Around 1000 a synod was held in the Rhineland in which French and German communities participated under the guidance of Rabbenu Gershom, the leading rabbinic authority of the region. The late Middle Ages saw the rise in eastern Europe of the Waʿad Arba' Aratzot (Council of the Four Lands) composed of communal representatives from Great Poland, Little Poland, Russian Poland (Volhynia), and Lithuania. At the beginning of the modern era Napoleon (1806) summoned an Assembly of Notables—representatives of communities under French dominion—to deal with questions arising from the dissolution of the older status of the Jews and their naturalization as individuals into the new national states. Those decisions of the Assembly that involved questions of Jewish law were subsequently submitted to a Grand Sanhedrin called into being by Napoleon to provide some sort of Halakhic justification for the acts the French imperial government had required of the Jewish communities. During the 19th century the demand for the reform of Jewish life—principally the liturgy of the synagogue, but many other aspects as well—evoked a series of rabbinical conferences and synods that debated the questions and sought to guide the changes thought to be necessary. A similar procedure was followed on the American scene. In both instances, after an initial period in which radicals, moderates, and conservatives argued their respective cases in the same forum, polarization set in and intellectual differences were transformed into competing organizations. In the 1970s the several tendencies within the Jewish communities in North America were institutionalized in rabbinical conferences and congregational unions—Orthodox, Conservative, and Reform—whose influence was in large measure limited to their adherents. In the United States the Synagogue Council of America claims to be the "united voice" of American Jewry in common concerns. There is also a worldwide body in Reform or Liberal Judaism—the World Union for Progressive Judaism.

Modern variations. The above sketch of basic practices and institutions has attempted to describe the so-called traditional situation, although it has been indicated that even here there are variations—actually more than have been noted. In addition, reference has been made to some shifts and changes that represent a giving up of traditional practices on the basis of intellectual decisions about the nature of Judaism, its beliefs, practices, and institutions. Such changes are far too numerous to describe in detail. What is more important is to indicate their motivation. Basically, it is the view that the Halakhic system is not, as a whole and in all of its parts, divinely revealed but is rather a human process that seeks to expose in mutable forms the meaning of the divine-human encounter. Thus viewed, the practices and institutions are understood to be historically determined, reflecting the multifaceted experience of the people Israel as it has sought to live in the presence of God. Historical scholarship has, from this point of view, disclosed the origins, rise, development, and decline of these structures in the past and thus authorizes such changes in the present and future as appear to fulfill the needs of the community and its members. An examination of the specific deviations from the traditional forms makes clear that the application of this position, or attitude, has been subject to wide variation during the 19th and 20th centuries, in which it has operated. Some have seen it as a call for the disengagement from much if not all of the traditional pattern, and a recognition that only the spiritual essence is of importance or consequence for Judaism. Others have argued that an indiscriminate use of historicism (the explanation of values and forms in terms of their historical conditions) is unjustified and that the burden of proof is always upon those who would introduce changes. In the post-World War II period, the question has been whether a reconstituted Halakhic system might not be a requirement of the day.

Reinterpretations of the Halakhic system

ART AND ICONOGRAPHY

The anti-iconic principle and its modifications. Although the Second Commandment (Ex. 20:4; Deut. 5:8), "You shall not make yourself a graven image, or any likeness of anything that is in heaven above, or that is in the earth beneath, or that is in the water under the earth," has indeed been understood as absolutely prohibiting any and all artistic representation, this is not the only way in which these words may be interpreted. What is intended is a prohibition against the construction of such likenesses as were the object of worship in the cultural area in which the Israelites dwelt. Even in the Bible there are reports of artistic productivity in the construction of the tent sanctuary and its ritual vessels (Ex. 25–31) and of the Temple in Jerusalem (I Kings 6–7). The literalness and rigour with which the commandment was interpreted depended upon the larger situation of the community, so that during periods of external pressures toward religious conformity, such as the reign of Antiochus IV Epiphanes in Antioch (175–164 BCE), the anti-iconic attitude sharpened. Similarly, during the Roman occupation, the presence of the battle standards of the legions with their animal representations was looked upon as an affront, while extreme Pietists would not even handle Roman coinage because of the images stamped on it. On the other hand, the walls of a 3rd-century-CE synagogue in Doura-Europus in Syria are covered from floor to ceiling with biblical scenes with human representations, and a number of synagogues in Palestine had elaborate mosaic floors with the signs of the zodiac, representations of the seasons, and the like. Further, illuminated manuscripts from the medieval period in Europe were frequently decorated with biblical figures, some quite clearly copied from Christian prototypes. A fascinating mediating position is to be seen in a Haggada, in which the human figures have bird heads. Synagogues from a later, although pre-emancipation, period (before the 18th century) were often decorated with animal figures. In the modern period the avoidance of human figures has not been entirely accomplished, although nothing like the decorations of Doura-Europus has appeared.

The synagogue at Doura-Europus

Ceremonial objects and symbols. Nonetheless, given this general anti-iconic attitude, much of Jewish artistic endeavour has been directed toward the creation of ceremonial objects: *kiddush* goblets, candlesticks and candelabra, spice boxes for the *havdala* ceremony at the end of the sabbath, ornamented containers for the *mezuza* (a parchment on which is written Deut 6:4–9 and 11:13–21, fastened to the doorpost on the right side as one enters), the silver crowns placed on the Torah scrolls, together with the mantles and breastplates for the same, and many other objects designed to embellish the performance of the large number of ritual acts of the individual and the community. All of these vary in artistic quality, from the work of simple artisans to exquisitely produced works of master craftsmen.

Architecture. The building of synagogues, too, is an expression of artistic interest and concern, as well as of religious and social function. Nothing is known of these edifices, if indeed there were any, until the Greco-Roman period. Then the Roman basilica often provided the appropriate model. What was required was a spacious hall for assembly, and galleries for the women, and this form served that purpose very well. However synagogues were furnished before the destruction of the Second Temple, after that event some attempt seems to have been made to transfer some of the latter's appurtenances to the former, a move that was successfully resisted. When possible, the synagogue stood on a hill. Before it stood a walled entrance court with a fountain for ablutions. Before the Temple destruction, the building may have been oriented

Synagogal sites and structures

with its doors facing eastward, but afterward they faced Jerusalem; still later, when the holy ark containing the Torah scrolls was placed in a fixed position, the orientation was reversed so that the central gate would not be blocked; ultimately, the ark was placed in or against the east wall, without reference to the actual direction of Jerusalem. As the Diaspora grew larger, the new communities adapted the architectural forms of the enveloping culture. The surviving buildings of the Muslim period in Spain are often built with the horseshoe arches and decorated with the exquisite stucco arabesques that mark the era. The medieval period in Christian Europe saw a revival of a very strict anti-iconic attitude and a gradual rejection of the church edifice in favour of secular buildings as a model for the synagogue. The increasingly limited role of the Jew in that society and the enlargement of restrictions by church and state made it necessary to modify the synagogal structure. The doors no longer were in the wall facing the ark; the courtyard grew smaller; galleries were discontinued (side rooms now serving as the women's section); and a double- rather than a triple-aisled construction was largely favoured. Similar developments took place in eastern Europe with the building of fortress-synagogues and the remarkable wooden synagogues of Poland. In the early postemancipation period, Baroque style had its day, followed by Greek temples, Romanesque, Gothic, and pseudo-Byzantine churches, and pseudo-Moorish mosques. In the most recent period, the various schools of functionalism and their commercial descendants have come to the fore. The best of these have brought together fine architectural design and beautifully conceived and executed decoration. The interior arrangement, even in some traditional synagogues, has been influenced by the Protestant sermon-centred form of worship, so that some of the unique forms that marked older structures are absent. The holy ark is, however, still a centre of attention and has often been treated in interesting and striking ways.

Paintings and illustrations. As noted above, the use of paintings in the decoration of synagogues goes back to at least the 3rd century CE and is found in the late pre-emancipation and modern synagogues as well. Manuscripts, too, were illuminated with miniatures and the Renaissance period saw the appearance of beautifully decorated Scrolls of Esther and *ketubbot* (marriage contracts). Nonetheless, the appearance of Jewish artists in painting and sculpture is a modern phenomenon. Beginning in the 19th century, interest grew apace and more and more Jews are to be found, often in the avant-garde of these fields. Some, such as Marc Chagall and Jacques Lipchitz, have done specifically religious art.

Liturgical cantillation

Music. The description of the synagogue service above noted the role of the *hazzan*, or cantor. It is he who reads the service and declaims the scriptural lessons to certain set musical modes that vary with the season and occasion. Many of these call for melodic responses on the part of the congregation. The origins and varying developments of these chants are ancient, often obscure, and equally complicated. Whatever the basic materials, these were enlarged, varied, corrupted, and reworked over the centuries in the various environments in which the Jewish communities have lived. In modern times musicologists have begun to examine with great care the history of synagogal music, analyzing its basic structures and its relationship to the music of Christian liturgical traditions. In the 19th century in Western Europe much of the traditional music was either discarded or re-worked under the influence of western forms and styles. In addition the pipe-organ was introduced and was the centre of stormy controversy.

Literature. Literature has been throughout the ages the home of Jewish artistic activity. The Hebrew Bible is a work of monumental artistry, exhibiting grandeur of form and language in historical narrative, poetry, rhetoric, and aphorism. The extrascriptural writings of the period, although their originals have often vanished, still disclose literary genius of a high order in translation. The documents of the rabbinic tradition are not often looked at with an eye to their literary worth but much of the material, particularly the Haggadic portions of the Midrashim, reveals a noteworthy sensitivity to the uses of language. In the medieval period much attention was given to the production of *piyyuṭim*, liturgical poetry with which to embellish the Siddur (prayer book), itself a collection containing much imaginative, as well as pedestrian, writing. In the Islamic world, under the influence of Arabic poetry, Hebrew poetry rose to a high peak in both liturgical and secular forms. The Middle Ages in the Rhineland also saw the beginnings of the Jewish form of Middle High German that was, over the centuries, to develop into an autonomous Jewish language, Yiddish, which, in the 19th century, became a literary vehicle of very high order. The same period saw the beginnings of the recreation of Hebrew into a literary language that has become the basis of the spoken vernacular of the State of Israel and of a flourishing literature encompassing every branch of the field. Since the emancipation at the end of the 18th century, Jews in western Europe and later in the United States have turned to literature in the vernaculars of their countries, and have produced writers of note dealing with both Jewish and general themes.

RELATION WITH NON-JUDAIC RELIGIONS

Exclusivist and universalist emphases. The biblical tradition out of which Judaism emerged was predominantly exclusivist ("no other gods"). The gods of the nations were regarded as "no gods" and their worshippers as deluded, while the God of Israel was acclaimed as the sole lord of history, and the Creator of heaven and earth. The unexpected universalist implications of this exclusivism are most forcibly expressed in an oft-quoted verse from Amos (9:7):

God's rule over all the nations

> "Are you not like the Ethiopians to me, O people of Israel?" says the Lord. "Did I not bring up Israel from the land of Egypt, and the Philistines from Caphtor and the Syrians from Kir?"

Here the universal rule of the God of Israel is unmistakably proclaimed. Yet in the same book (3:1–2), after referring to the deliverance from Egypt—an act recognized as similar to that occurring in the affairs of other peoples—the prophet, speaking for God, says: "You only have I known of all the families of the earth." Thus the exclusivism has two focuses, one universal, the other particularistic. The ultimate claim of the universalistic position is found in Malachi 1:11: "For from the rising of the sun to its setting my name is great among the nations." This, however, in no way negates the special covenantal relationship between God and his people; indeed, it is this universalistic theme that underscores that special relation. To interpret Judaism's stance toward other religious systems in any other way is to fail to do justice to its inner dialectic. It is neither a bland latitudinarianism that admits any or all viewpoints and practices, nor a fanatical intolerance but rather a subtle interplay of affirmation and rejection. The latter is directed primarily against the worship of finite things or aspects—idolatry—the basic failure of the peoples who are the objects of the same divine solicitude as is Israel. If the religions of the nations are rejected because of their failure fully and truly to know God, the peoples themselves are not. Living under the covenant with Noah (see above), their fulfillment of such responsibilities provides for their acceptance, for they are not expected to live within the realm of Torah (see also *Relations with other religions* below).

Relation to Christianity. Judaism's relation to Christianity is a complicated one because of the close historical interconnections between them. From a Judaic standpoint, Christianity is or was a Jewish "heresy" and as such may be judged somewhat differently than other religions. Its claims over against Judaism as the true fulfillment of the covenant and, thus, as the true Israel have given rise throughout the centuries to polemics of varying intensity. The rise to power of the church and the embodiment of its anti-Judaic sentiments and attitudes in the political structures and processes of Christian nations made sharply negative Jewish responses inevitable. Nevertheless, during the Middle Ages Jewish thinkers attempted to avoid designating Christianity as idolatry and

Christian and Jewish polemic and counter-polemic even to argue that, in a special way, being derived from Judaism, it was fulfilling—at least on the moral plane—the divine purpose.

In modern times the relation has undergone changes necessitated by the newer situations into which the Jewish community has moved. This does not mean that the polemical-apologetic stance has come entirely to an end. The rejection of Judaism as a living religion by Christians continued and continues, argued not so much on dogmatic as on scholarly grounds. The Jewish response to this has often been countercriticism. Beyond this, however, there has been a growing inclination within the Jewish community to respond to the development of an affirmative theology of Judaism in both the Roman Catholic and Protestant churches by providing a theology of Christianity within Jewish thought. Occasional formulations in this direction have appeared, but it is far too early to know exactly what will emerge. At the same time, it must be noted, there are those who see no need for such a movement, arguing that the failure of the Christian churches in recent years to respond adequately to the tragedies of Jewish existence precludes any real engagement one with the other.

Relation to **Islām.** The emergence of Islām in Arabia in the 7th century CE brought Judaism face to face with a second religious movement that derived some of its ideas and structures from the older tradition. In this case, as in that of Christianity, the new religion claimed a special relation with Judaism. Muhammad held that the faith he proclaimed was none other than the pristine religion of Abraham, the father of Ishmael—progenitor of the Arabs—as well as of Isaac, from whom the people of Israel descended. That religion had been distorted both by Judaism and Christianity; and Muḥammad, the "seal" of the prophets, had been called by God to restore it to its purity. The confrontation between Judaism and Islām, as that with Christianity, was coloured by political and social considerations both before and after Islām moved out of Arabia to build a world empire (including the conquest and settlement of Palestine). During the subsequent period, the intellectual development of the Islamic world and the emergence of theologians and philosophers of the highest order challenged Judaism and had considerable influence on the rise of similar thinkers within that community. Given the strong monotheism and the anti-iconic attitude of Islām, many of the questions that arose between Judaism and trinitarian and iconic Christianity were not an issue between Judaism and Islam. The crucial point of dispute here was the nature of prophecy, given Muhammad's claim concerning his culminating role in the prophetic tradition. The medieval period thus saw polemics directed against that claim and, as in the case of the theological work of Moses Maimonides, *More nevukhim* ("The Guide of the Perplexed"), an exposition of the nature of prophecy that, without directly dealing with Muḥammad's claim, may be understood to undercut it. Nonetheless, Islām, too, was understood to contribute to the fulfillment of the divine purpose. From the late medieval period onward, the intellectual engagement between the two religions diminished with the general decline in the Turkish Empire that then embraced the Muslim world. In modern times it has not yet been renewed for many reasons. Once the political problems in the eastern Mediterranean between the State of Israel and the Arab world have been meliorated, the contiguity of the two communities suggests an inevitable renewal of conversations on the religions as on many other levels.

Relations with other religions. Judaism's encounters with religions other than Christianity and Islam have been in large measure limited to the past. In the Hellenistic world, it confronted and rejected the varieties of syncretistic cults that grew up. Within the Sāsānian Empire it was forced to deal with Zoroastrianism, but the outlines of its response have not yet been entirely disentangled from the literature of the period. In the modern world, particularly in the most recent period, it has come face to face with the religions of the Middle and Far East, but beyond a few tentative explorations nothing tangible has appeared. What seems certain is that, considering the growing interest in and exchange between East and West, Jewish thinkers will not be able to rest with older formulations concerning the nature of other religious systems. Without compromising its own faith or falling into an uncritical relativism, Judaism may indeed in the future seek a new way of understanding and relating to the varieties of religious systems facing it on the world scene.

Judaism confronts the religions of mankind

THE ROLE OF JUDAISM IN WESTERN CULTURE AND CIVILIZATION

Its historic role. Given the relationship between Judaism and Christianity—the dominant religious force in the development of Western culture—the role of Judaism in that development was significant. Although the church drew from other sources as well, its retention of the sacred Scriptures of the synagogue (the "Old Testament") as an integral part of its Bible—a decision sharply debated in the 2nd century CE—was crucial. Not only was the development of its ideas and doctrines deeply influenced, but it received as well an ethical dynamism that constantly overcame an inclination to withdraw into world-denying isolation. It was, however, not only Judaism's heritage but its persistence that touched Western civilization. The continuing existence of the Jews, even as a pariah people, was both a challenge and a warning; and ultimately, at the beginning of the modern era, their liberation from the shackles of discrimination, segregation, and rejection was understood by many to be the touchstone of all human liberty. Until the final ghettoization of the Jew—it is well to remember that the term "ghetto" belongs in the first instance to Jewish history—at the end of the Middle Ages and the beginning of the Renaissance, intellectual contact between Judaism and Christianity, and thus with Western culture, did not cease. Jerome translated the Hebrew Bible into Latin with the aid of Jewish scholars; Luther, into German with the aid of commentaries beholden to Jewish authors. Jewish thinkers mediated the remarkable intellectual achievements of the Islāmic world to Christian Europe and added their own contributions as well. Even heresies within the church found, on occasion, their inspiration or prototype in Judaism.

Its present role. In the modern world, while the influence of Jews has increased in almost every realm of cultural life, the impact of Judaism has diminished. The reason for this is not difficult to find. The Gentile leaders who extended emancipation to the Jews at the end of the 18th and beginning of the 19th centuries, while eager to grant political equality to the individual Jew, did so with the implicit and explicit requirement that conformity through reforms of Judaism be agreed to. With the transformation of Judaism into an ecclesiastical institution, largely on the model of German Protestant churches, its ideas and structures took on the cast of its environment in a way quite unlike what had ensued in its earlier confrontations with various philosophical systems. Indeed, for some, Judaism and 19th-century European thought were held to be not merely congruent but identical. Thus, while numerous contributors to diverse aspects of Western culture and civilization are to be found among Jews of the 20th century—scientists, politicians, statesmen, scholars, musicians, artists—their activities cannot, except in specific instances, be considered as deriving from Judaism as it has been sketched above.

Future prospects. Two events of the 20th century have, however, confronted Judaism in such ways as to suggest that its wrestling with them and their profound challenge to it may presage a new role and a new influence for Judaism: "Auschwitz" and the establishment of the State of Israel. The premeditated murder of some 6,000,000 European Jews by the Nazis for no other reason than that they were Jews, has shaken Jewish thinkers to their very core. Indeed, so traumatic was this event, that for almost two decades following it, no substantial attempt was made to plumb its meaning. At the same time, the reappearance of the State of Israel, viewed for the most part from outside the Jewish community as nothing more than a political event, has set in motion an entirely different chain of theological inquiry. These two happenings have

The two key events: Auschwitz and the establishment of Israel clearly, but in as yet unpredictable ways, begun to work and to move within the thought of contemporary Judaism. Out of this working and moving there may emerge an inescapable spiritual impact upon Western culture and civilization, which have, as yet, resolutely refused to face the realities these fateful occurrences represent. If contemporary Judaism is able to say what they mean, however haltingly, it will have renewed its potent relationship to the Western world, and, given the nature of contemporary society, established a similar bond with the Eastern world as well.

PRESENT-DAY FORMS OF JUDAISM

Orthodoxy. Statistical evidence in the field of religious beliefs and behaviour being notoriously inaccurate, to write that the majority of the Jewish religious world is Orthodox is to state a fact—if such it is—that is irrelevant, for it falls victim of the difficulties with which this article began. What may be noted is that there are many Jews who, in one way or another, think of themselves as traditionalists—whatever the content of their beliefs and the scope of their behaviour. This group is composed of many subgroups that include upper-class American scientists and intellectuals at one extreme and impoverished Oriental Kabbalists at the other, with a host of intermediate stages. Their understanding of the affirmations that this article suggests are central to traditional Judaism varies from rationalistic to mystical approaches. The breadth of their observances too, discloses a variation. Yet, despite such caveats, there is a large, fervent, and devoted section of the Jewish community that sees itself affirming--often in the most unselfconscious manner—what in its several ways it understands to be the unchanged and unchanging faith of Israel, so that for many, the foreign designation "orthodox" (used first by Reform Jews for traditionalist Jews) makes little if any sense.

Reform or Liberal. On the other hand, far more self-consciously, there stands the Reform (not "Reformed," except in a few minor and inconsequential instances) or Liberal group, although here, too, variety rather than uniformity is its badge. Although it originated in large measure in 19th-century Germany, its centre and more radical form is now in North America, with smaller groupings in western Europe, Israel, South Africa, and South America. At the moment in considerable intellectual flux, because of the impact of some forms of Existentialism, in the past it built upon a base of post-Kantian ethical philosophy and historical criticism, although theological speculation was not---on the American scene —its métier. In the matter of observance, wide variances are also to be seen—from a relatively conservative traditionalism to most radical topical Reformism.

Conservatism. The Conservative movement—primarily an American phenomenon, although the attempt has been made to connect it integrally with the moderate historical-positive wing of the German Liberal movement —is sociologically a Reform, or Liberal, development among the eastern European immigrants and their descendants in the United States. Originally less active in the field of social issues than the Reform movement, it retained a firmer hold upon the reality of the Jewish people than did most of the reformers. Theoretically, according to some of its intellectual leaders, it reaffirms traditional Judaism, but, in fact, it has included a very wide spectrum touching "Orthodoxy" on the right and Reform on the left. Like the latter, in the 1970s it was in a period of intellectual change, buffeted by the same intellectual winds and seeking new formulations.

Reconstructionism. Reconstructionism, for a long time the left wing of Conservatism, emerged as an independent entity in American Jewish life. It represents the attempt to build an intellectual base compounded of a sociological analysis of the American Jewish community, a statement about the nature of religion derived from the French anthropologists Lévi-Bruhl and Durkheim, and John Dewey's exposition of the meaning of the religious in human experience.

Hasidic groups. The various Hasidic groupings represent in large measure the self-conscious right wing of "Orthodoxy." Ḥasidism is rooted in the mystical Pietist revival in the eastern reaches of 18th-century Poland, Hungary, and western Ukraine, where it was a radical response both to the dangers of the pseudomessianic movements at the end of the 17th and the beginning of the 18th centuries and to the aridity that had enveloped rabbinism. It has now, for the most part, become a highly stylized, introverted sectarianism avoiding contact with the rest of the Jewish community, which it barely recognizes.

Regional or ethnic groups. There are, in addition to these "doctrinally" defined groupings, several that may be described as regional or ethnic. Among them are the Bene-Israel, descendants of Jewish settlers in the Bombay region of India, whose deviation in some Halakhic matters from the present Orthodox consensus has raised problems for those among them who have migrated to Israel; the Falashas, the Jewish community of Ethiopia whose development has been almost entirely outside the mainstream described in this article; and the Black Jews of the United States whose place in, and relation to, the rest of the community is unclear but not for racial reasons.

BIBLIOGRAPHY

Basic beliefs and doctrines: KAUFMANN KOHLER, *Jewish Theology: Systematically and Historically Considered* (1918, reprinted 1928); YEHEZKEL KAUFMANN, *The Religion of Israel from Its Beginnings to the Babylonian Exile* (1960; orig. pub. in Hebrew, 1937–56), an abridgement and translation of the work of one of the most influential Jewish biblical scholars of modern times; GEORGE FOOT MOORE, *Judaism in the First Centuries of the Christian Era,* 3 vol. (1927–30), a masterful work by a distinguished Christian student of Judaism; SOLOMON SCHECHTER, *Some Aspects of Rabbinic Theology* (1909, reprinted 1936 and 1961), an insightful presentation of the basic doctrines; CLAUDE G. MONTEFIORE and HERBERT LOEWE, *A Rabbinic Anthology* (1960), a collection of materials from rabbinic sources, arranged under theologic headings, with ample notes and discussions; JULIUS GUTTMANN, *Philosophies of Judaism: The History of Jewish Philosophy from Biblical Times to Franz Rosenzweig* (1964), a philosophy of Judaism in the form of a history of philosophy in Judaism; ARTHUR HERTZBERG (ed.), *Judaism* (1961); JACOB NEUSNER, *The Way of Torah: An Introduction to Judaism* (1970), a very helpful statement using a history-of-religions approach; LEO BAECK, *Dieses Volk; jüdische Existenz,* 2 vol. (1955–57; Eng. trans., *This People Israel: The Meaning of Jewish Existence,* 1964), a masterful interpretation of Jewish affirmations set within an historical context; ABRAHAM E. MILLGRAM (ed.), *Great Jewish Ideas* (1964), a collection of essays by various scholars setting forth classic positions, covering a wide range of theological ideas.

Ethics and society: SIMON BERNFELD, (comp.), *The Foundations of Jewish Ethics,* 2nd rev. ed. (1968), source materials covering Jewish history, with brief introductions by several notable scholars; MORITZ LAZARUS, *Die Ethik des Judenthums,* 2 vol. (1898–1911; Eng. trans., *The Ethics of Judaism,* 1900), a classic presentation from a 19th-century perspective; SAMUEL S. COHON, *Judaism: A Way of Life* (1948), written from the Reform point of view but deeply sympathetic to a wide range of ideas.

Basic practices and institutions: LEWIS N. DEMBITZ, *Jewish Services in Synagogue and Home* (1898); HAYYIM SCHAUSS, *The Jewish Festivals* (1938; orig. pub. in Hebrew, 1933), and *The Lifetime of a Jew Throughout the Ages of Jewish History* (1950); ABRAHAM Z. IDELSOHN, *Jewish Liturgy and Its Development* (1932, reprinted 1967).

Art and iconography: FRANZ LANDSBERGER, *A History of Jewish Art* (1946); ABRAHAM Z. IDELSOHN, *Jewish Music in Its Historical Development* (1929, reprinted 1967); CECIL ROTH (ed.), *Jewish Art: An Illustrated History* (1961).

Relation with non-Judaic religions: LEO BAECK, *Judaism and Christianity* (1958); SAMUEL SANDMEL, *We Jews and You Christians: An Inquiry into Attitudes* (1967); J.B. SOLOVEITCHIK, "Confrontation," *Tradition,* 6:5–29 (1964); S.D. GOITEIN, *Jews and Arabs: Their Contacts Through the Ages* (1955).

The role of Judaism in Western culture and civilization: CECIL ROTH, *The Jewish Contribution to Civilization* (1938).

The present-day forms of Judaism: LEON D. STITSKIN (ed.), *Studies in Torah Judaism* (1969); ALFRED JOSPE (ed.), *Tradition and Contemporary Experience: Essays on Jewish Thought and Life* (1970); ARNOLD JACOB WOLF (ed.), *Rediscovering Judaism: Reflections on a New Theology* (1965);

The Condition of Jewish Belief: *A* Symposium, Compiled by the Editors of *Commentary* Magazine (1966); MORDECAI M. KAPLAN, *Judaism As* a Civilization, 2nd ed. (1957); SOLOMON POLL, The Hasidic Community of *Williamsburg: A* Study in *the* Sociology of *Religion* (1969); BERNARD MARTIN (ed.), Contemporary *Reform* Jewish Thought (1968); CHARLES LIEBMAN, "Orthodoxy in American Jewish Life," American Jewish Yearbook (1965); MORDECAI WAXMAN (ed.), Tradition *and* Change: The Development of Conservative Judaism (1958).

(L.H.S.)

Judaism, History of

The history of Judaism in this article includes the development of the religion of the Jews in the broadest and most complete sense, from their early ancestral beginnings down to contemporary times.

The article is divided into the following sections:

- I. General observations
  - Nature and characteristics
  - Periodization
- II. Biblical Judaism (20th–4th centuries BCE)
  - The ancient Near Eastern setting
  - The pre-Mosaic period: the religion of the patriarchs
  - The Mosaic period: foundations of the Israelite religion
  - The period of the conquest and settlement of Canaan
  - The period of the united monarchy
  - The period of the divided kingdom
  - The period of classical prophecy and cult reform
  - The exilic period
  - The period of the restoration
- III. Hellenistic Judaism (4th century BCE–2nd century CE)
  - The Greek period (332–63 BCE)
  - The Roman period (63 BCE–135 CE)
- IV. Rabbinic Judaism (2nd–18th centuries)
  - The age of the tannaim (135–c. 200)
  - The age of the *amoraim:* the making of the Talmuds (3rd–6th centuries)
  - The age of the *geonim* (c. 640–1038)
  - Medieval European Judaism (950–1750)
- V. Modern Judaism (c. 1750–c. 1970)
  - The new situation
  - The Haskala, or Enlightenment
  - Religious reform movements
  - Orthodox developments
  - Developments in scholarship
  - Developments in philosophy
  - Jewish–Christian relations
  - Zionism
  - American Judaism
  - Judaism in other lands
- VI. Judaism today

I. General observations

NATURE AND CHARACTERISTICS

In nearly 4,000 years of historical development, the Jewish people and their religion have displayed both a remarkable adaptability and continuity. In their encounter with the great civilizations, from ancient Babylonia and Egypt down to Western Christendom and modern secular culture, they have assimilated foreign elements and integrated them into their own socio-religious system, thus maintaining an unbroken line of ethnic and religious tradition. Furthermore, each period of Jewish history has left behind it a specific element of a Judaic heritage that continued to influence subsequent developments, so that the total Jewish heritage at any time is a combination of all these successive elements along with whatever adjustments and accretions are made imperative by the conditions of each new age.

The fundamental teachings of Judaism have often been grouped around the concept of an ethical (or ethical–historical) monotheism. Belief in the one and only God of Israel has been adhered to by professing Jews of all ages and all shades of sectarian opinion. By its very nature monotheism ultimately postulated religious universalism, although it could be combined with a measure of particularism. In the case of ancient Israel (see below Biblical Judaism), particularism took the shape of the doctrine of election; that is, of a people chosen by God as "a kingdom of priests and a holy nation" to set an example for all mankind. Such an arrangement presupposed a covenant between God and the people, the terms of which the chosen people had to live up to or be severely punished. As the 8th-century-BCE (before the Common Era, or BC) prophet Amos expressed it: "You only have I known of all the families of the earth; therefore I will punish you for all your iniquities." Further, it was a concept that combined with the messianic idea, according to which, at the advent of the Redeemer, all nations would see the light, give up war and strife, and follow the guidance of the Torah (divine guidance, teaching, or law) emanating from Zion (a hill in Jerusalem that has a special spiritual significance). With all its variations in detail, messianism has, in one form or another, permeated Jewish thinking throughout the ages and, under various guises, has coloured the outlook of many secular-minded Jews (see also MESSIAH AND MESSIANIC MOVEMENTS).

Law became the major instrumentality by which Judaism was to bring about the reign of God on earth. In this case law meant not only what the Romans called *jus* (human law) but also fas, the divine or moral law that embraces practically all domains of life. The ideal, therefore, as expressed in the Ten Commandments, was a religio-ethical conduct that involved ritualistic observance as well as individual and social ethics, a liturgical–ethical way constantly expatiated on by the prophets and priests, rabbinic sages, and philosophers., Such conduct was to be placed in the service of God, as the transcendent and immanent Ruler of the universe, and as such the Creator and propelling force of the natural world, and also as the One giving guidance to history and thus helping man to overcome the potentially destructive and amoral forces of nature. According to Judaic belief, it is through the historical evolution of man, and particularly of the Jewish people, that the divine guidance of history constantly manifests itself and will ultimately culminate in the messianic age. Judaism, whether in its "normative" form or its sectarian deviations, never completely departed from this basic ethical–historical monotheism.

PERIODIZATION

The division of the millennia of Jewish history into periods—a procedure frequently dependent on individual preferences—has not been devoid of theological or scholarly presuppositions. The Christian world long believed that until the rise of Christianity the history of Judaism was but a "preparation for the Gospel" (preparatio *evangelica*) followed by the "manifestation of the Gospel" (demonstratio *evangelica*) as revealed by Christ and the Apostles. This formulation could be theologically reconciled with the assumption that Christianity had been preordained even before the creation of the world.

On the other hand, 19th-century biblical scholars moved the decisive division back into the period of the Babylonian Exile and restoration of the Jews to Judah (6th–5th centuries BCE). They asserted that after the first fall of Jerusalem (586 BCE) the ancient "Israelitic" religion gave way to a new form of the "Jewish" faith, or Judaism, as formulated by Ezra the Scribe and his school (5th century BCE). A German historian, Eduard Meyer, in 1896 published Die Entstehung *des* Judentums ("The Origin of Judaism"), in which he placed the origins of Judaism in the Persian period (see below Biblical *Judaism*) or the days of Ezra and Nehemiah (5th century BCE) and actually attributed to Persian imperialism an important role in shaping the new emergent Judaism.

These theories have been discarded by most scholars, however, in the light of a more comprehensive knowledge of the ancient Middle East and the abandonment of a theory of gradual evolutionary development that was dominant at the beginning of the 20th century. Most Jews share a long-accepted notion that there never was a real break in continuity and that Mosaic-prophetic-priestly Judaism was continued, with but few modifications, in the work of the Pharisaic and rabbinic sages (see below *Rabbinic Judaism*) well into the modern period. Even today the various Jewish groups, whether Orthodox, Conservative, or Reform, all claim direct spiritual descent from the Pharisees and the rabbinic sages. In actual historical development, however, many deviations have occurred from so-called normative or rabbinic Judaism; discontinuities, as well as continuity.

In any case, the history of Judaism here is viewed as falling into the following major periods of development: biblical Judaism (c. 20th–4th century BCE), Hellenistic Judaism (4th century BCE–2nd century CE), rabbinic Judaism (2nd–18th centuries CE), modern Judaism (c. 1750–1970), and the Judaism of today (1970s).

(S.W.B.)

II. Biblical Judaism (20th–4th centuries BCE)

THE ANCIENT NEAR EASTERN SETTING

The family of the Hebrew patriarchs (Abraham, Isaac, and Jacob) is depicted in the Bible as having had its chief seat in the northern Mesopotamian town of Harran —then (mid-2nd millennium BCE) belonging to the Hurrian kingdom of Mitanni. From there Abraham, the founder of the Hebrew people, is said to have migrated to Canaan (comprising roughly the region of modern Israel and Lebanon)—throughout the biblical period and later ages a vortex of west Asian, Egyptian, and east Mediterranean ethno-culture. Thence the Hebrew ancestors of the people of Israel (named after the patriarch Jacob, also called Israel) migrated to Egypt, where they lived in servitude, and a few generations later returned to occupy part of Canaan. The Hebrews were seminomadic herdsmen and occasionally farmers, ranging close to towns and living in houses as well as tents.

Early Israelite culture

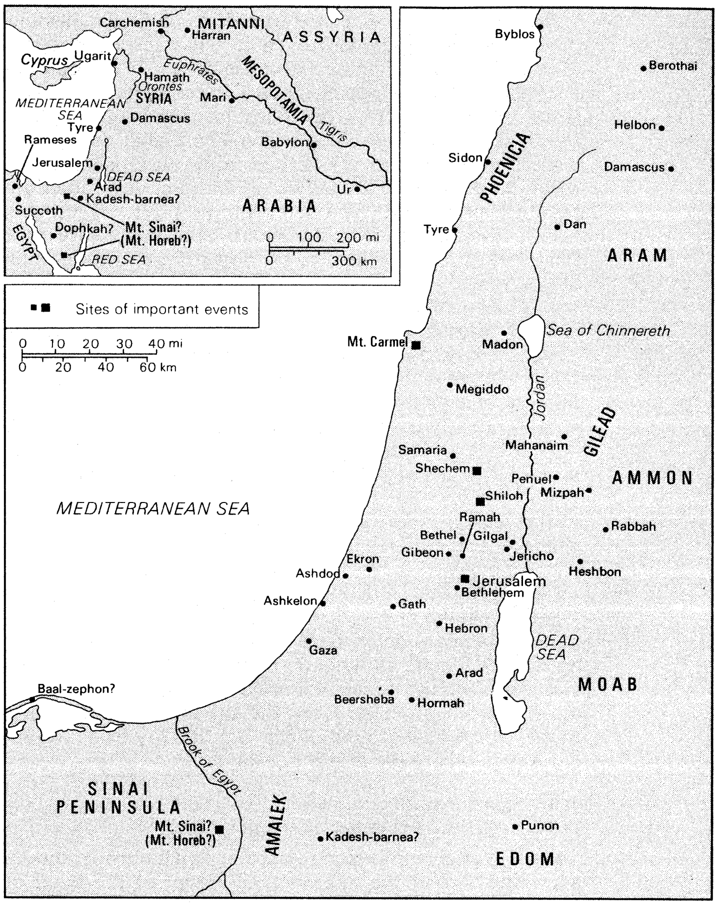
The initial level of Israelite culture resembled that of its surroundings; it was neither wholly original nor primitive. The tribal structure resembled that of West Semitic steppe dwellers known from the 18th-century-BCE tablets excavated at the north central Mesopotamian city of Mari; their family customs and law have parallels in Old Babylonian and Hurro-Semite law of the early and middle 2nd millennium. The conception of a messenger of God that underlies biblical prophecy was Amorite (West Semitic) and found in the tablets at Mari. Mesopotamian religious and cultural conceptions are reflected in biblical cosmogony, primeval history (including the Flood story in Gen. 6:9–8:22), and law collections. The Canaanite component of Israelite culture consisted of the Hebrew language and a rich literary heritage—whose Ugaritic form (which flourished in the northern Syrian city of Ugarit from the mid-15th century to the 12th century BCE) illuminates the Bible's poetry, style, mythological allusions, and religio-cultic terms. Egypt 'provides many analogues for Hebrew hymnody and wisdom literature. All the cultures among which the patriarchs lived had cosmic gods who were believed to have fashioned the world and preserved its order, including justice; all had a developed ethic expressed in law and moral admonitions; and all had religious rites and myths that were sophisticated.

Though plainer when compared with some of the learned literary creations of Mesopotamia, Canaan, and Egypt, the earliest biblical writings are so imbued with contemporary ancient Near Eastern elements that the once-held assumption that Israelite religion began on a primitive level must be rejected. Late-born amid high civilizations, the Israelite religion had from the start that admixture of high and low features characteristic of all the known religions of the area. Implanted on the land bridge between Africa and Asia, it was exposed to cross-currents of foreign thought throughout its history.

THE PRE-MOSAIC PERIOD: THE RELIGION OF THE PATRIARCHS

The God of the patriarchs

Israelite tradition identified YHWH (by scholarly convention pronounced Yahweh), the God of Israel, with the Creator of the world, who had been known to and worshipped by men from the beginning of time. Abraham (perhaps 19th or 18th–17th centuries BCE) did not discover this God, but entered into a new covenant relation with him, in which he was promised the land of Canaan and a numerous progeny. God fulfilled that promise through the actions of the 13th-century-BCE Hebrew leader Moses: he liberated the people of Israel from Egypt, imposed Covenant obligations on them at Mt. Sinai, and brought them to the promised land.



Important sites and regions of biblical Judaism.

Historical and anthropological studies present formidable objections to the continuity of YHWH worship from Adam (the biblical first man) to Moses, and the Hebrew tradition itself, moreover, does not unanimously support even the more modest claim of the continuity of YHWH worship from Abraham to Moses. Against it is a statement in chapter 6, verse 3, of Exodus that God revealed himself to the patriarchs not as YHWH but as El Shaddai —an epithet (of unknown meaning) the distribution of which in patriarchal narratives and Job and other poetical works confirms its archaic and unspecifically Israelite character. Comparable is the distribution of the epithet El Elyon (God Most High). Neither of these epithets appears in postpatriarchal narratives (excepting the Book of Ruth). Other compounds with El are unique to Genesis: El Olam (God the Everlasting One), El Bethel (God of Bethel), and El Ro'i (God of Vision). An additional peculiarity of the patriarchal stories is their use of the phrase "God of my [your, his] father." All of these epithets have been taken as evidence that patriarchal religion differed from the worship of YHWH that began with Moses. A relation to a patron god was defined by revelations starting with Abraham (who never refers to the God of his father) and continuing with a succession of "founders" of his worship. Attached to the founder and his family, as befits the patron of wanderers, this unnamed deity (if indeed he was one only) acquired various Canaanite epithets (El, Elyon, Olam, Bethel, *qone eretz* [possessor of the Land]) only after their immigration into Canaan. Whether the name of YHWH was known to the patriarchs is doubtful. It is significant that while the epithets Shaddai and El occur frequently in pre-Mosaic and Mosaic-age names, YHWH appears as an element only in the names of Yehoshua' (Joshua) and perhaps of Jochebed—persons who were closely associated with Moses.

The patriarchs are depicted as objects of God's blessing, protection, and providential care. Their response is loyalty and obedience and observance of a cult whose ordinary expression is sacrifice, vow, and prayer at an altar, stone pillar, or sacred tree. Circumcision was a distinctive mark of the cult community. The eschatology (doctrine of ultimate destiny) of their faith was God's promise of land and a great progeny. Any flagrant contradictions between patriarchal and later mores have presumably been censored; yet distinctive features of the post-Mosaic religion are absent. The God of the patriarchs shows nothing of YHWH's "jealousy"; no religious tension or contrast with their neighbours appears, and idolatry is scarcely an issue. The patriarchal covenant differed from the Mosaic Sinaitic Covenant in that it was modelled upon a royal grant to favourites and contained no obligations, the fulfillment of which was to be the condition of their happiness. Evidently not the same as the later religion of Israel, patriarchal religion prepared the way for it in its familial basis, its personal call by the deity, and its response of loyalty and obedience to him.

Little can be said of the relation of the religion of the patriarchs to the religions of Canaan. Known points of contact between the two are the divine epithets mentioned above. Like the God of the fathers, El, the head of the Ugaritic pantheon was depicted both as a judgmental and a compassionate deity. Baal (Lord), the aggressive young agricultural deity of Ugarit, is remarkably absent from Genesis. Yet the socio-economic situation of the patriarchs was so different from the urban, mercantile, and monarchical background of the Ugaritic myths as to render any comparisons highly questionable.

THE MOSAIC PERIOD: FOUNDATIONS OF THE ISRAELITE RELIGION

The Egyptian sojourn. According to Hebrew tradition, a famine caused the migration to Egypt of the band of 12 Hebrew families that later made up a tribal league in the land of Israel. The schematic character of this tradition does not impair the historicity of a migration to Egypt, an enslavement by Egyptians, and an escape from Egypt under an inspired leader by some component of the later league of Israelite tribes. To disallow these events would make their centrality as articles of faith in the later religious beliefs of Israel inexplicable.

The importance of Moses, the Exodus, and the Sinai Covenant

Tradition gives the following account of the birth of the nation. At the Exodus from Egypt (13th century BCE), YHWH showed his faithfulness and power by liberating Israel from bondage and punishing their oppressors with plagues and drowning at the sea. At Sinai, he made Israel his people and gave them the terms of his Covenant, regulating their conduct toward him and each other so as to make them a holy nation. After sustaining them miraculously during their 40-year wilderness trek, he enabled them to take the land that he had promised to their fathers, the patriarchs. Central to these events is God's apostle, Moses, who was commissioned to lead Israel out of Egypt, mediate God's Covenant to them, and bring them to Canaan.

Behind the legends and the multiform law collections, a historical figure must be posited to whom the legends and the legislative activity could be attached. And it is precisely Moses' unusual combination of roles that makes him credible as a historical figure. Like Muhammad at the birth of Islām, Moses fills oracular, legislative, executive, and military functions. The main institutions of Israel are his creation: the priesthood and the sacred shrine, the Covenant and its rules, the administrative apparatus of the tribal league. Significantly, though Moses is compared to a prophet in various texts in the Pentateuch (the first five books of the Bible), he is never designated as one—the term being evidently unsuited for so comprehensive and unique a figure.

Mosaic religion. The distinctive features of Israelite religion appear with Moses. The proper name of Israel's God, YHWH, was revealed and interpreted to Moses as meaning ehye *asher* ehye—an enigmatic phrase (literally meaning "I am/shall be what I am/shall be") of infinite suggestiveness. The Covenant, defining Israel's obligations, is ascribed to Moses' mediation. Although it is impossible to determine what rulings go back to Moses, the Decalogue, or Ten Commandments, presented in chapter 20 of Exodus and chapter 5 of Deuteronomy, and the larger and smaller Covenant codes in Ex. 20:22–23:33; 34:11–26) are held by critics to contain early covenant law. From them, the following features may be noted: (1) The rules are formulated as God's utterances —*i.e.*, expressions of his sovereign will. (2) They are directed toward, and often explicitly addressed to, the people at large; Moses merely conveys the sovereign's message to his subjects. (3) Publication being of the essence of the rules, the people as a whole are held responsible for their observance (see also COVENANT).

Moses breaking the tablets of the Law on Mt. Sinai in anger over the Israelites' worship of the golden calf. Etching by Marc Chagall (1887–).

By courtesy of the Societe des Editions Verve, from "La Bible illustree par Chagall." Verve, no. 33/34

The liberation from Egypt laid upon Israel the obligation of exclusive loyalty to YHWH. This meant eschewing all other gods—including idols venerated as such—and the elimination of all magical recourses. The worship of YHWH was aniconic (without images); even such figures as might serve in his worship were banned—apparently owing to the theurgic overtones (the implication that through them men may influence or control the god by fixing his presence in a particular place and making him accessible). Though a mythological background lies behind some cultic terminology (*e.g.*, "a pleasing odor to YHWH," "my bread"), sacrifice is rationalized as tribute or (in priestly writings) is regarded purely as a sacrament; *i.e.*, as a material means of relating to God. Hebrew festivals also have no mythological basis; they either celebrate God's bounty (*e.g.*, at the ingathering of the harvest) or his saving acts (*e.g.*, the festival of unleavened bread, which is a memorial of the Exodus).

Social values and concerns

The values of life and limb, labour, and social solidarity are protected in the rules on relations between man and man. The involuntary perpetual slavery of Hebrews is abolished, and a seven-year limit is set on bondage. The humanity of slaves is defended: one who beats his slave to death is liable to death; if he maims a slave he must set the slave free. A murderer is denied asylum and may not ransom himself from death, while for deliberate and severe bodily injuries the lex talionis ("an eye for an eye" principle) is ordained. Harm to property or theft is punished monetarily, never by death.

Moral exhortations call for solidarity with the poor and the helpless, for brotherly assistance to fellows in need. Institutions are created (*e.g.*, the sabbatical, or seventh, fallow year, in which land is not cultivated) to embody them in practice.

Since the goal of the people was the conquest of a land, their religion had warlike features. Organized as an army (called "the hosts of YHWH" in Ex. 12:41), they encamped in a protective square around their palladium—the tent housing the ark in which the stone "tablets of the Covenant" rested. When journeying, the sacred objects were carried and guarded by the Levites (a tribe serving religious functions), whose rivals, the Aaronites, had a monopoly on the priesthood. God, sometimes called "the warrior," marched with the army; in war, part of the booty was delivered to his ministers.

THE PERIOD OF THE CONQUE AND SETTLEMENT OF CANAA

The conquest of Canaan was remembered as a continuation of God's marvels at the Exodus. The Jordan River was split asunder, Jericho's walls fell at Israel's shout; the enemy was seized with divinely inspired terror; the sun stood still in order to enable Israel to exploit its victory. Such stories are not necessarily the work of a later age; they reflect rather the impact of these victories on the actors in the drama, who felt themselves successful by the grace of God.

A complex process of occupation, involving both battles of annihilation and treaty arrangements with the natives, has been simplified in the biblical account of Joshua's wars. Gradually, the unity of the invaders dissolved (most scholars believe that the invading element was only part of the Hebrew settlement in Canaan; other Hebrews, long since settled in Canaan from patriarchal times, then joined the invaders' covenant league). Individual tribes made their way with more or less success against the residue of Canaanite resistance. New enemies, Israel's neighbours to the east and west, appeared, and the period of the judges (leaders, or champions) began.

The religious situation during the time of the Book of Judges

The Book of Judges, the main witness for the period, does not speak with one voice on the religious situation. Its editorial framework describes repeated cycles of apostasy, oppression, appeal to God, and relief through a God-sent champion. The premonarchic troubles (before the kingship of Saul; see below) caused by the weakness of the disunited tribes were thus accounted for by the covenantal sin of apostasy. The individual stories, however, present a different picture. Apostasy does not figure in the exploits of the judges Ehud, Deborah, Jephthah, and Samson; YHWH has no rival, and faith in him is periodically confirmed by the saviours he sends to rescue Israel from their neighbours.

This faith is shared by all the tribes; and it is owing to their common cult that a Levite from Bethlehem could serve first at an Ephraimite and later also at a Danite sanctuary. The religious bond, preserved by the common cult, was enough to enable the tribes to act more or less in concert under the leadership of elders or an inspired champion in time of danger or religious scandal.

To be sure, both written and archaeological testimonies point to the Hebrews' adoption of Canaanite cults—the Baal worship of Gideon's family and neighbours in Ophrah in Judges, chapter 6, is an example. The many cultic figurines (usually female) found in Israelite levels of Palestinian archaeological sites also give colour to the sweeping indictments of the framework of the Book of Judges. But these phenomena belonged to the private, popular religion; the national God, YHWH, remained *one*—Baal sent no prophets to Israel—though YHWH's claim to exclusive worship was obviously not effectual. Nor did his cult conform with later orthodoxy; Micah's idol in Judges, chapter 17, and Gideon's ephod (priestly or religious garment) were considered apostasies by the editor, in accord with the dogma that other than orthodoxy there is only apostasy—heterodoxy (nonconformity) being unrecognized and simply equated with apostasy.

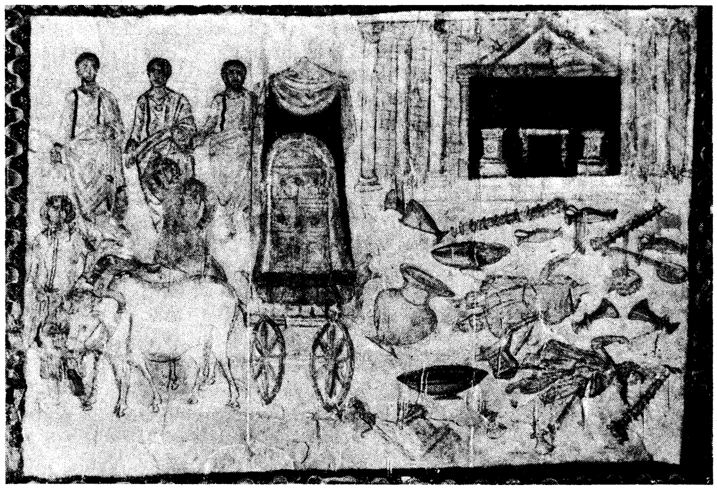
To the earliest sanctuaries and altars honoured as patriarchal foundations—at Shechem, Bethel, Beersheba, and Hebron in Cisjordan (west of the Jordan); at Mahanaim, Penuel, and Mizpah in Transjordan (east of the Jordan)—were now added new ones at Dan, Shiloh, Ramah, Gibeon, and Gilgal, among others. A single priestly family could not operate all these establishments, and Levites rose to the priesthood; at private sanctuaries even non-Levites might be consecrated as priests. The ark of the Covenant was housed in the Shiloh sanctuary, staffed by priests of the house of Eli, who traced their consecration back to Egypt. But the ark remained a portable palladium in wartime; Shiloh was not regarded as its final resting place. The law in Exodus, chapter 20, verses 24–26, authorizing a plurality of altar sites and the simplest forms of construction (earth and rough stone) suited the plain conditions of this period.

THE PERIOD OF THE UNITED MONARCHY

Conflicting views regarding the monarchy

The religio-political problem. The loose, decentralized tribal league could not cope with the constant pressure of external enemies—camel-riding desert marauders who pillaged harvests annually or iron-weaponed Philistines (an Aegean people settling coastal Palestine c. 12th century BCE) who controlled key points in the hill country occupied by Israelites. In the face of such threats to the Israelites, local, sporadic, God-inspired saviours had to be replaced by a continuous central leadership that could mobilize the forces of the entire league and create a standing army. Two attitudes were distilled in the crisis, one conservative and antimonarchic, the other progressive and promonarchic. The conservative appears first in Gideon's refusal, in Judges, chapter 8, verse 23, to found a dynasty: "I will not rule you," he tells the people, "my son will not rule over you; YHWH will . . .!" This theocratic view pervades one of the two contrasting accounts of the founding of the monarchy fused in chapters 8–12 of the First Book of Samuel: the popular demand for a king was viewed as a rejection of the kingship of God, which was embodied in a series of inspired saviours from Moses and Aaron, through Jerubbaal, Bedan, and Jephthah, to Samuel. The other account depicts the monarchy as a gift of God, designed to rescue his people from the Philistines (I Sam. 9:16). Both accounts represent the seer-judge Samuel as the key figure in the founding of Israel's monarchy, and it is not unlikely that the two attitudes struggled in his breast.

By courtesy of the Bollingen Foundation; photograph, Fred Anderegg



The triumph of the ark of the Covenant over paganism, a representation of the antagonism between Judaism and Hellenistic paganism. It was inspired by the biblical story (I Sam. 5:1-5) of the ark, which was captured in battle by the Philistines and which toppled the cult images of Dagon. Mural painting from the synagogue at Doura-Europus, Syria. 3rd century CE.

The Benjaminite Saul was made king (*c.* 1020 BCE) by divine election and by popular acclamation after his victory over the Ammonites (a Transjordanian Semitic people), but his career was clouded by conflict with Samuel, the major representative of the old order. Saul's kingship was bestowed by Samuel and had to be accommodated to the ongoing authority of that man of God. The two accounts of Saul's rejection by God (through Samuel) involve his usurpation of the prophet's authority. King David, whose forcefulness and religio-political genius established the monarchy (*c.* 1000 BCE) on an independent spiritual footing, resolved the conflict.

The Davidic monarchy. The essence of the Davidic innovation was the idea that, in addition to divine election through Samuel and public acclamation, David had God's promise of an eternal dynasty (a conditional, perhaps earlier, and an unconditional, perhaps later, form of this promise exist in Psalms, 132 and II Samuel, chapter 7, respectively). In its developed form, the promise was conceived of as a covenant with David, parallelling the Covenant with Israel and instrumental in the latter's fulfillment; *i.e.*, that God would channel his benefactions to Israel through the chosen dynasty of David. With this new status came the inviolability of the person of God's anointed (a characteristically Davidic idea) and a court rhetoric—adapted from pagan models—in which the king was styled "the [firstborn] son of God." An index of the king's sanctity was his occasional performance of priestly duties. Yet the king's mortality was never forgotten—he was never deified; prayers and hymns might be said on his behalf, but they were never addressed to him as a god.

David captured the Jebusite stronghold of Jerusalem and made it the seat of a national monarchy (Saul had never moved the seat of his government from his home town, the Benjaminite town of Gibeah, about three miles north of Jerusalem). Then, fetching the ark from an obscure retreat, David installed it in his capital, asserting his royal prerogative (and obligation) to build a shrine for the national God—at the same time joining the symbols of the dynastic and the national covenants. This move of political genius linked the God of Israel, the chosen dynasty of David, and the chosen city of Jerusalem in a henceforth indissoluble union.

The chosen dynasty and the chosen city

David planned to erect a temple to house the ark, but the tenacious tradition of the ark's portability in a tent shrine forced postponement of the project to his son Solomon's reign. As part of his extensive building operations, Solomon built the Temple on a Jebusite threshing floor, located on a hill north of the royal city, which David had purchased to mark the spot where a plague had been halted. The ground plan of the Temple—a porch with two free-standing pillars before it, a sanctuary, and an inner sanctum—followed Syrian and Phoenician sanctuary models. A bronze "sea" resting on bulls and placed in the Temple court has a Babylonian analogue. Exteriorly, the Jerusalem Temple resembled Canaanite and other Near Eastern religious structures, but there were differences; *e.g.*, no god image stood in the inner sanctum, but rather only the ancient ark and the new large cherubim (hybrid creatures with animal bodies, human or animal faces, and wings) whose wings covered it, symbolizing the presence of YHWH who was enthroned upon celestial cherubim.

Alongside a brief, ancient inaugural poem in I Kings, chapter 8, verses 12–13, an extensive (and, in its present form, later) prayer expresses the distinctively biblical view of the temple as a vehicle of God's providing for his people's needs. Since, strangely, no reference to sacrifice is made, not a trace appears of the standard pagan conception of the temple as a vehicle of man's providing for the gods.

That literature flourished under the aegis of the court is to be gathered from the quality of the preserved narrative of the reign of David, which gives every indication of having come from the hand of a contemporary eyewitness. The royally sponsored Temple must have had a library and a school attached to it (in accord with the universally attested practice of the ancient Near East), among whose products were not only royal psalms but also such liturgical pieces intended for the common man as eventually found their way into the book of Psalms.

The latest historical allusions in the Torah literature (the Pentateuch) are to the period of the united monarchy; *e.g.*, the defeat and subjugation of the peoples of Amalek, Moab, and Edom by Saul and David, in Numbers, chapter 24, verses 17–20. On the other hand, the polity reflected in the laws is tribal and decentralized, with no bureaucracy. Its economy is agricultural and pastoral, class distinctions apart from slave and free are lacking, and commerce and urban life are rudimentary. A premonarchic background is evident, with only rare explicit reflections of the later monarchy; *e.g.*, in Deuteronomy, chapter 17, verses 14–20. The groundwork of the Torah literature may thus be supposed to have crystallized under the united monarchy.

The emergence of Torah and wisdom literature

It was in this period that the traditional wisdom cultivated among the learned in neighbouring cultures came to be prized in Israel. Solomon is represented as the author of an extensive literature comparable to that of other Eastern sages. His wisdom is expressly attributed to YHWH in the account of his night oracle at Gibeon (in which he asked not for power or riches but for wisdom), thus marking the adaptation to biblical thought of this common Near Eastern genre. As set forth in Proverbs, chapter 2, verse 5, "It is YHWH who grants wisdom; knowledge and understanding are by his command." Patronage of wisdom literature is ascribed to the later Judahite king, Hezekiah, and the connection of wisdom with kings is common in extrabiblical cultures as well.

Domination of all of Palestine entailed the absorption of "the rest of the Amorites"—the pre-Israelite population that lived chiefly in the valleys and on the coast. Their impact on Israelite religion is unknown, though some scholars contend that a "royally sponsored syncretism" arose with the aim of fusing the two populations. That popular religion did not meet the standards of the biblical writers and that it incorporated pagan elements—and that such elements may have increased as a result of intercourse with the newly absorbed "Amorites"—is likely and required no royal sponsorship. On the other hand, the court itself welcomed foreigners—such as the Philistines, Cretans, Hittites, and an Ishmaelite, among others—and made use of their service. Their effect on the court religion may be surmised from what is recorded concerning Solomon's many diplomatic marriages: foreign princesses whom Solomon married brought along with them the apparatus of their native cults, and the King had shrines to their gods built and maintained on the Mount of Olives. Such private cults, while indeed royally sponsored, did not make the religon of the people syncretistic.

Such compromise with the pagan world, entailed by the widening horizons of the monarchy, violated the sanctity of the holy land of YHWH and turned the king into an idolator in the eyes of zealots. Religious opposition, combined with grievances against the organization of forced labour for state projects, led to the secession of the northern tribes (headed by the Joseph tribes) after Solomon's death.

THE PERIOD OF THE DIVIDED KINGDOM

Religion in the northern kingdom (Israel)

Jeroboam I (10th century BCE), the first king of the north (now called Israel, in contradistinction to Judah, the southern Davidic kingdom), appreciated the inextricable link of Jerusalem and its sanctuary with the Davidic claim to divine election to kingship over all Israel (the whole people, north and south). He therefore founded rival sanctuaries at Dan and at Bethel—ancient cult sites—and manned them with non-Levite priests whose symbol of YHWH's presence was a golden calf—a pedestal of divine images in ancient iconography and the equivalent of the cherubim of Jerusalem's Temple. He also moved the autumn ingathering festival one month ahead so as to foreclose celebration of this most popular of all festivals in common with Judah.

For the evaluation of Jeroboam's innovations and the subsequent official religion of the north down to the mid-8th century, one must rely almost exclusively on the book of Kings (later divided into two books). This work has severe limitations as a source for religious history. The material of this book, in good part contemporary, is subjugated to a dogmatic historiography that regards the whole enterprise of the north as one long apostasy ending in a deserved disaster. The culmination of Kings' history with the exile of Judah shows its provenience to have been Judahite. Yet the evaluation of Judah's official religion is subject to an equally dogmatic standard, namely, the royal adherence to the Deuteronomic rule of a single cult site. The author considered the Solomonic Temple to be the cult site chosen by God, according to Deuteronomy, chapter 12, the existence of which rendered all other sites illegitimate. Every king of Judah is judged according to whether or not he did away with all extra-Jerusalemite places of worship. (The date of this criterion may be inferred from the indifference toward it of all persons [*e.g.*, the 9th-century-BCE prophets Elijah and Elisha and the Jerusalemite priest Jehoiada] prior to the late-8th-century-BCE Judahite king Hezekiah.) Another serious limitation is the restriction of Kings' purview: excepting the Elijah–Elisha stories, it notices only the royally sponsored cult; notices of the popular religion are very few. From the mid-8th century the writings of the classical prophets, starting with Amos, set in. These take in the people as a whole, in contrast to Kings; on the other hand their interest in theodicy (justification of God) and their polemical tendency to exaggerate and generalize what they deem evil must be taken into consideration before approving their statements as sober history.

Religion in the southern kingdom (Judah)

For a half-century after the north's secession (c. 922 BCE), the religious situation in Jerusalem was unchanged. The distaff side of the royal household perpetuated, and even augmented, the pagan cults. King Asa (reigned *c.* 908–867 BCE) is credited with a general purge, including the destruction of an image made for the goddess Asherah by the queen mother, granddaughter of an Aramaean princess. He also purged the *qedeshim* ("consecrated men"—conventionally rendered as "sodomites," or "male sacred prostitutes").

Foreign cults entered the north with the marriage of the 9th-century-BCE king Ahab to the Tyrian princess Jezebel. Jezebel brought with her a large entourage of sacred personnel to staff the temple of Baal and Asherah that Ahab built for her in Samaria, the capital of the northern kingdom of Israel. In all else, Ahab's orthodoxy was irreproachable, though others of his court may have joined the worship of the foreign princess. That fierce opposition to the non-YHWH cults sprang up must he supposed in order to account for Jezebel's persecution of the prophets of YHWH, conduct untypical of a polytheist except in self-defense. Elijah's assertion that the whole country apostatized is a hyperbole based on the view that whoever did not actively fight Jezebel was implicated in her polluted cult. Such must have been the view of the prophets, whose fallen were the first martyrs to die for the glory of God. The quality of their opposition may be gauged by Elijah's summary execution of the foreign Baal cultists after they failed the test at Mt. Carmel, where they vied against him in a contest over whose god was truly God. A three-year drought (attested also in Phoenician sources), declared by Elijah to be punishment for the sin, must have done much to kindle the prophets' zeal.

To judge from the Elisha stories, the Baal worship in the capital city, Samaria, was not felt in the countryside. There the religious tone was set by the popular prophets and the prophetic companies ("the sons of the prophets") who attached themselves to them. In popular consciousness these men were wonder-workers—healing the sick and reviving the dead, foretelling the future, and helping to find lost objects. To the biblical narrator they witness the working of God in Israel. Elijah's rage at the Israelite king Ahaziah's recourse to the pagan god Baalzebub, Elisha's cure of the Syrian military leader Naaman's leprosy, and anonymous prophets' directives and predictions in matters of peace and war all serve to glorify God.

The issue of "true" and "false" prophecy

Indeed, the equation of Israel's prosperity with God's interest generated the issue of "true" and "false" prophecy that made its first appearance at this time. That prophecy of success could turn out to be a snare is exemplified in a story of conflict between Micaiah, the lone 9th-century-BCE prophet of doom, and 400 unanimous prophets of victory who lured Ahab to his death. The poignancy of the issue is highlighted by Micaiah's acknowledgment that the 400 were also prophets of YHWH—but inspired by him deliberately with a "lying spirit."

THE PERIOD OF CLASSICAL PROPHECY AND CULT REFORM

The emergence of the literary prophets. By the mid-8th century a hundred years of chronic warfare between Israel and Aram had finally ended—the Aramaeans having suffered heavy blows from the Assyrians. King Jeroboam II (8th century BCE) was able to undertake to restore the imperial sway of the north over its neighbour, and a prophecy of Jonah that he would extend Israel's borders from the Dead Sea to the entrance to Hamath (Syria) was borne out. The well-to-do expressed their relief in lavish attentions to the institutions of worship and their private mansions. But the strain of the prolonged warfare showed in the polarization of society between the wealthy few who had profited from the war and the masses whom it had ravaged and impoverished. Dismay at the dissolution of Israelite society animated a new breed of prophets who now appeared—the literary or classical prophets, first of whom was Amos, an 8th-century-BCE Judahite who went north to Bethel.

Prophetic denouncements of social injustices

That apostasy would set God against the community was an old conception of early prophecy; that violation of the socio-moral injunctions of the Covenant would have the same result was first proclaimed by Amos. Amos almost ignored idolatry, denouncing instead the corruption and callousness of the oligarchy and rulers. The religious exercises of such villains he proclaimed were loathsome to God; on their account Israel would be oppressed from the entrance to Hamath to the Dead Sea and exiled from its land.

The westward push of the Neo-Assyrian Empire in the mid-8th century BCE soon brought Aram and Israel to their knees. In 733–732 Assyria took Gilead and Galilee from Israel and captured Aramaean Damascus; in 721 Samaria, the Israelite capital, fell. The northern kingdom sought to survive through alliances with Assyria and Egypt; its kings came and went in rapid succession. The troubled society's malaise was interpreted by Hosea, a prophet of the northern kingdom (Israel), as a forgetting of God. As a result, in his view, all authority had evaporated: the king was scoffed at, priests became hypocrites, and pleasure seeking became the order of the day. The monarchy was godless; it put its trust in arms, fortifications, and alliances with the great powers. Salvation, however, lay in none of these, but in repentance and reliance upon God.

Prophecy in the southern kingdom. Judah was subjected to such intense pressure to join an Israelite–Aramaean coalition against Assyria that its 8th-century-BCE king Ahaz chose to submit himself to Assyria in return for relief. Ahaz introduced a new Aramaean-style altar in the Jerusalem Temple and adopted other foreign customs that are counted against him in the book of Kings. It was at this time that Isaiah prophesied in Jerusalem. At first (under Uzziah, Ahaz' prosperous grandfather), his message focussed on the corruption of Judah's society and religion, stressing the new prophetic themes of indifference to God (which went hand in hand with a thriving cult) and the fateful importance of social morality. Under Ahaz, the political crisis evoked Isaiah's appeals for trust in God, with the warning that the "hired razor from across the Euphrates" would shave Judah clean as well. Isaiah interpreted the inexorable advance of Assyria as God's chastisement; Assyria was "the rod of God's wrath." But since Assyria ignored its mere instrumentality and exceeded in an insolent manner its proper function, God, when he finished his purgative work, would break Assyria on Judah's mountains. Then the nations of the world, who had been subjugated by Assyria, would recognize the God of Israel as the lord of history. A renewed Israel would prosper under the reign of an ideal Davidic king, all men would flock to Zion (the hill symbolizing Jerusalem) to learn the ways of YHWH and submit to his adjudication, and universal peace would prevail (see also MESSIAH AND MESSIANIC MOVEMENTS).

The prophecy of Micah (8th century BCE), also a Judahite, was contemporary with that of Isaiah, and touched on similar themes (*e.g.*, the vision of universal peace is found in both their books). Unlike Isaiah however, who believed in the inviolability of Jerusalem, Micah shocked his audience with the announcement that the wickedness of its rulers would cause Zion to become a plowed field, Jerusalem a heap of ruins, and the Temple mount a wooded height. Moreover, from the precedence of social morality over the cult, Micah drew the extreme conclusion that the cult had no ultimate value and that God's requirement of men can be summed up as "to do justice, and to love kindness, and to walk humbly with your God."

The reform of King Hezekiah

Reforms in the southern kingdom. According to Jeremiah (about 100 years later), Micah's prophetic threat to Jerusalem had caused King Hezekiah (reigned c. 715–c. 686 BCE) to placate God—possibly an allusion to the cult reform instituted by the King in order to cleanse Judah from various pagan practices. A heightened concern over assimilatory trends resulted in his also outlawing certain practices considered legitimate up to his time. Thus, in addition to removing the bronze serpent that had been ascribed to Moses (and that had become a fetish), the reform did away with the local altars and stone pillars, the venerable (patriarchal) antiquity of which did not save them from the taint of imitation of Canaanite practice. Hezekiah's reform, part of a restorational policy that had political, as well as religious, implications, appears as the most significant effect of the fall of the northern kingdom on official religion. The outlook of the reformers is suggested by the catalog in II Kings, chapter 17, of religious offenses that had caused the fall, the affinity to which the objects of Hezekiah's purge is manifest. Hezekiah's reform is the first historical evidence for Deuteronomy's doctrine of cult centralization. Similarities between Deuteronomy and the Book of Hosea lend colour to the supposition that the reform movement in Judah, which culminated a century later under King Josiah, was sparked by attitudes inherited from the north.

Hezekiah was the leading figure in a western coalition of states that coordinated a rebellion against the Assyrian king Sennacherib with the Babylonian rebel Merodach-Baladan, shortly after the Assyrian's accession in 705 BCE. When Sennacherib appeared in the west in 701 the rebellion collapsed; Egypt sent a force to aid the rebels, but it was defeated. Hezekiah saw his kingdom overwhelmed and offered tribute to Sennacherib; the Assyrian, however, pressed for the surrender of Jerusalem. In despair, Hezekiah turned to the prophet Isaiah for an oracle. Though the prophet condemned the King's reliance upon Egyptian help, he stood firm in his faith that Jerusalem's destiny precluded its fall into heathen hands. The King held fast, and Sennacherib, for reasons still obscure, suddenly retired from Judah and returned home. This unlooked-for deliverance of the city may have been regarded as a vindication of the prophet's faith and was doubtless an inspiration to the rebels against Babylonia a century later. For the present, while Jerusalem was intact, the country had been devastated and its kingdom turned into a vassal state of Assyria.

Foreign influences and the Deuteronomic reform

During Manasseh's peaceful reign of 55 years in the 7th century BCE, Judah was a submissive ally of Assyria. Manasseh's forces served in the building and military operations of the Assyrian kings Esarhaddon and Ashurbanipal. Judah benefitted from the upsurge of commerce that resulted from the political unification of the whole Near East. The prophet Zephaniah attests to heavy foreign influence on the mores of Jerusalem—merchants who adopted foreign dress, cynics who lost faith in the efficacy of YHWH to do anything, people who worshipped the pagan host of heaven on their roofs. Manasseh's court was the centre of such influences. The royal sanctuary became the home of a congeries of foreign gods—the sun, astral deities, and Asherah (the female fertility deity) all had their cults there alongside YHWH. The countryside also was provided with pagan altars and priests, alongside the local YHWH altars that were revived. Presumably, at least some of the blood that Manasseh is said to have spilled freely in Jerusalem must have belonged to YHWH's devotees. No prophecy is dated to his long reign.

With Ashurbanipal's death in 627, Assyria's power faded quickly; the young Judahite king Josiah (reigned c. 640–609 BCE) had already set in motion a vigorous movement of independence and restoration, a cardinal aspect of which was religious. First came the purge of foreign cults in Jerusalem, under the aegis of the high priest Hilkiah; then the countryside was cleansed. In the course of renovating the Temple, a scroll of Moses' Torah (by scholarly consensus an edition of Deuteronomy) was found. Anxious to abide by its injunctions, Josiah had the local YHWH altars polluted to render them unusable and collected their priests in Jerusalem. The celebration of the Passover that year was concentrated in the Temple, as it had not been "since the days of the judges who judged Israel," according to II Kings 23:22, or since the days of Samuel, according to II Chron. 35:18; both references reflect the unhistorical theory of the Deuteronomic (Josianic) reformers that the Shiloh sanctuary was the precursor of the Jerusalem Temple as the sole legitimate site of worship in Israel (as demanded by Deuteronomy, chapter 12). To seal the reform, the King convoked a representative assembly and had them enter into a covenant with God over the newfound Torah. For the first time, the power of the state was enlisted on behalf of the ancient covenant and in obedience to a covenant document. It was a major step toward the fixation of a sacred canon.

Josiah envisaged the restoration of Davidic authority over the entire domain of ancient Israel, and the retreat of Assyria facilitated his program—until he became fatally embroiled in the struggle of the powers over the dying empire. His death in 609 was doubtless a setback for his religious policy as well as his political aspirations. To be sure, the royally sponsored syncretism of Manasseh's time was not revived, but there is evidence of recrudescence of unofficial local altars. Whether references in Jeremiah and Ezekiel to child sacrifice to YHWH reflect post-Josianic practices is uncertain. There is stronger indication of private recourse to pagan cults in the worsening political situation.

That Assyria's fall should have been followed by the yoke of a harsh new heathen power dismayed the devotees of YHWH who had not been prepared for it by prophecy. Their mood finds expression in the oracles of the prophet Habakkuk in the last years of the 7th century BCE. Confessing perplexity at God's toleration of the success of the wicked in subjugating the righteous, the prophet affirms his faith in the coming salvation of YHWH, tarry though it might. And in the meantime, "the righteous must live in his faith."

Prophecies of Jeremiah and Ezekiel

But the situation in fact grew worse as Judah was caught in the Babylonian–Egyptian rivalry. Some attributed the deterioration to the burden of Manasseh's sin that still rested on the people. For the prophet Jeremiah (active c. 626–c. 580 BCE), the Josianic era was only an interlude in Israel's career of guilt that went back to its origins. His pre-reform prophecies denounced Israel as a faithless wife and warned of imminent retribution at the hands of a nameless northerner. After Nebuchadrezzar's decisive defeat of Egypt at Carchemish (605 BCE), Jeremiah identified the scourge as Babylon. King Jehoiakim's attempt to be free of Babylonia ended with the exile of his successor, Jehoiachin, along with Judah's elite (597); yet the court of the new king, Zedekiah, persisted in plotting new revolts, relying—against all experience—on Egyptian support. Jeremiah now proclaimed a scandalous doctrine of the duty of all nations, Judah included, to submit to the divinely appointed world ruler, the Babylonian monarch Nebuchadrezzar. In submission lay the only hope of avoiding destruction; a term of 70 years had been set to humiliate all men beneath Babylon. Imprisoned for demoralizing the populace, Jeremiah persisted in what was viewed as his traitorous message; the leaders, on their part, persisted in their policy, confident of Egypt and the saving power of Jerusalem's Temple, to the bitter end.

Jeremiah also had a message of comfort for his hearers. He foresaw the restoration of the entire people—north and south—in the land, under a new David. And since events had shown that man was incapable of achieving a lasting reconciliation with God on his own, he envisioned the penitent of the future being met halfway by God, who would remake their nature so that to do his will would come naturally to them. God's new covenant with Israel would be written on their hearts, so that they should no longer need to teach each other obedience, for young and old would know YHWH.

Among the exiles in Babylonia, the prophet Ezekiel, Jeremiah's contemporary, was haunted by the burden of Israel's sin. He saw the defiled Temple of Manasseh's time as present before his eyes, and described God as abandoning it and Jerusalem to their fates. Though Jeremiah offered hope through submission, Ezekiel prophesied an inexorable, total destruction as the condition of reconciliation with God. The majesty of God was too grossly offended for any lesser satisfaction. The glory of God demanded Israel's ruin, but the same cause required its restoration also. For Israel's fall disgraced YHWH among the nations; to save his reputation he must therefore restore Israel to its land and make it prosper as never before. The dried bones of Israel must revive, that they and all the nations should know that he was YHWH (Ezek. 37). Ezekiel, too, foresaw the remaking of human nature, but as a necessity of God's glorification; the concatenation of Israel's sin, exile, and consequent defamation of God's name must never be repeated. In 586 BCE the doom prophecies of Jeremiah and Ezekiel came true. Rebellious Jerusalem was reduced by Nebuchadrezzar, the Temple was burnt, and much of Judah's population dispersed or deported to Babylonia.

THE EXILIC PERIOD

Postexilic religious forms

The survival of the religious community of exiles in Babylonia demonstrates how rooted and widespread the religion of YHWH was. Abandonment of the national religion as an outcome of the disaster is recorded of a minority only. There were some cries of despair, but the persistence of prophecy among the exiles shows that their religious vitality had not flagged. The Babylonian Jewish community, in which the cream of Judah lived, had no sanctuary or altar (in contrast to the Jewish garrison of Elephantine in Egypt); what developed in their place can be surmised from new postexilic religious forms: fixed prayer: public fasts and confessions; and assembly for the study of the Torah, which may have developed from visits to the prophets for oracular edification. The absence of a local or territorial focus must also have spurred the formation of a literary–ideational centre of communal life—the sacred canon of Covenant documents that came to be the core of the present Pentateuch. Observance of the sabbath—a peculiarly public feature of communal life—achieved a significance among the exiles virtually equivalent to all the rest of the Covenant rules together. Notwithstanding its political impotence, the spirit of the exiles was so high that foreigners were attracted to their ranks, hopeful of sharing their future glory.

Assurance of that future glory was given not only in the consolations promised by Jeremiah and Ezekiel (the fulfillment of whose prophecies of doom lent credit to their consolations); the great comforter of the exile was the writer or writers of what is known as Deutero-Isaiah (Isa. 40–66), who perceived in the rise and progress (from *c.* 550) of the Persian king Cyrus II the Great the instrument of God's salvation. Going beyond the national hopes of Ezekiel, animated by the universal spirit of the pre-exilic Isaiah, Deutero-Isaiah saw in the miraculous restoration of Israel a means of converting the whole world to faith in Israel's God. Israel would thus serve as "a light for the nations, that YHWH's salvation may reach to the end of the earth." In his conception of the vicarious suffering of God's servant—through which atonement is made for the ignorant heathen—Deutero-Isaiah found a handle by which to grasp the enigma of faithful Israel's lowly state among the Gentiles. The idea was destined to play a decisive role in the self-understanding of the Jewish martyrs of the Syrian king Antiochus IV Epiphanes' persecution in the 2nd century BCE (in, for example, Daniel) and later again in the Christian appreciation of the death of Jesus.

THE PERIOD OF THE RESTORATION

After conquering Babylon, Cyrus so far justified the hopes put in him that he allowed those Jews who wished to do so to return and rebuild their Temple. Though, in time, some 40,000 made their way back, they were soon disillusioned by the failure of the glories of the restoration to materialize and by the controversy with the Samaritans, and left off building the Temple. (The Samaritans were a judaized mixture of native north Israelites and Gentile deportees settled by the Assyrians in the erstwhile northern kingdom.) A new religious inspiration came under the governorship of Zerubbabel, a member of the Davidic line, who became the centre of messianic expectations during the anarchy attendant upon the accession to the Persian throne of Darius I (522). The prophets Haggai and Zechariah perceived the disturbances as heralds of an imminent overthrow of the heathen Persian Empire and a worldwide manifestation of God and glorification of Zerubbabel. Against that day they urged the people quickly to complete the building of the Temple. The labour was resumed and completed in 516; but the prophecies remained unfulfilled. Zerubbabel disappears from the biblical narrative, and the spirit of the community flagged again.

Role of repentance and the Torah

The one religious constant in the vicissitudes of the restored community was the mood of repentance and the desire to win back God's favour by adherence to his Covenant rules. The anxiety that underlay this mood produced a hostility to strangers, which encouraged a lasting conflict with the Samaritans, who asked permission to take part in rebuilding the Temple of the God they too worshipped. The Jews, however, rejected them on ill-specified grounds—apparently ethno-religious; *i.e.*, they felt the Samaritans to be alien to their historical community of faith, especially to its messianic hopes. Nonetheless, intermarriage occurred and precipitated a new crisis when, in 458, the priest Ezra arrived from Babylon, intent on enforcing the regimen of the Torah. By construing ancient and obsolete laws excluding Canaanites and others so as to make them apply to their own times and neighbours, the leaders of the Jews brought about the divorce and expulsion of several dozen non-Jewish wives and their children. Tension between the xenophobic (fear of strangers) and xenophilic (love of strangers) in postexilic Judaism was finally resolved some two centuries later with the development of a formality of religious conversion, whereby Gentiles who so wished could be taken into the Jewish community by a single, simple procedure.

The decisive constitutional event of the new community was the covenant subscribed to by its leaders in 444, making the Torah the law of the land: a charter granted by the Persian king Artaxerxes I to Ezra—scholar and priest of the Babylonian Exile—empowered him to enforce the Torah as the imperial law for the Jews of the province Avar-nahra (Beyond the River), in which the district of Judah (now reduced to a small area) was located. The charter required the publication of the Torah and the publication, in turn, entailed its final editing—now plausibly ascribed to Ezra and his circle. Survival in the Torah of patent inconsistencies and disaccords with the postexilic situation indicate that its materials were by then sacrosanct, to be compiled but no longer created. But these survivals made necessary the immediate invention of a harmonizing and creative method of text interpretation to adjust the Torah to the needs of the times. The Levites were trained in the art of interpreting the text to the people; the first product of the creative exegesis later known as Midrash is to be found in the covenant document of Nehemiah, chapter 9—every item of which shows development, not reproduction, of a ruling of the Torah. Thus, with the publication of the Torah as the law of the Jews the basis of the vast edifice of the Oral Law characteristic of Judaism was laid.

Origins of the Oral Law

Concern over observance of the Torah was fed by the gap between messianic expectations and the gray reality of the restoration. The gap signified God's continued displeasure, and the only way to regain his favour was to do his will. Thus it is that Malachi, the last of the prophets, concludes with an admonition to be mindful of the Torah of Moses. God's displeasure, how-ever, had always been signalized by a break in communication with him. As time passed and messianic hopes remained unfulfilled, the sense of a permanent suspension of normal relations with God took hold, and prophecy died out. God, it was believed, would some day be reconciled with his people, and a glorious revival of prophecy would then occur. For the present, however, religious vitality expressed itself in dedication to the development of institutions that would make the Torah effective in life. The course of this development is hidden from view by the dearth of sources from the Persian period. But the community that emerged into the light of history in Hellenistic times is one made over radically by this momentous, quiet process.

(Mo.Gr.)

III. Hellenistic Judaism (4th century BCE–2nd century CE)

THE GREEK PERIOD (332–63 BCE)

Hellenism and Judaism. Actual contact between Greeks and Semites goes back to Minoan and Mycenaean times and is reflected in certain terms in Homer and in other early Greek authors. It is not until the end of the 4th century, however, that Jews are first mentioned by Greek writers, who praise the Jews as brave. self-disciplined, and philosophical.

After being conquered by Alexander the Great (332 BCE), Palestine became part of the Hellenistic kingdom of Ptolemaic Egypt, the policy of which was to permit the Jews considerable cultural and religious freedom.

Hellenizing under Seleucid rule

When in 198 BCE Palestine was conquered by King Antiochus III (247–187 BCE), of the Syrian Seleucid dynasty, the Jews were treated even more liberally, being granted a charter to govern themselves by their own constitution, namely, the Torah. Greek influence, however, was already becoming manifest. Some of the 29 Greek cities of Palestine attained a high level of culture. The mid-3rd century-BCE Zenon papyri—containing the correspondence of a business manager of a high Ptolemaic official—present the picture of a wealthy Jew, Tobiah, who through commercial contact with the Ptolemies acquired a veneer of Hellenism, to judge at least from the pagan and religious expressions in his Greek letters. His son and especially his grandsons became ardent Hellenists. It has been argued that the Hellenic influence was so strong among the Jews of Judaea by the beginning of the 2nd century that if the process had continued without the forcible intervention of the Seleucids in Jewish affairs (see below) Judaean Judaism would have become even more syncretistic than that of Philo, the Hellenistic Jewish philosopher of Alexandria (*c.* 15 BCE–c. 40 CE). The apocryphal writer Jesus ben Sirach so bitterly denounced the Hellenizers in Jerusalem (c. 180 BCE) that he was forced by the authorities to temper his words.

The Maccabean revolt

In the early part of the 2nd century BCE, 'Hellenizing Jews came into control of the high priesthood itself. Jason as high priest (175–172 BCE) established Jerusalem as a Greek city, Antioch-at-Jerusalem, with Greek educational institutions. His ouster by an even more extreme Hellenizing faction, which established Menelaus (died 162 BCE) as high priest, occasioned a civil war, with the wealthy aristocrats supporting Menelaus and the masses Jason. The Syrian king Antiochus IV Epiphanes, who had initially bestowed exemptions and privileges upon the Jews, intervened upon the request of Menelaus' party. Antiochus' promulgation of decrees against the practice of Judaism and the offensive and cruel measures to enforce them led to the revolt of an old priest, Mattathias, and his five sons—the so-called Maccabees or Hasmoneans. It has been conjectured that one of the Dead Sea Scrolls, the War of *the* Sons of *Light* Against the Sons of Darkness, mirrors the fierceness of this struggle. In any case, the figure of the martyr, as known in Judaism and Christianity—the person who bears witness to the faith through his suffering and death—dates from this event.

The tactics employed both in the countryside and in Jerusalem by the Hasmoneans in their counterattack against Hellenizing Jews, whose children they forcibly circumcised, indicate the inroads that Hellenism had already made. On the whole, however, the chief strength of the Hellenizers lay among the wealthy urban population, while the Maccabees derived their strength from the peasants and urban masses. Yet, there is evidence that the ruthlessness exhibited by the Hasmoneans toward the Greek cities of Palestine had political rather than cultural origins, and that, in fact, they were fighting for personal power no less than for the Torah. In any case, some of those who fought on the side of the Maccabees were idol-worshipping Jews. The Maccabees soon found a modus vivendi with Hellenism: Jonathan (160–142), according to the Jewish historian Josephus (c. 38–c. 100 CE), negotiated a treaty of friendship with Sparta; Aristobulus (104–103 BCE) actually called himself Philhellene (a lover of Hellenism); Alexander Jannaeus (103–76) hired Greek mercenaries and inscribed his coins with Greek as well as with Hebrew. The Greek influence reached its height under King Herod I of Judaea (37–4 BCE), who built a Greek theatre, amphitheatre, and hippodrome in or near Jerusalem.

Social, political, and religious divisions. During the Hellenistic period the priests were both the wealthiest class and the strongest political group among the Jews of Jerusalem. The wealthiest of all were the Oniad family, who held the hereditary office of high priest until they were replaced by the Hasmoneans; the Temple that they supervised was, in effect, a bank, where the Temple wealth was kept and where private individuals also deposited their money. Hence, from a social and economic point of view, Josephus is justified in calling the government of Judaea a theocracy (rule by those having religious authority). Opposition to the priests' oppression arose among an urban middle class group known as scribes (soferim), who were interpreters and instructors of the Torah on the basis of an oral tradition probably going back to the time of the return from the Babylonian Exile (538 BCE and after). A special group of the scribes known as Hasidim (Greek, Hasideans), or "Pietists," became the forerunners of the Pharisees (middle-class liberal Jews who reinterpreted the Torah and the prophetic writings to meet the needs of their times) and joined the Hasmoneans in the struggle against the Hellenists, though on religious rather than on political grounds.

Josephus held that the Pharisees and the other Jewish parties were philosophical schools, and some modern scholars have argued that the groupings were primarily along economic and social lines; but the chief distinctions among them were religious and go back well before the Maccabean revolt. The equation of Pharisaic with "normative" Judaism can no longer be supported, at any rate not before the destruction of the Temple in 70 CE. The fact that in 70 CE, according to the Palestinian Talmud (see below Rabbinic Judaism), there were 24 types of "heretics" in Palestine indicates that there was, in fact, much divergence among Jews; and this picture is confirmed by Josephus, who notes numerous instances of religious leaders who claimed to be prophets and who obtained considerable followings.

The Pharisees and the Sadducees

Some other modern scholars have sought to interpret the Pharisees' opposition to the Sadducees—wealthy, conservative Jews who accepted the Torah alone as authoritative—as based on an urban–rural dichotomy; but a very large share of Pharisaic concern was with agricultural matters. To associate the rabbis with urbanization seems a distortion. The chief support for the Pharisees came from the lower classes, whether in the country or in the city.

The chief doctrine of the Pharisees (literally "Separatists") was that the Oral Law had been revealed to Moses at the same time as the Written Law. In their exegesis and interpretation of this oral tradition, particularly under the rabbi Hillel at the end of the 1st century BCE, the Pharisees were liberal, and their regard for the public won them considerable support. That the Maccabean ruler John Hyrcanus I broke with them and that Josephus set their number at merely "more than 6,000" at the time of King Herod indicates that they were less numerous and influential than Josephus would have his readers believe. The Pharisees stressed the importance of performing all the commandments, including those that appeared to be of only minor significance; those who were particularly strict in their observance of the Levitical rules were known as *ḥaverim* (“companions”). They believed in the providential guidance of the universe, in angels, in reward and punishment in the world to come, and in resurrection of the dead, in all of which beliefs they were opposed by the Sadducees. In finding a modus vivendi with Hellenism, at least in form and in terminology, however, the Pharisees did not differ greatly from the Sadducees. Indeed, the supreme council of the Great Synagogue (or Great Assembly) of the Pharisees was modelled in its organization on Hellenistic religious and social associations. Because they did not take an active role in fostering the rebellion against Rome in 66–70 CE, they were able, through their leader Johanan ben Zakkai, to obtain Roman permission to establish an academy at Jabneh (Jamnia), where, in effect, they replaced the cult of the Temple with study and prayer.

The Sadducees and their subsidiary group, the Boethusians (Boethosaeans), who were identified with the great landowners and priestly families, were more deeply influenced by Hellenization. The rise of the Pharisees may thus be seen, in a sense, as a reaction against the more profound Hellenization favoured by the Sadducees, who were allied with the philhellenic Hasmoneans. From the time of John Hyrcanus (135–104 BCE) the Sadducees generally held a higher position in comparison with the Pharisees and were in favour with the Jewish rulers. Religiously more conservative than the Pharisees, they rejected the idea of a revealed oral interpretation of the Torah, though, to be sure, they had their own tradition, the *sefer gezerot* ("book of decrees" or "decisions"). They similarly rejected the inspiration of the prophetic books of the Bible, as well as the Pharisaic beliefs in angels, rewards, and punishments in the world to come, providential governance of human events, and resurrection of the dead. For them Judaism centred on the Temple; but about ten years before the destruction of the Temple in 70 CE, the Sadducees in effect disappeared from Jewish life when the Pharisees excluded them from entering the Temple.

The "people of the land" and converts

Not constituting any particular party were the unlearned rural masses known as *ʿamme ha-'aretz* ("people of the land"), who were to be found among both the Pharisees and Sadducees and even among the Samaritans, descendants of the northern Israelites who had their own Torah and their own sanctuary. The *ʿamme ha-'aretz* did not give the prescribed tithes, did not observe the laws of purity, and were neglectful of the laws of prayer; and so great was the antagonism between them and the learned Pharisees that to their daughters was applied the biblical verse, "Cursed be he who lies with any kind of beast." The antipathy was reciprocated, for in the same passage in the Babylonian Talmud (*Pesaḥim*) are added the words, "Greater is the hatred wherewith the *ʿamme ha-ʾaretz* hate the scholar than the hatred wherewith the heathens hate Israel." That there was, however, social mobility is clear from the Talmudic dictum, "Heed the sons of the *'am ha-'aretz,* for they will be the living source of the Torah." That there is little evidence that the early church was particularly successful in converting *ʿamme ha-'aretz* suggests that their position was not unbearable.

Proselytes (converts) to Judaism, though not constituting a class, became increasingly numerous both in Palestine and especially in the Diaspora (the Jews living beyond Palestine). Scholarly estimates of the Jewish population of this era range from 700,000 to 5,000,000 in Palestine ana from 2,000,000 to 5,000,000 in the Diaspora, with the prevailing opinion being that about one-tenth of the population of the Mediterranean world at the beginning of the Christian Era was Jewish. Such numbers represent a considerable increase from previous eras and must have included large numbers of proselytes. Already in 139 BCE the Jews of Rome were compelled by the praetor (civil administrator) to go back to their own homes, charged with attempting to contaminate Roman morals with their religion, presumably an allusion to attempts at proselytism. The first large-scale conversions were by John Hyrcanus and Aristobulus, who, in 130 and 103 BCE, respectively, forced the people of Idumaea in southern Palestine and of Ituraea in northern Palestine to become Jews. The eagerness of the Pharisees to win converts is seen in a statement in Matthew that the Pharisees would "traverse sea and land to make a single proselyte." To be sure, some of the proselytes, according to Josephus, did return to their pagan ways, but the majority apparently remained true to their new religion. In addition, there were many "sympathizers" with Judaism who observed one or more Jewish practices without being fully converted.

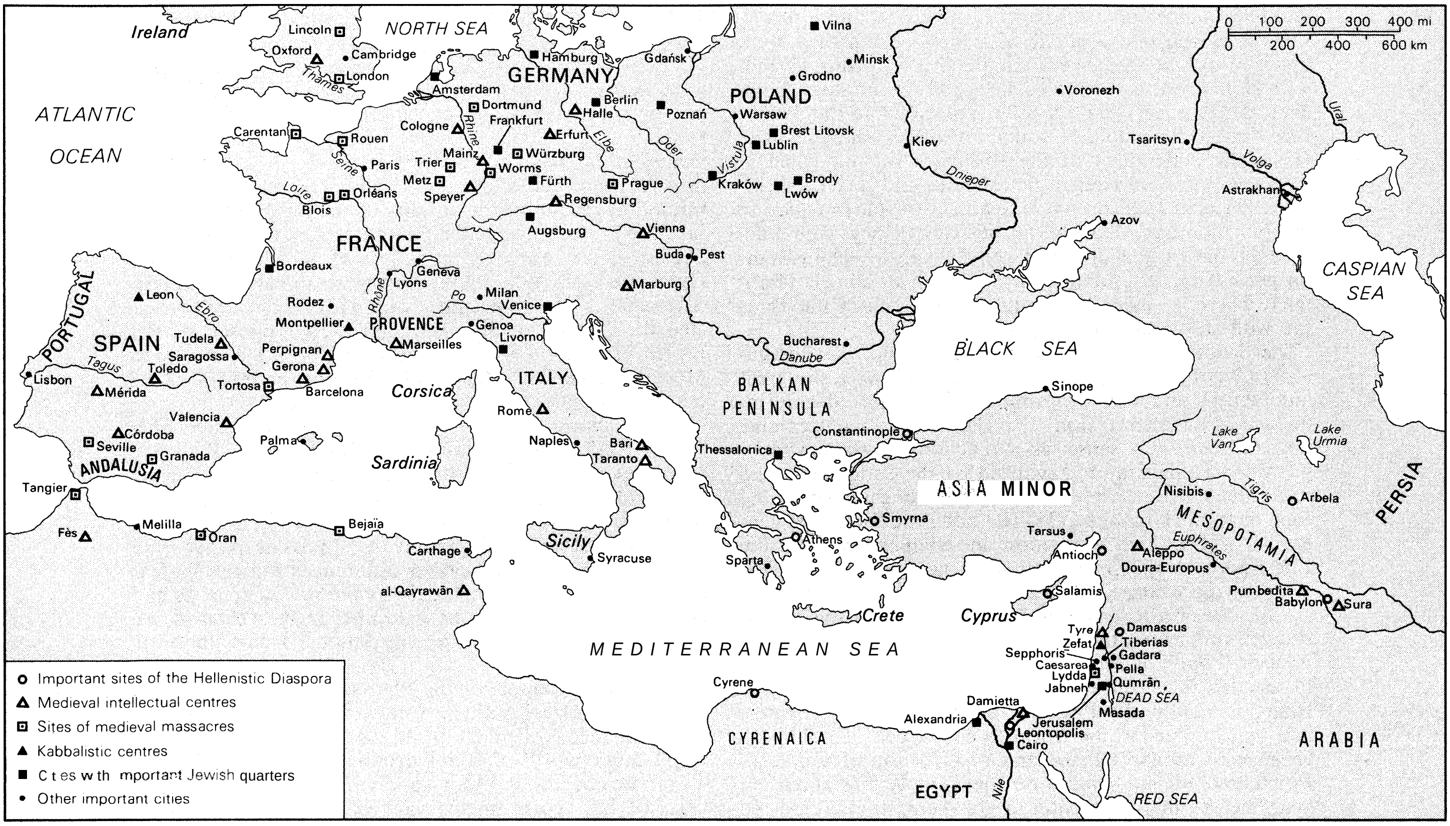
Outside the pale of Judaism in most, though not all, respects were the Samaritans, who, like the Sadducees, refused to recognize the validity of the Oral Law; and, in fact, the break between the Sadducees and the Samaritans did not occur until the conquest of Shechem by John Hyrcanus (128 BCE). Like the later so-called Qumrān covenanters (the monastic group with whom are associated the Dead Sea Scrolls), they were opposed to the Jewish priesthood and the cult of the Temple, regarded Moses as a messianic figure, and forbade the revelation of esoteric doctrines to outsiders.

Scholars have recently revised an older conception of a "normative" Pharisaic Judaism dominant in Palestine and a deviant Judaism dominant in the Diaspora. On the one hand, the picture of "normative" Judaism is broader than at first believed, and it is clear that there were many differences of emphasis within the Pharisaic party; and, on the other hand, supposed differences between Alexandrian and Palestinian Judaism were not as great as had been formerly thought. In Palestine, no less than in the Diaspora, there were then deviations from Pharisaic standards.

Hellenistic influences on Pharisaic Judaism

Despite the attempts of the Pharisaic leaders to restrain the wave of Greek influence, they themselves showed at least a surface Hellenization. In the first place, as many as 2,500–3,000 words of Greek origin are to be found in the Talmudic corpus, and they supply important terms in the fields of law, government, science, religion, technology, and everyday life, especially in the popular sermons preached by the rabbis. When preaching, the Talmudic rabbis often gave the Greek translation of biblical verses for the benefit of those who understood Greek only. The prevalence of Greek in ossuary (burial) inscriptions and the discovery of Greek papyri in the Dead Sea caves confirm the widespread use of the language, though few Jews, it seems, really mastered Greek. Again, there was a surface Hellenization in the frequent adoption of Greek names, even by the rabbis; and there is evidence (Talmud, *Soṭa*) of a school at the beginning of the 2nd century that had 500 students of "Greek wisdom." Even after 117 CE, when it was prohibited by the rabbis to teach one's son Greek, Rabbi Judah the Prince, the editor of the Mishna (authoritative compilation of the Oral Law) at the end of the 2nd century, remarked, "Why talk Syriac in Palestine? Talk either Hebrew or Greek." Even the synagogues of the period have the form of Hellenistic–Roman basilicas, have frequent inscriptions in Greek, and often have pagan motifs. Many of the anecdotes told about the rabbis have Socratic and Cynic (Greek philosophical) parallels. There is evidence of discussions of rabbis with Athenians, Alexandrians, and Roman philosophers, and even with the emperor Antoninus; but in all of these discussions there is evidence of only one rabbi, Elisha ben Abuyah, who became a Gnostic heretic—a person accepting certain esoteric religious dualistic views. The rabbis never mention the Greek philosophers Plato or Aristotle or the Hellenistic Jewish philosopher Philo, and they never use any Greek philosophical terms; the only Greek author whom they name is Homer. Again, the parallels between Hellenistic rhetoric and rabbinic hermeneutics are in the realm of terminology rather than of substance, and those between Roman and Talmudic law are inconclusive. Part of the explanation of this may be that, although there were 29 Greek cities in Palestine, none was in Judaea, the real stronghold of the Jews.

Religious rites and customs in Palestine: Temple and synagogues. The most important religious institution of the Jews until its destruction in 70 was the Temple in Jerusalem—the Second Temple, erected 538–516 BCE. Though services were interrupted for three years by Antiochus Epiphanes (167–165 BCE) and though the Roman general Pompey desecrated the Temple (63 BCE), Herod lavished great expense in rebuilding it. The high priest-



Important historical sites of Hellenistic and medieval Judaism.

hood itself became degraded by the extreme Hellenism of such high priests as Jason and Menelaus; and the institution declined when Herod began the custom of appointing the high priests for political and financial considerations. That not only the multitude of Jews but the priesthood itself suffered from sharp divisions is clear from the bitter class warfare that ultimately erupted in 59 CE between the high priests on the one hand and the ordinary priests and the leaders of the populace of Jerusalem on the other.

Though the Temple remained central in Jewish worship, synagogues may already have emerged during the Babylonian Exile in the 6th century BCE. In any case, in the following century, Ezra stood upon a pulpit of wood and read from the Torah to the people (Nehemiah). According to the interpretation of some scholars, a synagogue existed even within the precincts of the Temple; and certainly by the time of Jesus, to judge from the references to Galilean synagogues in the New Testament, synagogues were common in Palestine. Hence, when the Temple was destroyed in 70, the spiritual vacuum was hardly as great as it had been after the destruction of the First Temple (586 BCE).

The Sanhedrin

The chief legislative, judicial, and educational body of the Palestinian Jews during the period of the Second Temple was the Great Sanhedrin (council court), consisting of 71 members, among whom the Sadducees were an important party. The members shared the government with the king during the early years of the Hasmonean dynasty, but beginning with Herod's reign their authority was restricted to religious matters. In addition, there was another Sanhedrin, set up by the high priest, which served as a court of political council, as well as a kind of grand jury.

Religious and cultural life in the Diaspora. During the Hellenistic–Roman period the chief centres of Jewish population outside Palestine were in Syria, Asia Minor, Babylonia, and Egypt, each of which is estimated to have had at least 1,000,000 Jews. The large Jewish community of Antioch—which, according to Josephus, had been given all the rights of citizenship by the Seleucid founder-king, Seleucus Nicator (died 280 BCE)—attracted a particularly large number of converts to Judaism. It was in Antioch that the apocryphal book of Tobit was probably composed in the 2nd century BCE to encourage wayward Diaspora Jews to return to their Judaism. As for the Jews of Asia Minor, whose large numbers were mentioned by Cicero (1st century BCE), their not joining in the Jewish revolts against the Roman emperors Nero, Trajan, and Hadrian would indicate that they had sunk deep roots into their environment. In Babylonia, in the early part of the 1st century CE, two Jewish brothers, Asinaeus and Anilaeus, were able to establish an independent minor state; their followers were so meticulous in observing the sabbath that they assumed that it would not be possible to violate the sabbath even in order to save themselves from a Parthian attack. In the early part of the 1st century CE, according to Josephus, the royal house and many of their entourage in the district of Adiabene in northern Mesopotamia were converted to Judaism; some of the Adiabenian Jews distinguished themselves in the revolt against Rome in 66 (see below).

The Egyptian Diaspora

The largest and most important Jewish settlement in the Diaspora was in Egypt. There is evidence (papyri) of a Jewish military colony at Elephantine (Yeb), Upper Egypt, as early as the 6th century BCE. These papyri reveal the existence of a Jewish temple—which most certainly would be considered heterodox—and some syncretism (mixture) with pagan cults. Alexandria, the most populous and most influential Hellenistic Jewish community in the Diaspora, had its origin when Alexander the Great assigned a quarter of the city to the Jews. Until about the 3rd century BCE the papyri of the Egyptian Jewish community were written in Aramaic; after that, with the exception of the Nash papyrus in Hebrew, all papyri until 400 CE were in Greek. Similarly, of the 116 Jewish inscriptions from Egypt, all but five are written in Greek. The process of Hellenistic acculturation is, thus, obvious.

The most important work of the early Hellenistic period, dating, according to tradition, from the 3rd century BCE, is the Septuagint, a translation of the Pentateuch into Greek. (The translation of the whole Hebrew Bible was completed during the next two centuries.) The fact that, in the *Letter* of *Aristeas* (see below) and the works of Philo and Josephus, this translation was itself regarded as divinely inspired led to the neglect of the Hebrew original. The translation shows some knowledge of Palestinian exegesis and the tradition of Halakha (the Oral Law); but the rabbis themselves, noting that the translation diverged from the Hebrew text, apparently had ambivalent feelings about it, as is evidenced in their alternate praise and condemnation of it. The fact that such a concept as Torah was translated as *nomos* ("law") and *tzedaqa* as *dikaiosynē* ("justice") opened the way to antilegalism in early Christianity and to Platonic interpretations; and the introduction of such Greek mythological terms as "Titans" and "Sirens" helped to pave the way for the syncretism of Judaism and paganism.

The establishment of a temple at Leontopolis in Egypt (c. 145 BCE) by a deposed high priest, Onias IV, indicates that the temple was clearly heterodox; but this temple never really offered a challenge to the one in Jerusalem and was merely the temple of the military colony of Leontopolis. It is significant that the Palestinian rabbis ruled that a sacrifice intended for the temple of Onias might be offered in Jerusalem. That the temple of Onias made little impact upon Egyptian Jewry can be seen from the silence about it on the part of Philo, who often mentions the Temple in Jerusalem. The temple of Onias, however, continued until it was closed by the Roman emperor Vespasian in 73 CE.

The chief religious institutions of the Egyptian Diaspora were synagogues. As early as the 3rd century BCE there were inscriptions mentioning two *proseuchai*, Jewish prayerhouses. In Alexandria there were numerous synagogues throughout the city, of which the largest was so famous that it is said in the Talmud (see below Rabbinic Judaism) that he who has not seen it has never seen the glory of Israel.

Egyptian Jewish literature. In Egypt the Jews produced a considerable literature (most of it now lost), intended to inculcate in Greek-speaking Jews a pride in their past and to counteract an inferiority complex that some of them felt about Jewish cultural achievements. In the field of history, Demetrius, near the end of the 3rd century BCE, wrote a work *On* the Kings in Judaea—perhaps intended to refute an anti-Semitic Egyptian priest and author—showing considerable concern for chronology. In the 2nd century BCE a Jew who used the name of Hecataeus wrote On the Jews (or On Abraham). Another, Eupolemus (c. 150 BCE), like Demetrius, wrote On *the* Kings in Judaca; an indication of its apologetic nature may be seen from the fragment asserting that Moses taught the alphabet not only to the Jews but also to the Phoenicians and to the Greeks. Artapanus (c. 100 BCE), in his book On the *Jews*, went even further in romanticizing Moses by identifying him with the Greek Musaeus and the Egyptian Hermes-Thoth (god of Egyptian writing and culture) and by asserting that Moses was the real originator of Egyptian civilization and that he even taught the Egyptians the worship of the deity Apis (the sacred bull) and the ibis (sacred bird). In his history, Cleodemus (or Malchus), in an obvious attempt to win for the Jews the regard of the Greeks, asserted that two sons of Abraham had joined Heracles in his expedition in Africa and that the Greek hero had married the daughter of one of them. On the other hand, Jason of Cyrene (c. 100 BCE) wrote a history, of which II Maccabees is a summary, glorifying the Temple and violently attacking the Jewish Hellenizers; but his manner of writing history is typically Hellenistic, with emphasis on pathos. *III* Maccabees (1st century BCE) is a work of propaganda intended to counteract those Jews who sought to win citizenship in Alexandria. The Letter of *Aristeas*, though ascribed to a pagan courtier, Ptolemy II Philadelphus, was probably composed by an Alexandrian Jew about 100 BCE to defend Judaism and its practices against detractors.

Egyptian Jews also composed poems and plays, now extant only in fragments, to glorify their history. Philo the Elder (c. 100 BCE) wrote an epic On *Jerusalem* in Homeric hexameters. Theodotus (c. 100 BCE) wrote an epic On *Shechem*, quite clearly apologetic, to judge from the fragment connecting the name of Shechem with Sikimios, the son of the Greek god Hermes. At about the same time, a Jewish poet wrote a didactic poem, ascribing it to the pagan Phocylides, though closely following the Bible in some details; the author disguised his Jewish origin by omitting any attack against idolatry from his moralizing. A collection known as *The* Sibylline Oracles, containing Jewish and Christian prophecies in pagan disguise, includes some material composed by a 2nd-century-BCE Alexandrian Jew who intended to glorify the pious Jews and perhaps to win converts; it is possible that the Oracles were known to the Roman poet Virgil when he wrote his fourth Eclogue.

A Jewish dramatist of the period, Ezekiel (*c.* 100 BCE), composed tragedies in Greek. Fragments of one of them, The Exodus, show how deeply he was influenced by the Greek dramatist Euripides. Whether such plays were actually presented on the stage or not, they edified Jews and showed the pagans that the Jews had as much material for drama as they did.

Alexandrian Jewish philosophical achievements

The greatest achievement of Alexandrian Judaism was in the realm of wisdom literature and philosophy. In a work entitled An Interpretation of the Law of Moses, Aristobulus in the 2nd century BCE anticipated Philo in attempting to harmonize Greek philosophy and the Torah, in using the method of allegory to explain anthropomorphisms in the Bible, and in asserting that the Greek philosophers were indebted to Moses. The Wisdom of Solomon, dating from the 1st century BCE, shows an acquaintance with the Platonic doctrine of the pre-existence of the soul and with a method of argument known as sorites that was favoured by the Stoics (Greek philosophers). During the same period the author of IV Maccabees showed an intimate knowledge of Greek philosophy, particularly of Stoicism.

By far the greatest figure in Alexandrian Jewish literature is Philo, who has come to be recognized as a major philosopher. His synthesis of Greek philosophy, particularly that of Plato, and of the Torah, and his formulation of the Logos (Word, or Divine Reason) as an intermediary between God and the world, helped lay the groundwork for Neoplatonism (a philosophy dealing with levels of being), Gnosticism (a dualistic religious movement teaching that matter is evil and that spirit is good), and the philosophical framework of the early Church Fathers. Philo was a devotee of Judaism neither as a mystic cult nor as a collateral branch of Pharisaic Judaism; he was a Diaspora Jew with a profound knowledge of Greek literature who, though almost totally ignorant of Hebrew, tried to find a modus vivendi between Judaism and secular culture.

Mention may be made of the Jewish community of Rome. Numbering perhaps 50,000, it was, to judge from the inscriptions in the Jewish catacombs, predominantly Greek-speaking and almost totally ignorant of Hebrew. References in Roman writers, particularly Tacitus and the satirists, have led scholars to conclude that the community—which was influential, to judge from the pagan jibes—observed the sabbath and the dietary laws and was active in seeking converts.

Modifications of Jewish traditions by Diaspora Jews

The Hellenization of the Diaspora Jews is, however, to be seen not merely in their literature but even more in the papyri and art objects that have recently been studied at great length. As early as 290 BCE, Hecataeus of Abdera, a Greek non-Jew living in Egypt, had remarked that under the Persians and Macedonians the Jews had greatly modified the traditions of their fathers. The fact that—to judge from other papyri—at least three-fourths of the Egyptian Jews had personal names of Greek, rather than Hebrew, origin is significant. That the only schools of which mention is made are sabbath schools intended for adults and that, on the contrary, Jews were extremely eager to gain admittance for their children to Greek *gymnasia*—where quite obviously they would have to make compromises with their Judaism—indicates their scale of values. Again, there are a number of violations from the norms of Halakha (which precluded the charging of interest for a loan), most notably in the fact that of 11 known extant loan documents only two are without interest. There are often striking similarities between the documents of sale, marriage, and divorce of the Jews and of the Greeks in Egypt, though some of this, as with the documents of the Elephantine Jewish community, may be due to a common origin in the cuneiform law of ancient Mesopotamia. The charms and apotropaic (designed to avert evil) amulets are often syncretistic, and the Jews can hardly have been unaware of the religious significance of symbols that were still very much filled with meaning in pagan cults. The fact that the Jewish community of Alexandria was preoccupied in the 1st century BCE and the 1st century CE with obtaining rights as citizens—which certainly involved compromises with Judaism, including participation in pagan festivals and sacrifices—shows how far they were ready to deviate. Philo mentions Jews who scoffed at the Bible, which they insisted on interpreting literally, and of others who failed to adhere to the biblical laws that they regarded as mere allegory; he writes too of Jews who observed nothing of Judaism except the holiday of Yom Kippur. But despite such deviations, the pagan writers constantly accuse the Diaspora Jews of being "haters of mankind" and of being absurdly superstitious; and Christian writers later similarly attack the Jews for refusing to give up the Torah. At least they were loyal Jews in their contributions of the Temple tax and in pilgrimages to Jerusalem on the three festivals. Actual apostasy and intermarriage were apparently not common, but the virulent anti-Semitism and the pogroms perpetrated by the Egyptian non-Jews must have served as a deterrent.

Apocalyptic and wisdom literature

Palestinian Literature. During this period literature was composed in Palestine in Hebrew, Aramaic, and Greek, with the exact language still a subject of dispute among scholars in many cases and with the works often apparently composed by more than one author over a considerable period of time. Most of the works composed in Hebrew, many of them existing only in Greek—Ecclesiasticus, I Maccabees, Judith, *Testaments of the Twelve Patriarchs,* Baruch, *Psalms of Solomon*, Prayer of Manasseh—and many of the Dead Sea Scrolls are generally conscious imitations of biblical books, often reflecting the dramatic events of the Maccabean struggle and often with an apocalyptic tinge (involving the dramatic intervention of God in history). The literature in Aramaic consists of the following: (1) biblical or Bible-like legends or midrashic (interpretive) additions—Testament *of Job, Martyrdom of Isaiah, Paralipomena of Jeremiah, Life of Adam and Eve,* the Dead Sea *Genesis Apocryphon*, Tobit, Susanna, Bel and the Dragon; and (2) apocalypses—Enoch (perhaps originally written in Hebrew), *Assumption of Moses,* the Syriac Baruch, II (IV) Esdras, and *Apocalypse of Abraham*. In Greek the chief works by Palestinians are histories of the Jewish War against Rome and of the Jewish kings by Justus of Tiberias (both are lost) and the history of the Jewish War, originally in Aramaic, and the *Jewish Antiquities* by Josephus (both written in Rome).

Of the wisdom literature composed in Hebrew, the book of the Wisdom of Jesus the Son of Sirach, or Ecclesiasticus *(c.* 180 BCE), modelled on the book of Proverbs, identified Wisdom with the observance of the Torah. The *Testaments of the Twelve Patriarchs*, probably written in the latter half of the 2nd century BCE, patterned on Jacob's blessings to his sons, are now thought to belong to eschatological literature related to the Dead Sea Scrolls. The identification of Wisdom and Torah is stressed in the Mishnaic tract *Pirqe Avut* ("Sayings of the Fathers"), which, though edited 200 CE, contains the aphorisms of rabbis dating back to 300 BCE.

Books such as the *Testament of Job,* the Dead Sea Scroll *Genesis Apocryphon,* the *Book of Jubilees* (now known to have been composed in Hebrew, as seen by its appearance among the Dead Sea Scrolls), and *Biblical Antiquities,* falsely attributed to Philo (originally written in Hebrew, then translated into Greek, but now extant only in Latin), as well as the first half of Josephus' *Jewish Antiquities*, often show affinities with rabbinic Midrashim (interpretive works) in their legendary accretions of biblical details. Sometimes, as in *Jubilees* and in the Pseudo-Philo work, these accretions are intended to answer the questions of heretics, but often, particularly in the case of Josephus, they are apologetic in presenting biblical heroes in a guise that would appeal to a Hellenized audience.

Apocalyptic trends, given considerable impetus by the victory of the Maccabees over the Syrian Greeks, were not—as was formerly thought—restricted to Pharisaic circles. They were (as is clear from the Dead Sea Scrolls) found in other groups as well, and are of particular importance for their influence on both Jewish mysticism and early Christianity. These books, which have a close connection with the biblical Book of Daniel, stress the impossibility of a rational solution to the problem of theodicy—how to reconcile the righteousness of God with observable evil. They also stress the imminence of the day of salvation, which is to be preceded by terrible hardships, and presumably reflected the current historical setting. In the book of *Enoch* there is stress on the terrible punishment inflicted upon sinners in the Last Judgment, the imminent coming of the Messiah and of his kingdom, and the role of angels.

The sole Palestinian Jewish author writing in Greek whose works are preserved is Josephus. His account of the war against the Romans in his *Life* and, to a lesser degree, in the *Jewish War* are largely a defense of his own questionable behaviour as the commander of the Jewish forces in Galilee. But these works and more especially *Against Apion* and the *Jewish Antiquities* are largely defenses of Judaism against anti-Semitic attacks. Josephus' *Jewish War* is often quite deliberately parallel to Thucydides' *History of the Peloponnesian War;* and his *Jewish Antiquities* is quite deliberately parallel to Dionysius of Halicarnassus' *Roman Antiquities,* dating from earlier in the same century.

THE ROMAN PERIOD (63 BCE–135 CE)

New parties and sects. Under Roman rule a number of new groups, largely political, emerged in Palestine. Their common aim was to seek an independent Jewish state. All were zealous for, and strict in their observance of, the Torah.

The Herodians and the Zealots

The Herodians were a political group that after the death of Herod—whom they apparently regarded as the Messiah—sought the re-establishment of the rule of Herod's descendants over an independent Palestine as a prerequisite for Jewish preservation. Unlike the Zealots, however (see below), they did not refuse to pay taxes to the Romans.

The Zealots' party, founded c. 6–9 CE, refused to pay tribute to the Romans and advocated overthrowing them on the ground that they should acknowledge God alone as their master. A priestly, eschatologically oriented resistance movement, the Zealots were particularly dedicated to keeping the Temple and its cult pure and used guerrilla tactics toward that end. The Sicarii (Assassins), so-called because of the dagger (*sica*) they carried, arose c. 54, according to Josephus, as a group of bandits who kidnapped or murdered those who had found a modus vivendi with the Romans. It was they who made a stand at the fortress of Masada near the Dead Sea, committing suicide rather than be captured by the Romans (73).

The Essenes, New Covenanters, and Dead Sea groups

A number of other parties—various types of Essenes, Damascus Covenanters, and the Qumrān Dead Sea groups—were distinguished by their pursuit of an ascetic monastic life, disdain for material goods and sensual gratification, sharing of material possessions, concern for eschatology, strong apocalyptic views in anticipation of the coming of the Messiah, practice of ablutions to attain greater sexual and ritual purity, prayer, contemplation, and study. The Essenes were like the Therapeutae, a Jewish religious group that had flourished in Egypt two centuries earlier, but the latter actively sought "wisdom" whereas the former were anti-intellectual. Only some of the Essenes were celibate. The Essenes have been termed Gnosticizing Pharisees because of their belief, shared with the Gnostics, that the world of matter was evil; some have seen in them the influence of a quasi-monasticism.

The Damascus sect (New Covenanters) were a group of Pharisees who went beyond the letter of the Pharisaic Halakha. Like the Essenes and the Dead Sea sect, they had a monastic type of organization and opposed the way in which sacrifices were offered in the Temple.

The continuing recent discoveries of scrolls in caves of the Dead Sea area have focussed attention on the groups that lived there. On the basis of paleography, carbon-14 testing, and the coins discovered there, most scholars accept a 1st-century date for them. A theoretical relationship of the communities with John the Baptist and the nascent Christian groups remains in dispute, however. The sectaries have been identified variously as Zealots, an unnamed anti-Roman group, and especially Essenes; but a major difference between the Qumriin groups and the Essenes is that the former were militarily activist (the discovery of hymns and a calendar at Masada—a stronghold of the Sicarii—that had previously been found at Qumrān, may indicate a connection between the groups), while the latter were, for the most part, pacifist. That the groups had secret, presumably apocalyptic, teachings is clear from the fact that among the scrolls are some in cryptographic script and reversed writing; and yet, despite their extreme piety and legalistic conservatism, they apparently were not unaware of Hellenism, to judge from the presence of Greek books at Qumriin.

It has long been debated whether the Gnostic systems of the 1st and 2nd centuries go back to the collapse of the apocalyptic strains in Judaism—which expected a final transforming catastrophic event—when the Temple was destroyed in 70. It is doubtful that there is any direct Jewish source for this Gnosticism, though some characteristic Gnostic doctrines are found in certain groups of particularly apocalyptic 1st-century Jews—the dichotomy of body and soul and a disdain for the material world, a notion of esoteric knowledge, and an intense interest in angels and in problems of creation.

Primitive Christianity a part of the 1st-century Jewish religious spectrum

Origin of Christianity: the early Christians and the Jewish community. Though it attracted little attention among pagans and Jews at the beginning, the rise of Christianity was by far the most important "sectarian" development of the Roman period. With the revision, largely due to the discoveries at Qumrān, of the view that Pharisaic Judaism was to be considered normative, primitive Christianity, with its apocalyptic and eschatological interests, has come to be viewed by many scholars as no longer "sectarian" or peripheral to Jewish development but, at least initially, as part of a broad spectrum of attitudes within Judaism. Jesus himself, despite his criticisms of Pharisaic legalism, may now be classified as a Pharisee with strong apocalyptic inclinations; he proclaimed that he had no intention of abrogating the Torah, but of fulfilling it. It is possible to envision a direct line between Jewish currents, both in Palestine and the Diaspora in the Hellenistic Age, and Christianity, particularly in the traditions of martyrdom, proselytism, monasticism, mysticism, liturgy, and religious philosophy, especially the doctrine of the Logos (Word) as an intermediary between God and the world and the synthesis of faith and reason. The Septuagint, in particular, played an important role both theoretically, in the transformation of Greek philosophy into the theology of the Church Fathers, and practically, in converting Jews and Jewish "sympathizers" to Christianity. The connection of nascent Christianity with the Qumrān groups may be seen in their dualism and apocalypticism; but there are differences, notably in the conception of the Incarnation, in the relationship of the Son and the Father, and in Jesus' vicarious suffering for sinners as against the direct suffering of the Qumrān Teacher of Righteousness. Again, the Qumran group constituted an esoteric movement, militant, with enforced community of goods, concerned with strict observance of the Torah, especially with its calendar, whereas Christianity was pacifist, was open to all, and represented a New Covenant, with stress away from the Torah ritual and with voluntary community of possessions. In general, moreover, Christianity was more positively disposed toward Hellenism than was Pharisaism, particularly under the leadership of Paul, a thoroughly Hellenized Jew.

When Paul proclaimed his antinomianism (against Torah observance as a means of salvation) many Jewish followers of Jesus became Jewish Christians and continued to observe the Torah. Their two main groupings were the Ebionites—probably to be identified with those called minim, or "sectaries," in the Talmud—who accepted Jesus as the Messiah but denied his divinity, and the Nazarenes, who regarded Jesus as both Messiah and God, but regarded the Torah as binding upon Jews alone.

The break between Judaism and Christianity

The percentage of Jews converted to any form of Christianity was extremely small, as can be seen from the frequent criticisms of Jews for their stubbornness by Christian writers. In the Diaspora, despite the strong influence of Hellenism, there were relatively few Jewish converts, though the Christian movement had some success in winning Alexandrian Jews.

There were four major stages in the final break between Christianity and Judaism: (1) the flight of the Jewish Christians from Jerusalem to Fella across the Jordan in 70 and their refusal to continue the struggle against the Romans; (2) the institution by the patriarch Gamaliel II of a prayer in the Eighteen Benedictions against such heretics (c. 100), and (3 and 4) the failure of the Christians to join the messianic leaders Lukuas-Andreas and Bar Kokhba in the revolts against Trajan (115–117) and Hadrian (132–135), respectively.

Judaism under Roman rule. When Pompey entered the Temple in 63 BCE as an arbiter both in the civil war between Hyrcanus and Aristobulus and in the struggle of the Pharisees against both Jewish rulers, Judaea in effect became a puppet state of the Romans. During the civil war between Pompey and Julius Caesar, the Idumaean Antipater had ingratiated himself with Caesar by aiding him and was rewarded by being made governor of Judaea; the Jews were rewarded through the promulgation of a number of decrees favourable to them, which were reaffirmed by Augustus and later emperors. His son Herod, king of Judaea, an admirer of Greek culture, supported a cult worshipping the Emperor and built temples to Augustus in non-Jewish cities. Since he was by origin an Idumaean, he was regarded by many Jews as a foreigner. (The Idumaeans, or Edomites, were forcibly converted to Judaism by John Hyrcanus; see above.) On several occasions during and after his reign, Pharisaic delegations sought to convince the Romans to end the quasi-independent Jewish government. After the death of Herod's son and successor Archelaus in 6 CE, his realms were ruled by Roman procurators, the most famous or infamous of whom, Pontius Pilate (26–36), attempted to introduce busts of the Roman emperor into Jerusalem and discovered the intense religious zeal of the Jews in opposing this measure. When Caligula ordered the governor of Syria, Petronius, to install a statue of himself in the Temple, a large number of Jews proclaimed they would suffer death rather than to permit such a desecration. Petronius in response succeeded in getting the Emperor to delay. The procurators of Judaea, being of equestrian (knightly) rank and often of Oriental Greek stock, were more anti-Semitic than the governors of Syria, who were of the higher senatorial order. The last procurators in particular were indifferent to Jewish religious sensibilities; and various patriotic groups, to whom nationalism was an integral part of their religion, succeeded in polarizing the Jewish population and bringing on an extremely bloody war with Rome in 66–70. The climax of the war was the destruction of the Temple in 70, though, according to Josephus, the Roman general (and later emperor) Titus sought to spare it. The war was not ended, however, until 73, when the Sicarii at Masada committed suicide rather than submit to the Romans.

Jewish wars against the Romans

The papyri indicate that the war against Trajan (115–117), involving the Jews of Egypt, Cyrenaica, Cyprus, and Mesopotamia (though only to a minor degree those of Palestine), was a widespread revolt under a Cyrenian king-messiah, Lukuas-Andreas, aimed at freeing Palestine from Roman rule. The same spirit of freedom impelled another messiah, Bar Kokhba, who had the support of the greatest rabbi of the time, Akiba, in his spontaneous uprising (132–135). The result was Hadriax's decrees prohibiting circumcision and public instruction in the Torah, though these were soon revoked by Antoninus Pius. Having suffered such tremendous losses on the field of battle, Judaism turned its dynamism to the continued development of the Talmud (see below ***Rabbinic** Judaism*). (*L.H.F.*)

IV. Rabbinic Judaism (2nd–18th centuries)

THE AGE OF THE TANNAIM (135–C. 200)

The role of the rabbis. With the defeat of Bar Kokhba and the ensuing collapse of active Jewish resistance to Roman rule (135–136), politically moderate and quietist rabbinic elements remained the only cohesive group within Jewish society. With Jerusalem off limits to the Jews, rabbinic ideology and practice, which were not dependent on Temple, priesthood, or political independence for their vitality, provided a viable program for autonomous community life and thus filled the vacuum created by the suppression of all other Jewish leadership. The Romans, confident that the will for insurrection had been shattered, soon relaxed the Hadrianic prohibitions of Jewish ordination, public assembly, and regulation of the calendar and permitted rabbis who had fled the country to return and re-establish an academy in the town of Usha in Galilee.

The rabbinic way

The strength of the rabbinate lay in its ability to represent simultaneously the interests of the Jews and the Romans, whose religious and political needs, respectively, now chanced to coincide. The rabbis were regarded favourably by the Romans, as a politically submissive class, which, with its wide influence over the Jewish masses, could translate the Pax Romana (the peace imposed by Roman rule) into Jewish religious precepts. To the Jews, on the other hand, the rabbinic ideology gave the appearance of continuity to Jewish self-rule and freedom from alien interference. The rabbinic program fashioned by Johanan ben Zakkai's circle (see above *Hellenistic Judaism*) had replaced sacrifice and pilgrimage to the Temple with study of Scripture, prayer, and works of piety, thus eliminating the need for a central sanctuary (in Jerusalem) and making of Judaism a religious association capable of fulfillment anywhere. Judaism was now, for all intents and purposes, a Diaspora religion even on its home soil. Any sense of real break with the past was mitigated by continued adherence to purity laws (dietary and bodily) and by assiduous study of Scripture, including those legal sections that historical developments had now made obsolete. The reward held out for scrupulous study and fulfillment was the promise of messianic deliverance; *i.e.*, divine restoration of all those institutions that had become central in Jewish notions of national independence—the Davidic monarchy, Temple service, the ingathering of Diaspora Jewry—and, above all, the assurance of personal reward to the righteous through resurrection and participation in the national rebirth.

The Mishna

Apart from the right to teach Scripture publicly, the most pressing need felt by the surviving rabbis was for the reorganization of a recognized body that would reactivate the functions of the former Sanhedrin and pass on disputed questions of law and dogma. A high court was, accordingly, organized under the leadership of Simeon ben Gamaliel (reigned *c.* 135–c. 175), the son of the previous patriarch (the Roman term for the head of the Palestinian Jewish community) of the house of Hillel, in association with rabbis representing other schools and interests. In the ensuing struggle for power, the patriarch managed to concentrate all communal authority in his office. The dominating role of the patriarchate reached its zenith in the days of his son and successor, Judah the Prince, whose reign (*c.* 175–c. 220) marked the climax of this period of rabbinic activity, otherwise known as the "age of the *tannaim*" (teachers). Armed with wealth, Roman backing, and dynastic legitimacy (which the patriarch now traced to the house of David), Judah sought to standardize Jewish practice through a corpus of legal norms that would reflect recognized views of the rabbinate on every aspect of life. The Mishna (collection of rabbinic law) that soon emerged became the primary source of reference in all rabbinic schools and constituted the core around which the Talmud (commentary on Mishna, literally "teaching") was later compiled. It thus remains the best single introduction to the complex of rabbinic values and practices as they evolved in Roman Palestine.

The making of the Mishna. Although the promulgation of an official corpus represented a break with rabbinic precedent, Judah's Mishna did have antecedents. During the 1st and 2nd centuries CE, rabbinic schools had compiled for their own reference collections in which the results of their exegesis and application of Scripture to problematic situations (Midrash, "investigation" or "interpretation"; plural Midrashim) had been recorded in terse legal form. By 200 CE several such compilations were circulating in Jewish schools and were being utilized by judges. While adhering to the structural form of these earlier collections, Judah compiled a new one in which universally accepted views were recorded alongside those still in dispute, thereby largely reducing the margin for individual discretion in the interpretation of the law. Although his action aroused opposition, and some rabbis continued to invoke their own collections, the authority of his office and the obvious advantages of a unified system of law soon outweighed centrifugal tendencies, and his Mishna attained quasi-canonical status, becoming known as "The Mishna" or "Our Mishna." For all its clarity and comprehensiveness, its phraseology was often obscure or too terse to satisfy all needs, and a companion known as the Tosefta ("Additions") was compiled shortly thereafter in which omitted traditions and explanatory notes were recorded. Since, however, neither compilation elucidated the processes by which their decisions had been elicited, various authorities set about collecting the midrashic discussions of their schools and recording them in the order of the verses of Scripture. During the 3rd and 4th centuries the tannaitic Midrashim on the Pentateuch were compiled and introduced as school texts.

Fundamentally legal in character, this literature was designed to regulate every aspect of life—the six divisions of the Mishna on agriculture, festivals, family life, civil law, sacrificial and dietary laws, and purity encompass virtually every area of Jewish experience—and, accordingly, also recorded the principal Pharisaic and rabbinic definitions and goals of the religious life. One tract of the Mishna, Avot ("Sayings of the Fathers"), treated the meaning and posture of a life according to Torah, while other passages made reference to the mystical studies into which only the most advanced and religiously worthy were initiated; *e.g.*, the activities of the *Merkava*, or divine "Chariot," and the doctrines of creation (see JEWISH MYSTICISM). The rabbinic program of a life dedicated to study and fulfillment of the will of God was thus a graded structure in which the canons of morality and piety were attainable on various levels, from the popular and practical to the esoteric and metaphysical. Innumerable sermons and homilies preserved in the midrashic collections, liturgical compositions for daily and festival services, and mystical tracts circulated among initiates all testify to the deep spirituality that informed rabbinic Judaism.

THE AGE OF THE AMORAIM: THE MAKING OF THE TALMUDS (3RD–6TH CENTURIES)

Palestine (c. 220–c. 400). The promulgation of the Mishna initiated the period of the *amoraim* (lecturers or interpreters), those teachers who made the Mishna the basic text of legal exegesis. The curriculum now centred on the elucidation of the text of the standard compilation, harmonization of its decisions with extra-Mishnaic traditions recorded in other collections, and the application of its principles to new situations. The records of these amoraic studies have been preserved in the form of two running commentaries on the Mishna known as the Palestinian (or Jerusalem) Talmud ("Teaching") and the Babylonian Talmud, reflecting the study and legislation of the academies of the two principal centres of Jewish concentration in the Roman and Persian empires of that time. (Talmud is also the comprehensive term for the whole collections, Palestinian and Babylonian, containing Mishna, commentaries, and other matter.)

The principal agencies mediating the rabbinic way of

The rabbinic educational system

life and literature to the masses were the schools, ranging from the primary school to the advanced "house of study" and more formal academy (yeshiva), the synagogue, and the Jewish courts, which not only adjudicated litigations but also decided on ritual problems. Primary schools had long been available in the villages and cities of Palestine, and tannaitic law made education of male children a religious duty. Introduced at the age of five or six to Scripture, the student advanced at the age of ten to Mishna and finally in mid-adolescence to Talmud or the processes of legal reasoning. Regular reading of Scripture in the synagogue on Mondays, Thursdays, the sabbaths, and festivals, coupled with concurrent translations into the Aramaic vernacular and frequent sermons, provided for lifelong instruction in the literature and the values elicited from it. The amoraic emphasis on the moral and spiritual aims of Scripture and its ritual is reflected in their midrashic collections, which are predominantly homiletical (sermonic) rather than legal in content.

An amoraic sermon conceded that of every thousand beginners in primary school only one would be expected to continue as far as Talmud. In the 4th century, however, there were enough advanced students to warrant academies in Lydda, Caesarea, Sepphoris, and Tiberias (in Palestine), where leading scholars trained disciples for communal service as teachers and judges. In Caesarea, the principal port and seat of the Roman administration of Palestine, where pagans, Christians, and Samaritans maintained renowned cultural institutions, the Jews, too, established an academy that was singularly free of patriarchal control. The outstanding rabbinic scholar there, Abbahu (c. 279–320), wielded great influence with the Roman authorities and, because he combined learning with personal wealth and political power, attracted some of the most gifted students of the day to the city. In c. 350 the studies and decisions of the authorities in Caesarea were compiled as a tract on the civil law of the Mishna. Half a century later, the academy of Tiberias issued a similar collection on other tracts of the Mishna, and this compilation, in conjunction with the Caesarean material, constituted the Palestinian Talmud.

The role of the patriarchs

Despite increasing tensions between some rabbinic circles and the patriarch, his office was the agency providing a basic unity to the Jews of the Roman Empire. Officially recognized as a Roman prefect, a government official, the patriarch at the same time delegated apostles to Jewish communities to inform them of the Jewish calendar and of other decisions of general concern and to collect an annual tax of a half shekel paid by male Jews for his treasury. As titular head of the Jewish community of the mother country and as a vestigial heir of the Davidic monarchy, the patriarch was a reminder of a glorious past and of a hope for a brighter future. How enduring these hopes were may be seen from the efforts to gain permission to rebuild the Temple in Jerusalem. Although the emperor Julian (reigned 361–363) actually authorized the reconstruction, the project came to naught as a consequence of a disastrous fire on the sacred site and the subsequent death of the Emperor.

The adoption of Christianity as the religion of the empire had no direct effect on the religious freedom of the Jews; *i.e.*, on their freedom to worship and observe their life rules. The ever-mounting hostility between the two religions, however, resulted in severe curtailment of Jewish disciplinary rights over their coreligionists, interference in the collection of patriarchal taxes, restriction of the right to build synagogues, and, finally, upon the death of the patriarch Gamaliel VI in c. 425, the abolition of the patriarchate and the diversion of the Jewish tax to the imperial treasury. Though Mediterranean Jewry was now fragmented into disjointed communities and synagogues, the principles of the regulation of the Jewish calendar had been committed to writing c. 359 by the patriarch Hillel II, and this, coupled with the widespread presence of rabbis, ensured continuity of Jewish adherence. Even the emperor Justinian's (reigned 527–565) restrictions on synagogal worship and preaching apparently had no devastating effect. A new genre of liturgical poetry, combining ecstatic prayer with didactic motifs, developed in this period of political decline and won acceptance in synagogues in Asia Minor as well as beyond the Euphrates.

Babylonia (200–650). In the increasingly unfriendly climate of Christendom, Jews drew consolation in the knowledge that in nearby Babylonia (then under Persian rule) a vast population of Jews continued to live under a network of effective and autonomous Jewish institutions and officialdom. Steadily worsening conditions in Palestine had drawn many Jews to Persian domains, where economic opportunities and the Jewish communal structure enabled them to gain a better livelihood while living in accordance with their ancestral traditions. To regulate internal Jewish affairs and ensure the steady flow of taxes, the Parthian, or Arsacid, rulers (247 BCE–224 CE) had appointed c. 100 CE an exilarch, or "head of the [Jews in] exile"—who claimed more direct Davidic descent than the Palestinian patriarch—to rule over the Jews as a quasi-prince. In c. 220 two Babylonian disciples of Judah the Prince, Abba Arika and Samuel bar Abba, began to propagate the Mishna and related tannaitic literature as the yardsticks of normative practice. As heads of the academies at Sura and Nehardea, respectively, Abba and Samuel cultivated a native Babylonian rabbinate, which increasingly provided the manpower for local Jewish courts and other communal services. While the usual tensions between temporal and religious arms frequently erupted in Babylonia, too, the symbiosis of exilarchate and rabbinate endured uninterruptedly until the middle of the 11th century.

The model for Diaspora Judaism

Paradoxically, Babylonian rabbinism derived its ideological strength from its fundamentally unoriginal character. As a transplant of Palestinian Judaism it claimed historic legitimacy to the Sāsānid rulers (224–651), who protected Jewish practices against interference from fanatical Magian priests, and to native Jewish officials, who argued for the validity of indigenous Babylonian deviations from Palestinian norms. But ultimately the historic importance of this transplantation lay in Babylonia's serving as the proving ground for the adaptability of Palestinian Judaism to a Diaspora situation. Legal and theological adaptations generated by needs of the new locale and times inevitably effected changes in the religious tradition. The laws of agriculture, purity, and sacrifices all of necessity fell into disuse. The values embodied in these laws, however, and the core of the legal–theological system—ranging from doctrinal faith in the revelation and election of Israel, to the requirement that the individual live by the canons of Jewish civil and family law, and the establishment of a network of communal institutions modelled on those of the mother country—remained intact, thereby ensuring a basic continuity and uniformity to rabbinically oriented communities everywhere. The real contribution of the Babylonian rabbinate to Jewish religion lay, accordingly, in its demonstration of how Palestinian Judaism was to be implemented on Gentile soil. Since historic circumstances made Babylonia the mediator of this tradition to all Jewish communities in the High Middle Ages (9th–12th centuries), the Babylonian version of Jewish religion became synonymous with normative Judaism and the measure of Judaic authenticity everywhere.

"The law of the [Gentile] government is binding," the principle formulated by Samuel, head of the academy at Nehardea (died 254), summarizes the essential novelty in rabbinic reorientation to life on foreign soil. Whereas Palestinian rabbis had perforce to comply with imperial decrees of taxation de facto—and this was all that Samuel had in mind—Babylonian teachers now rationalized the legitimacy of governmental authority in this respect de jure and thus enjoined upon the Jews political quietism and submissiveness as part of their religious theory. In all other areas of civil law, the Jews were instructed by their rabbis to bring their litigations to Jewish courts and thus to conduct their businesses as well as their family lives by rabbinic law.

While the rabbis could obviously more effectively impose their discipline in matters of public law than in private religious practice, the density of the Jewish population in many areas of Parthia (northeastern Iran) and Babylonia facilitated the application of moral and disciplinary pressures. The most effective vehicle for the dissemination of their teachings was the academies, of which those of Sura and Pumbedita remained pre-eminent, where judges and communal teachers were trained. Frequent public lectures in the synagogues of the academies on sabbaths and festivals were capped by public *kalla* (study-course) assemblies for alumni of the schools during the two months, Adar (February–March) and Elul (August–September), when the lull in agricultural work freed many to attend semi-annual refresher instruction. These meetings were followed by regular popular lectures during the festival seasons that soon followed. Thus, while rabbis constituted a distinct class within the community, their efforts were oriented toward making as much of the community as possible members of an elite of learning and religious scrupulosity. The harmonious relations that obtained with but few interruptions over the centuries between the Sāsānian rulers and their Jewish subjects gave the Jewish population the air of a quasi-state, which the Jewish leadership frequently extolled as superior to the Jewish community of Palestine.

Redaction of the Babylonian Talmud, c. 500–650

The dissemination of the Palestinian Talmud probably stimulated the Babylonians to follow suit by collecting and arranging in similar fashion the records of study and decisions of their own academies and courts. The Babylonian Talmud, which apparently underwent several stages of redaction (c. 500–650) on the basis of the proto-Talmuds—the early collections of commentaries on the Mishna—used in the academies, accordingly became the standard of reference for judicial precedent and theological doctrine for all of Babylonian Jewry and all those communities under its influence. As had been the case with the Mishna, the redaction of the Babylonian Talmud was later designated by authorities as marking the end of a period in Jewish history, and the scholars who put the finishing stylistic touches, known as *savoraʾim* ("explicators"), were classified as a transitional stage between the *amoraim* and *geonim* (see below).

The enduring vigour of Jewish faith throughout these centuries is graphically demonstrated by the missionary activity of Jews throughout the ancient Near East, especially in the Arabian Peninsula. Proud Jewish tribes living in close proximity to each other in the vicinity of Yathrib (later Medina, Muhammad's home city), engaged in agriculture and commerce and also in proclaiming the superiority of their monotheistic ethos and eschatology (doctrine of last things). In Yemen (southwestern Arabia) the last of the Himyarite rulers (reigned from c. 2nd century CE), Dhu Nuwas, proclaimed himself a Jew and finally suffered defeat (c. 525) as a consequence of Christian influence on the Abyssinian armies. Jewish missionaries, however, continued to compete with Christian missionaries and thus helped lay the groundwork for the birth of an indigenous Arabic monotheism—Islām—that was to alter the course of world history.

THE AGE OF THE GEONIM (c. 640–1038)

Triumph of the Babylonian rabbinate. The lightning conquests in the Middle East, North Africa, and the Iberian Peninsula by the armies of Islam (7th–8th centuries) provided the environmental framework for the basically uniform (*i.e.*, Babylonian) character of medieval Judaism. As a "people of the Book" (*i.e.*, of the Bible), the Jews were permitted by the Muslims to live under the same autonomous structure that had developed under Arsacid and Sāsānian rule. The heads of the two principal academies were now formally recognized by the exilarch, and through him by the Muslim caliphate (religio-political rulers), as the official arbiters of all questions of religious law and as the religious heads of all Jewish communities that came under Muslim sway. Known as *geonim* (plural of *gaon*, "excellency"), and conducting high courts manned by scholars assigned graded ranks, these presidents drew their financial support from Jewish communities assigned to them by the exilarch. Religious questions and contributions were solicited from all Jewish communities, and these along with formal gaonic replies (*responsa*) were regularly publicized at the semi-annual *kalla* convocations. Under the strong leadership of Yehudai, *gaon* of Sura (presided 760–763), the Babylonian rabbinate exerted vigorous efforts to replace Palestinian usage wherever it was still in vogue, including the study of Palestinian amoraic legal literature, by Babylonian practice and texts, thus making the Babylonia Talmud the unrivalled standard of Jewish norms everywhere. The success of this campaign is evidenced by the fact that the term Talmud, when unqualified, has ever since meant the Babylonian Talmud. Indeed, even in Palestine the Babylonian corpus displaced its older rival and caused the study of Palestinian Talmudic literature to be confined to circles of legal specialists.

Antirabbinic reactions. The firm, and on occasion oppressive, tactics of exilarchs and *geonim* generated antirabbinic reactions, especially in outlying areas where enforcement was difficult, in the form of sectarian and messianic revolts. Inspired in part by ancient Palestinian sectarian doctrines and in part by Muslim usage, the sects were by and large quickly and forcefully suppressed. In the 9th century, however, a moderate group under the leadership of Anan ben David, a disaffected member of the exilarchic family, successfully organized a dissident movement that soon developed into a formidable challenger of Rabbinite (a term first used for the Talmudic adherents by the dissidents) supremacy. Known as Karaites (Scripturalists), the new sect advocated a threefold program of (1) rejection of rabbinic law as a human fabrication and therefore an unwarranted, unauthoritative addition to Scripture, (2) a return to Palestine to hasten the messianic redemption, and (3) a re-examination of Scripture to retrieve authentic law and doctrine. Under the leadership of Daniel al-Qumisi (c. 850?), a Karaite settlement prospered in the Holy Land, from which it spread as far as northwestern Africa and Christian Spain. A barrage of Karaite treatises arguing new views of scriptural exegesis stimulated renewed study of the Bible and Hebrew language in Rabbinite circles as well. The most momentous consequence of these new studies was the invention of several systems of vocalization for the text of the Hebrew Bible (Christian Old Testament) in Babylonia and Tiberias in the 9th and 10th centuries. The annotation of the Masoretic (traditional, or authorized) text of the Bible with vocalic, musical, and grammatical accents in the Tiberian schools of the 10th-century scholars Ben Naftali and Ben Asher fixed the Masoretic text permanently and through it the morphology (basic form and structure) of the Hebrew language for Karaites as well as Rabbinites.

The Karaite movement

In the face of sectarian challenges, the *geonim* intensified their efforts against any deviation from Rabbinite norms and began to issue handbooks of Jewish law that set forth in concise and unequivocal terms the standards for correct practice. A number of these codes, notably the *Halakhot gedolot* ("Great Laws"), *Siddur Rav Amram Gaon* (on liturgical practice), and *She'eltot* ("Disquisitions") by Aha of Shabha (c. 680–c. 752), attained authoritative status in local schools and further helped give a unitary stamp to medieval Judaism.

Secular culture and philosophy

The *geonim*, however, were powerless to halt several social developments in the 9th century that progressively undermined their hold even over Rabbinite communities. A renascence of Greek philosophy and sciences in Arabic translation, coupled with the progressive urbanization of the upper classes of all religio-ethnic groups in the centres of political, commercial, and cultural activity, generated a new intelligentsia that cut across religio-ethnic lines. Widespread skepticism in basic doctrines of faith such as creation, revelation, and retribution was most poignantly represented by latitudinarianism (the tendency to be flexible and tolerant about deviations from orthodox beliefs and doctrines) and by antinomian (anti-Mosaic-law) Gnostic groups that negated divine providence and omniscience. Hiwi al-Balkhi, a 9th-century skeptical Jewish pamphleteer, scandalized the faithful by an open attack on the morality of Scripture and by an expurgated edition of the Bible for schools that omitted "offensive" materials (*e.g.*, alleged stories of God acting dishonestly). A mystifying Hebrew tract entitled *Sefer yetzira* ("Book of Creation") posited in terse and enigmatic epigrams a novel theory of creation that betrayed unmistakable Neoplatonic influence. Karaites joined philosophically oriented intellectuals in heaping scorn on popular Rabbinite customs that smacked of superstition and, above all, on Talmudic homilies that referred to God in anthropomorphic terms.

Gaonic difficulties were compounded by the rise in North Africa and Spain of populous and wealthy Jewish communities that, thanks to the development of their own local schools and native talent, ignored the Babylonian academies or favoured one over the other with religious queries and, in consequence, with financial contributions. To the delight of dissidents and the chagrin of the faithful, competition between the Babylonian academies turned to internecine hostility. Occasional revolts against exilarchic taxation and administration in outlying areas of Persia had to be quelled with armed force. The Palestinian Rabbinites had revived their own academies, and their presidents now not only appealed for support in other Diaspora lands but challenged the authority of the Babylonians to serve as final arbiters on such matters of public import as the regulation of the calendar. By 900 the Rabbinite community of Babylonia was in a state of chaos and dissolution.

The gaonate of Sa'adia ben Joseph. In a bold effort to restore discipline and respect for the gaonate, an able exilarch, David ben Zakkai (916/917–940), bypassed the families from whom the *geonim* had traditionally been selected and in 928 appointed Sa'adia ben Joseph al-Fayyumi to head the academy of Sura. Of Egyptian birth, Sa'adia had gained wide acclaim for his scholarly retorts to Karaites, heretics, and Palestinian Rabbinites. Politically, Sa'adia's brief presidency was a fiasco and aggravated the chaos by a communal civil war. His gaonate, however, gave an official stamp to his many works, which responded to the ideological challenges to Rabbinism by restating traditional Judaism in intellectually cogent terms. Sa'adia thus became the pioneer of a Judeo-Arabic culture that was to come to full flower in Andalusian Spain a century later (see below *Sefardic developments*). His translation of the Bible into Arabic and his Arabic commentaries on Scripture made the rabbinic understanding of the Bible accessible to masses of Jews. His poetic compositions for liturgical use provided the stimulus for the revival of Hebrew poetry. Above all, his rationalist commentary on the puzzling "Book of Creation" and his brilliant philosophic treatise on Jewish faith, *Beliefs and Opinions,* synthesized Torah (the divine law in the Five Books of Moses and the rabbinic understanding of this revelation) and "Greek wisdom" in accordance with the dominant Muslim philosophical school of Kalām and thus made Judaism philosophically respectable and the study of philosophy a religiously acceptable pursuit.

Gaonates of Sherira and Hai

Far from tightening the gaonic hold over the Jewish communities of the Arabic world, Sa'adia's works actually provided the wherewithal for ever-greater intellectual and religious self-sufficiency. While economic, political, and military upheavals progressively weakened all institutional fabrics in the Middle East, concurrent prosperity and consolidation in the West stimulated the maturation of indigenous leadership in Egypt, al-Qayrawān (Kairouan; in present-day Tunisia), and Muslim Spain. To be sure, able *geonim* such as Sherira and his son Hai exercised enormous influence over the Judeo-Arabic world through hundreds of legal *responsa* issued in the course of their successive terms (968–1038) at Pumbedita. Circumstances beyond anyone's control, however, were bringing the curtain down on the effectiveness of exilarchate and gaonate. But by 1038, the year of Hai's death, the consequences of four centuries of gaonic activity had become indelible: the Babylonian Talmud had become the agent of basic Jewish uniformity; the synthesis of philosophy and tradition had become the hallmark of the Jewish intelligentsia; and the Hebrew classics of the past had become the texts of study in Jewish schools everywhere:

MEDIEVAL EUROPEAN JUDAISM (950–1750)

The two major branches. Despite the fundamental uniformity of medieval Jewish culture, the cultural–political divisions within the Mediterranean basin, in which Arabic-Muslim and Latin-Christian civilizations coexisted as discrete and self-contained societies, shaped the character of the Jewish subculture of the area. Two major branches of rabbinic civilization developed in Europe, the Ashkenazic, or Franco-German, and the Sefardic, or Andalusian-Spanish. Distinguished most conspicuously by their varying pronunciation of Hebrew, the numerous differences between them in religious orientation and practice derived, in the first instance, from the geographical fountainheads of their culture—the Ashkenazim (plural of Ashkenazi) tracing their cultural filiation to Italy and Palestine and the Sefardim (plural of Sefardi) to Babylonia—and from the influences of their respective immediate milieus. While the Jews of Christian Europe wrote for internal use almost exclusively in Hebrew, those of Muslim areas regularly employed Arabic for prose works and Hebrew for poetic composition. Whereas the literature of Jews in Latin areas was overwhelmingly religious in content, that of the other branch was well endowed with secular poetry and scientific works inspired by the cultural tastes of the Arabic literati. Most significantly, the two forms of European Judaism differed in their approaches to the identical rabbinic base that both had inherited from the East and in their radically different attitudes to Gentile culture and politics.

The Golden Age of Jewish Arabic and Hebrew culture

Sefardic developments. In Muslim Spain, Jews frequently served the government in official capacities and, therefore, not only took an active interest in political affairs but also engaged in considerable social and intellectual intercourse with influential circles of the Muslim population. Since the support of letters and scholarship was part of state policy in Muslim Spain, and since Muslim savants traced the source of Muslim power to the vitality of the Arabic language, scripture, and poetry, Jews looked at Arabic culture with undisguised admiration and unabashedly attempted to adapt themselves to its canons of scholarship and good taste. The hallmark of the cultured Jew accordingly became a polished command of Arabic style and the ability to display the beauty of his own heritage through a philological mastery of the text of the Hebrew Bible and through the composition of new Hebrew verse, now set to an alien Arabic metre. Since Arabic philosophers and scientists promulgated syntheses of Greek philosophy with the revelation to Muhammad, rationalist study of the Jewish classics and defense of rabbinic faith in philosophic terms became dominant motifs in the Andalusian Jewish schools (in southern Spain).

The atmosphere generated a fever of literary creativity in classical Jewish disciplines as well as in the sciences cultivated by the Arabs that has gained for the period the title of "the Golden Age of Hebrew literature" (c. 1000–1148). What distinguished the Jewish culture of this age was not only the supreme literary merit of its Hebrew poetry, the new spirit of relatively free and rationalist examination of hallowed texts and doctrines, and the extension of Jewish cultural perspectives to totally new horizons—mathematics, astronomy, medicine, philosophy, political theory, aesthetics, belles-lettres—but also the frequent overlapping of the Sefardic religious leadership with the new Jewish courtier class. The unprecedented heights which the latter attained—Hisdai ibn Shaprut as counsellor to the caliphs of Córdoba, the Ibn Nagrelas as viziers of Granada, the Ibn Ezras, Ibn Megashs, and Ibn Albalias as high officials in Granada and Seville—and the distinctions of these men and of their protégés in Jewish and worldly letters restored the ancient integration of culture and practical life and generated a neoclassicism ("classicism" here meaning biblicism) that expressed the identification of the Jewish elite with the biblical age of Jewish power and artistic creativity. The effort to recapture the vitality and beauty of biblical poetry stimulated comparative philological and fresh exegetical research that yielded new insights

Influence of Greco-Arabic philosophy

into the morphology of the Hebrew language and into the historical soil of biblical prophecy. Judah ibn Ḥayyuj and Jonah ibn Janah (c. 1000) produced manuals on biblical grammar that applied the results of Arabic philology to their own tongue and that have, accordingly, provided the principles of Hebrew grammatical study down to modern times. The anticipations of modern higher biblical criticism by Judah ibn Bala'am and Moses ibn Gikatilla (flourished 11th century) were popularized in Hebrew a few generations later by Abraham ibn Ezra. In the revival of Hebrew poetry, liturgical as well as secular, that translated the new preoccupation with language and beauty into art, Andalusian Jewry saw its greatest achievements. Solomon ibn Gabirol, Moses ibn Ezra, and Judah ha-Levi were but the acknowledged supreme geniuses of a form of expression that became a passion with thousands the length and breadth of Spain. But by far the most enduring consequence of the new temper was their redefinition of religious faith in the light of Greco-Arabic philosophical theories. Solomon ibn Gabirol's exposition of faith in Neoplatonic terms, Abraham ibn Daud's defense of Rabbinism by Aristotelian categories, Judah ha-Levi's attack on philosophy as religiously bankrupt, and Moses Maimonides' epoch-making synthesis of Judaism and medieval Aristotelianism fixed philosophic inquiry as an enduring subject on the agenda of rabbinic concerns. A new class of philosophers that emerged in the 13th century and sponsored the translation of Arabic literature into Hebrew and of Hebrew and Arabic literature into Latin brought Jews and their thought into the mainstream of Western philosophy and gained for them the position of middlemen of culture between East and West.

The salient trends of Sefardic Judaism did not imply relegation of the rabbinic class to a second place. Rather they shaped a fresh approach to rabbinic texts that paralleled in many respects those adopted in biblical exegesis. Strict adherence to consistency, systematization, and philological exactitude yielded new codes that often diverged from gaonic judgments. A digest of Talmudic law by Isaac Alfasi placed the Sefardic rabbinate on a self-reliant footing and epitomized its ideal of getting at the essentials of Talmudic law by sidestepping contingent discussions. In this area, too, it was Moses Maimonides who through his code of Jewish law, *Mishne Torah*, brought the Sefardic principles of comprehensiveness, lucidity, and logical arrangement to their apex. Written in Mishnaic Hebrew, the work remains to this day the only comprehensive treatment of all of Jewish law, including those fields that are not applicable in the Diaspora (agriculture, purity, sacrifices, Temple procedure).

With Maimonides, however, the pure Sefardic tradition came to an end, for the Almohad (Berber Muslim reformers) invasion of Spain in 1147–48 wiped out the Jewish communities of Andalusia and drove thousands either to northern Spain and Provence or, as in the case of Maimonides' family, to North Africa and Egypt. Sefardic Jewry suddenly encountered a discrete, mature, Jewish culture that for centuries had been developing independently and along quite different lines.

Ashkenazic developments. The spokesmen of Ashkenazic Jewry, into whose communities the Sefardim had been thrust by political events, regarded their own heritage and the Christian world in which they lived from a perspective shaped exclusively by rabbinic categories. From the world of the Talmud and Midrash they drew their school texts and the values that determined their judgments. Sensing no intellectual challenge in Christian faith, which they regarded with thinly concealed contempt, they constituted for the most part a merchant class that lived in urban centres under the protection of ecclesiastical and temporal rulers but under their own complex of laws and institutions. Except for mercantile relations, Christian society was closed to them, thanks largely to age-old ecclesiastical prohibitions forbidding all social intercourse with them. With the Arab conquest and the rise of the Carolingians (the 8th–10th-century dynasty that ruled France and Germany), the 12-decade interlude of suppression by the Visigoths (589–711) came to an end, and the Roman precedent of toleration and autonomy again became the rule. Merchants and rabbis moved from Italy to France and the Rhineland and infused new energies into the Jewish communities there. A native religious leadership began to emerge at the very time that Andalusian Jewry was entering its Golden Age. The bloody upheavals of the First Crusade (1096–99) in the communities of the Rhineland, although unleashing a tide of hatred, periodic violence, and progressive restrictions on Jewish activities, struck Jewish communities that had attained sufficient resilience to re-establish their communal institutions shortly afterward and continue the cultivation of their deeply ingrained traditions.

Ashkenazic cultural and religious patterns

By 1150 Ashkenazic Jewry had generated a culture pattern of its own with an indigenous literature that ranged from the popular homily to the esoteric tract on the nature of the divine glory. Study of the Bible and Talmud was oriented toward a mystical pietism in which prayer and contemplation of the secrets embedded in the liturgy were to lead to religious experience. Significantly, the fathers of the Ashkenazic tradition were remembered as liturgical poets and initiates into divine mysteries, and the early codes of the Franco-German schools were heavily weighted with discussions of liturgical usage. After the Second Crusade (1147–49), the German Jewish mystics (also called Hasidim, or pietists) placed heavy emphasis on the merits of asceticism, martyrdom and lifelong disciplines of penitence, thus adapting to Jewish idiom the features of saintliness celebrated in the universe of discourse of which they were a part. For the masses of Jews the cultural fare consisted principally of biblical tales and instruction, as interpreted by rabbinic Midrash, the lives of scholars and saints, and liturgical poetry reaffirming the election of Israel and faith in messianic redemption. The chief vehicle of popular instruction consisted of anthologies from the Rabbinic writings and commentaries on Scripture, of which the most popular was that of Rabbi Solomon ben Issac of Troyes, known as Rashi, the acronym formed from the initials of his name in Hebrew. For the more advanced student, Rashi composed a succinct commentary on the Talmud that, unmatched for compact thoroughness and lucidity, achieved an authority approaching that of the text itself.

As living sources of law and values, the Bible and Talmud had an impact that was apparent in communal decision and in the bearing of the leadership at home, in the marketplace, and in the synagogue. Taking their cue from Talmudic precedent and from Christian ecclesiastical procedures of their own times, the Ashkenazic rabbis occasionally gathered in regional synods to enact legislation on problems of a general nature for which there was no adequate precedent in the literature. Among the most enduring of these measures were the prohibition of bigamy and arbitrary divorce and severe economic penalties for abandonment of wives. Of far more immediate concern to the average Jew were the circumvention of Talmudic prohibitions against usury, relaxation of prohibitions regarding traffic with Gentiles in wines, and adoption of severe disciplinary measures, such as excommunication, against informers or those appealing, in cases involving Jews, to the Gentile authorities.

The rise and spread of Kabbala

A new religious trend began in Provence (a province of southeastern France) in the 13th century with the introduction into the Talmudic academies of a novel form of mystical study known as Kabbala (literally, "tradition"), which soon spread to northern Spain. Expressing Gnostic-type doctrines in rabbinic guise, the devotees of Kabbala devised an esoteric vocabulary that reinterpreted the Bible and rabbinic law as allegories of the various modes in which God is manifested in a spiritual universe, access to which was reserved for initiates. The most renowned literary product of this new circle was the *Zohar* ("The Book of Splendour"), a vast mystical commentary on the Pentateuch by Moses de León (c. 1275), which with later additions became the Bible of Jewish mystics everywhere. Although some of the theological notions of the Kabbalists deviated from basic postulates of Jewish monotheism, the insistence of the mystics on unflagging ritual orthodoxy and on a nominal acceptance of the biblical text as

Jews, longing for a return to the Holy Land, point to a visionary Jerusalem, which is depicted in the Gothic style of Christian church architecture. The Jews are shown with the pointed hats they were required to wear to distinguish them from Christians, and are represented with birds' heads, since they felt it was sacrilegious to depict the human form in sacred objects. Illustration from the Birds' Head Haggada, an illuminated manuscript from southern Germany, c. 1300.

By courtesy of the Israel Museum, Jerusalem

divine revelation helped them avert the suspicions aroused by Jewish Aristotelians and Averroists (followers of the 12th-century Arabic Aristotelian philosopher Averroes) and, in time, even won for them the status of a rabbinic elite. Indeed, some of the mystics lent their support to the antiphilosophic campaign that began in Montpelier, in southern France, *c.* 1200 and condemned the study of philosophy as generating skepticism, latitudinarianism, and disrespect for traditional literature. (For a fuller discussion of Kabbala see JEWISH MYSTICISM.)

Conflicts, disasters, and new movements. Basically, the conflict between "fundamentalist" and philosopher in Provence and northern Spain represented a clash between two mature Jewish subcultures of diverse geographic origins, the Sefardic and Ashkenazic, each of which had in the course of centuries developed different esoteric doctrines to transcend the legalistic formalism and confining dogmas of normative Judaism. Both forms of speculation sought salvation for exceptional individuals through knowledge and thus provided an immediate substitute for messianic deliverance from exile and servitude. Each group charged the other with distortion of tradition, and each issued apologias (defenses or justifications) and excommunications characteristic of medieval doctrinal controversy. While the rifts within communities attained bitter proportions, the common threat posed by ecclesiastical attacks on the Talmud in public disputations and by the expulsion of the Jews from France in 1306 prevented open rupture or resolution of the conflict. Ever since that time, two strands of orthodoxy representing the two forms of medieval metaphysical speculation have lived side by side in an uneasy truce.

Most rabbinic circles of the 14th and 15th centuries displayed a progressive dogmatism and insistence on uniformity of practice. The great legal code of Jacob ben Asher of Toledo, ***Arba'a ṭurim*** (*c.* 1335; "Four Rows"), which sought to level differences in usage between Ashkenazim and Sefardim, bespoke the dominant trend of the rabbinate. The increasing hardening of ideological lines, however, did not eliminate independent thinking. Levi Gersonides gave Jewish Aristotelianism a new and comprehensive formulation, while Isaac Albalag propounded an Averroist (rationalistic) interpretation of the Bible predicated on a theory of double truth (of reason and revelation). In Muslim areas, the Maimonidean regimen of philosophic contemplation was extended by Maimonides' son Abraham to a quest for pietist ecstasy that betrayed many features of Siifism (Islāmic mysticism).

Adversity and response: mysticism, messianism, rationalism

Anti-Jewish riots and massacres of 1391 and a wave of apostasy in the wake of the disputation of Tortosa (1411–14)—which ended with a papal bull forbidding Talmud study, compelling attendance at Christian sermons, and other onerous measures—struck catastrophic blows in the Spanish communities and fed the anti-intellectualism of the rabbinate. Hasdai Crescas, while conceding the philosophic untenability of traditional belief in freedom of the will, launched a scathing attack on Aristotelian approaches to religion, and his disciple Joseph Albo issued a compendium on dogma that reaffirmed the traditional postulates of divine creation, revelation, and retribution as axioms of Judaism. But these reassertions of traditional faith could not overcome the ideological and social flagmentation that had split the Spanish communities into congealed strata that were often in open conflict with each other. Widespread marranism (ostensible conversion to Christianity) polarized the community and left deposits of bitterness that extended to those returning to the fold. The expulsions from Spain (1492) and Portugal (1497 and 1506) dealt the final blow and drove the escaping leadership into intensified pursuits of mystical escape from, and rationalization of, the endless calamities that befell their flocks. In Italy and the Ottoman Empire (Asia Minor, northeastern Africa, and southeastern Europe), the two principal centres of refuge for the exiles of the Iberian Peninsula, legalistic Kabbalism, which insisted on strict observance of the law as precondition of mystical practice and study, became the dominant spirit of a rabbinic leadership that in the face of terribly adverse circumstances continued to produce works of encyclopaedic proportions and staggering erudition in every field of Jewish learning.

Inspired by the Jewish tradition that the messianic era —when the messiah would come to bring in the rule of God—would be preceded by horrendous catastrophes, a group of single-minded rabbis established a community in Ẓefat (Safed), Palestine, where in anticipation of the new dawn all of life was to be conducted on principles of saintliness and mystical contemplation. Under the leadership of one Jacob Berab, the ancient practice of ordination was reinstituted in 1538 to form the nucleus of a revived Sanhedrin so as to administer ritual procedures requiring ordained authorities. While the effort failed because of rabbinic opposition, it reflected a widespread temper and further fanned messianic hopes sparked shortly before by the campaigns of tragic consequences by David Reubeni and Solomon Molkho in Italy, which ended in their being burned at the stake by the Christian authorities. In Ẓefat itself Kabbalism soon entered a new phase under the inspiration of Isaac Luria and Ḥayyim Vital, who confided to their disciples that the calamities of Israel were but a mirror of the captivity into which many sparks of the Godhead itself had fallen. Liturgical innovations and a novel mystical theology were formulated to redeem the imprisoned elements of divinity and thus restore creation to the harmony intended for it (see also JEWISH MYSTICISM).

That the Almighty himself was not quite omnipotent, at least with respect to the fate of his chosen people, was cautiously hinted in a Hebrew work of history (1550) by Solomon ibn Verga, who saw the Jewish problem as a socio-political one to which theological answers were futile. Such guarded rationalism was entertained by a number of courageous thinkers in 16th-century Italy, where, despite the policy of ghettoization (the segregation of the Jewish community in a restricted quarter) begun by Venice in 1516 and soon extended to all major Italian cities, the spirit of the Renaissance and the passion for historical criticism had captivated many Jews. Catholic scholars and prelates occasionally employed rabbis to instruct them in the Hebrew language and in the secrets of the Kabbala, which some Christians believed actually verified the postulates of their own faith. Contacts with Christian scholars in turn introduced Jews like Azariah dei Rossi, whose *Meor ʿenayim* ("Enlightenment of the Eyes") inaugurated critical textual study of rabbinical texts, to new bodies of literature that had been lost to the Jewish community, such as the works of Philo and Josephus (see above Hellenistic Judaism).

The Shabbetaian debacle

Such phenomena, however, were decidedly in the minority and contrary to the dominant trend. Dogmatic Kabbalism spread progressively and finally came to social expression in 1666 with the widespread acceptance of the views of the pseudo-messiah Shabbetai Tzevi (Sabbatai Zevi). Most of European and Ottoman Jewry was swept into a hysterical pitch in the belief that the end was now finally at hand. When the pseudo-messiah converted to Islām after being apprehended by the Ottoman government, mass despondency took the form of crypto-Shabbetaianism in which the apostasy of the messiah was explained as a form of voluntary crucifixion for the sake of the Jews. A witch-hunt on the part of traditionalists to uncover the cells of heresy unsettled Jewish communities everywhere by an emphasis on greater rigidity than before.

The century that followed (to *c.* 1750) was the darkest in the history of rabbinic Judaism. Scholarship reached an ebb of quality and popular religion a mechanical state such as Jews had never before experienced. The massacres and impoverishment of Polish Jewry after 1648 brought a pall over the growing eastern European centres of Jewish life. Antinomian eruptions on the part of extreme Shabbetaians under the leadership of the self-proclaimed messiah and later Catholic convert Jacob Frank (1726–91) alarmed Gentile authorities almost as much as they did Jews. But the fossilization referred to above was only apparent. Beneath the surface many were restlessly searching for new avenues of faith, and the 18th century saw fresh responses that set the history of the Jews and of Judaism on new directions and spelled the beginnings of a new era. (G.D.C.)

V. Modern Judaism

THE NEW SITUATION

The various criteria used to mark off dividing points in the history of the Jews and Judaism (see above General *observations*) are especially notable when it comes to setting a starting date for the modern period. Historians of thought put it in the late 17th century with the appearance of men, such as the philosopher Benedict de Spinoza, who ceased, in part or in toto, to believe in the inherited faith without at the same time ceasing to be Jews (*i.e.*, to consider themselves and be considered as Jews). Some Israeli scholars set it at about 1700 with the first stirrings of that new and continuing emigration from the Diaspora to the Holy Land that culminated in the mid-20th century in the creation of the State of Israel. Political and social historians set it in the mid- and late-18th-century processes that led to the American and French revolutions and to the results that flowed from these two epochal events, among them the emancipation of Jews from discriminatory and segregative laws and customs, the attainment of legal status as citizens, and the freedom of individual Jews to pursue careers appropriate to their talents. These varying approaches appear to have one thing in common —the view that these postmedieval forms of Jewish experience assume the end of the doctrine of the Exile, whereby Jews saw themselves as a people waiting out centuries of woe in alien lands until the moment of divine redemption. Jewish modernity for most scholars, then, is marked by the end of a passive waiting on the Messiah and the beginning of an active pursuit of personal or national fulfillment on this earth and preferably in one's lifetime.

Although the 18th century Haskala (Enlightenment) among the Ashkenazim of central and eastern Europe is often taken as the starting point of Jewish modernity, the process of Westernization had begun a good deal earlier among the Sefardim in western Europe and Italy. The Marranos who went to such communities as Amsterdam and Venice in the 17th century to declare themselves as Jews carried with them the Western education that they had acquired while living as Christians in the Iberian Peninsula, and the habits of criticism that had kept them from assimilating into the majority during their Marrano years and that some, such as Spinoza, a son of Marranos, used in analyzing all of the biblical tradition, including especially their own religion. In Italy there was an older Jewish community that had never been sealed off culturally from the influence of its environment; some of its figures were influenced by, and participated in, the main currents of the Renaissance (see above Rabbinic Judaism).

Increased contact with Western languages, manners, and modes of living came to the Ashkenazim only in the 18th century when new economic opportunities created such possibilities and needs. Jewish bankers and factors in various German principalities, army provisioners in most of the European countries, capitalists who were permitted to live in such places as Berlin because they opened new factories or were otherwise helpful to the expansion of the economy—all were in increasing contact with Gentile society, and most of them began to look upon the goal of their lives as the winning of full acceptance. Around this wealthy element there arose a number of intellectuals who agitated for the end of ghetto ways as the necessary preamble to the emancipation of the Jews.

THE HASKALA, OR ENLIGHTENMENT

The role of Moses Mendelssohn

In central Europe. By far the most outstanding figure of the 18th-century Jewish Enlightenment was the philosopher Moses Mendelssohn, who, while remaining a devoted adherent of Orthodox Judaism, turned away from the traditional Jewish preoccupation with the Talmud and its literature to the intellectual world of the European Enlightenment, of which he became the foremost Jewish representative. Mendelssohn did not attempt a philosophical defense of Judaism until pressed to do so by Christians who questioned how he could remain faithful to what they saw as an unenlightened religion. In his response, *Jerusalem*, published in 1783, Mendelssohn defended the validity of Judaism as the inherited faith of the Jews by defining it as revealed divine legislation and declared himself at the same time to be a believer in the universal religion of reason, of which Judaism was but one historical manifestation. Aware that he was accepted by Gentile society as an "exceptional Jew" who had embraced Western culture, Mendelssohn's message to his own community was to become Westerners, to seek out the culture of the Enlightenment. To that end he joined with a poet, Naphtali Herz (Hartwig) Wessely, in translating the Torah into German, combining Hebrew characters with modern German phonetics in an effort to displace Yiddish, and wrote a modern biblical commentary in Hebrew, the Be'ur ("Commentary"). Within a generation, Mendelssohn's Bible was to be found in almost every literate Jewish home in central Europe and had served to introduce its readers to German culture. Through his personal example and his life's work Mendelssohn made it possible for his fellow Jews to join the Western world without sacrificing their Judaism; he had convinced them that their intellectual processes were those of universal reason, with which Judaism accorded.

Mendelssohn's work was carried on by a group of Jewish intellectuals who had gathered around him in his lifetime, forming the nucleus of the Berlin Haskala, which was most active in the 20 years following their mentor's death. In the pages of their Hebrew-language periodical, ha-nze'assef ("The Collector"), they preached the virtues of secular culture and used Hebrew as a vehicle by which to introduce that culture. To achieve their goal of an enlightened Judaism, the leaders of the Berlin Haskala publicized the need for secular education. In response to the Holy Roman emperor Joseph of Austria's Edict of Toleration of 1781, Naphtali Wessely welcomed the efforts and issued an urgent call for the

Educational reform

reform of Jewish education as a prelude to full emancipation. Purely secular subjects — mathematics, German, and world history and literature — were to take precedence over the traditional Jewish studies. The study of the Bible, since it was generally acknowledged to be a fundamental part of Western culture, was to be emphasized at the expense of the more traditional learning in the Talmud. Following this model, modern Jewish schools were established by Jewish intellectuals and businessmen in several German cities, among them Frankfurt and Hamburg. As its educational activities began to bear fruit in the wide dissemination of secular culture, the Berlin Haskala abandoned the use of Hebrew for German and gradually disintegrated. Unlike Mendelssohn himself, the immediate descendants of his circle and his own children were unable to strike a balance between Jewish and secular culture; their Western education undermined their religious faith and they perceived their identity as Europeans rather than as Jews.

One of Mendelssohn's disciples, David Friedlander, offered to convert to Christianity without accepting Christian dogma or Christian rites; he felt that both Judaism and Christianity shared the same religious truth but that there was no relation at all between Judaism's ceremonial law and that truth. The offer was refused unless Friedlander would acknowledge the superiority of Christianity and make an unconditional commitment to it, which he was not prepared to do. Unlike Friedlander, many others who began by following Mendelssohn chose to leave the Jewish faith as the only way to win full acceptance in the European community of which they felt themselves a part.

The modern Hebrew Renaissance

In eastern Europe. The Haskala, thus, was quickly played out in central Europe; as an idea its further career was to continue in eastern Europe, particularly in the Russian Empire, where it flourished in the middle third of the 19th century until, as a result of the pogroms of 1881, Jews lost faith in the goodwill of Russians to accept "enlightened" Jews. It was a tenet of the Russian Haskala that the tsar was a benevolent leader who would bestow emancipation upon his Jewish subjects as soon as they proved themselves worthy of it; and that it was the task of the Jews, then, to transform themselves into model citizens, enlightened, unsuperstitious, devoted to secular learning and productive occupations. Following the example of the Berlin Haskala, a Russian Hebrew-language writer, Isaac Baer Levinsohn, published a pamphlet, Te'uda *be-Yisrael* ("Testimony in Israel") citing the benefits of secular education. At the same time, such writers as Joseph Perl and Isaac Erter, though Orthodox Jews themselves, in virulent satire attacked the superstitious folk customs of the masses and opened the way to the anti-clericalism which was to become characteristic of the Russian Haskala. In the 1840s and 1850s the emphasis shifted from satire and attack on the cultural parochialism of the Pale of Settlement (the regions to which the Jews were restricted) to romanticization of life outside the Pale, including periods of the Jewish past. Thus, Hebrew poets and novelists, such as Michal Levensohn and Abraham Mapu, arose on Russian soil to contribute their talents to the creation of a modern Hebrew literature. With the climate of government reforms in the 1860s, the Russian Haskala entered a "positivist" phase, calling for practical social and economic reforms. Hebrew-language journals were established and the Hebrew essay and didactic poetry, calling for religious and cultural reforms, came into their own, particularly at the hands of such stylists as the poet Y.L. Gordon and the essayist Moses Leib Lilienblum. Abandoning the original Hebrew and German orientation of the Russian Haskala, a number of Jewish intellectuals, the most prominent of whom were Yoachim Tarnopol, Osip Rabinovich, and Lev Levanda, became Russifiers, founding Russian-language Jewish weeklies devoted to "patriotism, emancipation, modernism." Like their contemporary fellow Jews in western Europe, they declared themselves to be Russians by nationality and Jews by religious belief alone. In 1863 a group of wealthy Jews in St. Petersburg and Odessa created the Society for the Promotion of Culture among the Jews of Russia for the purpose of educating Jewry into "readiness for citizenship." The goal of all segments of the Russian Haskala in the 1860s and 1870s was to turn Jews into good Russians and to make their Jewishness a matter of personal idiosyncrasy alone. The period of reaction that set in with the pogroms (massacres) of 1881 was to prove how deluded the hopes of the Haskala had been.

RELIGIOUS REFORM MOVEMENTS

German Reform Judaism

One element of Westernization that the Haskala had championed was the reform of religion. It began in western Europe during the Napoleonic period (1800–15) when certain aspects of Jewish belief and observance were seen as incompatible with the new position of the Jew in Western society. Napoleon convoked a Sanhedrin (Jewish legislative council) in 1807 to create a new, modern definition of Judaism in its renunciation of Jewish nationhood and national aspirations, its protestations that rabbinic authority was purely spiritual, and its recognition of the priority of civil over religious authorities even in the matters of intermarriage. In areas other than France, the rationale for reform, at least in its early years, was more aesthetic than doctrinal. The external aspects of worship—*i.e.*, the form of the service — appeared unacceptable to the newly Westernized members of the Jewish bourgeoisie in both Germany and the United States, whose standards of cultural acceptability had been shaped by the surrounding society, and who desired above all to resemble their Gentile peers. Thus, the short-lived Reform temple established in Seesen, by the pioneer German reformer Israel Jacobson, in 1810 enshrined order and dignity of a Protestant type in the service and introduced an organ, sermon, and prayers in German, in place of Hebrew, to create an uplifting spiritual experience. The more radical temple in Hamburg (established 1818) adopted all of Jacobson's reforms and published its own much-abridged prayer book, which deleted almost all the references to the long-awaited restoration of Zion. Reformers in Charleston, South Carolina, introduced similar changes in the synagogue ritual in 1824, for they sought a non-national Judaism similar in form to Protestantism and adapted to the surrounding culture. It was apparent to the reformers that in Western society Judaism would have to divest itself of its alien customs and conform to the cultural and intellectual standards of the new "age of reason."

German Reform in the 1840s became institutionalized, a matter of organized, formal belief and practice, and, at a series of synods held at Brunswick (1844), Frankfurt (1845), and Breslau (1846), it created the first theological rationalization for changes introduced in the previous generation. Judaism, it was declared, had always been a developmental religion that conformed to the demands of the times, and, since the Jews were not now a nation, they were no longer bound by their entire religious–political code of law, but only by the dictates of moral law. Thus, those rituals which stood in the way of full Jewish participation in German social and political life were no longer considered valid expressions of Jewish religious truth. The use of Hebrew in religious services was limited; practices such as the dietary laws and circumcision and all national messianic hopes were discarded upon the altar of the "spirit of the times." Messianism in Reform Judaism was transmuted into active concern for social welfare in the present, and the Jewish role in history became Diaspora-centred, a mission to the Gentiles.

Reform Judaism in the United States

Although Reform was initiated in Europe, it did not enjoy a successful career there, for many central European governments that regulated the existence of religious communities would not recognize more than one form of Judaism in any one locale. Reform did not achieve its greatest success until it was imported into the United States along with the massive German-Jewish immigration of the 1840s and coalesced with earlier American trends toward reform. By 1880 almost all of the 200 synagogues in the United States had become Reform, amalgamating in the Union of American Hebrew Congregations (formed 1873). In 1885 the Reform philosophy was given its most comprehensive formulation in the so-called Pittsburgh Platform, drawn up by a conference of Reform rabbis. This manifesto declared that Judaism was an evolutionary faith, and no longer a national faith, and that it was now to be de-orientalized. While the preservation of historical identity was considered to be beneficial, the maintenance of continuity of tradition was not; the Talmud was to be considered merely as religious literature, and not as legislation. The rationalist principles of the Pittsburgh Platform remained the official philosophy of the American Reform movement until a later generation, seeking to meet different emotional and intellectual needs, reintroduced the concept of Jewish peoplehood into the Columbus Platform of 1937, which also re-emphasized Hebrew and traditional liturgy and practices. Classical (19th-century) Reform was very much a late child of the Enlightenment, and by mid-20th century its Enlightenment philosophy appeared antiquated to many Jewish thinkers.

Conservative Judaism

If Reform was a child of Enlightenment rationalism, Conservative Judaism was a child of historical romanticism. It began in 1845 when Zacharias Frankel and a group of followers seceded from a second Reform synod at Frankfurt over the issue of the limitation of the use of Hebrew to a small core of prayers. For Frankel, Hebrew represented the spirit of Judaism and the Jewish people, and Judaism itself was not merely a theology of ethics but the historical expression of the Jewish experience; this definition he called "positive-historical Judaism." Although Conservative Judaism conceived of Judaism as a developmental religion, it charted its course through close study of the tradition and the will of the people, and thus came to largely traditional conclusions about religious observance.

ORTHODOX DEVELOPMENTS

In western and central Europe. The bulk of the official Jewish establishment in western and central Europe, though affected by the efforts at religious reform, remained Orthodox (a term first used by Reform leaders to designate their traditionalist opponents). Under the leadership of Samson Raphael Hirsch, a more modern and militant form of Neo-Orthodoxy arose, based in Frankfurt am Main, which asserted its right to break with any Jewish community that contained Reform elements and to form an independent community. The thinking of this group was profoundly influential, for it indicated the possibility of living a ritually and religiously full life while being totally integrated into Western society. It accomplished this by positing a theoretical division between religion and culture; the Jews were to remain Orthodox in religion (although deferring their messianic aspirations to the unforeseeable future) while becoming Western in manners and culture. This form of Orthodoxy, which became the intellectual model for Western Orthodoxy, continued into the early 1970s in the United States in a variety of religious and academic institutions (such as the Yeshiva University and the bulk of English-speaking Orthodox synagogues), coexisting in substantial tension with a number of Orthodox groups, most notably the Lubavitcher and Satmar Hasidim (for Hasidism, see below), and some Talmudic academies that saw the Western world itself as the enemy and chose to recreate the ghetto.

In eastern Europe. By the mid-18th century Orthodoxy in eastern Europe, having been convulsed by frantic messianism and stifled by the sterility of purely legalistic scholarship, was ripe for revival. The experience of Shabbetaianism (the first messianic movement to excite virtually all of world Jewry) had revealed in the mid-17th century the pervasiveness of Jewish exhaustion with the Exile and fervent longing for messianic redemption, while the nihilistic sect of Frankists (the followers of Jacob Frank, see above Rabbinic Judaism) in the 18th century had transmuted that messianism into a this-worldly hysteria. Talmudic piety and study, sunk in excessive pilpul (acute logical distinctions that often became mere hairsplitting), was refreshed by the new critical methods of Elijah ben Solomon, the gaon of Vilna. Although essentially a legal rigorist, he was open to Western scientific learning insofar as it helped him to elucidate Talmudic texts. Orthodox religious expression also was raised to a new level with the development of Ḥasidism (Pietism) by Israel Baal Shem Tov in the mid-18th century. Although Ḥasidism contained elements of social protest, being at least in part a movement of the poor against the wealthy communal leadership and of the unlearned against the learned—though many of its leaders, among them Rabbi Baer, the *maggid* ("preacher") of Mezhirich and Rabbi Levi Isaac of Berdichev, were well-versed in Talmudic learning—it was essentially a non-messianic outcry in the name of religious emotion, emphasizing prayer and personal religious devotion here and now. Contemporary scholarship is investigating the linkage between Ḥasidism and eastern European Christian pietistic movements. The major innovation that Ḥasidism introduced into Jewish religious life was the charismatic leader, the rebbe who served as teacher, confessor, wonder-worker, God's vicar on earth, and occasionally as atoning sacrifice. Although the earliest *rebbes* were democratically chosen, the position of leadership passed to their descendants on the presumption that they had inherited their fathers' charisma and thus created spiritual dynasties. Ḥasidism spread throughout eastern Europe and enjoyed its greatest success in Poland.

Ḥasidism

Ḥasidism was notably unsuccessful, however, in Lithuania, where the traditional rabbinic class, under the leadership of Elijah, the Vilna gaon, was able to stave off its influence by issuing a ban of excommunication (*ḥerem*, "anathema") against the new movement. The tactic (a complete boycott and cutting off of communication) was widely embraced by non-Hasidic rabbis, who earned for themselves among the Hasidim the title of Mitnaggedim (Opponents), but it proved largely ineffective in areas where the rabbis had lost the respect of the masses, and it called forth a round of counter-excommunications by the Hasidic rebbes. With the passage of time, Hasidim and Mitnaggedim abandoned their conflict and came to see each other as allies against the threat to all Orthodox Jewish religion of Haskala and secularization. The impact of Ḥasidism on eastern European Jewry cannot be overemphasized; even in Lithuania, where it did not take firm hold, it stimulated the growth of a home-grown pietism in the Musar (ethicist) movement of the mid-19th century, and it renewed the Talmudic energies of its opponents.

DEVELOPMENTS IN SCHOLARSHIP

The Science of Judaism

As Jews moved into Western society in central Europe, there arose a group of young Jewish intellectuals who devoted themselves to Jewish scholarship of a far different type from traditional Talmudic learning or medieval philosophy. In 1819 Leopold Zunz and Moses Moser founded the Society for Jewish Culture and Learning for the study of Jewish history and literature. Although the original group quickly dissolved, Zunz became the unofficial leader of a generation of scholars dedicated to the Wissenschaft des Judentums (Science of Judaism). Under its carefully objective and scholarly facade, the Wissenschaft movement embraced a variety of nonacademic motives and goals. All of its members sought to prove that the Jewish past was intellectually respectable and worthy of study, and hence that the Jews deserved an equal place within European societies. Jewish scholarship was enlisted as a weapon in the battles for change. Thus, Isaac M. Jost wrote a general history of the Jews to promote Reform; Zunz's Gottesdienstliche Vortrage *der Juden, historisch* entwickelt (1832; "The Worship Sermons of the Jews, Historically Developed") served to legitimize the modern innovation of the sermon in the vernacular; and Abraham Geiger, the outstanding leader of German Reform in the 1840s and 1850s, interpreted the Pharisees as the forerunners of the reformers of his own day. In their work these intellectuals presented archetypes of what modern Jews should become. To support their claims of academic respectability, the Wissenschaft figures highlighted those aspects of the Jewish past that were closely integrated with general fields of study. In particular, Moritz Steinschneider, who owes his fame to towering achievements in bibliography, was concerned above all with the contribution of Jews to science, medicine, and mathematics. Nineteenth-century Jewish scholarship set out to praise Judaism as one of the cofounders of the Western tradition, and thus to argue that whenever the Jews were not excluded from European society they have produced great culture, and that they would repeat such accomplishments under conditions of social and political equality.

The Wissenschaft movement stimulated the critical study of the Jewish past, and great works of synthesis, written from a variety of perspectives, began to appear: Heinrich Graetz's multivolumed *Geschichte der Juden von den ältesten Zeiten bis auf die Gegenwart* (1853–75; *History of the Jews)*, written from a romantic-national point of view; Isaac Halevy's *Dorot ha-rishoizinz* (1897–*1932;* "The First Generations") and Ze'ev Jawitz's *Toldot Yisrael* (1894; "History of Israel"), from an orthodox standpoint; and Simon Dubnow's *Weltgeschichte* des *jüdischen Volkes* (1925–29; "World History of the Jewish People"), reflecting his belief in secular, nationalistic communal autonomy. Since the 1920s this tradition of great synthesis has been carried on in the United States by Salo W. Baron, who by the early 1970s had produced 14 volumes of his *Social and Religious History of the Jews* (1952), and in Israel by Benzion Dinur, the most prominent of the Zionist historians, who was still at work on his *Yisrael ba-gola* (2nd ed. 1958; "Israel in the Exile"). Many other first-rank scholars in Europe, Israel, and the United States have made notable contributions to the study of Jewish history, rabbinics, and mysticism. This great emphasis on historical research and knowledge from a wide variety of perspectives tended to propose veneration for the Jewish past as a substitute for waning religious faith.

DEVELOPMENTS IN PHILOSOPHY

Krochmal's philosophy of history

Modern Jewish philosophers have been faced with the special problem of attempting to define the reason for Jewish continuity now that the Halakha ("oral law") is no longer seen by the majority of Jewry as divinely ordained. While Moses Mendelssohn first presented the problem in *Jerusalem* with his declaration that pure reason can be taught by all religions, his own solution—that the uniqueness of Judaism was its status as revealed divine legislation—proved unsatisfactory to succeeding generations. Jewish philosophers in the period following Mendelssohn, then, had to find new rationales for Jewish uniqueness. Nachman Krochmal, whose *More nevukhe ha-zman* ("Guide for the Perplexed of Our Time") was published posthumously in 1851, posited Judaism as the bearer of the historical process. In his cyclical theory of history the Jews were the sole nation to arise anew after each period of decline, for they alone enjoyed a direct link to the Absolute Spirit. The Jews thus constituted a constant source of creativity, ever evolving to higher levels of self-realization, and their spiritual creativity became their raison d'être. For most central European Jewish philosophers of Krochmal's generation and the one following, Judaism became the bearer of the moral process. Following contemporary trends in philosophy, Solomon Formstecher, Samuel Hirsch, and Samuel Holdheim identified Judaism with what they posited to be the absolute good, ethical and moral reason. This tendency in Jewish philosophy reached its fruition in Hermann Cohen, who defined Judaism as the one true philosophy, which was, for him, Neo-Kantianism. In Cohen's work Judaism became the incarnation of ultimate moral teaching raised to the level of religious piety. However, in his last works, Cohen's personal piety, his sense of the reality of a personal God of forgiveness, contended with his philosophy of rationalism, as he attempted to find place for the God of Abraham, whom he had never really left. in the reasonable world he could not abandon. In the 20th century, Jewish philosophy has been renewed by the original contributions of Franz Rosenzweig and Martin Buber. A Neo-Orthodox existentialist, who emphasized the existing individual's relation to God, Rosenzweig conceived of Judaism as true revelation and valued its traditions as well as its ethical teachings. Yet he saw no reason to preach Judaism's superiority to Christianity; the two religions were merely different, but complementary, means of reaching the same truth. For Buber as well, Judaism was one very intense form of personal encounter with God, which is the height of religious experience. In the United States. Jewish religious philosophy has, as in Europe, proceeded into "modernity" by attempting to define a version of Judaism in terms of the dominant philosophy of the day. Perhaps the most noteworthy of American Jewish philosophers is Mordecai Menahem Kaplan, whose philosophy—reconstructionism—draws heavily upon John Dewey's Pragmatism. For Kaplan, Judaism is religious nationalism, an evolving religious civilization with a humanistic core. Like all modern religious philosophy, modern Jewish philosophy deals with the defense not merely of Judaism but of all religion. Jewish existence in particular is defended as the source of personal authenticity and creativity, the Jewish way of experiencing universal moral content.

JEWISH–CHRISTIAN RELATIONS

Christianity and Russian anti-Semitism

Christianity and the Nazi persecution

Jewish–Christian relations in the 19th century, strained at best, often erupted into open conflict. Established Christianity, and Roman Catholicism in particular, staunch upholders of the old order, identified the Jews as the major beneficiaries of the French Revolution and as the bearers of a liberal, secular, anticlerical, and often revolutionary threat. Clerical anti-Semitism was thus in France allied with the anti-Semitism of the traditional right, and these movements contended with those who affirmed the results of the French Revolution in the great convulsion of the "Dreyfus affair" in the last years of the 19th century. In Russia the conflict of the Jews and the Orthodox Church released the most open and virulent manifestation of religious anti-Semitism. To the church, the Jews were the enemy seeking to undermine Russian Orthodoxy and the tsar, the very foundations of Russian tradition. The church and the tsarist authorities went so far as to condone, and even encourage, the violent pogroms that were perpetrated against the Jews in 1881–82 and again in 1905. Russian Orthodoxy was active as well in spreading the so-called blood libel, a superstitious belief in Jewish ritual murder which had re-emerged even in the 19th century, in Damascus in 1840 (in which instance the French Consul in Syria initiated the accusation) and in Hungary in the Tiszaeszlar affair in 1882. In both cases torture was used to obtain false confessions but the accused were ultimately cleared. The most infamous recurrence of the blood libel in modern times, however, was the Beilis case of 1911–13, in which the tsarist government, with church complicity, sought unsuccessfully to convict a Jewish bookkeeper in Odessa of ritual murder. From Russian Orthodox circles, too, arose the "Protocols of the Elders of Zion," a fraudulent documentation of an alleged international Jewish conspiracy to conquer the world by subverting the social order through Liberalism, Freemasonry, and other modern movements; the concoction appeared around the turn of the century and enjoyed a phenomenal success in anti-Semitic propaganda. In spite of the fact that much of modern anti-Semitism was not Christian but racial, pagan, and often left-wing, Jews have attributed even secular anti-Semitism to older Christian teachings, which they assert persisted as a case of time lag. In the 20th century Jews and Christians have moved toward mutual understanding. In the early decades of the century, some liberal Christian voices were raised against anti-Semitism; in the United States the National Conference of Christians and Jews was founded (1928) as a response to anti-Semitism propagated in Henry Ford's *Dearborn Independent.* Elements of the Church spoke out during the 1930s against the Nazi persecution of the Jews, but the majority of Christian religious figures in Europe remained silent, even during the Holocaust (near extermination of European Jews). In response to the Holocaust, however, the World Council of Churches denounced anti-Semitism in 1946, and in 1965 the Roman Catholic Church's Schema on the Jews and other non-Christian religions, adopted by the Second Vatican Council, revised its traditional attitude toward the Jews as the killers of Christ. A growing sense of ecumenism (of fellowship and common concerns) has been shared by Jews and Christians alike. Although there remain many difficulties related to the question of the place that Zionism and the State of Israel hold within Judaism, the older forms of official church anti-Semitism have radically lessened.

ZIONISM

The most striking of the new phenomena in Jewish life is Zionism, which, insofar as it has focussed on the return to Zion (the poetic term for the Holy Land), is a re-echo of older religious themes. Insofar as it has stressed the national concentration of the Jews in a secular state, however, it is yet another example of the secularization of Jewish life and of Jewish messianism. In its secular aspects Zionism attempted to complete the emancipation of the Jews by transforming them into a nation like all other nations. Although it drew upon the general currents of 19th-century European nationalism, its major impetus came from the revival of a virulent form of racist anti-Semitism in the last decades of the 19th century, as noted above. Zionism reacted to anti-Semitic contentions that the Jews were aliens in European society and could never hope to be integrated into it in any numbers, and transformed this charge into a basic premise of a program of national regeneration and resettlement. Zionism has come to occupy roughly the same place in Jewish life as the "social gospel"—according to which the Kingdom of God is to be achieved in economic and social life—for Christians; the involvement in Israel as the new centre of Jewish energies, creativity, and renewal serves as the secular religion of many Diaspora Jews.

AMERICAN JUDAISM

The story of Judaism in the United States is the story of several fresh beginnings. In the colonial period the style of the tiny American Jewish community was shaped by the earliest Sefardic immigrants; the community was officially Orthodox but, unlike European Jewish communities, was voluntaristic, and by the early 19th century there was a significant drift of the younger generation from Judaism. By the mid-19th century a new wave of central European immigrants revived the declining American Jewish community and remade it to serve its own needs. Primarily petty shopkeepers and traders, the new immigrants migrated westward, founding new Jewish centres which were almost entirely controlled by laymen. The exigencies of life on the frontier within an open society created a predisposition for religious reform, and it is significant that the greatest American Reform Jewish leader of the 19th century, Isaac Mayer Wise, was based in Cincinnati, Ohio. Wise sought to unite all of American Jewry in the new nontraditional institutions that he founded, Hebrew Union College (1875), the Union of American Hebrew Congregations (1873), and the Central Conference of American Rabbis (1889); but his ever more radical reforming spirit ultimately drove the traditionalist elements within the American Jewish community into opposition. The head of the traditionalist circles had been Isaac Leeser, a native of Germany, who had attempted to create an indigenous American community on the lines of a modernized traditionalism. Conservative forces, after his death, became disorganized, but in reaction to Reform they defined themselves by their attachment to the sabbath, the dietary laws, and especially to Hebrew as the language of prayer. Under the leadership of Sabato Morais, an Orthodox Jew of Italian birth, Conservative circles in 1886 founded a rabbinic seminary of their own, the Jewish Theological Seminary of America.

The central and eastern European immigrant styles of Judaism

The eastern European immigrants who moved in large numbers to the American shores in the years from 1881 to 1914 were profoundly different in culture and manners from the older elements of the American Jewish community, and it is they and their descendants who made American Judaism as it is today. The bridge between the existing Jewish community led by German Jews of Reform persuasion and the new immigrant masses was the traditionalist element among the older settlers. Cyrus Adler, traditionalist himself, cooperated with a German Reform circle of Jacob Schiff in reorganizing the Jewish Theological Seminary (1902) and other institutions for the purpose of Americanizing the eastern European immigrants. Enough eastern European rabbis and scholars had immigrated, however, to create their own synagogues, which reproduced the customs of the old world. Whereas in 1880 almost all of the 200 Jewish congregations in the United States were Reform, by 1890 there were 533 synagogues, and most of the new ones founded by immigrant groups were Orthodox. The Union of Orthodox Jewish Congregations, which was established in 1898 by elements associated with the Jewish Theological Seminary, was soon taken over by Yiddish-speaking recent immigrants for whom the seminary was much too liberal. In 1902 immigrant rabbis also formed their own body, the Union of Orthodox Rabbis of the United States and Canada (the Agudath ha-Rabbanim), which fostered the creation of yeshivas (rabbinic academies) of the old type. In 1915 two small yeshivas, Etz Chaim and Rabbi Isaac Elhanan Theological Seminary, added Yeshiva College of secular studies in 1928, and became Yeshiva University in 1945. The eastern European Orthodox elements concentrated primarily on Jewish education and it was they who introduced the movement for Jewish day schools, analogous to Christian parochial schools. Gradually an American version of Orthodoxy developed on the Neo-Orthodox model of Samson Raphael Hirsch, which combined institutional separatism and cooperation with other Jewish groups in umbrella organizations.

American Orthodoxy

The immigrants and their children had three desires; to upgrade themselves socially by joining older congregations or forming their own in an Americanized image; to affirm an unideological commitment to Jewish life; and to maintain their ties to the overseas Jewish communities of their origin. With their strong sense of Jewish peoplehood, they introduced Zionism into American Jewish life, and accepted the basic ideas of Mordecai Kaplan's Reconstructionism, which was committed to Zionism. A small group of anti-Zionists remained a significant force in the 1930s and 1940s, but their central organization, the American Council for Judaism, represented the descendants of earlier German-Jewish immigrants. The later immigrants took over all the earlier institutions of the Jewish community and imbued them with their own spirit.

The religious continuum

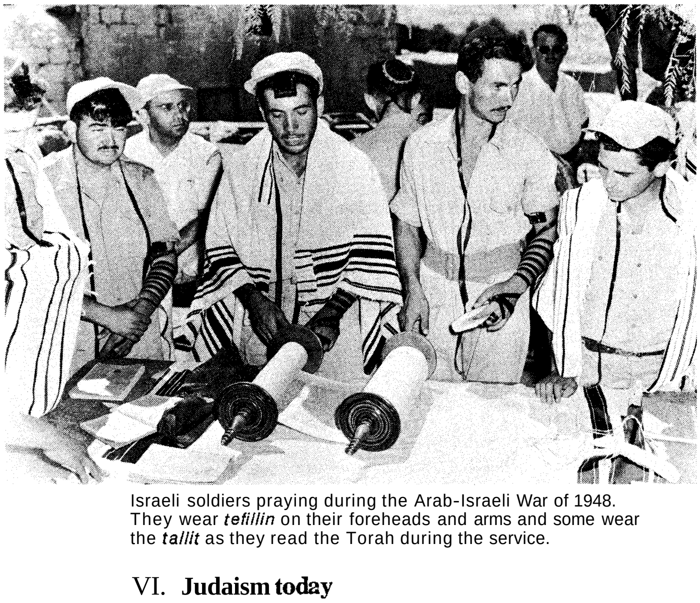
American Jewish religious life is a continuum—from the most traditional Orthodoxy to the most radical Reconstructionism. In theory, all of the Orthodox groups agree on the revealed nature of all of Jewish law; for the Reform groups, the moral doctrine of Judaism is divine and its ritual law is man made; the Conservatives see Judaism as the working out in both areas of a divine revelation that is incarnate in a slowly changing human history; and the Reconstructionists (who include both Conservative and Reform Jews) view Judaism as the evolving civilization created by the Jewish people in the light of its highest conscience. What really marks the various bodies in the mind of the Jewish community is their difference in ritual practices, but the ritual variations shade from one group into the other. The role of the rabbi is substantially the same in all three groups; he is no longer a Talmudic scholar but a preacher, pastor, and administrator, a cross between a parish priest and the leader of an ethnic group. Although there was some cooperation among the three major Jewish denominations—Orthodoxy, Reform, and Conservatism—the real effort of organized Jewish religion in America in the early 1970s revolved around the individual synagogue and the denomination to which it belonged. As religious identification became increasingly respectable in American life, the Jews followed the American norm, affiliating in greater numbers with synagogues, though often for ethnic or social, rather than religious reasons.

JUDAISM IN OTHER LANDS

Modernity came first to the Jewish people in Europe and it was, therefore, within the European context that representatives of important non-Ashkenazi communities such as the proto-Zionist Sefardi Judah ben Solomon Hai Alkalai of Sarajevo, the Luzzatto family in Italy, and Elijah ben Abraham Benamozegh in France, participated in variations of Jewish modernity. In England and France, more than in Germany or Russia, Wissenschaft des Judentums (see above Developments *in* scholarship), with its enlightenment ideology, was the central focus of Jewish experience; there the "republic of scholarship" became the synagogue of the Jewish intelligentsia. In neither country did Reform Judaism gain a major foothold, for the Orthodox establishment, which remained the official synagogue, liberalized its synagogue practice while retaining its essentially conservative outlook. In Anglo-Jewish life in the last decades of the 19th century the two most pronounced modernist tendencies were Solomon Schechter's moderate romantic traditionalism and the "renewed Karaism" of Claude Joseph Goldsmid Montefiore, whose version of religious reform was "back to the Bible."

Outside of Europe, in such places as South America and Canada, Jewish modernity appeared late, for European Jewry arrived in those places even later than in the United States, attaining a significant number only in the 20th century. These communities have been dependent on immigrant scholars and intellectuals for serious Jewish thought. Jews in the Arab lands, in North Africa and the Middle East, living within traditional societies, entered modernity even later than those on the peripheries of Europe. Many of them received their first introduction to the Western world in widespread schools set up by the Alliance Israélite Universelle (a Jewish defense organization centred in Paris), which combined Jewish education with the language and values of French civilization. Yet most of these communities remained traditionalist almost to the moment when they were expelled or felt compelled to relocate, since 1948, when the state of Israel was created. The ferment of modernity in all its forms is now being felt in their ranks. In Israel, which has received a large segment of Sefardic Jewry, the attention of these communities has turned to attaining equality with the more advanced Ashkenazim rather than developing some forms of modern Jewish thought.

Wide World Photos



Israeli soldiers praying during the Arab-Israeli War of 1948. They wear *tefillin* on their foreheads and arms and some wear the *tallit* as they read the Torah during the service.

VI. **Judaism today**

As a result of the Holocaust, Judaism has become a non-European religion; its three major centres, which together include more than three-fourths of world Jewry, are Israel, the Soviet Union, and the United States. Although Jews constitute only a small fraction of the population of the United States, Judaism occupies a role far surpassing its numerical importance and is regarded with Roman Catholicism and Protestantism as one of the major American faiths. Similarly, in the international realm of Western religion, Judaism has been welcomed as a partner able to deal with other major religions as an equal on such issues as anti-Semitism, human rights, and world peace.

The rights and needs of the world Jewish community, including Israel, have triggered deep conflicts with which Judaism has been involved with the Arab and Communist worlds. Friction between Israel and the Arab states has created tension with Islām, while the political stand of Israel and the treatment of the Jewish minority within the Soviet Union have led to open clashes with the Communist leadership. Some of the diatribes and charges that have issued from the Arabs and Communists in this struggle have at times re-evoked older forms of anti-Semitism. In the long-range view, the problems of Judaism and Islām seem more soluble than those of Judaism and such secular ideologies as Communism, for the major religions of the world are increasingly seeking accommodation with each other, as all are confronted by hostile secularist ideologies which have retained their conversionary élan.

Secularization and Jewish identity

Within its own community Jewry is faced with the increasing secularization of Jewish identity in its three major centres, each in its own way. In the United States the open society and the melting pot ideologies of past generations have fostered among many Jews a sense of Jewish identity increasingly devoid of concrete religious, national, or historical content; in the Soviet Union government policy since the 1930s has banned the teaching of Judaism and Jewish culture to the young and has severely discouraged any manifestation of Jewish identity as a sign of the disloyalty of "rootless cosmopolitans" to the Russian state; and in Israel a secular nationalism has taken root, raising questions as to the role that Judaism plays in the identity of the average Israeli. Nonetheless, underneath the external secularization there are signs of a persisting deep Jewish religious fervour, in which the sense of history, community, and personal authenticity figure as the intertwined strands of Jewish religious life, especially as it has been affected by the State of Israel. Some of the rituals of the Jewish tradition, especially the rites of passage at the crucial stages of individual existence, are almost universally observed, except in the Soviet Union; in the United States, for example, more than 80 percent of Jewish children receive some formal religious training. Among Jewish youth there is, in some circles, a recurring quest for tradition. In the United States, Jewish communes have been established that seek new forms of Jewish expression; in Israel, groups such as Mevaqshe Derekh (Seekers of the Way) have tried to bridge secular Israeli culture and Jewish tradition and to maintain traditional Jewish ethical standards of behaviour even in wartime; in the Soviet Union thousands of young people gather on several occasions of the year to dance and sing and express solidarity in the squares in front of the synagogues in Leningrad and Moscow. Still, signs of major weaknesses persist. The rate of intermarriage among Jews in the Diaspora increased, while regular synagogue attendance, at the very highest 20 percent in the United States, remained far below church attendance. Despite their lack of traditional piety, there is a general sense among Jews that they remain Jews not because of the force of anti-Semitism but because of the attractiveness of their own tradition and their sense of a common history and destiny.

In the largest Jewish community, that of the United States, of the close to 6,000,000 Jews, three-fifths are members of synagogues. In the Soviet Union, which contains the second largest community (about 3,000,000), despite decades of assimilation and cultural deprivation, an estimated 100,000 Jews persisted in pursuing Jewish studies and 1,000,000 Jews still listed Yiddish as their mother tongue. Amid growing anti-Semitism and a Soviet-Arab entente, a movement emerged in the early 1970s affirming Jewish identity and pressing, at great personal risk, for emigration to Israel. In Israel, where there is a Jewish population of about 2,500,000, 40 percent of the children attend schools of Orthodox religious orientation, though politically the strength of the (Orthodox) religious parties is not high, and non-Orthodox government schools also stress Jewish subjects. The 800,000 Jews of Central and South America express their sense of Jewish identity in secular and ethnic rather than religious terms, and have increasingly turned from Yiddish to Spanish. In western Europe the picture is mixed. Only Great Britain and France retain large Jewish communities, approximately 500,000 each. In England the community suffers from a declining birth rate, while the French community has been revivified by the recent influx of North African Jews in the 1950s and 1960s.

The enduring existence of the Jews and Judaism

If in 1945, the world Jewish community, decimated and horrified by the Holocaust, felt in danger of disappearing, there appeared to be no such despair in the early 1970s when there was an expectation that Jewish communal feeling would remain strong, especially, for many or most Jews, in the light of the existence of Israel. Judaism enjoyed a heightened dignity in the eyes of the world as well, not only as a result of the creation of the State of Israel, but also because of its close relations with other world religions. Although the recurring phenomenon of the alienation of young Jews from their tradition was a troubling prospect, it is no more so than in recent past generations. Along with other major religions, Judaism's most disturbing problem yet to be solved was how to deal with secular ideologies and with the growth of secularism within its own ranks. Thus, looking forward to the last decades of the 20th century, it appeared that Judaism would have to contend with as many problems as other major religions, but that it faced them with no less confidence than these, and with more confidence than it had felt earlier in the century. (Ar.H.)

BIBLIOGRAPHY

General observations: SALO W. BARON, *A Social and Religious History of the Jews,* 2nd ed., *15* vol. (1952–73), a comprehensive presentation of the intertwined social and religious history with copious bibliographical information critically evaluated; LOUIS FINKELSTEIN (ed.), *The Jews: Their. History, Culture and Religion,* 4th ed., *3* vol. (1970–71), critical essays by outstanding authorities on the major aspects of Jewish history and culture; JULIUS GUTTMANN, *Die Philosophic des Judentums (1933;* Eng. trans., *Philosophies of Judaism,* 1964), the best single volume on the history of Jewish thought from ancient times to the present, especially valuable for medieval Jewish philosophy; LEO W. SCHWARZ (ed.), *Great Ages and Ideas of the Jewish People* (1956), interpretive and highly readable essays by six historians on Jewish history, with emphasis on intellectual history, intended primarily for the layman; GERSHOM G. SCHOLEM, *Major Trends in Jewish Mysticism,* 3rd rev. ed. (1954), the classic treatment of mystical doctrines and schools in Judaism; MAX L. MARGOLIS and ALEXANDER MARX, *A History of the Jewish People (1927,* reprinted 1958), an excellent, readable, introductory survey.

Biblical Judaism: (General reference): The Interpreter's Dictionary of the Bible, 4 vol. *(1962). (Surveys of the culture and religion of ancient Israel):* JOHANNES PEDERSEN, *Israel: Its Life and Culture, 4* vol. in *2 (1926–40,* reprinted *1959);* SALO W. BARON, *A Social and Religions History of the Jews,* vol. *1, Ancient Times,* pt. *1 (1952);* W.F. ALBRIGHT, *From the Stone Age to Christianity,* 2nd ed. *(1957);* YEHEZKEL KAUFMANN, *The Religion of Israel, from Its Beginnings to the Babylonian Exile (1960)* and *The Babylonian Captivity and Deutero-Isaiah (1970);* ROLAND DE VAUX, *Les Institutions de l'Ancien Testament, 2* vol. *(1958–60;* Eng. trans., *Ancient Israel: Its Life and Institutions, 1961);* GERHARD VON RAD, *Theologie des Alten Testaments,* 2nd ed. *(1958;* Eng. trans., *Old Testament Theology, 2* vol., *1962–65);* HELMER RINGGREN, *Israelitische Religion (1963;* Eng. trans., *1966). (Annual bibliographic keys): Elenchus Bibliographicus Biblicus,* ed. by P. NOBER *(1920–); Book List* of the British Society for Old Testament Study. *(Special topics):* E.A. SPEISER, "The Biblical Idea of History in Its Common Near Eastern Setting," *Oriental and Biblical Studies: Collected Writings of E.A. Speiser,* ed. by J.J. FINKELSTEIN and MOSHE GREENBERG, pp. *187–210 (1967);* MENAHEM HARAN, "The Religion of the Patriarchs: Beliefs and Practices," *Patriarchs: The World History of the Jewish People,* vol. *2,* pp. *219–245 (1970);* MOSHE GREENBERG, "Crimes and Punishments," *Interpreter's Dictionary of the Bible,* vol. *1,* pp. *733–744 (1962);* "Some Postulates of Biblical Criminal Law," *The Jewish Expression,* ed. by JUDAH GOLDIN, pp. *18–37 (1970);* AELRED CODY, *A History of Old Testament Priesthood (1969);* JOHANNES LINDBLOM, *Prophecy in Ancient Israel (1962);* E.J. BICKERMAN, "The Historical Foundations of Postbiblical Judaism," *The Jews: Their History, Culture, and Religion,* ed. by LOUIS FINKELSTEIN, 4th ed., vol. 1, pp. 70–*114 (1960).*

Hellenistic Judaism: (Bibliographies): GERHARD DELLING (ed.), *Bibliographie zur jüdisch-hellenistischen und intertestamentarischen Literatur 1900–1965* (1969), extremely comprehensive bibliography on religion and literature of Diaspora Judaism, arranged according to topics, with separate bibliographies on every major Hellenistic Jewish author and on each book of the Apocrypha and Pseudepigrapha; RALPH MARCUS, "A Selected Bibliography *(1920–1945)* of the Jews in the Hellenistic-Roman Period," *Proceedings of the American Academy for Jewish Research,* 16:97–181 (1946–47), covers both Palestine and the Diaspora, helpful for noting works that are useful introductions and works that are indispensable to the specialist; LOUIS H. FELDMAN, *Scholarship on Philo and Josephus, 1937–1962* (1963), critical bibliography, arranged topically, with comments on both Diaspora and Palestinian Judaism generally. *(Papyrological and archaeological sourcebooks):* VICTOR A. TCHERIKOVER, ALEXANDER FUKS, and MENAHEM STERN (eds.), *Corpus Papyrorum Judaicarum, 3* vol. (1957–64), contains text, translation, bibliography, and commentary on all papyri and inscriptions pertaining to Jews from *323* BCE to *641* CE—thoroughly reliable; ERWIN R. GOODENOUGH, *Jewish Symbols in the Greco-Roman Period, 13* vol. (1953–68), a magnificent, exhaustive collection of the archaeological findings, with highly insightful, if controversial, commentary. *(Standard scholarly treatments of Hellenistic Judaism):* SALO W. BARON, *A Social and Religious History of the Jews,* 2nd ed. rev., vol. 1 and 2 *(1952).* VICTOR A. TCHERIKOVER, *Hellenistic Civilization and the Jews (1959;* orig. pub. in Hebrew, 1930), extremely meticulous and generally balanced (though with an anti-theological bias), particularly in dealing with the political and social factors in both Palestinian and Diaspora Jewry; ROBERT H. PFEIFFER, *History of New Testament Times, with an Introduction to the Apocrypha* (1949), a sane and useful, if unoriginal, survey of the political, religious, and literary history of Palestinian and especially Diaspora Judaism, from 200 BCE to 200 CE; MOSES HADAS, *Hellenistic Culture: Fusion and Diffusion* (1959), highly suggestive, though often extravagant, treatment of the interaction of Hellenism and other cultures, especially Judaism. *(Introductory popular treatments of Hellenistic Judaism):* RALPH MARCUS, "The Hellenistic Age," in LEO W. SCHWARZ (ed.), *Great Ages and Ideas of the Jewish People,* pp. 93–*139* (1956), extremely judicious and readable account by an eminent authority; for a readable guide to the literature, together with representative samples, see his "Hellenistic Jewish Literature," in LOUIS FINKELSTEIN (ed.), *The Jews: Their History, Culture and Religion,* 3rd. ed., vol. *2,* pp. *1077–1115 (1960). (Works on Palestinian Judaism):* GEORGE FOOT MOORE, *Judaism in the First Centuries of the Christian Era, 3* vol. *(1927–30,* reprinted 1966–67), a classic treatment based primarily on the Talmudic corpus, though the view of a Pharisaic "normative" Judaism has since been strongly challenged; SOLOMON ZEITLIN, *The Rise and Fall of the Judaean State: A Political, Social and Religious History of the Second Commonwealth, 2* vol. (1962–67), stimulating and often highly original survey of the period from *332* BCE to *66* CE, though the scholarly dogmatism is occasionally jarring; SAUL LIEBERMAN, *Greek in Jewish Palestine,* 2nd ed. *(1965)* and *Hellenism in Jewish Palestine,* 2nd ed. (1962), highly significant, ingenious, and learned illustrations of the influence of Greek culture on the language and exegetical format of the Palestinian rabbis. *(Works on Diaspora Judaism):* HARRY A. WOLFSON, *Philo: Foundations of Religious Philosophy in Judaism, Christianity, and Islam,* rev. ed., *2* vol. (1962), a great and seminal, though controversial, work that makes Alexandrian Judaism a collateral branch of Palestinian Pharisaic Judaism; LOUIS H. FELDMAN, "The Orthodoxy of the Jews in Hellenistic Egypt," *Jewish Social Studies,* 22:215–*237* (1960), a survey using literature, papyri, and art objects to examine the synthesis of Greek culture and Judaism in the upper and lower classes, respectively, of Hellenistic Egypt.

Rabbinic Judaism: (Palestinian Judaism): GEORGE FOOT MOORE, *Judaism in the First Centuries of the Christian Era, 3* vol. *(1927–30);* SOLOMON SCHECHTER, *Aspects of Rabbinic Theology* (1961), a concise, authoritative, and engaging treatment of classical (*i.e.,* rabbinic) Judaism. *(Babylonian Judaism):* JACOB NEUSNER, *A History of the Jews in Babylonia, 5* vol. *(1965–70;* vol. *1,* rev. 1969), the most comprehensive treatment of Babylonian Jewry during the Tannaitic and Amoraic periods. *(Judeo-Arabic culture):* S.D. GOITEIN, *Jews and Arabs: Their Contacts Through the Ages (1955* and 1964), a popular work by the ranking authority on all aspects of Jewish-Arabic symbiosis, particularly valuable for the medieval period; ABRAHAM IBN DAUD, *Sefer ha-Qabbalah (The Book of Tradition),* ed. and trans. by GERSON D. COHEN (1967), the classic medieval Hebrew chronicle with analytic essays on Spanish Jewry's "golden age." *(Jews of medieval*

Europe): ISRAEL ABRAHAMS, *Jewish Life in the Middle Ages*, new ed. rev. by CECIL ROTH (1932), delightful and erudite studies of social life and institutions in medieval Europe; MORITZ GUEDEMANN, *Geschichte des Erziehungswesens und der Cultur der Juden*, 3 vol. (1880–88), a social and intellectual history of medieval Ashkenazic Jewry (France, Germany, Italy); CECIL ROTH, *The Jews in the Renaissance* (1959), lucid and informative, but with little critical analysis, valuable on Jewish contact with Christian men of letters; SOLOMON SCHECHTER, "Safed in the Sixteenth Century," *Studies in Judaism*, 2nd Series, pp. 202–285 (1908), a scholarly *tour de force* on the great community of Jewish jurists and mystics; GERSHOM SCHOLEM, *Shabbethai Zevi*, 2 vol. (1957), a penetrating and comprehensive study (in Hebrew) of the great false Messiah as well as of his religious antecedents and legacy.

Contemporary Judaism: The most convenient summary for the study of modern Jewish history is HOWARD MORLEY SACHAR, *The Course of Modern Jewish History* (1958). Modern Jewish thought and movements are covered in a useful manual by JOSEPH BLAU, *Modern Varieties of Judaism* (1966). NATHAN ROTENSTREICH, *Jewish Philosophy in Modern Times* (1968); and the last (modern) section of JULIUS GUTTMAN, *Philosophies of Judaism* (1964), are more advanced. The very best book on Jewish religion in the modern age is unfortunately still untranslated: MAX WIENER, *Jüdische Religion im Zeitalter der Emanzipation* (1933). For Zionism, see the only attempt at a comprehensive reader in English, ARTHUR HERTZBERG (ed.), *The Zionist Idea* (1959). There are two excellent expositions of Judaism from a Reform–Liberal point of view: LEON ROTH, *Judaism: A Portrait* (1960); and LEO BAECK, *The Essence of Judaism* (1961). Conservative Judaism is well described in a book about its early history, MOSHE DAVIS, *The Emergence of Conservative Judaism* (1963); and by JACOB AGUS, *Dialogue and Tradition* (1969), the essays of a distinguished Conservative thinker. Reconstructionism is best understood in the words of its founder, MORDECAI M. KAPLAN, *Judaism As a Civilization*, 2nd ed. (1957). The standard modern single volume about Orthodox Judaism is ISIDORE EPSTEIN, *Judaism* (1935). Neo-Ḥasidism was best described by its greatest name, MARTIN BUBER, *The Origin and Meaning of Hasidism* (1960).

(S.W.B./L.H.F./G.D.C./Ar.H./Mo.Gr.)

Juglandales

This is a small but important order of flowering plants consisting of two families, Rhoipteleaceae, with one genus and one species, and Juglandaceae, the walnut family, with eight genera and about 57 species. The walnuts are common large forest trees, primarily of temperate areas but also of subtropical regions, mostly of eastern North America, Central America, western South America and eastern Asia. Walnuts, hickories, and pecans have been appreciated for many years for their edible oily nuts and their valuable wood, and they are now cultivated in large numbers in most temperate regions of the world.

Some of the trees, at least in the early days before extensive cutting for lumber, grew to be 150 feet or even 175 feet high (46–53 metres). The black walnut, *Juglans nigra*, has been reported to reach six feet (two metres) in diameter. The smallest member of the order is the Texas walnut, *J. microcarpa*, a shrubby tree of Texas and New Mexico only 20 to 30 feet (six to nine metres) high.

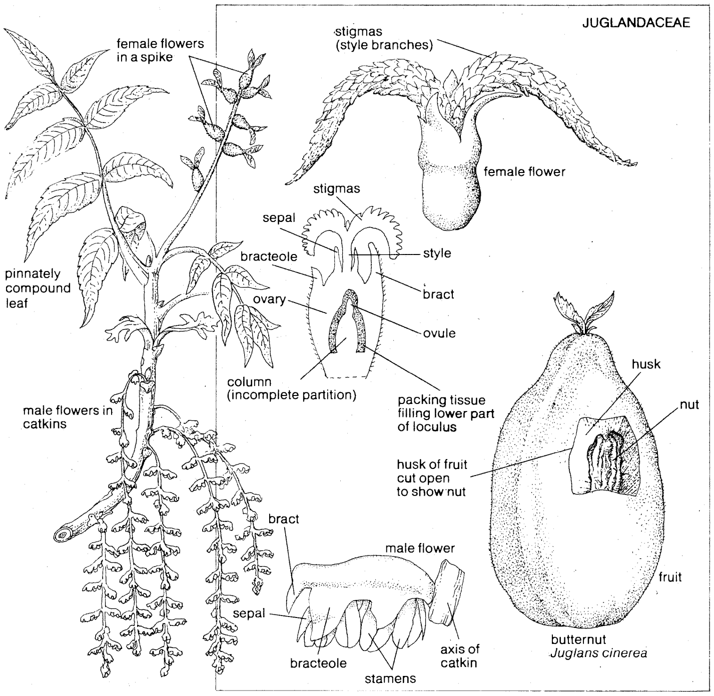
GENERAL FEATURES

Distribution and abundance. The trees are common throughout the deciduous forests of eastern United States and in parts of eastern Asia. In Mexico and Central America they are only locally common in certain canyons in the mountains or in fog belts on the slopes of the mountains. One species, *Oreomunnea (Engelkardia) pterocarpa*, occurs in only one valley in Costa Rica. The black walnut, *Juglans nigra*, yields such valuable timber that most of the native trees in the eastern United States have been cut down; hickories (*Carya* species), too, have been greatly reduced in numbers. The fall webworm attacks the hickory and walnut leaves in many regions, and insect damage also seems to be reducing the numbers of the butternut (*Juglans cinerea*) in the northeast. In the United States and in Europe many walnut and hickory trees have become naturalized in the regions where they have been planted.

Uses of the black walnut

Economic importance. The most important and best known member of the order is undoubtedly *Juglans regia*, the Persian walnut, also known as the English walnut and Circassian walnut. This was originally native to China, the Himalayan mountains, and from Iran to Greece. The wood and nuts have been used and prized for over 2,000 years; the Persian walnut is probably the most important of all nuts. Orchards and individual trees have been planted in warm temperate areas in many countries. The pecan, *Carya illinoensis*, native of the southeastern United States and northern Mexico, though not as valuable for its wood, is important, next to the Persian walnut, for its valuable nuts; more and more orchards are being planted not only in the southern United States but also in other countries. The black walnut, *Juglans nigra*, native in the eastern United States, is well known for its valuable wood used in fine furniture and in cabinetwork, interior panelling, and gun stocks. The tree has become so scarce as a native tree that lumbermen have been known to try to buy yard trees from private homes. Even single trees can be extremely valuable because the wood can be extended considerably if used as thin veneers. Black walnut meats are very tasty and are used in candy and cakes, but the hard-shelled nuts are difficult to crack. Important primarily locally for their nuts are the American butternut, *Juglans cinerea;* Japanese butternut, *J. ailantifolia;* shagbark hickory, *Carya ovata;* and the big shellbark hickory, *C. laciniosa*. The wood of the hickories is extremely valuable because of its toughness, hardness, strength, and resistance to shock and is used for tool handles, wagon axles, and agricultural implements. The fresh husks of the butternut have been used as a dye (khaki); the fruits of *Platycarya* have been used in China as a black dye for cotton cloth. The meats of the Persian walnut have been used for a drying oil. Several members of the temperate genera have been planted for dooryard trees or park trees. The Japanese butternut, a rapidly growing tree, is very hardy and is increasingly being planted in the United States; the nuts of the heartnut variety are easily cracked. The woods of the Asiatic members other than walnuts and hickories have local uses but none of these trees has edible nuts. Leaves of various species have been thrown into the water for use as fish poisons by primitive people. A toxic substance, juglone, is given off by the leaves and roots of the black walnut and under certain conditions injures vegetable crops growing near the trees,

From (left) *American Journal of Botany*, vol. 25 (1938) and (female flower, butternut) *American Journal of Botany*, vol. 27 (1940)



Vegetative, floral, and fruiting structures of the walnut order.

Natural history. The trees are adapted to wind pollination; thus the stamens or staminate (male) flowers are very numerous and the stigmas (female flower structures) are enlarged to catch the pollen. Adaptations for seed dissemination have developed in two different directions. All members of the order have one-seeded fruits and the seed remains inside the fruit for dissemination. Some genera have developed small fruits with wings for wind dispersal; others have developed large fruits, or nuts, for animal, primarily squirrel, dissemination. The seeds are typically viable for only one winter season in the northern regions and must not dry out; squirrels plant them underground, where it is moist. The nuts have an adherent husk for protection, but it is not known whether its principal function is to prevent drying out or injury in falling from the trees.

Many interesting natural hybrids are known among the hickories and the walnuts, but most of these are not used commercially. Some of the species of Carya (hickories) are diploids—*i.e.*, they have the normal two sets of chromosomes per cell; some are tetraploids, with four sets of chromosomes in the nucleus of each cell.

Ecology. Many of the trees in the walnut order grow in the forests of the temperate regions; others grow in the mountains of the tropics; a few are subtropical or even grow near the seashore in the tropics, but none of them is strictly alpine. Most of the members have deciduous leaves, even in the tropics, but a few are evergreen. All appear to require moisture in the soil. A few grow in swamps, some in rich bottomlands; others prefer dry hillsides, but only where rainfall is sufficient; a few grow in arid regions, but then only along streams in canyons or in the fog belts in the mountains. Members of the order rarely occur as solid forests but are mixed with other trees, especially oaks. An oak–hickory forest covers much of the deciduous forest of eastern United States and part of eastern Mexico. *Engelhardia* is occasionally dominant in eastern Asia.

Scattered occurrence of trees

Distinguishing features. All members of the order are trees. The alternate or, rarely, opposite leaves are pinnately compound (*i.e.*, each leaf is composed of several leaflets arranged along both sides of a central axis). The lower surfaces of the leaflets, and often the buds, flowers, and young fruits, are dotted with yellow or sometimes pale several-celled resinous scales. The inconspicuous flowers lack petals, and at least some occur in catkins or elongate slender spikes. Catkins are long and scaly, and usually drooping, spikes of unisexual flowers; that is, the male or staminate flowers occur in one catkin, the female or pistillate flowers in another. The ovary is two-celled or incompletely two-celled (that is, with one or two ovarian cavities or locules), but there is only one ovule. The one-seeded fruits are either winged nutlets or nuts, which are enclosed in husks. The seeds are oily and lack endosperm, a starchy nutrient tissue for the developing embryo.

EVOLUTION

Fossil record. There is no fossil record for the rare Chinese and Indochinese family, Rhoipteleaceae, but the walnut family, Juglandaceae, is abundantly represented in the fossil record in the form of leaves, nuts, and pollen grains. Members of this family appear first in the Upper Cretaceous Period (about 75,000,000 years ago), along with the earliest flowering plants, and by then had already evolved into essentially modern genera. This indicates a long previous history, which is also true of many other flowering plants. As early as the Eocene Epoch (about 50,000,000 years ago) the walnut family had spread over much of the Northern Hemisphere, with a much wider distribution than in modern times, and continued to spread through the rest of the Tertiary Period. Trees, therefore, occurred in regions where modern members are no longer native, such as Europe, North Africa, Alaska, Greenland, and the northwestern United States. Thus, in north central Oregon, in the Clarno Formation of the Eocene, of about 50,000,000 years ago, 75 recently discovered petrified fruits strongly resemble the living Arizona walnut, *Juglans major*. The species is, however, much more primitive than any modern one and apparently is intermediate between modern species. Furthermore, some of the genera now restricted to eastern Asia, such as Pterocarya, Platycarya, and Engelhardia, once were to be found over much of the United States as well as Europe.

Restriction of distribution during recent geologic time

The range has since become much restricted, due to a general cooling of the climate in the north and to mountain building, resulting in present-day drier conditions. Furthermore, glacial activities in the Pleistocene Epoch of the last 2,000,000 years completely wiped out the trees in Europe, where they could not survive south of the glaciers because of the extensive east-to-west oriented European mountains that blocked their southward range extensions as the glaciers advanced. In eastern United States and in China the mountains run largely from north to south, and during periods of approaching glaciation the trees were able to move south by means of seed dispersal along the valleys, encountering no physical barriers, and move back north after the glaciers had retreated.

Phylogeny. Juglandales is one order of a group formerly known as the Amentiferae. All members of this group are woody plants, with inconspicuous wind-pollinated flowers that lack petals and are arranged in elongated clusters known as catkins (aments) or spikes, such as are found in the willows, bayberries, birches, oaks, and beeches. The catkins superficially resemble the cones of the gymnosperms (pines, spruces, firs, etc.). Some botanists believe that all of these orders are related, are very primitive, and have evolved directly from ancient gymnosperms, perhaps even the seed ferns. This belief is due partly to the resemblance of catkins to cones, and especially to the separation on the same tree of the male catkins from the female ones; partly to wind pollination and lack of petals as in the gymnosperms; and partly to the early appearance of the plants in the fossil record. While most botanists no longer believe that the Amentiferae group gave rise to more advanced flowering plants, some still believe that they are primitive and possibly had a different origin and line of evolution from the rest of the flowering plants. One European botanist, for example, has found features about the ovule of Engelhardia, a primitive genus of the family Juglandaceae, to be different from other angiosperms and highly reminiscent of certain fossil seed ferns.

Evolutionary status of walnut order and its relatives

Most present-day botanists, however, now believe that the families of the Amentiferae are not primitive but advanced, having had ancestors with conspicuous, insect-pollinated flowers that had petals and bisexual flowers (with male stamens and female pistil together in the same flower). Thus, the presence of an abortive female pistil in a few male flowers in various species of the Juglandaceae family indicates that the ancestral flowers were bisexual. In addition, the compound pistil with inferior ovary (*i.e.*, located below the attachment point of the other flower parts) in several families of the Amentiferae group suggests an advanced rather than a primitive nature. The ancestral bisexual flowers have been reduced, the sexes separated, the petals lost, and the flowers crowded into catkins with specialization for wind pollination. Furthermore, the catkin and the cone are technically not equivalent; each flower of a catkin (for example, a staminate flower) is equivalent to a complete cone (for example, a staminate cone) in the gymnosperms. Thus, theoretically, a catkin could be considered equivalent to a large number of massed cones. There is also a difference of opinion among present-day botanists as to the nature and evolution of the primitive flower itself and as to the type of plant, and its reproductive structure, from which the angiosperms arose. Since truly primitive angiosperms have not yet been located in the decidedlv incomplete fossil record, the presence of early fossils of the Amentiferae group does not necessarily prove them to be the most primitive of flowering plants.

The Juglandales order itself, in turn, has stood apart from the other orders of the Amentiferae group because of its compound leaves, and the Juglandaceae family itself lacks the stipules (small, leafiike appendages at the base of leaf stalks) present in the other orders.

It has often been suggested that the Juglandaceae family has been derived from the Sapindales order, which includes the sumacs. These are woody plants with pinnately compound leaves also lacking stipules, some of them having resinous dots on the leaflets as do the members of the Juglandaceae family. They also have panicles of flowers but, unlike the walnut family, petals are present. A few members even have reduced crowded flowers arranged in spikes. The wood anatomy and pollen structure, however, indicate that not only is the Juglandaceae family unrelated to the Sapindales order, but it is even more primitive than the latter in several features. Thus, the walnut family was probably derived from some other more primitive group. It now seems likely, though not proved, that the family was derived with the bayberries, oaks, and birches through the Hamamelidales, the witch hazel order, from the Magnoliales, the magnolia order. It is generally agreed that this last order is the most primitive living group of flowering plants, having, as it does, extremely primitive gymnosperm-like wood structure as well as other primitive features.

Relatives of the walnut family

There have been differences of opinion about which families are related to the walnut family. The families Julianiaceae, Picrodendraceae, and Rhoipteleaceae have been considered to be close to the Juglandaceae family because of their compound leaves and general flower and inflorescence (*i.e.*, the flower cluster) structure. Studies on wood anatomy and pollen grain structure have shown that the first two are much more likely to be related to the Sapindales order. The stipules, flowers, and flattened winged fruits of the Rhoipteleaceae family strongly suggest the elms in the order Urticales, though the latter have simple leaves. Again, wood anatomy indicates that the Rhoipteleaceae family is more closely related to the walnut family, especially to the more primitive genera of that family, *Engelhardia* and *Alfaroa*. The flower and inflorescence structure of the genus *Rhoiptelea* fit exactly the proposed hypothetical ancestor of the Juglandaceae family. In the Amentiferae group the bayberry family, Myricaceae, considered by many to constitute a distinct order, has been placed by some in the order Juglandales. This family has catkins, and one-seeded nutlets that are often winged. Once more the bayberries have simple leaves and stipules, but they could still be close to the Juglandales order. The compound leaves of the Juglandaceae (walnut) family, however, separate it from the oaks and birches and even from witch hazels and magnolias.

CLASSIFICATION

Distinguishing taxonomic features. The two families of the order are quite distinct in position of the ovary (superior or inferior) and internal structure of the ovary (two-celled or one-celled; that is, two-loculed or one-loculed), and in having bisexual (perfect) flowers in long slender spikes or unisexual (imperfect) flowers in true catkins. Within the Juglandaceae family itself the subfamilies, tribes, and genera are separated on the basis of the modification of the primary and secondary bracts (scales) in the fruiting condition into wings or into husks for the nutlets or nuts. In addition, the presence or absence of sepals (the outer whorl of flower structures, similar to flower petals but usually smaller and coloured green), type and position of inflorescences, variation in position of the carpels (ovule-bearing leaves of the pistil), type of stigma, wood anatomy, and pollen structure are of value.

Annotated classification.

ORDER JUGLANDALES

Trees; leaves mostly deciduous, alternate or rarely opposite or whorled, equally or unequally pinnately compound; leaflets as few as 3 but usually 5 to 31, serrate (toothed) or entire (smooth margined), dotted beneath with yellow or pale several-celled resinous scales; inflorescences consisting of elongate spikes or catkins, the pistillate (female) spikes sometimes few-flowered and short; bracts (leaflike appendages) associated with the flowers usually as conspicuous as or more conspicuous than the floral parts; individual flowers bisexual or unisexual, inconspicuous, wind-pollinated, lacking petals, with 4 sepals or these lacking; filaments of stamens very short; pistil with 2 carpels, and 2 styles or style branches; ovary completely or incompletely 2-celled but with only one ovule in the middle of the fertile locule; fruit 1-seeded, a small winged nutlet or a large nut, the seed lacking endosperm. Two families, 9 genera, and about 58 species distributed in temperate and subtropical regions of eastern and southwestern United States, Mexico, West Indies, western South America to Argentina, Japan and Philippines to New Guinea and Java, Manchuria and China to Malay Peninsula to Himalayan mountains of India, and Iran (Persia) and from the Caucasus mountains to Greece. They are absent from Australia, New Zealand, Africa, most of Europe, and western Asia.

Family Rhoipteleaceae

Leaves alternate, with stipules, unequally pinnately compound; leaflets serrate; inflorescence consisting of a terminal panicle of identical elongate slender spikes resembling catkins; flowers 3 to a primary bract along the axis of the spike, the central flower bisexual, with 4 sepals, 6 stamens and a pistil, the lateral 2 flowers pistillate (female only) but sterile; ovary 2-celled, superior, the single ovule attached to the partition in the fertile locule, ovule anatropous (curved), unlobed, with 2 integuments, and without packing tissue below it; fruit a small flattened nutlet, 2-winged due to outgrowths of the ovary wall; bracts and sepals free from the ovary and stamens; staminate (male only) flowers none. One genus and 1 species, *Rhoiptelea chiliantha,* found in southwestern China and North Vietnam; unknown outside the Orient.

Family Juglandaceae (walnut family)

Leaves without stipules, alternate, opposite or whorled, unequally or equally pinnately compound; leaflets serrate or entire; buds naked or scaly; inflorescences of 2 types: many-flowered, usually drooping staminate catkins, or many-flowered, several-flowered, or few-flowered, drooping or erect pistillate spikes; plants monoecious (*i.e.*, flowers of both sexes occur on one plant) or rarely dioecious (*i.e.*, there are separate plants of each sex); flowers 1 to a primary bract, unisexual, with the primary and usually also the two secondary bracts fused with the staminate receptacle and in most cases also with at least the base of the ovary, thus appearing to be part of the floral envelope; staminate flowers with or without 4 sepals and 3 to 100 stamens; pistillate flowers with 4 sepals or these rarely absent; ovary inferior, 1-celled (1-loculed), but 2- to 4- (sometimes to 8-) celled near the base because of the presence of 1 or 2 (or 3, rarely) incomplete partitions in the lower half of the ovary; packing tissue at first filling the lower part of the locule; ovule orthotropous (straight), erect at the top of the primary incomplete partition with 1 integument; fruit a small winged nutlet or a nut, the latter enclosed in an adherent dehiscent (splitting) or indehiscent husk, the nut enclosed in the husk resembling a drupe (stoney-seeded fruit such as a cherry or peach), and this whole fruit sometimes called a tryma; wings and husks of the fruits developed from the bracts of the flower; in 1 genus, *Alfaroa,* the nut lacks both true husk and wings; seed with a straight embryo, 2- or 4- or rarely 8-lobed at the base; cotyledons ("seed leaves") 4-lobed, at germination remaining underground in the fruit (cotyledons hypogeous) or appearing aboveground as 4-lobed leaves (cotyledons epigeous). Three subfamilies, 3 tribes, 8 genera, and 57 species. Distribution as given for the order.

Subfamily Juglandoideae

Fruiting catkin or spike not conelike, either many-fruited and drooping or few-flowered and erect; fruiting primary bract remaining attached to nut or nutlet and falling off the tree with it; catkins or spikes many to few-flowered, at least the staminate catkins drooping; pistillate catkin or spike completely pistillate; secondary bracts present in the staminate flowers.

Tribe Juglandeae. Three genera: *Juglans* (walnut, butternut) 20 species, *Pterocarya* (wing-nut) 5 species, and *Cyclocarya,* 1 species.

Tribe Engelhardieae. Three genera *Engelhardia,* 5 species, *Oreomunnea* (in many books combined with *Engelhardia),* 2 species, and *Alfaroa,* 5 species.

Tribe Caryeae. One genus, *Carya* (hickory, pecan), 18 species.

Subfamily Platycaryoideae

Fruiting catkin condensed, many-flowered, conelike, erect, the unmodified unlobed primary bracts rigid, woody; fruiting primary bract separating from the nutlet, the bract persisting on the tree, the very small flattened slightly two-winged nutlet falling off the tree; all catkins elongate, erect, clustered, the cluster terminal; pistillate catkin central in the terminal cluster, many-flowered, tipped by a staminate catkin; secondary bracts absent in the staminate flowers; carpels lateral; pith solid; leaves alternate, with terminal leaflets. One genus and one species, *Platycarya strobilacea,* of Japan and eastern Asian mainland.

Critical appraisal. The classification used here for dividing the Juglandaceae (walnut family) into subfamilies and tribes is a recent one of recognized authority. Most of the former classifications divided the family merely on the basis of the appearance of the fruits; for example, one group of genera has small, winged fruits, the other has larger fruits, most of which have husks. Yet various studies in wood anatomy, pollen structure, inflorescence type and position, floral structure, and nature of primary and secondary bracts show that the traditional classification schemes bring together unrelated members.

Tremendous advances have been made within the past 50 years, primarily in the family Juglandaceae, in the knowledge of inflorescences, flower structure, wood anatomy, pollen-grain structure, seedling type and its anatomy, number of genera and species, number of fossils, relationship within the family, and relationship to other families and orders. Much still remains to be done, however. Active studies are still being made. Opinions differ as to what genera and how many species should be recognized. Relationships among the genera are still open to question. Knowledge of fruit and seedling structure is still incomplete. Better knowledge of the relationships of the order to other orders is still needed as is indicated in the lack of a satisfactory explanation for the compound leaf. Intensive studies are now going on in an attempt to obtain a better natural classification of the flowering plants as a whole. The location of the Juglandales order in this classification is still in doubt.

BIBLIOGRAPHY. E.W. BERRY, *Tree Ancestors* (1923), a semipopular book giving the modern and fossil distribution of trees, showing maps of these distributions and notes on fossil history and economic uses; C. HEIMSCH, JR. and R.H. WETMORE, "The Significance of Wood Anatomy in the Taxonomy of the Juglandaceae," *Am. J. Bot.,* 26:651–660 (1939), a technical study of wood structure in the walnut family, pointing out correlation between wood anatomy and morphology of inflorescences in distinguishing primitive from advanced members of the family; H. HJELMQVIST, "Studies on the Floral Morphology and Phylogeny of the Amentiferae," *Bot. Notiser, Suppl.,* 2:1–171 (1948), a detailed study of flower and inflorescence structure of all Amentiferae, including the Juglandaceae family, with a discussion of relationships between the genera and between the families of that group; J. HUTCHINSON, *Genera of Flowering Plants (Angiospermae),* vol. 2, *Dicotyledones* (1967), a book for professionals containing technical descriptions of plant orders, families, and genera of the world, including Juglandaceae and Rhoipteleaceae (with keys to families and genera and literature references); J.F. LEROY, "Etude sur les Juglandaceae," *Mém. Mus. Natn. Hist. Nat., Paris (Botanique),* n.s., ser. *B,* 6:1–246 (1955), a detailed and thorough study in French of all genera of the Juglandaceae with emphasis on floral anatomy but also treating relationships, fossils, origin of flowers, and taxonomy; W.E. MANNING, "The Morphology of the Flowers of the Juglandaceae," pt. 1, "The Inflorescence," *Am. J. Bot.,* 25:407–419 (1938), technical descriptions of all inflorescences of the six genera recognized by this author and most of the species in them found throughout the world; D.C. PEATTIE, *A Natural History of the Trees of Eastern and Central North America,* 2nd ed. (1966), a popular description of native walnut and hickory trees, and other trees, of the United States with notes on economic importance, interesting facts, and historical incidents; A. REHDER, *Manual of Cultivated Trees and Shrubs, Hardy in North America,* 2nd rev. ed. (1940), a basic book with technical descriptions of the Juglandaceae family and other families with keys for separation of all temperate genera, Asiatic as well as North American (including for each a technical description of the genus and most of its species); C.S. SARGENT, *Manual of the Trees of North America (Exclusive of Mexico)* (1933), a basic work containing a technical description of the walnut family, along with all other tree families of the region, with keys for separation of the two genera and all species native to the United States (including descriptions, exact ranges, and a small illustration of each species); D.R. WHITEHEAD, "Pollen Morphology in the Juglandaceae," pt. 2, "Survey of the Family," *J. Arnold Arbor.,* 46:369–410 (1965), a technical study and description of pollen grains of all genera and most species of the Juglandaceae family with a chart of evolution of pollen grains in family; C.L. WITHNER, "Stem Anatomy and Phylogeny of the Rhoipteleaceae," *Am. J. Bot.,* 28:872–878 (1941), a technical study of the wood structure of the genus *Rhoiptelea* with a comparison to that of the Juglandaceae family and Urticales order.

(Wa.E.M.)

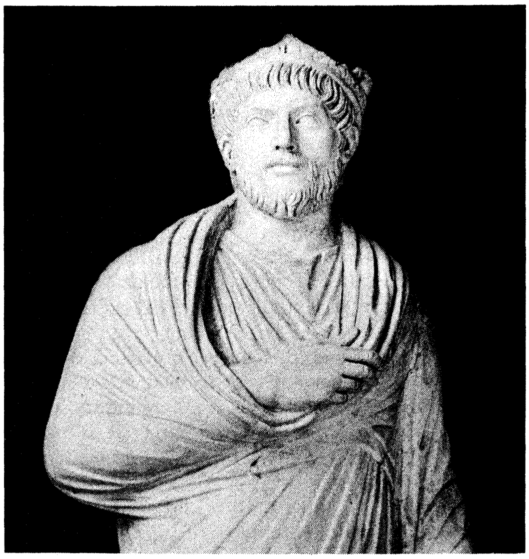
Julian the Apostate

Flavius Claudius Julianus, born *c.* AD 331/332, famous in history as the Apostate, was the last Roman emperor to attempt to replace Christianity by a revived polytheism of the Greco-Roman pantheon. Julian was a younger son of Julius Constantius, the half-brother of Constantine I (the Great), and his second wife, Basilina. When he was five, his father's nephew, also called Constantius, became emperor in the East as Constantius II and in 340, by the death of his brother Constantine II, sole emperor. The army, determined to have none but Constantine I's sons as his successors, murdered the latter's half-brothers and other possible aspirants. Constantine I's third son, Constantius II had Julian's father killed in or just after 337, and an elder brother of Julian was killed in 341. Basilina had died soon after the birth of Julian, who was thus early left an orphan. With his surviving half-brother, Gallus, seven years his senior, he was brought up in obscurity, first by Eusebius, bishop of Nicomedia, in Bithynia and later at the remote estate of Macellum in Cappadocia. By the patronage of Eusebia, wife of Constantius II, Julian, when 19, was allowed to continue his education, first at Como and later in Greece. His physical appearance is described thus by his contemporary and comrade-in-arms, Ammianus Marcellinus:

> He was of medium stature, his hair was soft, as if it had been combed, his beard rough and pointed. His eyes were fine and flashing, an indication of the nimbleness of his mind. He had handsome eyebrows, a straight nose, rather a large mouth with a drooping lower lip. His neck was thick and slightly bent, his shoulders broad and big. From top to toe he was well-knit, and so was strong and a good runner.

His statue in the Louvre generally confirms this description, showing him as a stocky, rather diffident-looking philosopher.

Giraudon



Julian the Apostate, detail of a marble statue. In the Louvre, Paris.

Julian's freedom as a student had a powerful influence on him and ensured that for the first time in a century the future emperor would be a man of culture. He studied at Pergamum, at Ephesus, and later at Athens. He was attracted by Neoplatonism and adopted the cult of the Unconquered Sun.

That his literary talent was considerable is demonstrated in his surviving works, most of which illustrate his deep love of Hellenic culture.

Julian had been baptized and raised as a Christian; but, although he outwardly conformed until he was supreme, Christianity in its official guise meant to him the religion of those who had murdered his father, his brother, and many of his relations and, as such, was hardly likely to commend itself to him. He found far more solace in his philosophic speculations. This reaction has sometimes been defended as natural but eccentric. Natural it certainly was; but it is a misinterpretation of the age to imagine that Julian was alone in preferring Hellenism to Christianity. Society, and particularly the educated society in which Julian was at home, was in fact still largely if not predominantly pagan. Even bishops were proud of their Greek culture; no one was proud of the exotic degeneracy and extravagance of the court of Constantius. It is not surprising that Julian's austerity, chastity, and enthusiasm for the heritage of Greece found a sympathetic response among many of his cousin's subjects.

Military successes in Gaul

Rise to supremacy. In 351 Constantius II, perturbed by the death of his brother Constans and subsequent disorders in the West, appointed Gallus as his caesar; that is, as his coadjutor and eventual successor. Gallus was a failure, and was executed near Pola in Italy (now Pula, Yugoslavia) in 354. Constantius, again in need of a caesar of his own house, after much hesitation summoned Julian from Greece, whence the latter arrived "still wearing his student's gown." In November 355, at the age of 23, he was duly proclaimed and invested as caesar, an honour which he accepted with justifiable foreboding. The Emperor gave Julian his sister Helena as wife. She died after five years of marriage—the fate of their issue, if any, is unknown. Julian was at once dispatched to Gaul, where he proved a resolute and successful commander. He defeated and expelled the Alemanni and the Franks, feats that aroused the jealousy of Constantius, who kept Julian short of funds and under secret surveillance. In 360, while Julian was wintering at Paris, his preferred headquarters, the Emperor sent a demand for a number of his best troops, ostensibly for service in the East but in reality to weaken Julian. Julian's army thereupon hailed him as Augustus. This naturally infuriated Constantius, who refused any accommodation. Julian, realizing that war between himself and Constantius was now inevitable, decided to move first. But, before the clash could come, Constantius died near Tarsus (November 361), having on his deathbed accepted the inevitable by bequeathing the empire to Julian.

Policies as emperor. Julian, now sole Augustus, greatly simplified the life of the palace and reduced its expenses. He issued proclamations in which he declared his intention to rule as a philosopher, on the model of Marcus Aurelius. All Christian bishops exiled by Constantius were allowed to return to their sees (although the purpose of this may have been to promote dissension among the Christians), and an edict of 361 proclaimed freedom of worship for all religions.

But this initial toleration of Christianity was coupled with a determination to revive paganism and raise it to the level of an official religion with an established hierarchy. Julian apparently saw himself as the head of a pagan church. He performed animal sacrifices and was a staunch defender of pagan orthodoxy, issuing doctrinal instructions to his clergy. Not surprisingly, this incipient fanaticism soon led from apparent toleration to outright suppression and persecution of Christians. Pagans were openly preferred for high official appointments, and Christians were expelled from the army and prohibited from teaching. The latter action led Ammianus, who admired Julian's virtues and was himself an adherent of the traditional religion, to censure the emperor:

> That was inhumane, and better committed to oblivion, that he forbade teachers of rhetoric and literature to practice their profession if they were followers of the Christian religion.

Julian wrote an attack on Christianity that is known today only by fragmentary citation. "The trickery of the Galileans"—his usual term—has nothing divine in it, he argues; it appeals to rustics only, it is made up of fables and irrational falsehoods. Here perhaps may be detected the sunset snobbery of the Athens of his day. Though professing to be a Neoplatonist and a sun worshipper, Julian himself was an addict of superstition rather than religion, according to Ammianus.

His project to rebuild the Jewish Temple in Jerusalem was designed rather to insult the Christians than to please the Jews, who, for long accustomed to the worship of the synagogue, would have found the revival of animal sacrifice acutely embarrassing. The plan was dropped when it was reported (as it was on both an earlier and a later occasion) that "balls of fire" had issued from the old foundations and scared away the workmen. Christian cities were penalized, and churches were burned in Damascus and Beirut. Bishops, including the great Athanasius, were banished. One was horribly tortured. Bacchus was installed in the Christian basilicas of Emesa (modern Homs, Syria) and Epiphameia (modern Ḥamāh, Syria). At Antioch, where Julian was preparing for a campaign against the Persians, his closing of the great basilica and the removal of the relics of the martyr Babylas from the sacred grove of Daphne annoyed the Christians. His priggish austerity did not endear him to the pagans, either, and both were equally incensed by his pamphlet entitled *Misopogon* ("Beard Hater"), in which he assailed the Antiochenes for the ridicule that they poured on him for his personal conduct, his religion, and his claim to be a philosopher on the strength of his beard.

The campaign against Persia.

The invasion of Persian territory was always a lure in antiquity and one to which Julian was not immune. Motivated by a desire for military glory and a decision to reassert Rome's pre-eminence in the East, he assembled, despite counsels of prudence from Rome and the Levant, the largest Roman army (65,000-strong and backed by a river fleet) ever to head a campaign against Persia. The Persians, aided by the desert, famine, treachery, and incompetence of the Romans, once again proved themselves superior. During a disastrous retreat from the walls of Ctesiphon, below modern Baghdad, Julian was wounded by a spear thrown "no one knew hence," which pierced his liver. He died the next night (June 363) at the age of 31, having been emperor for 20 months.

Julian's religious policy had no lasting effect. It had shown, however, that paganism, as a religion, was doomed. It is perhaps sad, in retrospect, that the odium of proving it should rest on Julian, who with a little less venom and more tact might have been remembered for his many virtues rather than for his two fatal blunders.

BIBLIOGRAPHY

Ancient: By far the most important source is the history of AMMIANUS MARCELLINUS. Ammianus was a contemporary and fellow-soldier of Julian in his Persian campaign. He was an orthodox pagan, but is outspokenly critical of Julian's anti-Christian policy. He devotes far more space to Julian than to any other character in the extant portion of his history. See books xv, 8 to xxv, 4, in the 3 vol. ed. trans. by J.C. ROLFE (1950–56), which are virtually a biography of Julian. See also Julian's own works, *Juliani imperatoris epistulae . . .*, ed. by J. BIDEZ and F. CUMONT (1922). For English readers the best edition is that of W.C. WRIGHT, 3 vol., "Loeb series" (1953–54), with a very helpful introduction. Of secondary ancient sources the most valuable and balanced is the BREVIARIUM OF EUTROPIUS. The *Decline of Rome* of ZOZIMUS, finished after 425, is the work of an avowed enemy of the Christian Church. ST. GREGORY NAZIANZEN is naturally hostile to Julian.

Modern: Many monographs have been devoted to this controversial prince. Among them may be cited G.H. RENDALL, *The Emperor Julian, Paganism and Christianity* (1879); G. NEGRI, *L'imperatore Giuliano l'Apostata*, 2nd ed. (1902; Eng. trans., *Julian the Apostate*, 1905); P. ALLARD, *Julien l'Apostat*, 3 vol., 3rd ed. (1906–10); A. ROSTAGNI, *Giuliano l'Apostata* (1919); and J. BIDEZ, *La Vie de l'empereur Julien* (1930). H. MATTINGLY's article on Julian in the OXFORD CLASSICAL DICTIONARY is to be commended.

(S.H.P.)

Julius II, Pope

The greatest art patron of the papal line and one of the most powerful rulers of his age, Julius II (Giuliano della Rovere) re-established the authority of the papacy in its ancient territory after the Borgia wars earlier in the century. As a politician and patron of the arts, he strove for a synthesis of church and state, of spirit and culture, with a grandeur unequalled by succeeding popes.

Career as cardinal and exile

He was born on December 5, 1443, in Albisola, Italy, the son of the impoverished Rafaello della Rovere, Pope Sixtus IV's only brother. In 1468 he became a Franciscan, and in 1471 Sixtus IV made him a cardinal. In this office Giuliano displayed all of the attributes of cupidity and corruption of an unscrupulous Renaissance prince. The Pope lavished on him six bishoprics in France and three in Italy along with an abundance of wealthy abbeys and benefices. The Cardinal, who lacked any interest in spiritual pursuits, became an outstanding patron of the arts. He is shown with his protégés in Melozzo da Forlì's superb fresco of Sixtus IV in the Vatican Museum.

Leonard von Matt—EB Inc.

Julius II, contemporary medallion. In the coin collection of the Vatican Library.

After the death of Sixtus IV, for whom Giuliano commissioned a bronze sepulchre by Antonio Pollaiuolo, now in the Vatican Grotto of St. Peter's, the Cardinal's candidate, the weak Innocent VIII, was elected through bribery. When Rodrigo Borgia, elected pope as Alexander VI in 1492, plotted Giuliano's assassination, Giuliano fled in 1494 to the court of Charles VIII of France. He accompanied the French king on his expedition against Naples in the hope that Charles would also depose Alexander VI. After accompanying Charles on his forced return to France, Giuliano took part in Louis XII's invasion of Italy in 1502. Alexander VI twice attempted to seize him.

Following the death of the Borgia pope in 1503, he returned to Rome, having been ten years in exile, and, after Pius III's brief pontificate, was, with the liberal help of simony, elected Pope Julius II in October 1503. Immediately after his election he decreed that all future simoniacal papal elections would be invalid and subject to penalty.

Political activities

Julius II viewed as the main task of his pontificate the restoration of the Papal States, which had been reduced to ruin by the Borgias. Large portions of it had been appropriated by Venice after Alexander VI's death. As a first step as pope, Julius subjugated Perugia and Bologna in the autumn of 1508. Then, in March 1509, he joined the League of Cambrai, an anti-Venetian alliance formed in December 1508 between Louis XII, who then ruled Milan, Emperor Maximilian I, and Ferdinand II of Spain, who had been king of Naples since 1503. The league troops defeated Venice in May 1509 near Cremona, and the Papal States were restored.

Having become an exponent of Italian national consciousness, Julius II proposed to drive the French from Italy, but his second war, which lasted from September 1510 to May 1511, was unsuccessful. Several cardinals defected to Louis XII and called a schismatic council, to which Julius responded by summoning the fifth Lateran Council. After concluding an alliance with Venice and Ferdinand II of Spain and Naples in October 1511, he opened the council in May 1512 at the Lateran palace. Louis XII had defeated the troops of the alliance at Ravenna in April 1512, but the situation changed when Swiss troops were sent to the Pope's aid. The territories in northern Italy occupied by the French revolted, the French left the country, and the Papal States were augmented by the acquisition of Parma and Piacenza. Toward the end of his life, he viewed with concern the replacement of French by Spanish efforts to attain supremacy in Italy. Julius II was Italy's saviour.

Patron of the arts

The enduring impact of the life of Julius II stemmed from his gift for inspiring great artistic creations. His name is closely linked with those of great artists such as Bramante, Raphael, and Michelangelo. With his wealth of visionary ideas, he contributed to their creativity. Following an overall plan, he added many fine buildings to Rome and laid the groundwork in the Vatican Museum for the world's greatest collection of antiquities. Among the innumerable Italian churches that benefited from his encouragement of the arts was Santa Maria del Popolo in Rome, for which he commissioned Andrea Sansovino to create sepulchres for a number of cardinals and Pinturicchio to paint the frescoes in the apse. Donato Bramante became the architect of Julius' fortifications in Latium, of the two galleries that form the Belvedere Court, and of other Vatican buildings. Around 1503 the Pope conceived the idea of building a new basilica of St. Peter, the first model of which Bramante created. Its foundation stone was laid on April 18, 1506.

The Pope's friendship with Michelangelo, begun in 1506, was enduring despite recurrent strains imposed on their relations by the two overly similar personalities. Their relationship was so close that the Pope became, in fact, Michelangelo's intellectual collaborator. Of Julius' tomb only the Moses in the church of San Pietro in Vincoli, in Rome, was completed; the Pope is, however, not interred there, but in St. Peter's, along with the remains of Sixtus IV. The famous bronze statue of the Pope for the church of San Petronio in Bologna, completed in 1508, was destroyed in 1511. In 1509 Michelangelo was prevailed upon by Julius to begin his paintings on the ceiling of the Sistine Chapel, which were unveiled in October 1512. The paintings, which represent a climax in Western art, were, in form and conception, a product of the artistic symbiosis of Michelangelo and the Pope.

In 1509 Raphael, introduced to Julius, began his masterpieces for the Pope, the frescoes in three rooms of the Vatican. Spiritual references to the person and the pontificate of Julius II are evident in one of the rooms (the Stanza della Segnatura), where earthly and celestial wisdom are juxtaposed in the "School of Athens" and the "Disputa," while the beauty of creativity is represented in the "Parnassus." The theme of another room (the Stanza d'Eliodoro), which could be called a transcendental "political" biography of the Pope, is still more personal. "The Expulsion of Heliodorus from the Temple" symbolizes the expulsion of the French and the subjugation of all the church's enemies, with Julius II depicted witnessing the scene from his portable throne. Closely related to this is the "Liberation of St. Peter," in which light and darkness serve to symbolize the historic events of the pontificate. The third great fresco in this room, the "Mass of Bolsena," shows the Pope kneeling, rather than enthroned, in commemoration of his veneration of the *corporale* (communion cloth) of Bolsena in the cathedral of Orvieto. In addition to these fresco portraits, there is one by Raphael in the Uffizi gallery in Florence, one of the masterpieces of portraiture, which shows the Pope not as the victorious Moses springing to his feet, as Michelangelo portrayed him, but as a resigned, pensive old man at the end of an adventurous, embattled life. Michelangelo's chalk drawing of the Pope in the Uffizi gallery approaches it in quality.

As cardinal, Julius II fathered three illegitimate daughters, Felice, Clarissa, and Giulia, none of whom achieved any particular distinction. He made four members of the della Rovere family cardinals, only one of whom achieved any importance. From the marriage of the Pope's only brother, Giovanni, to the daughter and heiress of Duke Federigo of Montefeltro descended the dukes of Urbino. Julius II died in Rome on February 21, 1513.

The Pope added wisely to the church's treasures. Although he had little of the priest in him, he was concerned toward the end only with the church's grandeur. He wished for greatness for the papacy rather than for the pope, and he wished for peace in Italy. The Swiss historian Jacob Burckhardt called him the "saviour of the papacy," because Alexander VI had greatly endangered its existence for the sake of his family interests.

Assessment

Julius had an extraordinarily violent temper, often lost his self-control, and could be rude and often even vulgar in manner. Yet, apart from the avarice and corruption inherent in his office and time as much as in himself, he was incapable of baseness and vindictiveness and despised informers and flatterers; no one was able to influence his decisions. Everywhere he saw and sought out greatness. He lacked the smooth manners of the servile. His faults arose from his relentless candour and uncontrollable temper. He was called *terrible,* an epithet suggesting that he was regarded as sublime, even superhuman.

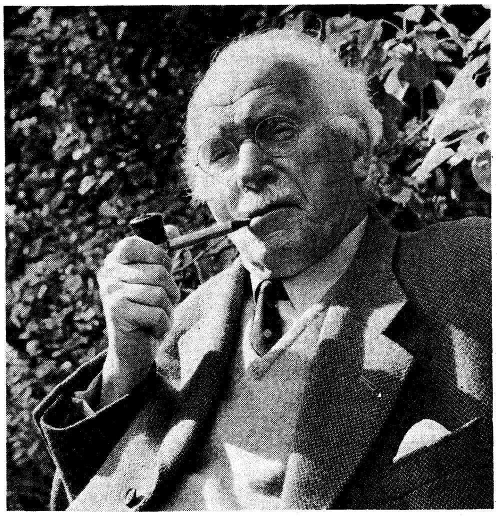
BIBLIOGRAPHY. LUDWIG VON PASTOR, *Geschichte der Päpste*, vol. 2 (1889; Eng. trans., *The History of the Popes,* vol. 6, 1898), is the standard work of papal history; E.P. RODOCANOCHI, *Histoire de Rome: Le Pontificat de Jules II, 1503–1513* (1928), the main French work on Julius; and G.B. PICOTTI, *La politica italiana sotto il pontificato di Giulio II* (1949), a study of Italian politics at the time of Julius. Two recent biographies are C. FUSERO, *Giulio II* (1965); and L. BINDI SENESI, *Giulio II* (1967).

(H.Ku.)

Jung, Carl

Carl Gustav Jung achieved international recognition through his psychological and psychiatric studies and therapeutic work. A contemporary and one-time colleague of Sigmund Freud, he was a highly original thinker and promoted what he termed analytic psychology rather than Freud's psychoanalysis. He advanced the concepts of the extroverted and introverted personality and of archetypes and the collective unconscious. His work has been influential in psychiatry and in the study of religion and related fields. The greater number of his books are long, technical, and intended for analysts and scholars. Later in his life he produced some shorter and more accessible works, including *Modern Man in Search of a Soul* (1933), an appraisal of modern religious dilemmas; *Aufsatze zur Zeitgeschichte* (1946; *Essays on Contemporary Events,* 1947), a study of the problems of World War II; *Antwort auf Hiob* (1952; *Answer to Job,* 1954), an essay attacking conventional views of God; and *Erinnerungen, Träume, Gedanken* (1962; *Memories, Dreams, Reflections,* 1963), an autobiographical essay published posthumously.

Leni Iseley–Nancy Palmer Agency



Jung.

Jung was born July 26, 1875, in Kesswil, Switzerland, the son of a philologist and pastor. His childhood was lonely, though enriched by a vivid imagination; and from an early age he observed the behaviour of his parents and teachers, which he tried to resolve. Especially concerned with his father's failing belief in religion, he tried to communicate to him his own experience of God. Though the elder Jung was in many ways a kind and tolerant man, neither he nor his son succeeded in understanding each other. Jung seemed destined to become a minister, for there were a number of clergymen on both sides of his family. In his teens he discovered philosophy and read widely, and this, together with the disappointments of his boyhood, led him to forsake the strong family tradition and to study medicine and become a psychiatrist. He was a student at the universities of Basel (1895–1900) and Zurich (M.D., 1902).

After beginning work at the University Psychiatric Clinic in Zurich, he was fortunate in joining the staff of the Bergholzli Asylum in Zurich at a time (1902) when it was under the direction of Eugen Bleuler, whose psychological interests had initiated what are now considered classical researches into mental illness. At Bergholzli, Jung began, with outstanding success, to apply association tests initiated by earlier researchers. He studied, especially, patients' peculiar and illogical responses to stimulus words and found that they were caused by emotionally charged clusters of associations withheld from consciousness because of their disagreeable, immoral (to them), and frequent sexual content. He used the now famous term complex to describe such conditions.

Early association with Freud

These researches, which established him as a psychiatrist of international repute, led him to understand Freud's investigations; his findings confirmed many of Freud's ideas, and, for a period of five years (between 1907 and 1912), he was Freud's close collaborator. He held important positions in the psychoanalytic movement and was widely thought of as the most likely successor to the inventor of psychoanalysis. But this was not to be the outcome of their relationship. Partly for temperamental reasons and partly because of differences of viewpoint, the collaboration ended. At this stage Jung differed with Freud largely over the latter's insistence on the sexual bases of neurosis. The formal break came in 1912, with the publication of Jung's *Wandlungen und Symbole der Libido (Psychology of the Unconscious,* 1916), which ran counter to many of Freud's ideas. Though Jung had been elected president of the International Psychoanalytic Society in 1911, he resigned from the society and was the cofounder (with A. Maeder) of a new school in Zurich.

His first achievement was to differentiate two classes of people according to attitude types: extroverted (outward-looking) and introverted (inward-looking). Later he differentiated four functions of the mind—thinking, feeling, sensation, and intuition—one or more of which predominate in any given person. The results of this study were embodied in *Psychologische Typen* (1921; *Psychological Types,* 1923). Jung's wide scholarship was well manifested here, as it also had been in *The Psychology of the Unconscious.*

As a boy Jung had remarkably striking dreams and powerful fantasies that had developed with unusual intensity. After his break with Freud, he deliberately allowed this aspect of himself to function again and gave the irrational side of his nature free expression. At the same time, he studied it scientifically by keeping detailed notes of his strange experiences. He later developed the theory that these experiences came from an area of the mind that he called the collective unconscious, which he held was shared by everyone. This much contested conception was combined with a theory of archetypes that Jung believed were of fundamental importance for the study of the psychology of religion. In Jung's terms, archetypes are instinctive patterns, having a universal character, expressed in behaviour and images.

Character of his psychotherapy

The rest of his life was given over to the development of his ideas, especially those on the relation between psychology and religion. In his view, obscure and often neglected texts of writers in the past shed unexpected light not only on Jung's own dreams and fantasies but also those of his patients; he thought it necessary for the successful prosecution of their art that psychotherapists become familiar with writings of the old masters.

Besides the development of new psychotherapeutic methods that derived from his own experience and the theories developed from them, Jung gave fresh importance to the so-called Hermetic tradition. He conceived that the Christian religion was part of a historic process necessary for the development of consciousness, but he thought that the heretical movements, starting with Gnosticism and ending in alchemy, were manifestations of unconscious archetypal elements not adequately expressed in the varying forms of Christianity. He was particularly impressed with his finding that alchemical-like symbols could be found frequently in modern dreams and fantasies, and he thought that alchemists had constructed a kind of textbook of the collective unconscious. He drove this home in four large volumes of his *Collected Works.*

His historical studies aided him in pioneering the psychotherapy of the middle-aged and elderly, especially those who felt their lives had lost meaning. He helped them to appreciate the place of their lives in the sequence of history. Most of these patients had lost their religious belief; Jung found that if they could discover their own myth as expressed in dream and imagination they would become more complete personalities. He called this process individuation.

In later years he became professor of psychology at the Federal Polytechnical University in Zurich (1933–41) and professor of medical psychology at the University of Basel (1943). His personal experience, his continued psychotherapeutic practice, and his wide knowledge of history placed him in a unique position to comment on current events. As early as 1918 he had begun to think that Germany held a special position in Europe; the Nazi revolution was, therefore, highly significant for him, and he delivered a number of hotly contested views that led to his being wrongly branded as a Nazi sympathizer.

To the end of his life he retained his love of simple things: the Swiss countryside in which he had been brought up, the peasant lore, and life in the open air. He and his wife Emma had built a beautiful home by the lake of Zurich, where she acted as a hostess to their many distinguished visitors; helped him in his work, and made original contributions of her own. Jung lived to the age of 85, dying on June 6, 1961, at Kusnacht, Zürich.

MAJOR WORKS

Jung wrote and published continually during an active life and revised and reissued many papers; his writings occasionally appeared first in English and then in German. The selection below does not take revision and reissue into account and is based upon publication in English, whether in translation or not. Further, certain books represent the gathering together of papers first published separately, and this is indicated, the dates, but not always the titles, of papers in German being cited.

Studies in Word Association, trans. by M.D. Eder (1918; in *Collected Works,* vol. 2; *Diagnostische Assoziationsstudien,* 2 vol., 1906–10); *The Psychology of Dementia Praecox* (1909; *Collected Works,* vol. 3; *Über die Psychologie der Dementia Praecox,* 1907); *The Psychology of the Unconscious,* trans. by Beatrice M. Hinkle (1916; rev. as *Symbols of Transformation* in *Collected Works,* vol. 5; *Wandlungen und Symbole der Libido,* 1911–12); *The Theory of Psychoanalysis* (1915; *Collected Works,* vol. 4; *Versuch einer Darstellung der psychoanalytischen Theorie,* 1913); *Collected Papers on Analytical Psychology,* ed. by Constance Long (1916); *Psychological Types,* trans. by *H.G.* Baynes (1923; *Collected Works,* vol. 6; *Psychologische Typen,* 1921); *Two Essays on Analytical Psychology,* trans. by C.F. and H.G. Baynes (1928), comprising "The Unconscious in the Normal and Pathological Mind" (German 1926) and "The Relation of the Ego to the Unconscious" (German 1928; rev. in *Collected Works,* vol. 7); with Richard Wilhelm, *The Secret of the Golden Flower,* trans. by C.F. Baynes (1931, in part as *Das Geheimnis der goldenen Blüte,* 1929), containing the following by Jung—"Commentary" (in *Collected Works,* vol. 13), "Examples of European Mandalas," and "In Memory of Richard Wilhelm" (German 1930); with Jung and Pauli, "On the Nature of the Psyche," trans. by R.F.C. Hull ("Theoretische Überlegungen zum Wesen des Psychischen" in *Votz den Wurzen des Rewusstseins,* 1954), in *The Structure and Dynamics of the Psyche, Collected Works,* vol. 8); *Psychology and Religion* (1938; somewhat expanded as *Psychologie und Religion,* 1940; in *Collected Works,* vol. 11); *The Integration of the Personality,* trans. by Stanley M. Deil (1939); *Psychology and Alchemy,* trans. by R.F.C. Hull (1949; *Psychologie und Alchemie,* 1944; in *Collected Works,* vol. 12); with Károly Kerényi, *Essays on a Science of Mythology,* trans. by R.F.C. Hull (1949; *Einführung in das Wesen in der Mythologie,* 1941); *Aion: Researches into the Phenomenology of the Self,* trans. by R.F.C. Hull (1959; *Collected Works,* vol. 9; *Aion: Untersuchungen zur Sytnbolgeschichte,* 1951); *Mysterium Conjunctionis,* trans. by R.F.C. Hull (German 1955–56; *Collected Works,* vol. 14); *The Undiscovered Self* (1958; *Collected Works,* vol. 10; *Gegenwart und Zukunft,* 1957); *Flying Saucers: A Modern Myth of Things Seen in the Sky* (1959; *Collected Works,* vol. 10; *Ein moderner Mythus: Von Dingen, die am Himmel gesehen werden,* 1958).

BIBLIOGRAPHY. The authoritative English collection of all Jung's published writings is HERBERT READ, MICHAEL FORDHAM, and GERHARD ADLER (eds.), *The Collected Works of C.G. Jung,* trans. by R.F.C. HULL, 2nd ed. (1966–). Examples of Jung's own collections of essays addressed to the general public are: *Modern Man in Search of a Soul* (1933, reprinted 1961), which covers a wide range of topics from dream analysis and literature to the psychology of religion; *Aufsatze zur Zeitgeschichte* (1946; Eng. trans., *Essays on Contemporary Events,* 1947; contained in *Collected Works,* vol. 10), which gives an account of his theory of the psychology of mass movement; *Antwort auf Hiob* (1952, rev. ed. 1961; Eng. trans., *Answer to Job,* contained in *Collected Works,* vol. 11), the best example of his thinking on religious history; and *Memories, Dreams, Reflections* (1962), fascinating semi-autobiographical reading, partly written by Jung himself and partly recorded by his secretary. FRIEDA FORDHAM, *An Introduction to Jung's Psychology,* 3rd ed. (1966), is the most clear and widely read introduction to Jung's theorizing and analytic practice. A fairly good critical assessment is provided by AVIS M. DRY in *The Psychology of Jung* (1961).

(M.S.M.F./F.Fo.)

Jungles and Rain Forests

The word jungle comes from the Hindi *jaṅgal,* originally from the Sanskrit *jaṅgala,* meaning desert. Later it came to mean scrub or scattered bush, and jungly became a word in India for something poor or savage. In modern usage, jungle has come to mean tropical forest and connotes luxuriant, tangled, impenetrable vegetation in a hot, steamy environment, teeming with wildlife. The rain forest exhibits a wide variety of forms but very seldom approaches its popular image; in fact, it is darkly shaded, with little ground cover, and is fairly easy to travel in. Where light can penetrate to the ground, however, as along tracks and riverbanks, a dense and tangled mass of vegetation appears, and because most people see the forest from a track or from a boat it is easy to believe in the impenetrable jungle. Tangled and dense bush often invades as a secondary growth after a rain forest has been cleared, and the word jungle is sometimes used in a more technical sense for such regrowth.

Rain forests, as the name suggests, occur in regions of high rainfall, commonly over 1,800 millimetres (70 inches) per year. It is overall wetness that forests require, however, and no single figure really expresses the most desirable conditions. This is because the effectiveness of precipitation depends on losses from runoff and evapotranspiration (the amount of water that is lost by direct evaporation and by transpiration through plants) that in turn depends on topography, soil, and temperature. Rain forests, therefore, span a wide range of environments, with widely varying geology, water regime, soils, landscapes, and potential use.

Commonly, most rain-forest areas are places of relatively low human-population density. Thus, such areas are possible sites of agricultural expansion thereby providing food for ever-expanding populations elsewhere or of settlement areas for the release of future population pressures.

This article treats the types and characteristics of rain forests, their topography and hydrology, weathering and soils, geological history, and plants and animals. For further information on processes of weathering and soil formation, see WEATHERING; SOILS. For relevant data on an allied environment, see SWAMPS, MARSHES, AND BOGS. For a discussion of the geological origins of swamps and rain forests, see CARBONIFEROUS PERIOD, LOWER; CARBONIFEROUS PERIOD, UPPER.

VARIETIES AND CHARACTERISTICS OF RAIN FORESTS

Some authorities divide the tropical rain forest into "equatorial" and "tropical" or "tropical" and "subtropical" forest, the former in each case being the more permanently wet. In this article, "equatorial" and "subtropical" are used for these divisions, and "tropical" is a general term for both. The term selva is synonymous with tropical rain forest. The world distribution of jungles and rain forests is shown in Figure 1.

Equatorial rain forests. Equatorial rain forests grow in areas of little varying climate with rain throughout the year, no frost, high average temperatures, and no very marked seasonal variations, though occasional dry periods of a month or two can occur. Average daily temperatures in equatorial-rain-forest regions usually range between a daytime maximum of about 30° C (86" F) and a nighttime minimum of about 20° C (68" F), with a monthly and yearly average of about 27° C (81" F). Wide variations from these temperatures are exceptional, and the average annual range is usually only one or two degrees. The average daily range, normally between five and eight degrees, is greater than the annual range, so it may truly be said that night is the winter of the tropics. Any seasons that can be recognized are dependent on variations in the amount of rainfall rather than in the amount of temperature.

Rainfall in equatorial forests

Annual rainfall generally amounts to between 1,500 and 3,500 millimetres (60 and 140 inches), generally well distributed throughout the year (see Figure 2). There may be a drier period, and some places have two periods of maximum rainfall that roughly coincide with the two periods of vertical noonday sun. Some equatorial rain originates from thermal convection, but most of the annual precipitation comes from organized, migrating atmospheric disturbances. The relative humidity in an equatorial rain forest is high at all times.

There is a large water surplus and thus a large runoff; soils are permanently moist, leaching is intense, and erosion is potentially severe. Chemical reactions affecting soil and vegetation take place rapidly because of the high temperatures and abundant rainfall. Plant growth is continuous, though individual species have their own seasons of leaf shedding. The light intensity on the forest floor is usually less than 1 percent of that just above the forest canopy. The ground is littered with fallen leaves and wood, and there are few low-growing plants except for the seedlings of the trees themselves.

The main areas of equatorial forest are the Amazon lowlands, the Congo lowlands (together with a coastal zone extending from Nigeria to Guinea), parts of Indonesia (especially Sumatra), and several Pacific islands.

Rain forest can sometimes extend into areas where rainfall appears to be inadequate if it is augmented by groundwater—forests often extend along valleys surrounded by other vegetation types—or if generally high humidity effectively reduces water losses, as in parts of southern Nigeria.

Equatorial forest grades into subtropical rain forest on windward coasts, into monsoon forests in parts of Southeast Asia and Australia, and into montane forest with increasing altitude.

Subtropical rain forests. Subtropical rain forest is located on the windward coasts, roughly from 10" N and 10° S to the tropics, but occasionally it extends farther. It differs from the equatorial rain forest by having a season of reduced rainfall or even drought, although the wet season is dominant. With distance from the Equator, there is greater variation in length of day and greater seasonal variation in temperature; as a result, this type of rain forest is slightly more open and lower and has a smaller number of species and fewer lianas.

Such forest is found in Central America and in the Caribbean—especially on the windward side of islands; in the Western Ghāts of India and coastal areas of Burma, where there is a short dry season; and in Vietnam, the Philippines, parts of the Brazilian coast, and Madagascar. In the Everglades of Florida there are small patches of such jungle, known as hammocks, which contain mahogany, strangler fig, and epiphytes (air plants).

Monsoon forests. Monsoon forest, also known as tropical deciduous forest, occurs in regions with a large total rainfall but a marked dry season. Monsoon climates are characterized by seasonal pressure and wind reversal resulting in heavy rainfall during the high-sun period of onshore winds and little or no rain during the low-sun period of offshore winds. Average temperatures are high, often over 25° C (77" F), but the annual range is greater than in areas of equatorial rain forest. This forest is lower than equatorial forest, trees have more spreading branches, and there is more light penetration. Because of the dry season, a high proportion of the trees are deciduous, and there is a general period of leaf shedding. Teak is a common species, lianas and epiphytes are locally abundant, and thickets of bamboo are common. Undergrowth is likely to be denser in monsoon forest than in equatorial rain forest because of the greater openness and light penetration, and there is more "jungle" formed after clearing than in tropical rain forest.

This sort of forest is found in monsoon Asia, Burma (behind the coastal tropical forest already mentioned), Thailand, Cambodia, Indonesia (especially Java and the Celebes), northeast Australia, and in parts of West Africa and South America bordering the tropical rain forest.

Mangrove forests. Mangrove forest grows along many tropical coasts, especially on low-lying muddy shores. Trees are commonly between two and ten metres (seven and 33 feet) high, but some reach 30 metres (100 feet). The tangle of stilt roots, air roots, and general debris unites to trap mud and thus build up the land along a low

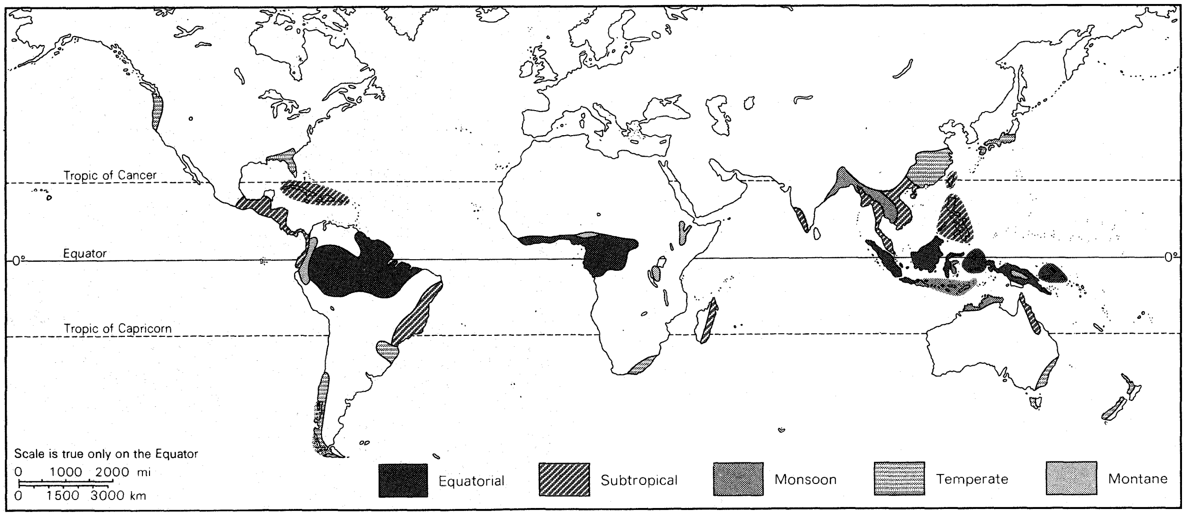


Figure 1: **Types** and locations of the world's major jungles and rain forests.

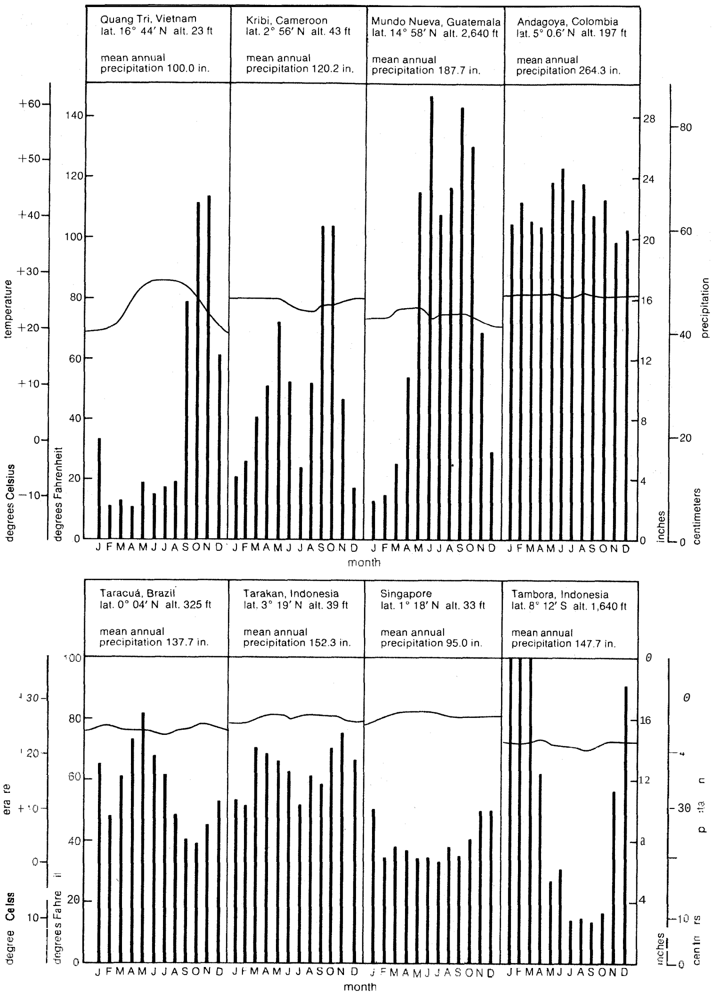


Figure 2: Representative temperatures and rainfall regimes in tropical areas.
Reprinted with permission of the Macmillan Company from *Climatology and the World's Climates* by George R. Rurnney. Copyright © by George R. Rurnney, 1968

coast. Mangrove can tolerate a wide range of salinity and often lines tropical estuaries and deltas, sometimes extending for tens of kilometres inland. Several species may grow in belts parallel to the coast. Mangrove has little use except as fuel, but it may be cleared for the planting of swamp rice.

Temperate rain forests. Temperate rain forest, also known as temperate evergreen forest and laurel forest, has fewer tree species than other types of rain forest, is lower and less dense, and commonly has species with leathery leaves. It is found in regions with an equable climate, abundant and well-distributed rainfall, and small or moderate temperature range. For these reasons it occurs in two main areas; namely, on west coasts between latitudes 35° and 55" and on east coasts between latitudes 25° and 40".

Areas of temperate rain forest

In the first group, the western coasts, are southern Chile, the west coast of New Zealand, Tasmania, and the Pacific coast of the United States. In the second group are the southeastern U.S., southeast China, southeast Brazil and neighbouring areas of Paraguay and Uruguay, the southeast part of South Africa, southeast Australia, and southern Japan.

In these temperate regions, a wider range of soils is present than in the tropics, and climatic controls of the environment are more variable. Evapotranspiration is low and decomposition is slow, so there is relatively deep humus accumulation. Despite its high productivity, temperate rain forest has low recuperative powers and can be permanently destroyed.

Montane forests. Montane forest occurs in the high altitude regions within or bordering the tropical rain forest and includes the Ruwenzori mass of Central Africa, the New Guinea highlands, the Amhara Plateau of Ethiopia, and the Monts Gotel (Gotel Mountains) of Cameroon. With its lower temperatures, this forest has many affinities with the temperate rain forest; but, although altitude brings lower average temperatures, other features, such as length of day, cloudiness, and seasonal variation, are unlike the temperate regions, and the two forest types are therefore dissimilar. Tropical forest gives way to submontane forest at an average altitude of 1,000

By courtesy of CSIRO, Australia

Rain forest in New Guinea.

metres (3,300 feet) and to true montane forest, or "mossy forest," at about 1,500 metres (4,900 feet).

TOPOGRAPHY AND HYDROLOGY

Rock type and landscape. Rain forest is found on a wide variety of rock types and in a wide range of topographic situations, but a few major varieties of landscape may be distinguished.

High plateau and gentle uplands are found in large areas of Africa and in the highlands of the Guianas, Brazil, and other parts of South America. Broad, shallow valleys are separated by broad, gently convex or flat interstream areas, on which erosion is very slight. Weathering to a depth of 100 metres (330 feet) is quite common, and detailed investigations usually show that the lower limit of weathering is irregular with respect to depth; rapid lateral changes in the depth of weathering occur. Occasional hills of fresh rock rise quite abruptly from a surrounding plain, to a height of tens or hundreds of metres. Such hills are called inselbergs, and some authorities regard them as typical of the humid tropical environment.

In areas where erosion has followed a period of widespread laterite formation (see below *Lateritic soils),* flat-topped hills capped by lateritic ironstone are common in the landscape. Lateritic ironstone requires seasonal drying out for its formation, however, and so is more frequent in savanna regions and around the edges of the tropical forest than in the pure equatorial forest.

A completely different kind of landscape is present in some rain forest of highland areas and even in areas of quite low hills, where V-shaped valleys with sharply angular interstream areas produce a landscape consisting entirely of slopes that are often steep and straight. This landscape is dominant in the New Guinea highlands, in the island arcs (*q.v.*) and mountain belts of Indonesia and Central America, and around the edges of high plateaus. Similar topography may also be found in mountainous areas of temperate forest, as in New Zealand.

Volcanoes in various stages of dissection by erosion are found in several parts of the humid tropics, including Central America and Indonesia. These often produce fertile soils. Limestone areas in tropical forest often give rise to a distinctive kind of landscape with "haystack hills," which rise abruptly from the surrounding plains. Broad alluvial plains, such as the lowlands of the Amazon, the Congo, and the Fly in New Guinea, make up large areas of rain forest. There is much variation in alluvial landforms, and these lowlands are by no means uniform with regard to sediments, landforms, or soils.

Precipitation, evapotranspiration, and runoff. In rain-forest areas there is an excess of precipitation over evapotranspiration for most, if not all, of the year (Figure 3).

Effects of plant cover

Only a small proportion of the raindrops fall directly to the ground except at canopy openings. Most are intercepted by the leaves in the canopy and either re-form as waterdrops that fall from the drip tips of the leaves or trickle down the tree trunks. Tropical forest trees generally have broad leaves with drip tips to help shed the water, but, even so, the rain forests probably intercept more rain than any other cover. Estimates of precipitation reaching the ground under various types of cover are: dense forest, 70–80 percent; dense, high grass, 80 percent, cereals and other crops, 80–85 percent; and bare soil, 100 percent.

These data ard based on an assumed average of four millimetres held (intercepted) per shower. This means that rains of less than four millimetres (0.2 inch) do not contribute much to the supply of soil moisture or runoff. The amount retained depends on the total rainfall per day; with heavier showers a greater proportion gets through.

The approximate quantity of rainfall retained by vegetation in a typical site in Indonesia was estimated to be:

Average rainfall	2,400 millimetres (95 inches)
Number of rainy days	160
Rainfall per rainy day	15 millimetres (0.6 inch)
Retained by vegetation (160 X 4)	640 millimetres (25 inches)
Amount reaching soil	1,760 millimetres (69 inches)

Similar results were obtained in Brazil, where it was estimated that about two-thirds of the total rainfall reaches the ground, one-third as raindrops and waterdrops and one-third by streaming down tree trunks.

Evaporation rates can be determined from various kinds of apparatus, but results are very difficult to correlate with evaporation from a ground surface. Available data indicate an annual evaporation of 542 millimetres (21 inches) at Djakarta; 1.139 millimetres (45 inches) at Surabaja; 2,300 millimetres (91 inches) from a pond in Madras, India; and 1,930 millimetres (76 inches) from a reservoir in Bombay.

Evaporation from soil is very hard to determine, but, in Indonesia where temperature is fairly uniform throughout the year, it was found empirically that evaporation (in millimetres for periods of a month) could be expressed by the formula $E = 60 + 0.125P$, in which E is evaporation and P is precipitation. This means that with less rainfall there will be less evaporation, and with no rain at all there will still be 60 millimetres (2 inches) of evaporation per month.

Transpiration usually accounts for a major share of the water lost by an area, and different species of plants exhibit a wide range in their rates of transpiration under similar conditions. Some types of vegetation, including many trees, may utilize moisture equivalent to nearly all of the rain falling even during the wettest months, whereas grasses or mixed forest may transpire only a fraction of the available moisture. This may account for a rise in the water table after land clearance. The average annual transpiration of plants in Java (in millimetres) has been estimated to be:

Mature forest— (trees 740) (undergrowth 130)	870 (34 inches)
Bamboo forest	1,500 (60 inches)
Jungle below 1,000 metres	1,200 (45 inches)
Jungle at 1,000 metres	1,200–500 (45–20 inches)
Jungle at 2.500 metres	600–500 (25–20 inches)

From *A.N.* Strahler, *Introduction* to *Physical Geography* (1965); John Wiley & Sons, Inc.

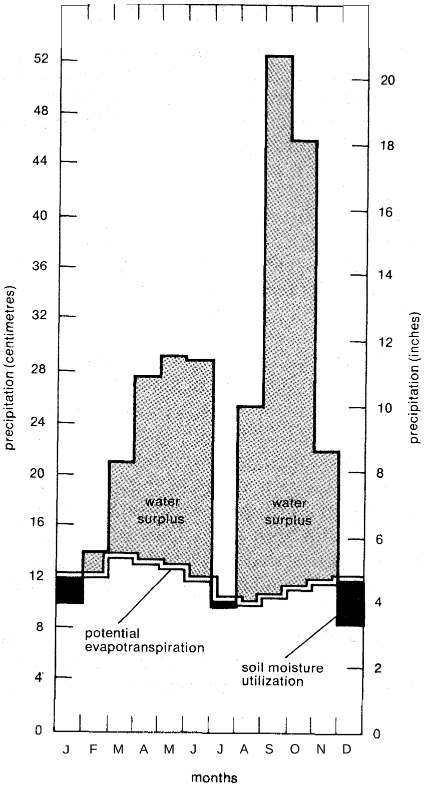


Figure 3: Water budget at Kribi, Cameroon, illustrative of an equatorial climate with two rainfall maxima.

The annual evapotranspiration in the forest zone of the Congo Basin is estimated to be 1,650 ± 100 millimetres (65 ± 4 inches).

Forest soils tend to be permanently moist and spongy, but their structure allows a great deal of infiltration. Nevertheless, tropical storms are so intense that runoff is common during rains, especially on steeper slopes. There are very great differences between the hydrological regimes in areas of different topography, especially between the plateau landscapes and the ridge and valley landscapes.

Sediment transport by streams

In ridge and valley areas large runoff produces high rates of erosion, and rivers carry a large sediment load. In flatter areas, minor streams are frequently blocked by fallen trees and associated plant growth to form swamps and lakes. This reduces the ability of the streams to transport debris and leads to sedimentation in temporary pools. Major streams, however, are open and often fast flowing, their sediment load depending on their catchments (drainage areas).

In the Amazon Basin there are "black water" streams, draining swamp-forest peats that carry black, acid water stained with organic matter but little sediment. This black water contrasts markedly with the muddy water of other streams and has the incidental advantage to explorers of being free from insects. In New Guinea, the Fly River has a comparatively small sediment load, although it is nevertheless murky and opaque; but its tributary, the Strickland, coming from a more mountainous interior, carries much greater quantities of sediment in its turbulent waters. The difference in the two streams is sufficient to give rise to different alluvial landforms and different vegetation patterns along the respective rivers.

Flooding is important in lowland rain forest. It is often of considerable range, such as the eight-metre (26-foot) variation of the Fly River, which is a great obstacle to development or permanent settlement of such areas. Such floods cover the ground and trees over many square miles. Over-bank floods spread out over the land, depositing new sediments, building levees, and adding new layers of silt to the soils of the lowland forest.

Change occurs in the water table after forest clearance. In Nigeria a rise of 60 metres (about 200 feet) has been noted in 50 years. Larger streams continued to flow in the dry season longer than they formerly did, new springs appeared, and one valley was drowned by a new lake.

Climatic changes also alter hydrologic regimes, but not always in easily predictable ways. In main valleys, the reduction in flow due to greater desiccation might be more than offset by increased flow from runoff and from accentuation of groundwater ridges.

WEATHERING AND SOILS

Tropical weathering. In hot, wet areas chemical weathering is most intense, and, in the intertropical zone, weathering of rock produces clay-rich weathering mantles up to 100 metres (330 feet) deep or even more. The distinction between soil and weathering mantle is hard to draw.

The weathered rock that is still in place retains features of the unweathered hard rock and is separated from an overlying loose mantle of more or less transported material by stone lines of varied composition and controversial origin. All of the rain forests have abundant rainfall, so on well-drained sites there is a washing out, or leaching, of dissolved chemicals. In all of the rain forests, bases such as calcium and magnesium tend to be leached out, but in the tropics there is also removal of silica, which leads to the formation of different soils under tropical and temperate rain forest.

On well-drained sites in the tropics the leaching of bases and silica leads to an accumulation of aluminum and iron oxides and hydroxides and the formation of the clay mineral kaolinite, which is relatively poor in silica (see further CLAY MINERALS). Many upland soils in the tropics consist of little more than kaolinite and iron hydroxides, which give the soil a red or yellow colour. There are two main divisions among the tropical red soils: tropical red loams and lateritic soils.

Tropical red loams

Tropical red loams (known also by other names, such as krasnozem and tropical red earth) are deep, friable, and easily tilled soils. They are very porous and freely drained at all depths, and there is little change in colour or texture of the soil material in the vertical soil profile. Clay is common, but it is so combined with the iron oxides that it behaves more like a loam and retains a porous, freely drained structure even when wet. It does not break into larger blocks, such as commonly happens in clay-rich soils of temperate regions. Soil profiles of this kind are not affected by water tables within the soil profile. In the tropical rain forest these soils are much more common than the lateritic soils.

Lateritic soils. In contrast to the uniformity of the tropical red-earth profiles, lateritic soils have profiles with well-marked horizons (layers). The uppermost horizon is quite variable but is commonly a red loam, lighter in texture than the underlying horizons and sometimes containing ironstone gravel. Beneath this is a layer rich in iron. So long as it is kept moist under a forest cover this material will remain soft, but if it is dried out it will irreversibly harden and may even be used for brickmaking. It was to such material that the term laterite was first applied, but the word has been used very loosely since then. There was in the early 1970s a tendency among soil scientists to return to this early usage, and the whole sequence of associated horizons may be referred to as a laterite. or lateritic, profile. In some profiles the laterite horizon is hard, even within the intact soil profile. Beneath this layer there may be a mottled zone and beneath that a bleached-white, kaolin-rich zone, known as the pallid zone. These zones may be many metres thick.

The formation and potential hardening of the iron-rich zone of lateritic soils are their most distinctive and important features. Hardening can come about in several ways: forest clearing by man, drying out of the soil due to climatic change, and drying of the soil due to valley downcutting and drainage change.

The iron-enriched layer occurs closer to the ground surface with increasing distance from the Equator, and north of 14" north only fossil crusts are found. In the Ivory Coast, for instance, there is a latitudinal sequence: in the south, under rain forest with over 1.5 metres (five feet) of rain, are soils that do not form crusts so long as the forest cover is intact; whereas to the north, under subtropical forest and with rainfall less than 1.5 metres, the soils are less deep, and ironstone crusts exist on dry sites.

With increasing distance from the Equator and longer dry seasons, less silica is removed, and the soils of the humid tropics give way to the ferruginous (iron-rich) soils of the savanna areas. In extremely wet and hot areas, on the other hand, even more silica is leached away, leading to the formation of aluminum-rich bauxite rather than laterite. In the humid tropics the influence of parent rock is less than in temperate areas, and lateritic soils can form on practically any parent rock.

Nevertheless, there is considerable variation. In South America, dark lateritic soils occur in humid areas or on rocks rich in calcium and magnesium, red lateritic soils on acid rocks (rich in soda and potash), and brown lateritic soils on volcanic ash or basalt. Volcanic soils are generally more fertile than others, and in regions of active volcanoes, such as parts of Indonesia, occasional eruptions produce layers of ash that act like a rich topsoil in restoring fertility. Terra rossa soils (clay-rich, red, granular soils) are common on limestone, including that of coral islands, which often have rain-forest cover.

On base-poor rocks, such as sandstones, podzols (leached soils with distinct sandy topsoils) may form. Tropical podzols are widespread in the catchment areas of the Rio Negro in Amazon, in Guyana, Malaya, Borneo, and Thailand. These are not under normal forest but under "heath forest," the padang of Southeast Asia.

On tropical mountains there is a general tendency for tropical red loam or lateritic soils to be produced on lower slopes and podzols to be formed higher up, but there are many complications due to parent material, erosion, aspect, drainage, and other factors.

Thus far, discussion has centred on the soils of leached sites, the well-drained areas on higher ground. A well-marked division is found in many tropical areas, however, with red, leached soils on the uplands and black clays, gray soils, and fresh alluvium in the lowlands. Furthermore, swamp forest with very acid soil consisting largely of peat is found in Java, Sumatra, Borneo, Celebes, Malaya, and Guyana, although it is apparently absent in Africa. This peat occurs in large areas, is often over six metres (20 feet) thick, and is composed almost entirely of the remains of trees and other woody plants of the swamp forest.

The common idea that all of the tropical soils are laterites is a gross oversimplification, and detailed investigations in New Guinea and elsewhere reveal a wide range of soil types.

Agricultural possibilities in the rain forest

Despite the luxuriant plant growth, tropical forest soils are not particularly fertile. Under intense weathering, leaf litter quickly decomposes and thus provides the nutrients essential for the trees, which absorb some of the released ions before they are leached from the soil profile. Most plant nutrients are thus recycled by the trees, and there is very little reserve of nutrients in the soil. The humus layer is normally only a few centimetres thick, and decomposition of vegetation is rapid.

Removal of tropical rain forest to allow cropping swiftly eliminates leaf litter, and the water-retaining, spongy soil structure is destroyed. The mineral cycle is broken and the soil productivity declines; laterite dries out because of increased evaporation and so accelerates the development of undesirable physical features. Clearing of lateritic soils also accelerates erosion of the meagre upper horizon, exposing the ironstone layer at the surface and facilitating the hardening process. Ironstone crusts have been known to form in as little as 30 years in West Africa from such causes.

Cleared forest is therefore of use for only a short time before it is exhausted, so primitive inhabitants of tropical forest practice shifting cultivation. If only small patches of forest are cleared, regeneration is possible, but on any large scale the changes of soil become progressively worse and irreversible, and the forest gives way to other vegetation types. In some areas there is tangled secondary regrowth, or jungle, while in others, especially in areas of climatic stress, savanna expands at the expense of forest; in New Guinea, for example, large areas of kunai grass appear to be of anthropogenic origin, replacing original forest.

Erosion in rain forests. Authorities are divided on the type and rate of erosion that occur under tropical forests. This state of affairs results partly because generalizations have been drawn from studies in restricted areas and partly because a paucity of factual data has led to a too-theoretical consideration of tropical erosional processes.

Theory of dominance of chemical weathering

One school of thought, derived largely from studies in South America and Africa, may be summarized as follows: in temperate forest it has been observed that erosion under forest is much slower than under grassland. Forest canopies intercept a considerable proportion of the rainfall, breaking the force of raindrop impact and increasing immediate loss by evaporation, and raindrops that reach the ground strike an absorbent layer of leaf litter. Some water reaches the ground quietly by trickling down the stems of trees.

Similarly, the dense, luxuriant vegetation of tropical forest should be adequate to protect the ground from mechanical erosion, enabling the buildup of thick mantles of weathering products. The thorough weathering of fine-textured rocks would leave few tools (large, hard particles) for stream erosion and would tend to produce mass movement such as landslides and soil creep. The essence of this argument is that humid tropical landscapes are dependent on the predominance of chemical weathering and mass action over fluvial erosion.

Peculiarities of river behaviour in tropical rain forests, in this view, result from the dominance of chemical weathering over mechanical weathering. Decay of the rocks involves removal of material in solution and leaves a residue consisting of fine particles, with no rock fragments and mineral particles to make river gravel and coarse sand. Running water is thus deprived of the tools of erosion. Streams fail to erode laterally or even vertically, and both small streams and large rivers flow alike in sluggish, swampy reaches separated by falls and rapids.

Quite different data and ideas have been produced by investigators in areas of steep ridges and valleys in New Guinea, Malaya, and elsewhere. This school of thought may be summarized as follows: In primary tropical forest, canopy openings due to tree fall are common, the shrub layer is sparse, and leaf litter is thin (averaging only one to three centimetres) because of the rapid decay rate. Despite the high porosity of the soils, runoff occurs because the rainfall intensity is greater than the rate of infiltration can cope with. These factors, it is held, lead to much greater runoff under tropical than under temperate forest cover.

Importance of waterdrop erosion

Much evidence exists to support this view. The effect of raindrop and waterdrop erosion is revealed by many small earth pillars (columns) on bare soil patches. In tropical rain forest many waterdrops will have a free fall of more than eight metres (26 feet) and will be close to terminal velocity when they strike the ground; these drops will be just as effective as raindrops with unimpeded fall. Waterdrop erosion is thus more important in tropical than in temperate rain forest because of the higher upper story, the greater rainfall, and the less abundant leaf litter.

Runoff floats away leaf litter, which helps the rapid decay to keep the forest floor free of debris. Although leaf litter may attain a depth of 250 millimetres (ten inches) on ridge crests, on hillslopes it is usually less than three millimetres, and often the ground has only a scattering of recently fallen leaves. A muddy suspension is carried from areas affected by erosion, and transported material is deposited at root barriers to form steps. Scoured surfaces and accumulations of lag gravel (deposits of particles too large to be moved under existing hydrological conditions) indicate that unconcentrated sheetwash is able to transport detached soil particles. Rainwater streaming down tree trunks has sufficient energy to effect erosion at the tree base, and lateral surface roots are often undermined on the downslope side. On very steep slopes, trees usually fall down before reaching maturity.

Rain-forest soils are always nearly saturated, and there is an immediate response to heavy rainfalls in streamflow. Streams rise rapidly and, armed with a coarse bed load, cut deep into bedrock. Despite the rapidity of weathering, streams carry unweathered gravels together with the detritus from landslides. In New Guinea and Malaya much gravel and rock in stream beds provide abundant erosive tools, and, in fact, only in stream beds are rock outcrops frequently found.

Sheetwash is the main form of erosion on slopes. Mass movements such as landslides and avalanches are restricted to steeper slopes; even there landslides may be rare as the soil is too thin. But in some areas, like parts of Hawaii, landslides may dominate the erosive processes and produce landforms with sharp ridges and gullies.

In the rugged country of the New Guinea rain forests, the straight slopes are mantled with shallow, loamy soils that contain a high proportion of unstable minerals from the parent rock, and there are no changes in depth of weathering up- or downslope or accumulation of soil or litter at the base of the slope. Weathering is therefore balanced by erosion, and the material eroded is transported from the area by rivers as fast as it is supplied. Such rivers carry abundant sediment and are well equipped by rocks to erode, and many valleys have a rock-cut niche at the base of the V-shaped valley.

GEOLOGICAL HISTORY OF RAIN FORESTS

Rain forests have occasionally been much more widespread in the world than they are at present. By the end of the Carboniferous Period (280,000,000 years ago), forests of tree ferns, giant horsetails, and club mosses covered vast areas, and their remains formed coal seams. Conifers made their appearance before the end of Carboniferous times, probably in tropical climatic conditions, and became dominant throughout Triassic and Jurassic times (225,000,000 to 136,000,000 years ago). During Cretaceous times (136,000,000 to 65,000,000 years ago) the angiosperms gained ground, and by Eocene times (54,000,000 to 38,000,000 years ago) trees similar to modern types were dominant in many areas.

Through the Tertiary Period (from 65,000,000 to 2,500,000 years ago) climates were generally warmer than now and forests more widespread. In early mid-Tertiary times midlatitude vegetation of North America, Europe, and Asia was like that of the humid subtropics, rather like the forest of northern New Zealand today. Indeed, the New Zealand forest may be a true descendant of the ancient and once almost universal midlatitude forests. In middle Pliocene time (about 5,000,000 years ago) deciduous trees from high latitudes invaded the middle latitudes, and conifers also spread; and in Pleistocene times (from 2,500,000 to 10,000 years ago) there was a complete transformation, culminating in the advance of the ice sheets.

In the last 1,000,000 years or so, the world's climate has been dominated by a number of ice ages. When ice caps grew at the poles, there was a general shift of climatic and vegetation belts. In the view of some authorities, however much the boundaries may vary in relation to the Equator, the actual equatorial belt tends to remain in place. In this view, the day length, insolation (incident solar radiation), and tropical convectional storm patterns remained constant, and the only variables were the total rainfall and the average temperature. It is possible that the temperature varied slightly through time but to a lesser degree than in other latitudinal zones. Some authorities have postulated pluvial (moist) periods, corresponding to glacial periods of the poles, in which rainfall was higher than at present; but in equatorial forest regions even large changes in total rainfall would make little difference because these areas already have abundant water. The forest belt may in the past have extended toward the desert by about two degrees of latitude.

Pleistocene aridity in present rain-forest areas

Other authorities have postulated drier periods, even arid periods, corresponding with the glacial periods outside the equatorial zone. Low effective-solar-radiation levels of the glacial periods should have resulted in greatly reduced evaporation from the tropical oceans, which in turn would lead to greatly reduced equatorial rainfall. Some studies in Africa, South America, and Australia, supported by radioactive-carbon dating, suggest that glacial maxima were, in fact, matched by widespread development of deserts in the tropics. During the cold–dry maxima the great rivers of central Africa almost dried up, and the now vegetated "fossil" dunes can be traced from the Sahara and from the Kalahari into the Congo. It is said that there is scarcely a jungle tree in the Congo Basin that does not have its roots in desert sand.

If there were simultaneous glaciation in both hemispheres, rain-forest belts would merely widen or narrow at the expense of adjacent deserts, but some authorities believe glaciation was asymmetrical, and the forest belt may actually have intermittently shifted north and south.

At the present time there is no doubt that the rain forests are shrinking, but this is due to the continued expansion of man's clearing and not to natural causes. In earlier times the tropical rain forests were occupied only by small numbers of primitive hunters and collectors who did not attempt to alter the forest environment. The slightly more advanced shifting cultivators clear patches of forests but do not maintain soil fertility, a wasteful and dangerous method of land use often leading to irreparable damage to the soil. More advanced communities destroy the forest at even greater rates. All too often, forests are regarded as mere "undeveloped areas" awaiting either utilization of the wood or clearance for occupation. It seems incredible that these vast, apparently interminable rain forests are in danger of significant destruction, but they must be considered a diminishing resource.

(C.D.O.)

LIFE IN THE RAIN FOREST

The tropical rain forest is the most complex ecological system (ecosystem) on earth, and its plant and animal life is richer and more diverse than that of any other type of forest. A five-acre (two-hectare) sample often has more than 100 different species of trees 30 centimetres (12 inches) or more in diameter (in the Malay Peninsula more than 200 species have been recorded), compared with about 25 tree species in the richest temperate broad-leaved forests (North America). Similar data for animals are difficult to obtain, but comparisons have been made of the number of bird species: more than 600 in a 300-mile-square area of rain forest in Panama, as against about 100 in a similar area of temperate forest of the eastern United States. The number of bird species, however, is small compared with the number of insects.

The environmental setting. *Climate.* One obvious reason for the enormous variety of plant and animal life in the rain forest is the climate, which is hot and moist throughout the year. The soil never dries out, and the humidity of the air is always high. This constantly warm and moist environment allows plants and animals to grow and reproduce all year long, though they may not in fact do so. In the rain forest the struggle for survival is, thus, less against a hostile physical environment, as in climates with a cold winter or summer drought, than against the competition of other organisms. This has led, among both plants and animals, to the evolution of a great variety of species filling many different niches, or ecological roles, some of which are highly specialized.

The constancy of physical conditions

Another and less obvious reason for the great diversity of species in the rain forest is probably its age. In the tropics the climate has changed little since the Cretaceous period, and therefore evolution has proceeded uninterrupted for some 60,000,000 years.

Forest structure and microclimates. The trees and other plants that grow together in a primary, or undisturbed, tropical rain forest may seem a chaotic mass of vegetation, but the forest, in fact, has an irregular pattern of structure not easily discernible at ground level. Tree crowns are arranged in several stories, or strata (Figure 4), called A, B, C, D, and E from above downward. These layers are generally poorly defined, so that only an arbitrary boundary can be drawn between one stratum and the next. Most of the trees in the lower strata are immature individuals of species that may later reach a higher stratum. The A stratum, of trees 30 to 50 metres (100 to 165 feet) high, is never continuous; their crowns rarely touch and may be widely separated. For this reason, the roof of the forest, as seen from the air, has a very uneven surface, the crowns of the largest trees standing out like gigantic, dark-green cauliflowers. The B-stratum trees, whose crowns are smaller and more closely spaced, likewise do not generally form a continuous layer. The trees of the A and B strata are often called the emergents; together they form the forest canopy.

From *Ecology*, vol. 19 (1933); reproduced by permission of the Duke University Press

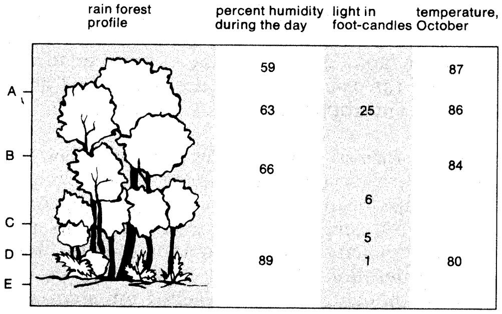


Figure 4: Stratification of a tropical rain forest (see text).

The C and D strata comprise the densest layers of the forest, where little space is unoccupied by trunks, branches, twigs, or foliage. In contrast, the lowest, or E, stratum is often thinly stocked, and much of the ground surface itself is bare except for a thin covering of dead leaves and litter. Contrary to popular belief, the undergrowth of the rain forest is seldom impenetrable: it is usually possible to walk about and to see another person at least a distance of 50 metres (about 165 feet). Only where sunlight

The scant undergrowth of jungle

reaches the ground in sufficient intensity, in clearings and on the edges of roads and rivers, does the undergrowth become the dense jungle commonly described. Similarly, the undergrowth of young secondary forests, which spring up in clearings left by timber operations or by agriculture. is much denser and more tangled than that of undisturbed primary forest.

The leaves and branches act as a filter to light and heat radiation and as a barrier to air movements, resulting in wide differences in temperature and humidity of the air in the lower stories, compared with the relatively open canopy. In the undergrowth there is very little air movement, and, apart from occasional flecks of sunlight, most of the light has passed through or been reflected from leaf surfaces. Within the forest, conditions are quite different from those in a clearing or in the open, and there is a range of "local climates" (micsoclimates).

At midday in sunny weather the temperature in the A and B strata might be 32° C (90" F), while in the undergrowth it could be 26° C (79" F). During cloudy or rainy weather the difference would be smaller, but the upper strata are always warmer than the undergrowth during the day. At night the position is reversed, and, especially on clear nights, the air around the treetops is cooler than that near the ground. Although at night the air in all strata is nearly saturated with moisture, during the day there are considerable differences in humidity between the upper strata, where relative humidity often falls to 60 percent in the afternoon of a sunny day, and the undergrowth, where the humidity rarely drops below 80 to 90 percent. The undergrowth of the tropical rain forest is in fact one of the most constant biological environments on earth, and it is not surprising that, as in caves, another unusually constant environment, some strange forms of animal life are found—some of which, such as the Ricinulei (a small group of arachnids) and the wormlike *Peripatus*, can be regarded as "living fossils."

The three tropical sectors compared

The biological component. The rain forests of tropical America, Africa, and Southeast Asia to Oceania (northern Australia and the western Pacific islands) are similar in many features, but each has a distinctive plant and animal life, with almost no species common to all three and not many common to any two of the tropical land sectors. Whole groups of organisms are confined to a single sector. For example, the Dipterocarpaceae, the most important family of trees in Indo-Malayan rain forests, is absent in the Americas and in Africa is represented by only a few small trees of the savannas. The pineapple family (Bromeliaceae), so common as air plants (epiphytes) in the American tropics, have only one species in the Old World tropics. Much the same is true of the animals: tropical America alone has monkeys with prehensile tails, sloths, toucans, and leaf-cutting ants, but it has no apes, hornbills, or very large mammals, such as elephants and rhinoceroses, all of which are confined to tropical Africa and Asia. Often a group in one sector replaces one of another; thus, the flower-visiting sunbirds of the Old World occupy a niche similar to that of the hummingbirds of tropical America.

Plants. The tropical rain forest is overwhelmingly dominated by trees and woody-stemmed vines (lianes, or lianas). Woody plants are the chief producers of the food on which all the animals depend and of the shelter in which they live. Most of the nonwoody plants are epiphytes living on the trunks and branches of the trees, but a few, including grasses, grow on the ground.

Rain-forest trees range from giants 60 metres (almost 200 feet) high down to treelets of not more than one metre (just over three feet). The great majority belong to the large group of flowering plants known as dicotyledons, but palms and other monocotyledons occur in the lower stories and in open areas. Conifers are found only in certain parts of the rain forests of Borneo, New Guinea, and neighbouring areas. Although belonging to a very large number of different families, rain-forest trees are remarkably uniform in appearance, the majority having thick, leathery leaves, like those of temperate laurels in size and shape. Most are evergreen, but some are deciduous, remaining leafless for a few days or weeks. Because the behaviour of the different plants is not closely synchronized, the forest as a whole always appears in full leaf. The young leaves of the trees are often bright red and hang limply, as if wilted.

Shallowness of root systems

Flangelike buttresses, extending up the trunk for three to four metres (ten to 13 feet) and outward for about the same distance, are characteristic of many of the larger trees, and stiltlike aerial roots are not uncommon among the smaller trees. Root systems in general are shallow—even those of the largest trees seldom penetrate deeper than a metre—and most of the fine nutritive roots are concentrated in the humus-rich layer within a few centimetres of the soil surface.

Large, brilliantly coloured flowers are not common in the rain forest; those of many trees are inconspicuous and white or greenish. The production of flowers on the main trunk, rather than on the twigs or branches, is characteristic of many small- to medium-sized trees; this feature may be related to pollination by bats or other animals that cannot easily reach flowers hidden within a mass of twigs and leaves.

Lianas are plentiful, especially where clearings are formed by the death of old trees. These woody vines reach the tops of all but the tallest trees, linking them together and competing with them for space and light. In the eastern tropics, climbing palms (rattans) are common and grow to a length of 100 metres (30 feet) or more. Strangling plants, which include many members of the fig genus (Ficus) and, in the American tropics, species of *Clusia*, also compete with the trees. These plants start life as epiphytes on tree branches to which their seeds are carried by birds; later they send aerial roots to the ground and enclose the host trunk in a network of stout, woody roots that may kill it, leaving the strangler as an independent tree.

Epiphytes are commonest on branches 30 metres (about 100 feet) or more above ground in positions fully exposed to the sun. Common rain-forest epiphytes include orchids, bromeliads, and specialized ferns as well as lichens, mosses, and liverworts that grow on the surface of leaves in the humid, shaded undergrowth. Most epiphytes do not harm the host tree but simply depend on it for physical support. They obtain mineral nutrients from rainwater and from the organic debris that collects among their roots, often brought there by ants. Many epiphytes have special adaptations, such as bracket- or pitcher-shaped leaves that collect water and humus. The most remarkable of such adaptations are the "tanks" of bromeliads, formed by closely overlapping leaf bases. These may hold several litres, or quarts, of water and provide a home for many kinds of aquatic organisms, including mosquito and other insect larvae, tadpoles, and even crabs.

Many rain-forest plants are not green and so cannot make their own food; they live either on decaying matter as saprophytes or on other living plants as parasites. Although most of these nongreen plants are fungi, a few are specialized flowering plants. One of the most spectacular of all rain-forest plants is the Malaysian monster flower (*Rafflesia*), which is a parasite of certain woody vines; its flowers may be a metre (over three feet) in diameter.

Animals. The variety of animals even in a small area of tropical rain forest is so vast that a large proportion of species of all of the groups, except the vertebrates, are still unnamed and unclassified, a formidable obstacle to research in rain-forest ecology. All groups of vertebrates except fishes are represented and almost all groups of terrestrial invertebrates. Even groups that are normally aquatic, such as planarians (flatworms) and polychaete worms, are able to live in the continually moist undergrowth of the rain forest. Land leeches that attach themselves to passing animals (including man) ale one of the most troublesome pests of the rain forests of the Philippines and other parts of the eastern tropics.

Wide representation of animal groups

Rain-forest mammals range from such large creatures as the elephant (of Africa and Asia) and forest rhinoceros (of Malaysia) down to tiny mice and shrews. In tropical America the range of size is less, because there are no mammals larger than jaguars and tapirs. A tendency to gigantism is shown among rain-forest invertebrates, which include the giant snail of Africa, "bird-eating" spiders, and the goliath beetle. The splendid ***Ornithoptera*** butterflies of Malaysian forests, the blue ***Morpho*** butterflies of tropical American forests, and the huge Atlas moth of Asian forests are among the largest Lepidoptera.

Although the animals of the tropical rain forest are abundant and many striking and colourful in appearance, often few can be seen except for some butterflies and bees, an occasional hummingbird, ground squirrel, or armadillo, and the omnipresent ants. One reason for this is that a large proportion of the animals, including most of the birds, live in the trees and rarely come down to ground level. Others are extremely well camouflaged; *e.g.*, the leaf butterflies, which are virtually invisible when they rest on dead leaves; green tree snakes, which hide among the foliage; and the numerous shy, brownish birds and mammals, which are well concealed until they move.

Another reason for the apparent scarcity of animal life is that many of the rain-forest animals are nocturnal, including such mammals as lemurs and tarsiers, such birds as nightjars and owls, most toads and frogs, and most moths, as well as a host of other insects of many different families. All of these animals become active only after dark, when the temperature is lower and the dangers from water loss and predators less than in the day. Many nocturnal animals are highly adapted to their way of life; the large saucer eyes of tarsiers, bushbabies, and owls and the luminous organs of fireflies and lightning bugs are a few examples of the ways in which nocturnal animals are adapted to their way of life.

Often more evident than the animals themselves are their habitations. Birds' nests are seldom easy to find, but the beautifully constructed nests of wasps, bees, ants, and termites are everywhere to be seen. Termites, which are among the most abundant of all of the forms of life in the rain forest, play a vital part in removing deadwood and plant debris. During the day they remain within their nests (termitaria), and in covered passages of cemented wood and soil particles. Termitaria are found both on trees and rising from the ground as fantastic, spired or conical structures sometimes a metre or more in height. Both types of nests often have highly developed rain-shedding devices: in the tree nests, chevron-like arrangements of ridges direct the water away, and in some free-standing termitaria superposed "hats," somewhat like the roof of an African grass hut, drain water away.

Animal stratification

The animals of the tropical rain forest, like the plants, are stratified, but, because of the difficulty of exploring the forest above ground level, zoologists usually divide the forest merely into canopy and undergrowth, or into top, middle, and lower layers, rather than into the five strata mentioned earlier.

In the canopy not only is there more open space but, because of the higher light intensity and greater photosynthetic activity of the foliage, more young leaves, fruits, and flowers are available as food. In the undergrowth there is less variety of plant food, though leaves and wood that continually fall from above provide a plentiful source of food for some animals and plants.

For the larger animals the canopy provides opportunities for flying, gliding, and leaping, as well as for climbing and running along branches; both sight and hearing are important senses in seeking food and escaping predators. The life of an undergrowth animal is quite different. Running, fluttering, hopping, and climbing are the chief modes of locomotion, and, because the animals tend to be camouflaged, communication between them is more by sound than by sight. Many undergrowth birds—for example, the cotingas and bellbirds—have dull, concealing colours but surprisingly loud, clear calls.

The American naturalist William Beebe was one of the first to study the stratification of animals in a tropical forest. In Guyana he recognized five height "zones," each mainly inhabited by different mammals and birds. More recent studies in Malaya, where the forest animals are similarly stratified, have shown that the animals of each level tend to differ in both their mode of locomotion and their feeding habits. For example, the mammals living on the ground include large species, such as the elephant and rhinoceros, with no climbing ability, and smaller species that climb only to a limited extent. This group includes herbivores, which browse on leaves and eat fallen fruits (elephant, deer), mixed feeders (Malayan bear), and carnivores (tiger, leopard). The inhabitants of the middle levels and canopy have little contact with the ground; they climb, leap, or swing from branch to branch (gibbons, monkeys) or glide for long distances (flying squirrels, flying lemur). Mostly they are fruit and insect eaters, though some are carnivorous (clouded leopard, marten). The birds of different strata show similar differences in their movements and feeding habits.

The animals of the rain forest show many striking structural adaptations to their mode of life, those of the treetops, in particular, having evolved special characteristics not found in those living nearer the ground; ***e.g.***, the winglike skin flaps of the flying squirrels and lemurs, the prehensile tails of the New World monkeys, the peculiar limbs of sloths (which allow them to hang from branches upside down), the stiff tails of woodpeckers (used as an aid to climbing), and the beaks of parrots (well-adapted to cracking nuts and to assist in climbing).

Adaptations to the requirements of climbing in different strata are particularly well shown among rain-forest primates. The only forest primates that live mainly on the ground are the chimpanzees and gorillas of Africa; they have limbs adapted to walking and climbing but not to venturing high in the trees. The tarsiers, galagos, and lemurs, found mostly in the small trees of the B and ***C*** strata, have short legs and arms; they climb well, run, and make short jumps. The primates of the treetops include the orangutan and the athletic gibbons of Malaysia and monkeys of many kinds. The orangutan and the gibbons have very long arms with which they swing from branch to branch, while the monkeys are better adapted to running along branches and leaping from tree to tree. In feeding habits the different groups of primates are also well adapted to the strata in which they mainly live.

Specialized tree-bound animals

Among the animals that rarely come down to earth are certain frogs and toads, which must obtain water for their tadpole stage. Some lay their eggs in water held in holes in trees or by epiphytic bromeliads; in others the tadpole stage is abbreviated, and the developing eggs are carried attached to a moist membrane on the mother's back. One Central American species lays its eggs in packets fixed to branches of trees overhanging streams, so that the tadpoles fall into the water when they hatch.

Rotholes and bromeliad "tanks" also provide treetop habitats for mosquitoes and other insects with aquatic larvae. Mosquitoes that breed in the treetops are of much practical importance because some are carriers of malaria, yellow fever, and other diseases of man and animals. Malaria-carrying ***Anopheles*** mosquitoes breeding in forested regions of tropical America sometimes come down to ground level and become a source of infection for people living in the area. Similarly, ***Aedes*** mosquitoes, which carry yellow fever, normally bite monkeys living in the treetops in the forests of Africa and the Amazon, but when trees are felled these mosquitoes may bite lumbermen and initiate epidemics among susceptible human populations.

There are many rain-forest inhabitants normally found below the soil surface to a depth of a metre or more (three or four feet). This underground community includes burrowing vertebrates, such as armadillos, which also spend much time above ground, and others, such as caecilians (wormlike amphibians), worm snakes, and many more, which are wholly subterranean. Much more important than these, however, is the host of small soil invertebrates (worms, mites, insects, etc.) that, together with fungi and micro-organisms, play a large part in decomposing dead vegetation and freeing the elements contained therein for recirculation in the forest ecosystem.

The year-round warm and moist conditions in the rain forest are of great significance for animals as well as for plants. In general, rain-forest animals are unrestricted in their activities by seasonal changes of climate. The environment demands no prolonged rest periods—neither hibernation nor estivation—and resting stages of insects are either absent or very short, with little relation to the time of year.

A very important aspect of the environment for animals is that food is always available. Foods such as leaves, flowers, and fruit may vary in abundance but are never quite unobtainable as in a temperate-zone winter. Many forest birds depend exclusively on fruits, and bats as well as hummingbirds and sunbirds have evolved flower-visiting habits.

Some animals actually reproduce at all times of year. Some African forest birds, such as fruit pigeons and flycatchers, breed at all of the seasons, and some tropical butterflies, such as the swallowtails, have a succession of broods throughout the year rather than the one or two as in temperate Europe and America. Although breeding is always possible, the majority of animals have fairly definite breeding periods; these periods, however, are not synchronized for different species, so there is no time at which some birds, mammals, and insects are not reproducing. The breeding season of many rain-forest birds and mammals seems to be related to variations in abundance of their chief foods.

Reproduction tends to take place more frequently or over a larger part of the year than in temperate animals, and there is a tendency for fewer young to be born in each brood. This is especially true of tropical birds, which lay fewer eggs at a time than do temperate birds.

Complex association forming a stable ecosystem

Relationships between organisms. The very large and diverse populations of organisms in a tropical rain forest interact with each other and with their environment to form an extremely complex but very stable ecosystem. As in any other natural community, the basic relationships between the constituent organisms are the food and energy links. Radiation from the sun provides energy for the photosynthesis of trees and other chlorophyll-containing plants. These are eaten by herbivorous animals, the most important of which in the rain forest are probably the insects. These primary consumers are the food of secondary consumers, such as insectivorous birds, which are in turn eaten by others, so that there are food chains with up to four or five "links" (trophic levels). The food chains end with such predators as jaguars, eagles, and owls, all of which are relatively scarce and, because of their size and large food requirements, rather unselective in their diet. Most rain-forest plants and animals have parasites, which may in turn be attacked by parasites (hyperparasites), establishing another kind of food chain.

An important difference between a tropical forest and other ecosystems is that the food chains (or food webs because they are interconnected) are so many and so complex. A further characteristic of the rain forest is that within the main ecosystem with its food webs, there are small subsidiary systems, which are partially isolated. One subsidiary system is that of the bromeliad tanks in tropical American rain forests. In these the chief sources of food are plant debris falling into the tanks or brought there by ants and production of food by populations of certain green algae. The consumers are worms, insect larvae, tadpoles, etc., some of which remain in the tanks, whereas others, such as frogs and mosquitoes, leave them when mature to find food elsewhere in the forest.

Another subsystem consists of the large number of organisms dependent in one way or another on ants. Many kinds of insects live in ants' nests in various relationships with their hosts, the nature of which is often not clear. Some large predators, such as anteaters, feed largely on ants and termites, while ant birds (a characteristic group of thrushes and other species) follow the migratory army ants of tropical America and the similar driver ants of Africa as they advance through the forest, eating not the ants themselves but the insects and small vertebrates disturbed by them.

There are also many kinds of ant–plant relationships, the most remarkable perhaps being that between green plants and the tropical American leaf-cutting ants. These ants bite out small semicircular pieces from leaves and carry them in procession into the recesses of their mound-like nests. Special fungi, which grow on the leaf bits, produce knoblike outgrowths on which the ants feed.

Another kind of ant–plant relationship found in all of the tropical rain forests is that between certain types of ants and certain species of plants in which they nest, known as myrmecophytes ("ant plants").

Stratification of pollinators and seed dispersers

Among rain-forest trees, in contrast to temperate trees, pollination by wind is uncommon, and insects, birds, or bats are generally the pollinators. Specialized coadaptations between flower and pollinator have often evolved, the most extraordinary being perhaps the fantastically elaborate pollination mechanisms in some tropical orchids. Several unusual features common among rain-forest trees, such as flowers borne on the main trunk or hanging from the branches on long cordlike stalks, seem to be adaptations to bird or bat pollination.

Animals of many kinds play a part in dispersing the seeds of many rain-forest trees and vines, though some, especially species colonizing gaps and clearings, have wind-borne fruits. Trees of the A and B strata often have very heavy fruits that crash through the foliage beneath and are subsequently distributed by rodents and other small animals. Other trees have edible nuts or berries, which are dispersed by parrots, monkeys, and other treetop inhabitants.

Many forest plants have evolved protective features that may make them distasteful or toxic. It is significant that the seeds of the Fabaceae (Leguminosae), a family to which many rain-forest trees and vines belong, often have bitter seeds avoided by insect larvae. Similarly, the leaves of the Rubiaceae, the largest family of tropical plants, often contain poisonous or bitter compounds; it can hardly be a coincidence that few caterpillars are known to feed on them.

Biological productivity. The luxuriance of the tropical rain forest and the speed with which clearings are invaded by secondary growth suggest that the rates of photosynthesis, growth, and other physiological processes in tropical plants must be high compared with those in temperate climates. To test such impressions, estimates of biomass (total quantity of organic material per unit of land area) and of the amount of new material added yearly (biological productivity) are required.

A few estimates for some primary rain forests indicate that from 300 to 400 metric tons (330 to 440 tons) of plant material occupy a hectare (2.5 acres), which is greater than the biomass of most temperate forests but less than that of the redwood forests of California. It is probable that the biomass of the forests of the Malay Peninsula, however, reaches as much as 500 tons per hectare.

In every case, over 80 percent of the biomass consists of woody stems and branches, leaves forming only a small proportion of the total. The organic material available as food for herbivorous insects and other leaf feeders is thus a small fraction of the whole. If the biomass can be regarded as the forest's capital, the annual increment of organic material is its income, and this latter depends on several factors, of which the most important is the amount of available energy. More radiation falls on the tropical zone than on any other part of the earth. In the humid tropics the radiation actually reaching the earth's surface is about double the amount received at 60° north, the latitude of Leningrad, and the northern tip of Newfoundland. It might, therefore, be expected that the tropical rain forest would produce more organic matter per unit of land surface than any other kind of vegetation, but this does not in fact seem to be so.

Surprisingly low organic productivity

A reliable estimate of organic production in the tropical rain forest of the Ivory Coast (West Africa) gave the figure of nine metric tons per hectare per year, not much higher than the productivity of a beech wood in Denmark. For several reasons this forest was perhaps untypical, but estimates for other tropical areas do not suggest much higher figures.

If it is indeed true that tropical forests are not much more productive than those in temperate climates, the explanation may lie mainly in the high respiration rates of tropical plants. The total amount of organic material produced by a tropical tree in a year may be prodigious (about two or three times that produced by a temperate tree), but much of it is unavailable for plant growth because it is lost in respiration. It seems that, generally, tropical plants lose through respiration a larger proportion of the carbohydrate material they build up during the day than do temperate plants.

The tall trees of the forest, whose crowns are more or less fully exposed to the Sun, photosynthesize more actively than those in the shaded undergrowth. In Cambodia the A stratum contributes more than half the productivity of the forest, though its foliage forms only about a fifth of the total leaf area.

The rate of growth in girth and height, like the productivity, of primary rain-forest trees does not seem to be unusually high. The evidence suggests that primary rain-forest trees do not, in general, grow as rapidly as the fastest growing temperate trees, such as poplars, eucalyptus, and some pines. Very high growth rates, however, are found in trees of young secondary rain forests, such as the *Cecropia* of tropical America and the African musanga (parasol tree). The latter can reach a height of over 20 metres (65 feet) in 20 years. Such trees do not, in fact, grow as fast as such temperate plants as the sunflower under comparable conditions. Temperate plants, however, grow fast only during a short summer season, whereas the growth rates of trees such as the *musanga* are maintained through the year.

Rapid turnover of nutrients

At tropical temperatures, decomposition is normally much faster than under temperate conditions. As a result, plant and animal remains do not accumulate as litter or as humus incorporated in the soil. Studies in the Thailand rain forest showed that more than 80 percent of the carbon present is in the form of living wood and leaves and less than 20 percent as dead material on and in the soil. By contrast, in a northern coniferous forest more than half the carbon may be present as organic matter in the soil or as litter on its surface.

Elements released from the dead organic material by decomposition are almost immediately reabsorbed by tree roots, often with help from fungi living in close association with them. The very small amounts of nutrients washed down into the soil by the rain or carried away by streams are replaced by additional nutrients set free by the weathering of soil minerals. There is thus an almost closed cycle of mineral nutrients, and a rain-forest soil is able to support a luxuriant vegetation and abundant animal life in a permanent equilibrium with its environment. Because the loss of nutrients is so small, the concentration of mineral elements in the water of rivers, such as the Amazon, that drain large rain-forest areas is scarcely higher than in rainwater.

In the undisturbed rain forest the greater part of the mineral resources is locked up in the biomass itself, but there is a drastic change if the forest is felled. When clearings are made by the slash-and-burn system—the commonest method of growing food crops in rain-forest areas—good harvests are seldom obtained after the first year. When the vegetation is burned, the minerals contained in it are suddenly released in soluble form, and a large proportion is washed away in surface erosion or carried down into the soil and drained away later to the rivers. The fertility of the soil can be restored only by a fallow period of 10 or more years, during which secondary forest covers the old field. Thus, new forest clearings have to be made almost every year.

(P.W.R.)

BIBLIOGRAPHY

Physical aspects: P. BIROT, *Le Cycle d'drosion sous les différents climats* (1960; Eng. trans., *The Cycle of Erosion in Different Climates*, 1968); J. TRICART and A. CAILLEUS, *Le Modelé des région chaudes, foréts et savanes* (1965); and B.P. RUXTON, "Slopewash Under Mature Primary Rain Forest in Northern Papua," in J.N. JENNINGS and J.A. MABBUTT (eds.), *Landform Studies from Australia and New Guinea* (1967), all discuss landforms and their origins in rain-forest areas; J. BUDEL, "Klima-Genetische Geomorphologie," *Geogr. Rdsch.*, 15:269–288 (1963), discusses the effect of past climatic changes on tropical geomorphology; R.W. FAIRBRIDGE, "Savanna Landscape," C . ~ COTTON, "Relict Landforms," D.S. SIMONETT, "Selva Landscapes," and H.F. GARNER, "Tropical Weathering and Relief," in R.W. FAIRBRIDGE (ed.), *The Encyclopedia of Geomorphology* (1968), several relevant articles; E.C.J. MOHR and F.A. VAN BAREN, *Tropical Soils* (1954), treats the soils of the rain forest.

Vegetation and animal life: E. AUBERT DE LA RUE, F. BOURLIERE, and J.P. HARROT, *Tropiques* (1954; Eng. trans., 1957), a beautifully illustrated account of plant and animal life in the tropics for the general reader; MARSTON BATES, *Where Winter Never Comes* (1952), this and several later works by the same author are excellent introductions for the general reader to rain-forest ecology: K. HOZUMI *et al.*, "Plant Biomass" and "Photosynthetic Production in an Evergreen Seasonal Forest," in *Nature and Life in Southeast Asia*, ed. by T. KIRA and K. IWATA (1969), these and other papers in the same periodical deal with biological productivity in the rain forests of Southeast Asia; NYE and GREENLAND, *The Soil Under Shifting Cultivation* (1960), valuable data on productivity and the mineral cycle in African forests with special reference to shifting agriculture; H.T. ODUM and R.F. PIGEON (eds.), *A Tropical Rain Forest* (1970), an account of the El Verde forest in Puerto Rico, including the effects of gamma radiation; P.W. RICHARDS, *The Life of the Jungle* (1970), a lavishly illustrated popular account of rain-forest ecology, and *Tropical Rain Forest* (1952, reprinted 1964), a standard work on the tropical rain forest; J.P. SCHULZ, *Ecological Studies on Rain Forest in Northern Suriname* (1960), a work that gives much information on the vegetation, microclimates, and soils of the Suriname rain forest.

(C.D.O./P.W.R.)

Jupiter

Jupiter, symbol ♃ in astronomy, is the most massive of the planets and the fifth in distance from the Sun. When ancient astronomers named the planet Jupiter for the ruler of the gods in the Greco-Roman pantheon, they had no idea of the planet's true dimensions, but the name is appropriate, for Jupiter is larger than all of the other planets combined. It has a narrow system of rings and 16 known satellites, one larger than the planet Mercury and three larger than our Moon. Jupiter also has an internal heat source; *i.e.*, it emits more energy than it receives from the Sun. This giant also has the strongest magnetic field of any planet, with a magnetosphere so large that it would exceed the apparent diameter of our Moon if it could be seen from Earth. Jupiter's system is the source of intense bursts of radio noise, at some frequencies occasionally radiating more energy than the Sun.

Knowledge about the Jovian system grew dramatically during the 1970s. The new information came in part from Earth-based observations, but especially from two sets of spacecraft—Pioneers 10 and 11 in 1974–75 and Voyagers

Table 1: Basic Data

Orbital characteristics		
Semimajor axis	5.20 a.u. (483.4 X 10^6 miles)	
Sidereal period	11.86 years	
Eccentricity	0.049	
Inclination	1°19′	
Physical elements		
Mean apparent diameter	47 seconds of arc	
Equatorial radius	71,400 km	= 11.2 (Earth = 1)*
Polar radius	66,770 km	= 10.4 (Earth = 1)
Ellipticity $\left(\frac{R_e - R_p}{R_e}\right)$	0.065	= 18 (Earth = 1)
Mass	18.99 X 10^{29}	= 317.8 X Earth's mass
		$= \frac{1}{1047.39}$ X Sun's mass
Mean density	1.33 g/cm^3	= 0.24 X Earth's density
Surface gravity	2,288 cm/s^2	= 2.65 X Earth's gravity
Escape velocity	59.5 km/sec	= 5.45 X Earth's escape velocity
Rotation periods		
System I (±10" from equator)	9h50m30s	
System II (higher latitudes)	9h55m40s	
System III (magnetic field)	9h55m29s	
Inclination of equator to orbit	3°04'	
Miscellaneous data		
Dimensions of Great Red Spot	26,000 km X 14,000 km	
Mean apparent visual magnitude	−2.55	
Magnetic field strength at equator	4.3 gauss	= 13.8 X Earth's surface field

*Radius of Earth = 6,378 km or 3,963 mi

1 and 2 in 1979 (see Figure 1). The Pioneer spacecraft served as scouts for the Voyagers, showing that the radiation environment of Jupiter was tolerable and mapping out the main characteristics of the planet and its environment. The more sophisticated instrumentation on the later spacecraft then filled in the details, providing so much new information that it was still being analyzed in the early 1980s.

Basic physical and orbital data

Table 1 shows physical and orbital data for Jupiter. Of special interest are the low mean density of 1.33 grams per cubic centimetre (g/cm^3)—in contrast with Earth's 5.5 grams per cubic centimetre—coupled with the large dimensions and mass, and the short rotational period. The low density and large mass suggest that Jupiter's composition and structure are quite unlike those of Earth and other inner planets (the planets between Jupiter and the Sun), a supposition supported by detailed investigations of its atmosphere and interior.

Three rotational periods have been established. The two periods labelled Systems I and II (Table 1) are mean values and refer to the speed of rotation at the equator and at higher latitudes, respectively, as exhibited by features observed in the planet's visible cloud layers. Jupiter has no solid surface; the transition from the atmosphere to the core occurs gradually at great depths. This variation in rotation period at different latitudes does not imply, therefore, that the solid body of the planet rotates with either of these mean velocities. In contrast, the apparent constancy of the period deduced from observations at radio wavelengths, System III, during 20 years of observation, suggests that it is associated with the planet's magnetic field, formed deep in Jupiter's liquid or semisolid interior.

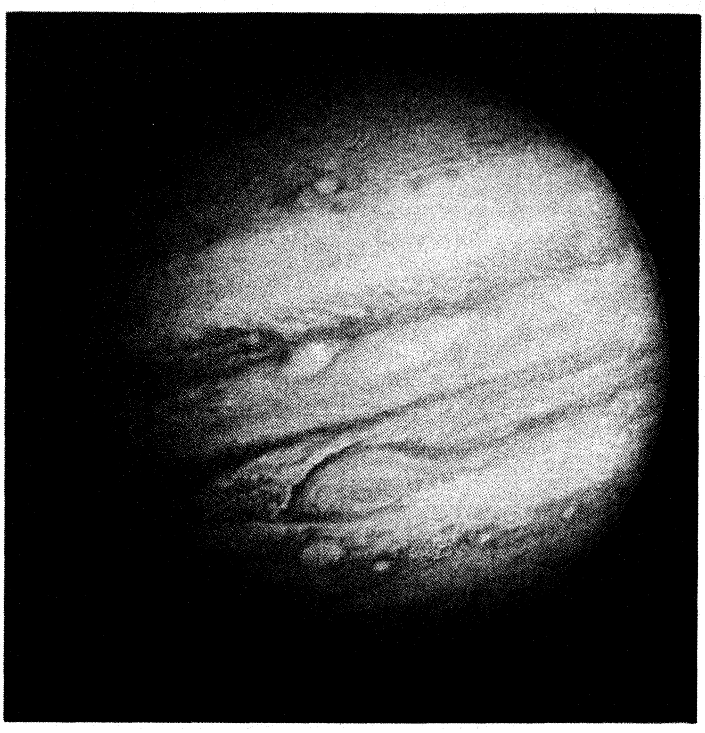


Figure 1: Photograph of Jupiter taken by Voyager 1 on February 1, 1979, at a range of 32,700,000 kilometres (20,300,000 miles).

By courtesy of the Jet Propulsion Laboratory/National Aeronautics and Space Administration

THE OUTER LAYERS

The clouds and the Great Red Spot. Even a modest telescope can show much detail on Jupiter. The region of the planet's atmosphere that is seen from Earth contains several different types of clouds that are separated both vertically and horizontally. Changes in these cloud systems can occur in a few hours, but an underlying pattern of latitudinal currents has maintained its stability for decades. (It has become traditional to describe the appearance of the planet in terms of a standard nomenclature for alternating dark belts and bright zones. The currents, however, seem to have greater persistence than this pattern.)

Various cloud forms

The close-up views of Jupiter from the Voyager spacecraft revealed a variety of cloud forms, with a predominance of elliptical features reminiscent of cyclonic and anticyclonic storm systems on Earth. All of these systems are in motion, appearing and disappearing on time scales dependent upon their sizes and locations. Also observed to vary are the pastel shades of various colours present in the cloud layers—from the tawny yellow that seems to characterize the main layer, through browns and grays, to the well-known salmon-coloured Great Red Spot, Jupiter's largest, most prominent, and longest-lived feature (see Figure 2). The vertical and horizontal segregation of the cloud systems is evidently accompanied by chemical differences as well.

Jovian meteorology can be compared with the global circulation of the Earth's atmosphere. On Earth huge spiral cloud systems often stretch over many degrees of latitude and are associated with motion around high- and low-pressure regions. These cloud systems are much less zonally confined than the cloud systems on Jupiter and move in latitude as well as longitude. Local weather on Earth is often closely tied to the local environment, which in turn is determined by the varied nature of the planet's surface.

Jupiter has no solid surface, hence no relief features, and the planet's large-scale circulation is dominated by latitudinal currents. The lack of physical boundaries on Jupiter's surface makes the persistence of these currents and their associated cloud patterns all the more remarkable. The Great Red Spot, for example, moves in longitude with respect to all three of the rotation systems, yet it does not move in latitude. The three white ovals found at a latitude just south of the Great Red Spot exhibit similar behaviour; white ovals of this size are found nowhere else on the planet. The dark brown clouds, evidently holes in the tawny cloud layer, are found almost exclusively at latitudes near +18°. The blue-gray or purple areas, from which the strongest thermal emission is detected, only occur in the equatorial region of the planet.

The nature of the Great Red Spot. The true nature of Jupiter's unique Great Red Spot was still unknown by the early 1980s, despite extensive observations by the instruments on the Voyager spacecraft. On a planet whose cloud patterns have lifetimes often counted in days, the Great Red Spot has survived as long as detailed observations of Jupiter have been made—at least 100, and perhaps 300, years. There is some evidence that the spot may be slowly shrinking, but a longer series of observations is needed to confirm this suggestion. Its present dimensions are about 26,000 by 14,000 kilometres (16,150 by 8,700 miles), making it large enough to accommodate, side by side, two planets the size of Earth. These huge dimensions are probably responsible for the feature's longevity and possibly for its distinct colour.

The rotation period of the Great Red Spot with respect to the rotation of Jupiter itself shows a variability that has not been successfully correlated with other Jovian phenomena. Earth-based observations in 1966 and 1967 revealed the counterclockwise circulation of the material within the spot itself to have a period of 12 days. This period was confirmed by the Voyager observations, which recorded a large number of interactions between the Great Red Spot and much smaller disturbances moving in the current at the same latitude (Figure 2). The Voyager pictures showed the interior of the spot to be remarkably tranquil, with no clear evidence for the expected upwelling (or divergence) of material from lower depths.

The Great Red Spot, therefore, appears to be a huge anticyclone. a vortex, or eddy, whose lateral dimensions are greater than the Earth's diameter. This lateral size is presumably accompanied by a huge vertical extent that allows the feature to extend well below and well above the main cloud layers. The extension above the main clouds can be observed directly, and it also is manifested by lower temperatures and by less gas absorption above the Great Red Spot than at neighbouring regions on the planet.

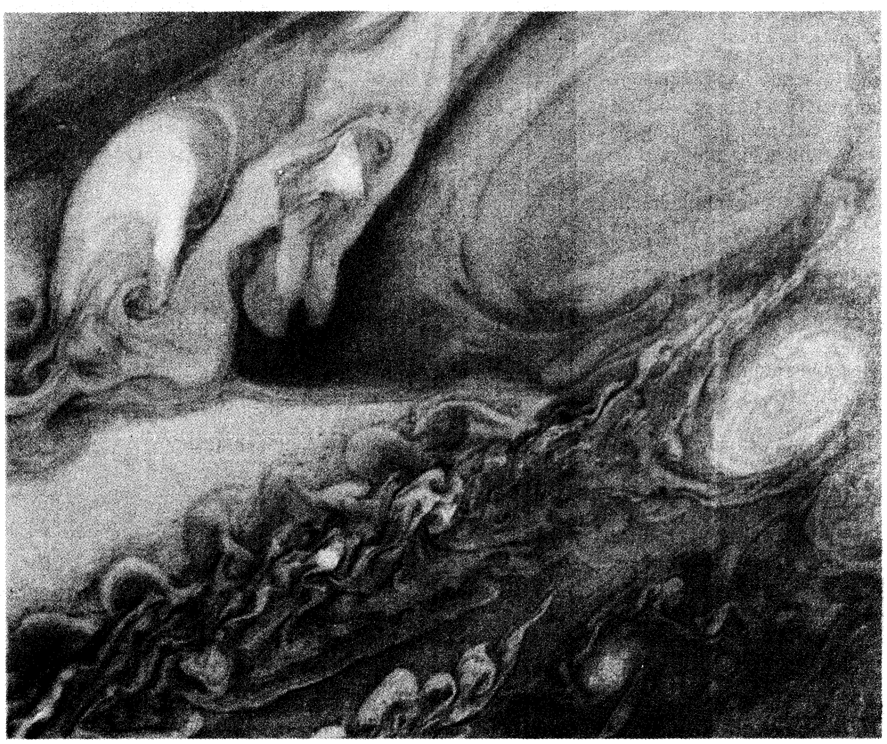


Figure 2: The Great Red Spot (top right) and the surrounding region photographed by Voyager 1 on March 1, 1979. At centre right is one of the white ovals visible from Earth

By courtesy of the **Jet** Propulsion Laboratory/National Aeronautics and Space Administration

The tawny clouds

Cloud composition. The clouds that are recorded in pictures of Jupiter are formed at different altitudes in the planet's atmosphere. Except for the top of the Great Red Spot, the white clouds are the highest, with cloud-top temperatures of about 140 kelvin (140 K; −133" C). The white clouds apparently consist of frozen ammonia crystals and are thus analogous to the water-ice cirrus clouds in the Earth's atmosphere. At lower levels occur the tawny clouds that are widely distributed over the planet. They appear to form at a temperature of about 220 K (−53° C), which suggests that they probably consist of condensed ammonium hydrosulfide (NH_4SH) and that their colour may be caused by other ammonia-sulfur compounds such as ammonium monosulfide, $(NH_4)_2S$, or ammonium polysulfides, $(NH_4)_2S_x$ (x may equal 2,3, . . .).

The reasons for invoking sulfur compounds as likely colouring agents (chromophores) lie in the relatively high cosmic abundance of sulfur and the absence of any evidence of hydrogen sulfide (H_2S) in Jupiter's atmosphere. Jupiter is primarily composed of hydrogen and helium. Under equilibrium conditions—allowing all of the elements to react with one another at an average Jovian pressure and temperature—a predominance of methane (CH_4), ammonia (NH_3), water (H_2O), and hydrogen sulfide over other compounds involving these abundant elements would be expected. This outcome is true for carbon, nitrogen, and oxygen, but no form of sulfur has been directly detected on Jupiter. This can be understood if the presence of sulfur compounds—formed in part because of the ease with which hydrogen sulfide can be broken down by solar ultraviolet light—in the lower clouds is postulated. In this hypothesis the sulfur made available by photodissociation is free to combine with the more abundant ammonia to form tawny chromophores.

Sulfur compounds have also been proposed to explain the dark brown colouration of the clouds detected at still lower levels, where the measured temperature is 260 K (−13" C). These clouds are seen through what are apparently holes in the otherwise ubiquitous tawny clouds. They appear bright in pictures of Jupiter that are made from the thermal radiation detected at a wavelength of five micrometres.

The colour of the Great Red Spot has been attributed to the presence of complex organic molecules, red phosphorus, or yet another sulfur compound. All of these ideas find support in laboratory experiments, but there are counter arguments in each case. Dark regions occur near the heads of white plume clouds near the planet's equator, where temperatures as high as 300 K (27° C) have occasionally been measured. Despite their blue-gray appearance, these dark features have a reddish tint. They may be clear gas exhibiting a blue colour (from Rayleigh scattering) overlain with a thin haze of reddish material. That these regions occur only at the equator, the elliptical dark brown clouds only near +18°, and the most prominent red colour on the planet only in the Great Red Spot implies a localization of cloud chemistry that is particularly puzzling in such a dynamically active atmosphere.

Localized cloud chemistry

At still lower depths in the atmosphere, astronomers expect to find water-ice clouds and water droplet clouds, both consisting of dilute solutions of ammonium hydroxide (NH_4OH). These cloud layers are expected to be reached by a probe to be sent into the Jovian atmosphere by the Galileo Project planned for the late 1980s.

The atmosphere. The proportions of constituents. Until a probe has entered Jupiter's atmosphere, studies of the planet's spectrum must be relied upon to provide information about the composition, temperature, and pressure of the atmosphere. In this technique light or thermal radiation from the planet is spread out in wavelengths (colours, in visible light, as in a rainbow) by the dispersing element in a spectrograph. The resulting spectrum indicates that there are discrete intervals at which energy has been absorbed by the constituents of the planet's atmosphere. By measuring the exact wavelengths at which this absorption takes place and comparing the results with spectra of gases obtained in the laboratory, the gases in Jupiter's atmosphere can be identified.

The presence of methane and ammonia in Jupiter's atmosphere was deduced more than 50 years ago, while hydrogen was detected for the first time in 1960. (Although 500 times more abundant than methane, hydrogen has much weaker absorption lines because it is a molecule of two identical atoms that interacts only very weakly with electromagnetic waves.) Subsequent studies led to a growing list of new constituents, including the discovery of hydrogen cyanide (HCN) in 1981. Table 2 includes a list of Jupiter's atmospheric constituents and their abundances as determined by early 1982.

If the condition of chemical equilibrium held rigorously in Jupiter's atmosphere, molecules such as carbon monoxide (CO), hydrogen cyanide, acetylene (C_2H_2), and ethane (C_2H_6) would not occur in the abundances shown in Table 2. Sources of energy other than the molecular kinetic energy corresponding to local temperatures are evidently available. Solar ultraviolet radiation is responsible for the breakdown of methane and subsequent reactions of its fragments into acetylene and ethane. In the convective region of the atmosphere, lightning discharges contribute to these processes and may be responsible for the production of hydrogen cyanide. Still deeper, at temperatures around 1200 K, carbon monoxide is made by a reaction between methane and water vapour. Vertical mixing must be sufficiently strong to bring this gas up to a region where it can be detected from outside the atmosphere.

Differences between Jupiter and the Sun

Table 2 includes a comparison of elemental abundances in Jupiter's atmosphere with the composition of the Sun. If the planet had formed by simple condensation from the primordial solar nebula (see below), the abundances of the elements should be the same in both Jupiter and the Sun. Instead, there is a real spread in the values for different elements. On the other hand, except for methane and helium, the abundances of the gases from which the elemental abundances are derived depend on dynamical phenomena in Jupiter's atmosphere—principally condensation and vertical mixing. Thus the apparent enrichment in carbon may be an indication of a real difference in composition between Jupiter and the Sun.

Another difference is indicated by the presence of deuterium (D) on Jupiter. This heavy isotope of hydrogen has disappeared from the Sun as a result of nuclear reactions in the solar interior. Because no such reactions occur on Jupiter, the ratio of deuterium to hydrogen there should be identical to the ratio of those isotopes in the cloud of interstellar gas and dust that collapsed to form the solar system 4,600,000,000 years ago. Since deuterium may have been made primarily in the "big bang" that has been postulated to have begun the expansion of the universe, an accurate measurement of deuterium/hydrogen on Jupiter would allow the calibration of expansion models.

Temperature and pressure. The best profiles for the relation between changes in temperature and pressure in the Jovian atmosphere have resulted from the Pioneer and Voyager measurements. Both passed behind the planet as viewed from Earth, and the attenuation of the signal as it passed through Jupiter's atmosphere before being completely extinguished provided a measure of the change in atmospheric density, which is a function of temperature and pressure. A self-consistent analysis of the data is in agreement with independent studies based on analyses of the thermal radiation escaping from the planet.

Table 2: Atmospheric Abundances

gas	mixing ratio*	element ratio	Jupiter/Sun
H_2	1		
He	0.12	He/H	1
CH_4	2×10^{-3}	C/H	2
NH,	2×10^{-4}	N/H	1
C_2H_6	4×10^{-5}		
H_2O	10^{-6}	O/H	10^{-2}
C_2H_2	8×10^{-7}		
PH_3	4×10^{-7}	P/H	1
CH_3D	3.8×10^{-7}	D/H	†
CO	3×10^{-9}		
HCN	2×10^{-9}		
GeH_4	6×10^{-10}	Ge/H	10^{-1}
C_3H_8	(detected)		

*The mixing ratio is the number of molecules of a given atmospheric constituent in a unit volume divided by the number of hydrogen molecules in that same volume.

†Deuterium is not present on the Sun because of nuclear burning. The value of D/H on Jupiter is 3.5×10^{-5} ($\pm 1.5 \times 10^{-5}$), which is approximately 2.3 times the present interstellar value.

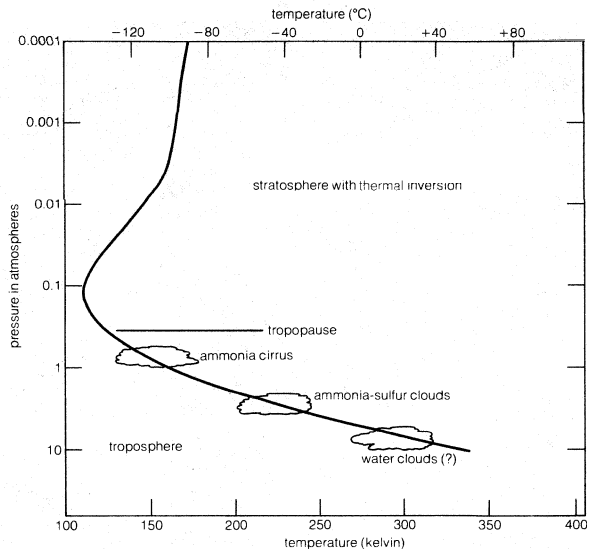


Figure 3: The structure of the atmosphere of Jupiter as deduced from Voyager measurements.

One of the profiles resulting from these studies is shown in Figure 3. From the figure it is possible to locate the positions of the principal cloud decks. It is interesting to notice that temperatures higher than the freezing point of water occur at pressures just a few times greater than the sea-level pressure on Earth. This is mainly a consequence of Jupiter's inernal energy source, although some warming would occur just by the trapping of infrared radiation by the atmosphere in the so-called greenhouse effect.

Temperature inversion

The increase in temperature above the tropopause is known as an inversion because temperature normally decreases with height. The inversion is caused by the absorption of solar energy at these altitudes by gases and aerosol particles.

Other likely atmospheric constituents. The list of atmospheric abundances in Table 2 is certainly not complete. For example, although the noble gas neon has about the same cosmic abundance as nitrogen, it, like helium, is very difficult to detect by spectroscopic observations. Neon had not been observed in the Jovian atmosphere by the early 1980s, even though it should be as abundant as ammonia.

Complex organic molecules in Jovian atmosphere

The formation of complex organic molecules in Jupiter's atmosphere is of great interest in the study of the origin of life. The initial chemical processes leading to the formation of living organisms may have occurred on Earth at a time when the composition of the terrestrial atmosphere was very similar to the present Jovian atmosphere, allowing for an appropriate depletion of hydrogen and helium. The active Jovian cloud system is a source of electrical discharges, while solar ultraviolet radiation, precipitation of charged particles, and the internal energy of the planet are also available to drive chemical reactions in the Jovian atmosphere. Thus, Jupiter may well represent an enormous natural laboratory in which the initial steps toward the origin of life are being pursued again and again. A determination of the degree of complexity reached under such conditions constitutes one of the most fascinating problems confronting any program of space exploration.

Radio emission. Jupiter was the first planet found (in 1955) to be a source of radiation at radio wavelengths. The radiation was recorded at a frequency of 22 megahertz (*i.e.*, a wavelength of 13.6 metres, or 1.36 decametres) in the form of noise bursts with peak intensities sometimes great enough to make Jupiter the brightest source at this wavelength, except for the Sun during its most active phase. The bursts of radio noise from three distinct areas constituted the first evidence for a Jovian magnetic field. Subsequent observations at shorter (decimetre) wavelengths revealed that Jupiter is also a source of steady radio emission. It has become customary to refer to these two types of emission in terms of their

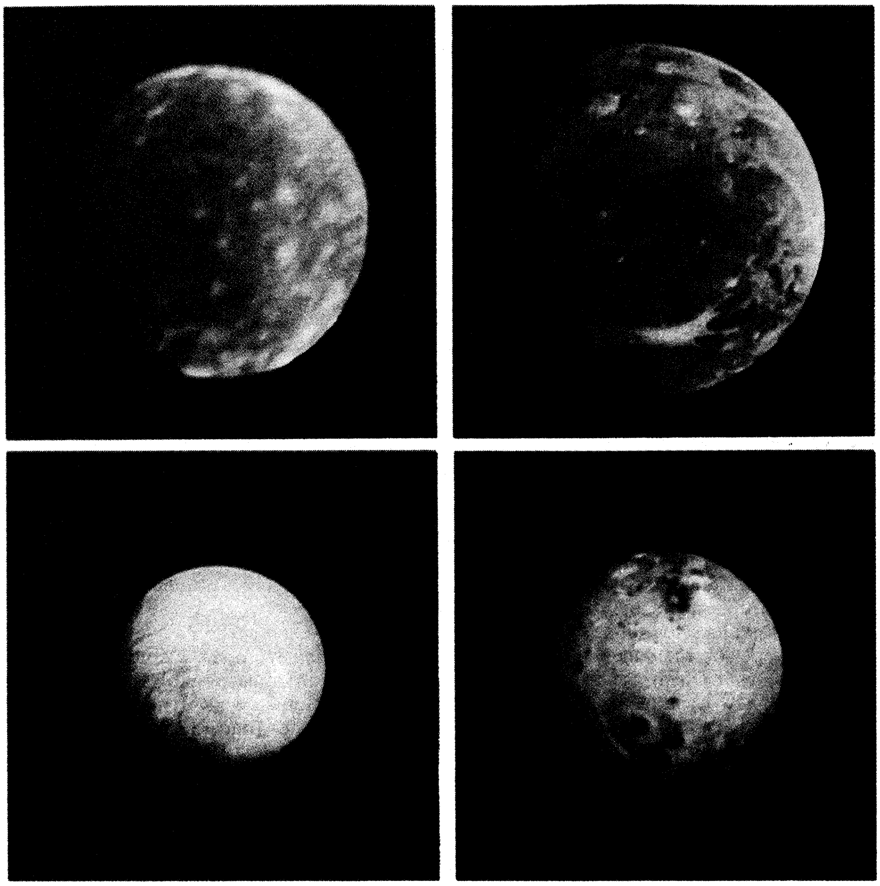


Figure *4:* The four Galilean satellites of Jupiter.
(Top) Callisto and Ganyrnede. (Bottom) Europa and Io. Photographs taken by Voyager 1 between March 1 and 3, 1979.

By courtesy of the Jet Propulsion Laboratory/National Aeronautics and Space Administration

characteristic wavelengths: decametre radiation (the erratic bursts) and decimetre radiation (the continuous source).

The nonthermal component of the decimetric radiation is interpreted as synchrotron emission; that is, radiation emitted by very high-speed electrons moving in a magnetic field within a toroidal, or doughnut-shaped, region surrounding Jupiter—a phenomenon closely analogous to that of the terrestrial Van Allen belts. The maximum emission occurs at a distance of two planetary radii from the centre of the planet and has been detected at 178–5,000 megahertz. The position of the plane of the polarization (vibrations of the radio emission preferentially being in a plane) and intensity of the radio emission vary with the same period. Both effects are explained if the axis of the planet's magnetic field is inclined by about 10" to the rotational axis. The period of these variations is the period designated as System III (see Table 1).

The radio emission at decametre (10-metre) wavelengths has been studied from Earth in the accessible range of 3.5 to 39.5 megahertz (3.5×10^6 to 39.5×10^6 cycles per second). Freed from the low-frequency cutoff established by the Earth's ionosphere, the radio-wave experiment on the Voyager spacecraft was able to detect emissions from Jupiter down to 60 kilohertz (60×10^3 cycles per second), corresponding to a wavelength of five kilometres. The strength of the radio signal and the frequency of noise storms show a marked time dependence that led to the early detection of three "sources," or emitting regions. The System III coordinate system was initially defined through the periodicity of these sources.

Effect of Io on noise storms

The noise storms are greatly affected by the position of the satellite Io in its orbit. For one source, events are much more likely to occur when Io is 90° from superior geocentric conjunction (*i.e.*, 90° from the position in which Earth, Jupiter, and Io are in a straight line) than otherwise. The noise sources appear to be regions that lie in the line of sight toward the optical disk of the planet (unlike the nonthermal decimetric radiation).

The most promising explanation of the effect of the orbital motion of Io on noise storms relates the emission to a small region of space linked to Io by magnetic field lines (a flux tube) that move with Io. Electrons moving in spirals around the magnetic field lines could produce the observed radiation. Interaction between these electrons and the Jovian ionosphere can be expected and were, in fact, observed by the Voyager spacecraft.

The Jovian magnetosphere. The nonthermal radio emissions described above are the natural result of trapped charged particles interacting with Jupiter's magnetic field and ionosphere. Interpretation of these observations led to a remarkably accurate definition of the basic characteristics of the planet's magnetosphere that was supported by the direct exploration of fields and particles in the vicinity of Jupiter by the Pioneer and Voyager spacecraft. The basic magnetic field of the planet is dipolar in nature, generated by a hydromagnetic dynamo that is driven by convection within the electrically conducting outer layers of Jupiter's interior. The magnetic moment is 19,000 times greater than that of Earth, leading to a field strength at the equator of 4.3 gauss, compared with 0.3 gauss at the Earth's surface. The axis of the magnetic dipole is offset by 0.1 R_J (where R_J is Jupiter's equatorial radius of 71,400 kilometres) from Jupiter's rotational axis, to which it is indeed inclined by 10°. The orientation of the Jovian magnetic field is exactly opposite to the present orientation of the Earth's field, such that a terrestrial compass taken to Jupiter would point south (relative to the planet's vector of angular momentum).

Magnetic links with Io

The magnetic field dominates the region around Jupiter to a distance of about 10 R_J to a point between the orbits of its satellites Europa and Ganymede (see below Satellites and rings). Within this region the most striking activity is generated by the satellite Io, whose influence on the decametric radiation has been discussed above. An electrical current of approximately 5×10^6 amperes flows in the magnetic flux tube linking Jupiter and Io. This satellite is also the source of a torus (doughnut-shaped) cloud of ions that surrounds its orbit.

THE INTERIOR

The atmosphere of Jupiter comprises only a very small fraction of the planet, much as the skin of a fruit compares with its contents. Nothing can be directly observed below this thin outer layer, so that indirect conclusions are drawn from the evidence in order to determine the composition of the interior of Jupiter.

The quantities derived directly from observation with which the astronomer can work are the atmospheric temperature and pressure, mass, radius, shape, rate of rotation, heat balance, and perturbations of satellite orbits and spacecraft trajectories. From these follow the ellipticity and the departure from an ellipsoid of the planet, quantities that may also be predicted using theoretical descriptions, or models, for the internal distribution of material. Such models can then be tested by their agreement with the observations.

Jupiter's central temperature and pressure

The basic difficulty in constructing a model that will adequately describe the internal conditions for Jupiter is the absence of extensive laboratory data on the properties of hydrogen and helium at pressures and temperatures appropriate to the conditions near the centre of this giant planet. The central temperature is estimated to be close to 25,000 K, to agree with an internal source of heat that allows Jupiter to radiate about twice as much energy as it receives from the Sun. The central pressure is in the range of 50,000,000–100,000,000 atmospheres (where one atmosphere is the sea-level pressure on Earth). The required extrapolation is not quite as dramatic as it may seem, since at such tremendous pressures hydrogen assumes a metallic state, the properties of which can be calculated with some confidence. The difficulty lies, first, in establishing the transition point at which the metallic state occurs and, second, in defining the properties of hydrogen between this transition point and the region where laboratory measurements are available. A third difficulty is posed by the solubility of helium in hydrogen at these pressures and temperatures, resulting in a multicomponent system.

Predominance of hydrogen and helium

Despite these problems, there has been a steady improvement in the precision of the models. Perhaps the most significant early conclusion from these studies was the realization that Jupiter cannot be composed entirely of hydrogen; if it were, it would have to be considerably larger than it is to account for its mass. On the other hand, hydrogen must predominate, constituting at least 78 percent of the planet by mass, no matter in what form—gas, liquid, or solid—it may occur. If it is assumed that the bulk of the rest of the planet is helium, the proportion of hydrogen to helium is on the order of 14:1, in close agreement with the atmospheric value and present ideas for the composition of the Sun (see Table 2). Current models agree on a transition from molecular to metallic hydrogen at approximately 0.75 R_J. It should be stressed that this is not a transition between a liquid and a solid, but rather between two liquids with different electrical properties. In the metallic state the electrons are no longer bound, thus giving the hydrogen the conductivity of a metal. No solid surface exists in any of these models, although most contain a rock and ice core with a radius of 0.03–0.10 R_J (0.33–1.1 Earth radii).

The source of internal heat has not been resolved. The favoured explanation in the early 1980s invoked the gradual release of primordial heat left over from the planet's formation. In other words, the conversion of gravitational energy to thermal energy that initially led to a very hot Jupiter is still progressing with a very gradual contraction of the planet (see below Theories of origin).

SATELLITES AND RING

The first objects in the solar system discovered by means of a telescope were the four brightest satellites of Jupiter. Galileo, who first observed them in 1610, proposed that the satellites be named the Medicean stars, in honour of his patron, Cosimo II de' Medici; but they soon came to be known as the Galilean satellites in honour of their discoverer. Galileo regarded their existence as a fundamental argument in favour of the Copernican model of the solar system in which the planets orbit the Sun.

In order of increasing distance from the planet, these satellites are called Io, Europa, Ganymede, and Callisto, for legendary figures closely associated with Jupiter (Zeus) in Greek mythology. The names were assigned by the German astronomer Simon Marius, Galileo's contemporary and rival. There proved to be a certain further aptness in the choice of Io's name: 10—"the wanderer" (Greek *iōn*, "going")— has an indirect influence on the ionosphere of Jupiter. The fifth satellite was also discovered by visual observation at a telescope, while the other 11 known satellites were found photographically (JVI–JXIII) or in pictures obtained by the Voyager spacecraft (JXIV–JXVI). Roman numerals are assigned to the satellites in order of their discovery.

Satellite groups. Jovian satellite data are summarized in Table 3. The orbits of the inner eight satellites have low inclinations and low eccentricities; *i.e.*, all of them are nearly circular. The orbits of the outer eight, in contrast, have much higher inclinations and eccentricities. The two innermost satellites (XIV and XVI) seem to be intimately associated with Jupiter's ring and may in fact be two of the sources of the fine particles within the ring itself. There are almost certainly additional members of all of these groups.

The Galilean satellites. Although approximate diameters and spectrophotometric characteristics of the four largest satellites of Jupiter had been determined from ground-based observations, it was the Voyager missions of 1979 that indelibly established these four bodies as worlds in their own right (see Figure 4). Before Voyager,

Table 3: Satellites of Jupiter

roman numeral	name	discoverer	year of discovery	mean distance from Jupiter (km)	sidereal period (days)	orbital inclination (degrees)	orbital eccentricity	radius (km)	mass (kg)	mean density (g/cm^3)
XVI	Metis	S. Synott	1979	127,949	0.295	0.0	0.0	?	?	?
XIV	Adrastea	D. Jewitt, E. Danielson	1979	129.020	0.299	0.0	0.0	?	?	?
V	Amalthea	E.E. Barnard	1892	181,300	0.489	0.455	0.003	120	?	?
XV	Thebe	S. Synott	1979	221,697	0.674	0.0	0.0	?	?	?
I	Io	Galileo, S. Marius	1610	412,600	1.769	0.027	0.000	1.816	8.916×10^{22}	3.55
II	Europa	Galileo, S. Marius	1610	670.900	3.551	0.468	0.000	1.563	4.873×10^{22}	3.04
III	Ganymede	Galileo, S. Marius	1610	1,070,000	7.155	0.183	0.001	2.638	1.490×10^{23}	1.93
IV	Callisto	Galileo, S. Marius	1610	1,880,000	16.689	0.253	0.007	2.410	1.064×10^{23}	1.81
XIII	Leda	C. Kowal	1974	11,110,000	240	27	0.147	1.7	?	?
VI	Himalia	C.D. Perrine	1904	11,470,000	250.6	28	0.158	85	?	?
X	Lysithea	S.B. Nicholson	1938	11,710,000	260	29	0.12	3–16	?	?
VII	Elara	C.D. Perrine	1905	11,740,000	260.1	26	0.207	40	?	?
XII	Ananke	S.B. Nicholson	1951	20,700,000	617	147	0.169	3–14	?	?
XI	Carme	S.B. Nicholson	1938	22,350,000	692	163	0.207	4–20	?	?
VIII	Pasiphae	P. Mellote	1908	23,300,000	735	147	0.40	4–23	?	?
IX	Sinope	S.B. Nicholson	1914	23,700,000	758	156	0.275	3–18	?	?

it was known that Callisto and Ganymede were both as large or larger than the planet Mercury; that they and Europa had surfaces covered with water ice; that Io was surrounded by a torus of atoms and ions that included sodium, potassium, and sulfur; and that the inner two satellites have mean densities much larger than those of the outer two. This density gradient resembles that found in the solar system itself and seems to result from the same cause (see below *Theories of origin*). The implication is that Io and Europa have a rocky composition similar to that of our Moon, whereas approximately **60** percent of Ganymede and Callisto must be a much less dense substance, with water ice as the most likely candidate.

Callisto. The surface of this satellite is so dominated by impact craters that there are no smooth plains such as the dark maria on the Moon. In other words, there seem to be no areas on Callisto where subsequent internal activity has obliterated any of the record of early bombardment. This record was formed by impacting debris (comet nuclei and asteroidal material) primarily during the first **500,000,000** years after the formation of the solar system in much the same way that the craters on Earth's Moon were produced.

Ganymede. Unlike Callisto, this equally icy satellite reveals distinct patches of dark and light terrain. This contrast is reminiscent of what can be seen on the Moon, but the association of chronology with albedo (reflected brightness) is exactly reversed. The dark regions on Ganymede are the oldest areas, showing the heaviest concentration of craters. The light regions are younger, revealing a complex pattern of parallel and intersecting ridges and grooves in addition to unusually bright impact craters typically surrounded by systems of rays. Like Callisto, Ganymede also exhibits relatively low topographic relief, indicating the action of viscous flow of the icy surface.

Smoothness of Europa

Europa. The surface of Europa is totally different from that of Ganymede and Callisto, despite the fact that the infrared spectrum of this object indicates that it, too, is covered with ice. There are few impact craters on Europa—the number per unit area is about comparable to that on the continental regions of Earth. This indicates that the surface that can be seen is relatively recent. Some scientists think the surface is young enough to suggest that significant resurfacing is still taking place on the satellite. This resurfacing evidently consisted of the outflow of water from the interior to form an instant frozen ocean. Its frozen surface is crisscrossed by a pattern of dark stripes and curved ridges. The relief is extremely low, with ridge heights perhaps a few hundred metres at most. Europa thus has the smoothest surface of any of the solid bodies examined in the solar system thus far.

Io's volcanoes

Io. Seen through a telescope, Io looks reddish-orange, while the other satellites are neutral in tint. Io's infrared spectrum shows no evidence of the absorptions of water ice. Scientists expected Io's surface to look different from those of Jupiter's other moons, but the Voyager pictures revealed a landscape even more unusual than anticipated. Volcanic calderas, instead of impact craters, dot the surface of Io, and **10** volcanoes were observed in eruption when the two Voyager spacecraft flew by in **1979.** This unprecedented level of activity makes Io the most tectonically active object in the solar system. The surface of the satellite is continually and completely replaced in spans of time considered short when compared with the age of the solar system. Various forms (allotropes) of sulfur appear to be responsible for the black, orange, and red areas on the satellite's surface, while solid sulfur dioxide is probably the main constituent of the white areas. Sulfur dioxide was detected as a gas near one of the active volcanic plumes by Voyager's infrared spectrometer and was identified as a solid in ultraviolet and infrared spectra obtained from Earth-orbital and ground-based observations. These identifications provide sources for the sulfur and oxygen ions observed in the Jovian magnetosphere.

The energy for this volcanic activity requires a special explanation, since radioactive heating is inadequate for a body as small as Io. The favoured explanation is based on the observation that orbital resonances with the other Galilean satellites perturb Io into a more eccentric orbit than it would assume if only Jupiter controlled its motion. The resulting tides developed by the contest between the other satellites and Jupiter may release enough energy to account for the observed volcanism.

Surface of Amalthea

Other satellites. The only other Jovian satellite that was close enough to the trajectories of the Voyager spacecraft in **1979** to allow surface features to be seen was Amalthea. So small that its gravitational field is not strong enough to deform it into a sphere, it has an irregular, oblong shape (see Table 3). Like Io, its surface exhibits a reddish colour that may result from a coating of sulfur compounds released by Io's volcanoes.

The ring. One of the tasks of the Pioneer II mission of **1974** was to monitor the charged particles around Jupiter. As the spacecraft sped toward its closest approach of **1.6** R_J, a sudden decrease in the density of charged particles was detected at a distance of **1.7–1.8** R_J. This led to the suggestion that a moon or a ring of material might be orbiting the planet at this distance. The existence of a ring was verified by the first Voyager spacecraft when it crossed the planet's equatorial plane, and the second spacecraft recorded additional pictures, including a series taken in the shadow of the planet looking back at the ring material. The ring is comprised of large numbers of micron-sized particles that produce strong forward scattering of incident sunlight. Submicron dust is also present, as indicated by a faint halo of material surrounding the ring plane. The halo particles are charged and move out of the equatorial plane in response to the rocking motion of the magnetic equatorial plane as the planet rotates.

Source of Jupiter's ring

The presence of such small particles requires a source. Indeed, the finest material extends all the way in to the planet itself. It seems likely that the source of this material is large boulders, or small moonlets, within the ring. Visible examples of what are presumably among the largest of such objects are satellites JXIV and JXVI. The ring particles are generated by impacts of micrometeoroids, cometary debris, and possibly volcanically produced material from Io. It seems plausible that the inner edge of Jupiter's ring is defined by the orbit of one of these moonlets, even as the outer edge appears to be defined by satellite JXIV.

THEORIES OF THE ORIGIN OF THE JOVIAN SYSTEM

The origin of Jupiter and its satellites is part of the problem of the origin of the entire solar system. Current thinking favours the gradual development of the Sun and planets from a huge cloud of gas and dust containing gravitational instabilities. Details of the later evolutionary picture vary, however, especially on the role of magnetic fields and on the relative importance of condensation and accretion.

Early history of Jupiter. Given the planet's large proportion of hydrogen and its huge mass, it has been traditional to assume that Jupiter formed by condensation from the primordial solar nebula. This hypothesis implies that the elements should all be present on Jupiter in the same proportions that they occur in the Sun. The most recent evidence (Table 2) suggests that this may not be true.

Accretion of dust and ices

Current models for Jupiter's origin suggest that a core formed first as a result of the accretion of grains of dust and ices into accumulating planetesimals. The heavier elements in the core would have been present in solar proportions, but as the core grew and began to attract more volatile material, a nonsolar composition could have resulted. In particular, models have been proposed with a core containing high proportions of water ice and an envelope of hydrogen and helium. Although these light gases, relative to the heavier elements in the core, do not reach their solar abundance, the ratio of hydrogen to helium is solar in all of these models, as appears to be the case on Jupiter itself.

Early history of the satellites. The inner eight satellites of Jupiter are commonly thought to have originated in much the same way as the planet itself. Just as the primordial solar nebula is believed to have broken up into accreting planetesimals with a central condensation that became the Sun, the accumulation of material into a proto-planetary cloud at Jupiter's orbit ultimately led to the formation of the planet and its inner satellites. The analogy goes further. The high temperature of the forming planet apparently prevented volatile substances from condensing at the distances of the innermost satellites. Hence Ganymede and Callisto represent the volatile-rich outer bodies in this system.

The origin of the outer eight satellites, with their orbits of high eccentricities and inclinations, is thought to be quite different. They are members of the population of irregular satellites in the solar system and have most likely been captured by Jupiter.

BIBLIOGRAPHY. A discussion of telescopic observations of Jupiter consisting of descriptions of motions, colours, and transformations observed in the Jovian cloud layers may be found in B.M. PEEK, ***The Planet Jupiter*** (1958). An extensive compilation of papers on all aspects of the Jovian system through 1975 is presented in T. GEHRELS (ed.), ***Jupiter*** (1976). The Pioneer missions are described in R.O. FIMMEL, W. SWINDELL, and E. BURGESS, ***Pioneer Odyssey: Encounter with a Giant*** (1974). An early summary of information about the satellites appeais in J.A. BURNS (ed.), ***Planetary Satellites*** (1977). A later account is given in D. MORRISON (ed.), ***Satellites of Jupiter*** (1981), which includes a discussion of the Voyager results pertaining to the satellites. Summary articles of other Voyager discoveries may be found in the following journals: ***Science***, 204:945–957, 960–1008 (June 1, 1979), and 206:925–996 (November 23, 1979); ***Nature***, 280:725–806 (August 30, 1979); ***J. Geophys. Res.***, 86 (September 30, 1981), the entire issue being devoted to "Voyager Missions to Jupiter." The Voyager spacecraft and their experiments are described in ***Space Science Reviews***, vol. 21, no. 2 and 3 (1977). An overview of this mission is given in D. MORRISON and J. SAMZ, ***Voyage to Jupiter*** (1980); and many of the results and their interpretation are presented in articles in J.K. BEATTY, B. O'LEARY, and A. CHAIKIN (eds.), ***The New Solar System*** (1981). (T.C.O.)

Jura Mountains

A major physical feature of the European continent, the Jura Mountains are smaller and somewhat younger companions of the Alps, stretching in an arc from the Rhône River gorge near the Isère River in France north and east to the Rhine Valley upstream from Basel. Consisting of plateaus with an average elevation of more than 2,800 feet (850 metres) and mountains up to about 5,650 feet, the Jura is about 225 miles long and up to 40 miles wide. It lies mostly in Switzerland, but a good part of the western sector lies in France.

Geologically, the Jura is a part of the Alpine foreland. Thus it is geologically related to the Swabian and Franconian Jura (Swabian and Franconian Alps), which form a separate region north and east of the Rhine.

Relation to the Alps

The Jura Mountains were named for their pronounced forested character, *jura* meaning "forest" (from the Gaulish *jor, juria;* ultimately related to Slavic *gora*, "mountain"). The Jurassic System of rocks was named for the mountains' fossil-bearing limestone formations, which date to the Jurassic Period of geologic time, about 190,000,000 to 136,000,000 years ago.

History. The Jura Mountains were inhabited by Celtic tribes, the Rauraci and Sequani, when they became known to the Romans, and, after the Roman conquests, they belonged to the Roman province of Germania Superior (50 BC). The area was shown as mountain country on maps of the 2nd century AD. The region was Christianized by St. Columban in the early 7th century, and it supported a series of monasteries.

Scientific study of the mountains began in the 18th century with the work of Alexander von Humboldt, who named the Jurassic System for the Jura's rocks. Definitive geological studies, however, did not begin until the end of the 19th century and the beginning of the 20th. The best analysis was given by the Swiss geologist Albert Heim.

The physical environment. The highest peaks of the Jura Mountains are in the south, in the Geneva area, and include Crêt de la Neige, 5,636 feet, and Le Reculet, 5,633 feet, both in France, and Mont Tendre and La Dôle, both more than 5,500 feet, in Switzerland.

Stratigraphically, the mountains include formations from the Mid-Triassic Period (preceding the Jurassic and up to 225,000,000 years old) in the plateaus to the Miocene and early Pliocene (from 7,000,000 to less than 26,000,000 years old). During the Jurassic Period the Jura landscape was a more or less flat, shallow maritime trough in which up to 3,000 feet of sedimentary rocks accumulated. The rocks are principally fossil-rich limestones interlayered with chalks, clays, and marls, with iron-ore concretions in places.

The formation of the Jura Mountains began in the Early Tertiary (about 65,000,000 to 50,000,000 years ago) as a part of the same Earth movements that formed the Alps. The main thrust from the southeast occurred in the Pliocene. Thus the mountain building process took place in at least two separate thrust phases: the first one formed the outer, French chain of ridges; and the second thrust formed the adjacent, higher inner ranges.

Classification by structure and form

On the basis of their structure and form, the Jura Mountains are classified as either the folded Jura, including the chain Jura and the plateau Jura, or the Jura tableland. The greater part of the range consists of the folded Jura, with its chain Jura constituting the inner side of the Rhône-Rhine arc and being characteristically folded into parallel anticlines (longitudinal upraised ridges) and synclines (longitudinal valley troughs). Cutting through the longitudinal ridges are transverse valleys (cluses). The area of less pronounced folds in the western outer zone of the arc is known as the plateau Jura. There the folds have been worn down by erosion, and their original direction is shown only by surface ribs along the sides of isolated hills. In the north the structure of the mountains has been complicated by faulting that produced upthrust blocks, or horsts, forming the tableland of the Swiss Jura. To the northeast, these same strata form the overmantle of the Black Forest (Schwarzwald) highlands and the terraced landscape of the Swabian and Franconian Jura. But it is the folded Jura that forms the characteristic landscape type. The folds are seldom simple and symmetrical. Rather, they are overturned, or at least compressed, by pronounced thrusting toward the northwest; nevertheless, they are clearly wavelike in form. Often, older rocks lie over younger ones, as in the nappe of the Alps.

The prevalence of limestone, alternating with softer layers of marl, results in great permeability; the fissured surfaces of the limestone beds permit the surface waters to percolate underground, producing a karst topography (*i.e.*, creating sinks and sinkholes and dissolving out caverns and underground drainage channels). As a result, there is a lack of running water, particularly in the central part (plateau Jura). Traditionally, each farmstead collected its own cistern water; today, supply networks bring water up from the gorges of the rivers. Lake Joux has an underground outlet reappearing as a river, the Orbe, about two miles farther down. Similar underground stream sources are numerous. The largest rivers are the Doubs, the Ain, and the Birs.

Landscape and population. Lack of water affects both vegetation and settlement. Annual precipitation increases to the south and west, reaching more than 80 inches (2,030 millimetres) on Le Mont Risouz and Mont Tendre; but the Delsberg Valley and the north-facing corridor of the Ergolz River (Liestal) receive less than 40 inches. The climate is of the maritime-continental transitional type: rawer on the Jura heights, milder in the protected valleys and on the south-facing slopes. Pools of cold air form in the high valleys in winter. The average summer temperature at Geneva is 66" F (19° C), while the winter equivalent is 34° F (1° C). At La Brévine, for example, more extreme conditions prevail: the corresponding figures are 55° F (13° C) and 25° F (−4° C). Maximum and minimum readings at the stations vary from highs of, respectively, 95° F (35° C) and 84° F (29° C) to lows of 1° F (−17° C) and −44° F (−42° C).

Climate, vegetation, uses, and settlement

Natural vegetation is of the forest type throughout the region. From lower to higher elevations there occur oak groves and beech forests (most handsomely manifested at 2,000–4,000 feet) interspersed at higher elevations with fir. In the southern Jura, the firs are replaced by sycamore maple. The crests above the tree line (about 5,300 feet) are covered with Alpine grasses. Only over deposits of marl and clay are there also isolated moors. Wildlife is practically extinct.

Pastures on the high plateaus are more characteristic of the Jura than are the cultivated fields in the basins and valleys. The dairy economy is based on cattle raising, and a number of types of cheese are produced. In the west the Jura reaches the vineyard zone of the Saône Valley, and grapes also are grown on the south face of the Jura. The Jura farmer, in addition, raises fruits and ample quantities of potatoes and vegetables but little grain. The most important mineral resource has been granular iron ore.

Villages are found in abundance, especially in the valleys, while larger cities are almost entirely on the margins—Chambéry, Geneva, Neuchâtel, Biel (Bienne), Solothurn, Aarau, Baden, Bourg-en-Bresse, Lons-le-Saunier, Besançon, Montbkliard-Porrentruy, and, on the Rhine, Basel and Schaffhausen.

Industries of the Jura area

Centres of the Swiss watchmaking industry in the Jura are La Chaux-de-Fonds, Le Locle, Saint-Imier, Geneva, Biel, Grenchen, and Schaffhausen. Other light industries include gem cutting and polishing, the manufacture of eye glasses, briar tobacco pipes, music boxes, phonographs, camera equipment, paper, and typewriters; and the processing of tobacco and chocolate. The Jura population has always been noted for its industrial activities.

The strong sense of nationality possessed by the inhabitants of the Jura was manifested by the controversy in Switzerland that ended with the creation in 1978 of Jura canton, Switzerland's 26th, which separated about 70,000 French-speaking residents of Bern canton from its predominantly German-speaking majority.

In the second half of the 20th century the beauty of the Jura has been the destination of many tourists, both in the summer and, increasingly, also in winter for skiing. A consequence has been the development of the transportation network. Now, every scene of natural beauty and every Jura valley is accessible over excellent roads, while for through traffic the mountains form an obstacle. The cluses (gaps in the transverse ranges), since the earliest settlement, have been transportation routes. The few railroad lines follow them, as did the oldest Roman roads.

The principal express rail lines (Basel–Geneva, Basel–Olten, Paris–Lausanne–Milan via Simplon Pass) cut through the chains in tunnels. New tunnels were made for the Swiss national road network, and cuts were made for roadway pass crossings (for example, Faucille, Belchen, and Bozberg). The construction of the tunnels and cuts has made communications easier and also has laid bare new sections of the mountains for study of their structure. (W.T.K.)

BIBLIOGRAPHY. The most comprehensive recent bibliography is the *Bibliographie* jurassienne, 1928–1972 (1973), published by the SOCIETE JURASSIENNE D'EMULATION; an annual survey of new publications covering the Swiss portions of the region may be obtained from the Bibliographie des montagnes neuchdteloises, issued by the city library of La Chaux-de-Fonds. (Ed.)

Jurassic Period

The Jurassic Period occupies an intermediate position in the Mesozoic Era between the Triassic Period and the Cretaceous Period. Its name refers to the chalk sequence in the rock strata of the Swiss–French Jura Mountains. Around 1800 William Smith recognized that a group of strata was characterized by a particular set of fossils, thus making it possible to establish a connection, or relation, between Swiss-Jura limestones and English oolitic limestones. In France, Alexandre Brongniart introduced the term terrain jurassique ("Jurassic ground"), which also included the underlying Lias rocks. In southern Germany, L. von Buch distinguished Lower, Middle, and Upper Jurassic, which were later designated Black Jurassic (Lias), Brown Jurassic (Dogger), and White Jurassic (Malm) by F.A. Quenstedt and A. Oppel. On the basis of the predominant fossils, Quenstedt named six component layers within each of these three divisions (Table 2).

Classification of stages

Alcide d'Orbigny classified the period into stages, which in general were named after geographic locations in western Europe or were names introduced by previous geologists. With some later additions, the succession is summarized in Table 1. Oppel named the uppermost' (Portlandian) Stage in the Alpine–Mediterranean area "Tithon" because of its differing facies. That name has been extended to include other Late Jurassic units outside Great Britain. Sergey Nikolayevich Nikitin (1881) called the Jurassic of the Volga Basin the Volgian, which has been retained as a stage name for the youngest Jurassic rocks. Characteristic and fascinating landscapes in Jurassic rocks may be seen in the Jura Mountains, the Swabian–Franconian Jura, the Krakow district, southern England, and the western United States.

In areas of continuous succession the Jurassic is underlain by Rhaetic rocks (the youngest Triassic) and overlain by the Lower Cretaceous. The Berriasian in the Mediterranean area and the Ryazanian in the Soviet Union have been considered lowermost Cretaceous, though their ammonite faunas are close to Jurassic types. According to radiometric dating, the Jurassic Period began about 190,000,000 years ago and ended 136,000,000 years ago.

There is a peculiar balance between ancient and modern traits in the Jurassic record. The submergence of considerable portions of the continents beneath the seas and the dominance of the now-extinct ammonites and dinosaurs provide the unfamiliar, ancient elements of the picture. Geosynclinal seas of the Jurassic were to be the birthplaces of many of the present-day mountain ranges, however, and the development of Jurassic birds and mammals began the process of the modern forms of life.

Of economic importance among the Jurassic deposits are chromite (in basic igneous rocks in the Balkans); oolitic and sideritic iron ores of marine origin in central and western Europe; alluvial deposits of uranium ores and gold in the U.S.; coal in Siberia, China, Korea, Australia, Iran, Hungary, and elsewhere; salt in underground salt domes in Arkansas, Louisiana, and northern Germany; building stone (limestone, sandstone), brick-making and cement-making materials in Europe; and oil-producing rock strata along the Gulf Coast and in the western interior of North America, in the Emba district of the Soviet Union, and in northern Germany. In northwestern Germany, asphalt is obtained from Upper Jurassic limestone strata. Near Oldenburg, also in northwestern Germany, there is an important celestite (strontium sulfate) deposit. Fuller's earth is still mined near Bath, England. The production of lithographic stones in the Franconian Jura was of substantial economic importance in the 19th century, and the stone is still prized by lithographers.

This article treats the rocks, life-forms, and environments of the Jurassic Period. For similar information on previous and subsequent time intervals, see TRIASSIC PERIOD; CRETACEOUS PERIOD. For the relationships of these three periods, see MESOZOIC ERA; and see FOSSIL RECORD for additional details on all life-forms.

JURASSIC ROCKS

General distribution

Rock strata of the Jurassic System occur on all continents (Table 2). Jurassic beds that are covered by younger sediments have been discovered in recent decades by borings in southeastern England, Denmark, northern Germany, Poland, Louisiana, and elsewhere.

The picture of the Jurassic world still is very problematical, but it particularly derives from the study of the sedimentary rocks (*q.v.*). The restriction of the marine Jurassic sediments to the border regions of the present continents indicates that these land areas must have existed as shields by the Jurassic Period. Only in North America and Asia did the Jurassic oceans extend far into the interior. The absence of any Jurassic sediments on the east coasts of North America and South America, on the coast of West Africa, and, in part, on the coast of Europe suggests that in Jurassic times the Atlantic Ocean did not yet exist in its full form. The Jurassic sequence of eastern Africa and western Madagascar is interpreted as resulting from deposits laid down in an ocean strait that was beginning to divide Gondwanaland, the supercontinent that extended from South America through Africa to India, during the Jurassic Period. Continental drift (*q.v.*)

must be considered in the genesis of the Atlantic and Indian oceans and in reconstruction of the paleogeography (*q.v.*) of the Jurassic (see Map). Except for their peripheral areas, the old shields remained free of the Jurassic sediments (Scandinavia and the Canadian Shield). In inland depressions, fresh water and wind laid down Jurassic deposits of thick terrestrial (lacustrian, fluviatile, eolian) sediments, as in Siberia and Australia (Walloon Lake) and before and after the marine flooding in the western interior of North America.

The sedimentary strata deposited in shallow Jurassic seas were generally from several hundred to a thousand metres thick. Around the Pacific Ocean, thicker layers associated with igneous rocks indicate belts of particularly marked subsidence (geosynclines); in Jurassic and later times those belts rose to form the folded mountains of today. The greatest thicknesses of Jurassic rocks occur in southern Alaska (5,000 metres), California and Oregon (5,000 to 8,000 metres [16,000 to 26,000 feet]), California Coast Ranges (16,000 metres [53,000 feet]), Daghestan, Caucasus (12,000 metres [40,000 feet]), Verkhoyansk Mountains (3,000 to 7,000 metres [10,000 to 23,000 feet]), and Ouda Basin, Amur (7,800 metres [25,600 feet]).

The geosyncline called Tethys provided similar sites for thick accumulations in southern Europe and southern Asia. At its eastern and western ends the Tethys Sea was linked, at least temporarily, with the circum-Pacific geosynclines; at the western end the link probably consisted of a Middle Atlantic Seaway between southern Europe and Central America.

Sediment types

Conditions that have recurred constantly during the earth's history permitted the deposition of sandstones, conglomerates, shales, limestones and dolomites, and evaporites (*qq.v.*) during Jurassic time. In various locations on the periphery of the Arctic Ocean and in the interior of the Soviet Union, lime-poor, clayey, and marly rocks and marginally distributed silt, sand, and conglomerates were deposited.

The sediments of the central and northwestern European shelf region (European Archipelago) contain a large proportion of carbonates that grade to sand in nearshore facies and to clay and evaporites in offshore facies. Rhythmic sedimentation (cyclothems) often appears.

The secondary clay minerals (*q.v.*) of quite a few of these rocks contain kaolin, indicating that copious amounts of material from the bordering land were deposited in the sea.

Parts of the Alpine geosynclines contain marly or pure-bedded limestones, siliceous limestones, and radiolarites of Jurassic age. A relatively small thickness indicates rapid subsidence. Because of distance from land and the solubility of lime at depth, the sedimentation rate was low. Among the secondary clay minerals, kaolin is lacking. Scarcity of terrigenous sediment material resulted in scarcity of nutrients, a scant bottom fauna, and consequently an oxygen-rich ocean floor with red colours in the rock. Breccias resulted from the action of density currents (*q.v.*) on undersea slopes.

Thick clays and marls and marly limestones were deposited in many geosynclinal areas and were intruded by basic magma. Metamorphosis of the thick sequence of layers into phyllites occurred. Such lustrous, slaty rocks are represented by the Biindner (Graubünden) schists of the Swiss Alps.

Other Jurassic sedimentary rock types include bedded limestones and reef limestones on sinking ridges within the geosynclines; thin limestones formed on stable ridges and submarine plateaus under conditions of sweeping bottom currents; breccias of great thickness at the foot of large submarine ridges (*e.g.*, Nappe de la Brèche, Swiss Alps); flysch sediments that indicate rapid deposition in orogenic. or mountain-building, areas such as the California Coast Ranges and the Caucasus; and terrestrial limestones, clays, sands, sandstones, and coal seams in inland depressions of Jurassic continents.

The output of magmatic rock was substantial in Jurassic times. The rocks of the inner geosynclinal region were metamorphosed by orogenesis. Granites were converted to orthogneisses; clayey–limy sediments to phyllite (*e.g.*, central Alps, Sumatra, and Java) and into paragneiss (Cordilleras); and the sandstones to quartzite. The dating of metamorphic rocks is still uncertain in many cases because of the scarcity of fossils.

JURASSIC LIFE

Invertebrates

Protista. In the Lower Jurassic of the Tethys Geosyncline, the coccolithophorids (flagellates, up to millions per cubic centimetre) appeared; they were major producers of calcareous material in the calcium economy of the warm seas.

Associated limestone-forming organisms in the Upper Jurassic were the *Calpionella* (Tintinnida, 1). Among other microfossils are the problematic single-celled hystrichosphaerids, which in Jurassic times had a second burst of development. Dinoflagellates are known since the Middle Jurassic. Among the Foraminiferida, the lime-shell types were most important. In the Tethys appeared complex, large Foraminifera (Lituolidae) and the first planktonic Globigerinidae. The Radiolaria formed siliceous radiolarite deposits in the absence of clastic or carbonate materials.

Sponges. The siliceous sponges, colonizing somewhat deeper waters than the calcareous sponges, reached their greatest prevalence in Jurassic times. In the Alpine sequences rock beds consist almost entirely of the spicules of sponges; starting in the Lower Jurassic, the latticed Hexactinellida and the Lithistida formed local undersea meadows. On the northern edge of the central European Tethys (Swiss Jura, southern Germany, Poland)—while

Table 1: Stage Names of the Jurassic Period

stage	author	derivation of name (locality)
Portlandian	Brongniart 1829	Portland Isle (southern England)
Kimmeridgian	d'Orbigny	Kimmeridge (southern England)
Oxfordian	Brongniart 1829 amended by d'Orbigny	Oxford (England)
Callovian	d'Orbigny	Kellaways (Gloucester, England)
Bathonian	d'Orbigny	Bath (England)
Bajocian	d'Orbigny	Bayeux (Normandy, France)
Aalenian	Meyer-Eymar 1864	Aalen (Wiirttemberg, southern Germany)
Toarcian	d'Orbigny	Thouars (Deux Sèvres, western France)
Pliensbachian (Liassic)	Oppel 1858 (d'Orbigny)	Pliensbach (Württemberg)
Sinemurian	d'Orbigny	Sémur (Côte d'Or, France)
Hettangian	Renevier 1864	Hettange-Grande (Lorraine, eastern France)

Table 2: Selected Rock Successions and Suggested Correlation with European Succession

Key: **A** ammonite correlations; ⊕ very thick deposits; ⊖ thin but significant deposits; ζ through *a*, Quenstedt's classical subdivisions of Swabian Jurassic; 〰 major orogeny; ∼ strong folding (N.F.H.).

	age/stage	England	Russian platform	Northwest Germany	Southern Germany (Swabia)	Jura Mountains	Southern France (Ardèche)
Late Jurassic	Tithonian/ Volgian	Lower Purbeck Beds and Lower Spilsby Sands **A** Portland Beds **A** Upper Kimmeridge Clay	**A** Upper Volgian greensand, marl ⊖ **A** Lower Volgian	Lower Serpulite Miinder Marls	Neuburg Beds ζ	**A** Upper Tithonian Limestones beds of Purbeck Facies limestones of Mont Salève, etc.	**A** Upper Tithonian Limestones (Ardtche) Middle and Lower Tithonian
	Kimmeridgian	**A** Middle and Lower Kimmeridge Clay	**A** clays	**A** Gigas Beds Limestone **A** clays **A** Corallian Oolite	Solnhofn-Plattenkalke	**A** Lower Tithonian Limestones **A** limestones with coral reefs	**A** Limestones **A** limestones of Crussol
	Oxfordian	**A** Corallian Beds **A** Oxford Clay	**A** clays	**A** Heersum Clays	Malm ε or δ **A** White γ Jura β **A** a	marls	**A** marly limestones
Mid-Jurassic	Callovian	**A** Kellaways Beds **A** Cornbrash Limestones ⊖	**A** clays	**A** clays	**A** ζ	**A** oolite ironstone and limestone ⊖	**A** marls and dolomite
	Bathonian	**A** Great Oolite Limestones, etc.	sands	**A** succession near Hildesheim Fuller's earth	Dogger or Brown Jura ε	**A** oolite limestone and **A** marls	**A** limestones ⊖
	Bajocian	**A** Inferior Oolite Limestones	Bajocian and Toarcian in Donets Basin	**A** Bielefeld Succession **A** clay succession near Hanover	δ **A** γ β *a*		**A**
Early Jurassic	Toarcian	**A** Upper Lias Shales, Sands		**A**	**A** ζ ε	**A** clays and oolite ironstone	**A** ferruginous oolite ⊖
	Pliensbachian	**A** Middle Lias Shales, Ironstone **A**		**A** Lias	γ δ **A** Lias or	**A** Middle and	**A**
	Sinemurian	**A** Lower Lias Shales		**A**	**A** Black Jura β	**A** Lower Lias Black Marls	
	Hettangian	**A**		**A**	*a* **A**	**A**	

heavy lime sedimentation was proceeding—those sponges and the calcareous algae (stromatolites) built up a great reef complex. Because of greater resistance to erosion, such reefs often appear in the present-day landscape as groups of crags. During or following the sedimentation, the silica concentrated by the sponge skeletons formed as chert nodules (see SILICEOUS ROCKS).

Coelenterates. In the Tithonian *Plattenkalke* (plate-like limestones) of the Franconian Jurassic magnificent impressions of large jellyfish (Scyphozoa) are found. In some locations hydrozoans made up a large part of the Upper Jurassic coral reefs of the Tethys, but the rich world of Jurassic corals belongs to the order of the Scleractinia; these anthozoans were dominant from the Triassic Period to the present. The conditions required for the life of these modern corals must have been the same in Jurassic time as today (see CORAL ISLANDS, CORAL REEFS, AND ATOLLS).

Worms. Concentrations of calcareous tubes of the marine worm *Serpula* constitute the Upper Jurassic Serpulite Formation of northwest Germany.

Bryozoa. The Cyclostomata, today characteristically are to be found in cold seas, were by far the predominant forms in the warm Jurassic seas. Cryptostomata lived in cavities excavated in the shells of dead mollusks. The modern Cheilostomata first became important in the Cretaceous.

Brachiopods. Brachiopods were most abundant and diverse in the Paleozoic. Stragglers from the Paleozoic Helicopegmata are known in the form of ***Spiriferina*** as late as the lowest Middle Jurassic beds. The little *Cadomella* (Upper Pliensbachian–Lower Toarcian), in spite of a helicopegmatic arm support converging toward the Spiriferida, appears to be the last representative of the Strophomenacea.

The great bulk of the Jurassic Brachiopoda consists,

Table 2: Selected Rock Successions and Suggested Correlation with European Succession (cont.)

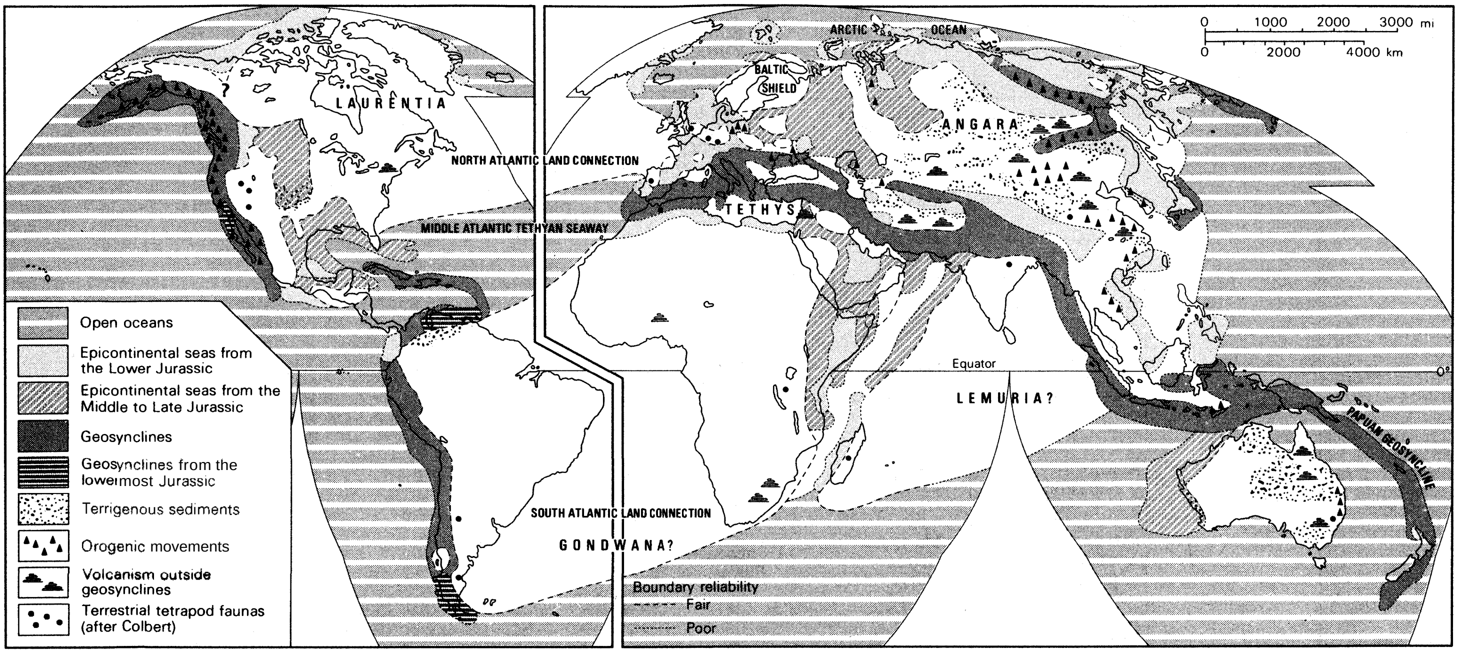
Caucasus Mountains	India (Kutch)	United States (California)	United States western interior (Wyoming)
	A Umia **Ammonite Beds**	Knoxville Group and ⊕	
Tithonian Limestones	A Upper Katrol Shales	A Franciscan Group (Coast Ranges) Nevadan	Morrison Formation (fresh water)
orogeny **A**	A Katrol Sandstones	**orogeny** (Sierra Nevada)	
A marine **shales**	**A Chari Group** **A**	**A** Mariposa	A **Sundance** Formation
A much volcanic activity	A **Patcham** Limestones	A **Amador** group ⊕ **much** volcanic activity	**A** shales sandstones limestones
throughout ⊕	**Kuar Bet** Beds	**A**	**Gypsum Spring Formation** **A shales limestone**
A **A**		**A Dunlap Formation (western Nevada)** **Dunlap** **orogeny**	**Nuggett Sandstone (nonmarine)**
A **shales and sandstones** **with coals** ⊕		**A** **A** **A Sunrise Formation (western Nevada)**	

however, of the modern groups Rhynchonellacea and Terebratulacea, which experienced a new evolutionary burst. The majority lived firmly attached by means of byssuses (silky filaments) to the bottoms of the epicontinental shallow seas and to the reefs. The form taken by *Pygope* in the Tithonian strata of the Mediterranean Tethys is particularly remarkable. The hole that extends through both its valves probably permitted a complete separation of the outflowing water from the nutrient-bringing inflowing water—possibly an adaptation to nutrient-poor deep water.

Bivalves. The Jurassic pelecypod fauna consisted of approximately 40 families. (Today there are 54 families.) Of those, 19 originated in the Paleozoic with all but 3 of the 19 still surviving today. Ten families began in the Triassic and about 12 in the Jurassic itself; none of these have become extinct. The Ostreidae, although originating in the Triassic, played an important role in the Jurassic, especially the *Liostrea*, *Gryphaea*, and *Lopha*. Thus the Jurassic was a time of especially marked development and modernization of the bivalves. The flourishing Trigoniidae were characteristic, as were *Inoceramus* (Jurassic–Cretaceous) and *Buchia* (Upper Jurassic–Lower Cretaceous). The curious spondylids *Lithiotis* and *Cochlearites* were limited to shallow-water facies of the Tethys.

The sediments which were richest in pelecypods were those of the epicontinental shallow seas and the geosynclinal ridge areas. It is thought to be probable that many Pectinidae already possessed a limited swimming ability, and the life-style of *Posidonia* (Silurian–Jurassic) is assumed to have been some type of constant swimming motions. Pelecypod groups characteristic of brackish and fresh water also have been studied for various correlation purposes.

Gastropods. Like the bivalves, the prosobranch gastropods, dominant since the Paleozoic, developed greatly


0 1000 2000 3000 mi
0 2000 4000 km
ARCTIC OCEAN
BALTIC SHIELD
LAURENTIA
ANGARA
NORTH ATLANTIC LAND CONNECTION
MIDDLE ATLANTIC TETHYAN SEAWAY
TETHYS
Equator
LEMURIA?
PAPUAN GEOSYNCLINE
SOUTH ATLANTIC LAND CONNECTION
GONDWANA?
Open oceans
Epicontinental seas from the Lower Jurassic
Epicontinental seas from the Middle to Late Jurassic
Geosynclines
Geosynclines from the lowermost Jurassic
Terrigenous sediments
Orogenic movements
Volcanism outside geosynclines
Terrestrial tetrapod faunas (after Colbert)
Boundary reliability
Fair
Poor

Depositional regions and tectonics during the Jurassic Period.

during the Jurassic. Among the primitive Archaeogastropoda, *Pleurotomaria, Discohelix* (Triassic?–Jurassic), and *Trochus* (the latter beginning with the Lias, or Early Triassic) are important. *Patella, Helcion,* and *Scurria* have a secondarily acquired dish-shape, an adaptation to their sessile, or anchored, existence. Among the Mesogastropoda (especially numerous throughout the Mesozoic), the Nerineacea, a group of extraordinary high-spired forms, originated in the Jurassic. Freshwater Pulrnonata genera found in the lacustrine Upper Jurassic (Purbeck Facies) are *Limnaea, Planorbis, Physa,* and *Auricula.* These Pulmonata, which had taken the path from sea to land beginning with the Carboniferous, had turned about and returned to water. Prosobranchiate-gilled snails of the Purbeck Facies comprise *Valvata, Bithynia,* and representatives of the Hydrobiidae, all beginning with the Jurassic.

Occurrence and evolution of ammonites

Cephalopods. The Ammonites, an order derived from the Nautiloidea in the Ordovician Period (*q.v.*), for unknown reasons almost became extinct toward the end of the Triassic. Only the genera *Phylloceras* and *Lytoceras* managed to survive and to initiate a new evolutionary surge.

The shell forms showed great diversity. The ammonite shell may be distinguished from the nautilus by a two-flap lid member (aptychus) and by the complexity of the sutures along which the septa (dividing walls) are attached to the inner walls of the shell tubes. These sutures exhibit lobes (rearwardly directed) and saddles (forward inflections); in adult shells both are folded, branched, and often serrated like moss. The suture is the prime distinguishing characteristic in establishing the family relationships of the ammonites.

The Phyllocerata and Lytocerata remained quite constant during the Jurassic and Cretaceous periodsr Most of the other Jurassic ammonites descended from a line that presumably arose from the Lytocerata toward the end of the Triassic and appeared at the beginning of the Jurassic in the form of the genus *Psiloceras* with an almost unsculptured shell. Numerous Jurassic superfamilies developed from *Psiloceras.* The greatest complexity of the suture is attained in the Stephanocerataceae in the Bajocian Age. Many regressions appear in the Perisphinctaceae of the Upper Jurassic. Nevertheless, the primary suture is always quinquelobate. Atavisms in the form of unrolled or rod-shaped shells, like those of the earliest Paleozoic ammonites, occur from the Bajocian to the Callovian and later in the Tithonian (Protancyloceras, from which the aberrant Cretaceous ammonites developed).

The Nautiloidea, like the ammonites, survived the crisis of the Triassic–Jurassic boundary in the form of only one genus, *Cenoceras,* which also experienced a new surge of development.

Coleoidea (cephalopods with internal skeletons). Of the pre-Jurassic Belemnoidea, *Atractites* (originating in the Triassic) still existed in the Lower Jurassic, particularly in the geosynclines. The Belemnoidea first attain some measure of importance in the Jurassic, beginning with small forms in the European Hettangian. Their pikelike calcite rostrums, including the distal part of the chambered cone, certainly have fewer distinctive characteristics than do shells of the ammonites. Early Jurassic Belemnoidea had rostrums with short distal furrows. In the Bajocian, *Megateuthis,* the largest belemnite form, grew to a length of perhaps two metres. A peculiarity of an older Jurassic group (Upper Pliensbachian to Toarcian) was the development of a hollow epirostrum of unknown function.

Crustacea. Minute, bivalved Ostracoda serve as important guide fossils, which in many cases allow a finer biostratigraphic classification than the ammonites. The Isopoda (wood lice) are known beginning with the Jurassic. The decapods developed more markedly than previously into swimming and benthonic (bottom-dwelling) forms. The earliest true lobster was found in the Upper Jurassic of Germany. Within the Anomura, *Gastrodorus* abandoned a protecting external skeleton on the rear portion of the body, suggesting the origin of the hermit crabs. In the form of primitive Dromiacea, the Brachyura (crabs) arose in the Jurassic. Close relatives of the Jurassic Anomura and Dromiacea still inhabit deep-sea waters today.

Echinoderms. The dominant order of the crinoids, Articulata, showed a tendency toward smaller and more motile calyx-forming calcareous platelets. The Isocrinida produced in *Seirocrinus subangularis* the largest crinoid forms ever to exist; in the Toarcian Age they led a pseudoplanktonic life attached to driftwood. Extraordinary preservations occurred in fine sediments of oxygen-poor basins where the remains were not disturbed by animal burrowing.

The Middle Jurassic *Paracomatula helvetica,* exhibiting atrophy of the stalk, was the probable ancestor of the modern nonstalked comatulid sea lilies, which first appeared in the Upper Jurassic. *Apiocrinus, Millericrinus,* and *Eugeniacrinites,* with jointed calyxes and strong stalks provided with rootlets, were adapted to the conditions of life found in moving waters over hard bottoms and in the reefs.

The genetic development of the echinoids during the Jurassic is centred in the emergence of asymmetrical sea urchin forms from the older symmetrical species. The anus moved from the centre of the upperside toward the rear on the underside; the mouth, on the underside, moved forward; and large club-shaped spines disappeared. Those transformations permitted residence in mud: the large spine would have been a hindrance, and a top-positioned anus would have interfered with the intake of clean water.

Vertebrate life

Fishes and amphibians. Among the Chondrichthyes (see SELACHII), are found hybodontids (seven-foot-long, rather primitive sharks) and less frequently Holocephali (ancestors of the ratfish or chimaera). Of the Osteichthyes (bony fish), crossopterygian and chondrost fishes (*q.v.*) were few, but holosts ("ganoid fishes") were quite numerous.

The first teleosts with fully ossified vertebrae appeared in the Jurassic. These small herringlike fish were the primitive members of the group that came to dominate the fish world in later times. According to some taxonomists, they are representatives of separate lines of development, which led polyphyletically from the holosts to the teleosts (see FISH).

Among the amphibians, the Labyrinthodonts and allied groups, which died out in the Late Triassic, were replaced by the first certainly identified Anura (frog and toad group). From the Bajocian strata of southern France a vertebra has been identified as the oldest find of a salamander (Urodeles). Very little fossil material bridges the gap between abundant Late Paleozoic and Triassic amphibian and the present-day orders.

Reptiles. On land the dominant forms were the archosaurians, successors to the Triassic Pseudosuchia. The best known of the reptiles are the dinosaurs (Saurischia and Ornithischia, differing in pelvis structure). The Saurischia included vegetarians and carnivores and the largest land animals in the history of life on the planet. The vegetarian sauropods were 20 to 30 metres (65 to 100 feet) long. Some carnivorous species attained ten metres (33 feet) (see REPTILIA). The Ornithischia included no carnivores.

The vegetarian therapsids, still numerous in the Triassic, had died out after several development lines split off from them, lines that led to the primitive mammals of Jurassic times.

The seas of the Jurassic were inhabited by Ichthyosauria and Plesiosauria, whose ancestors had secondarily adapted themselves to life in the water. This adaptation was further perfected in the Jurassic. Also, tortoises took to the water for the first time. The crocodiles (Archosauria) are known in the Jurassic from the Pliensbachian onward as armoured forms with land-type limbs, inhabiting shores and nearshore waters. In the Upper Jurassic the Metriorhynchidae developed modifications of limbs and tail adapted to marine life independent of the land; but the group died out in the Lower Cretaceous, before the adaptation process was complete.

Buds. The conquest of the air was accomplished in the Jurassic (presumably it had begun in the Triassic) by descendants of the Pseudosuchia. The flying, or gliding, saurians (with flight membrane on the elongated fourth finger) are known from remains of long-tailed forms, Rhamphorhynchoidea, from the earliest Jurassic onward; later in the period the short-tailed Pterodactyloidea appeared. Parallel with the flying saurians, and from similar roots, the birds emerged; the three oldest specimens known (*Archaeopteryx*) date from the Kimmeridgian of the Franconian Alps in southern Germany and might be regarded as feathered, flying saurians (with feathers as modifications of scales).

Lines of development

Plants. The terrestrial plant world, whose history is recorded in continental and marginal marine sediments, exhibits a smooth line of development from the Triassic through the Jurassic and the Lower Cretaceous. The plants are represented primarily by the pteropsid sporebearing and seed-bearing fern groups and the gymnosperms; the occurrence of angiosperms (*q.v.*) is still dubious. The Matoniaceae and the Dipteridaceae appear among the sporebearing ferns for the first time; a few still survive. A particularly large variety of the Osmundaceae (ferns) occurs. The lycopsids (club mosses) are limited to herbaceous growths. The Jurassic stand of bushes and trees consists of the gymnosperms, most numerous of which are the conifers; others include the gingkoes, cycads, Nilssoniae, and Caytonialae. Bennettitaceae also appears; it already possessed flower characteristics resembling those of the angiosperms.

GUIDE AND INDEX FOSSILS AND CORRELATIONS

Guide fossils, which are characterized by narrow vertical (time) and wide horizontal (space) distribution, plentifulness, and easy recognizability, are made use of for the definition of biostratigraphic zones (see STRATIGRAPHIC BOUNDARIES). Albert Oppel's first comprehensive zone scheme (1856–58) made use of the ammonites principally, because they underwent particularly rapid phylogenetic change. It has since become possible to structure the entire Jurassic into ammonite zones that furnish the time scale now accepted as standard.

In spite of the supposedly continuous phylogenetic development of all Jurassic ammonites from *Psiloceras* as the initial form, it is not possible to base the zonal division upon a continuous chain of ammonite species. A phylogenetic species series can never be followed for more than a finite length of time because of abrupt changes of characteristics, sudden reduction in numbers, and competition by other lines. The Jurassic zonal succession is therefore supported only by fragments of various species chains and by isolated species.

Because many guide ammonites are confined to certain areas, correlation of the ammonite zones of different regions has been achieved by matching ammonite fauna. Nevertheless, many uncertainties still exist and more studies need to be made.

Other fossils can be used as indices in regions where ammonites are lacking. Preferred fossils for such "para-indices" include brachiopods, pelecypods, snails (*Discohelix* in the Lias); and among the microfossils (important because of their great numbers) ostracods, foraminiferids, and, in the Upper Jurassic, the *Calpionella* (Ciliata). Their genera, on the average. show a longer life than those of the ammonites; yet even among them are found rapidly changing species chains and typical, short-lived associations.

Stratigraphic correlation and system boundaries

The chronological correlation of worldwide Jurassic profiles remains an uncompleted task. The most important means of global comparison is the biostratigraphic method, which involves the use of index fossils or fauna. For the dating of terrestrial Jurassic sediments, dovetailing them with marine sediments, as is possible in East Asia, is important.

The Jurassic Period as a whole corresponds to the rise of the Neoammonoidea. Its exact beginning is marked by the base of the Hettangian-type stratum, or the first ammonite zone of the Hettangian. Correlation is incomplete for the latest Jurassic stage, which zoogeographically consists of the Tithonic (Tethys), the Volaian (Boreal). and the Portland (northwest Europe). Correlation of those latest stage areas, as well as of the Jurassic–Cretaceous boundary, has not yet been done with consistency because of the gaps in the profiles of the epicontinental regression areas and because of the smooth development of the ammonites. Jost Wiedmann (1968) argues for adding the Beriassian (Lower Cretaceous) to the Jurassic. The establishment of an accepted boundary, however, can only be accomplished by explicit agreement, that also has not been reached for the still nonuniform boundaries between Lower, Middle, and Upper Jurassic.

In many areas during the Jurassic, sedimentation occurred only temporarily, during occasional marine invasion or subsidence of continental interior basins. Even in profiles that were formed under continuous ocean cover, interruptions in the series of sediments are often found. Such gaps are recognized by layers of shell fragments, seeming spurts of organic evolution, the absence of certain biozones, and the condensation of the index fauna of several zones into one horizon. Cases of extreme condensation over long periods in the ridge areas of the Tethys have been reported by Jobst Wendt from the Sicilian Jurassic.

OROGENIC ACTIVITY AND VOLCANISM

The varying distribution of land and sea primarily is caused by epeirogeny, or the relatively mild deformation and vertical movement of the earth's crust. A cyclically advancing invasion of the sea from the geosynclines onto the continental margins began at the outset of the Jurassic and, after a setback in the Bajocian, reached its peak in the Callovian–Oxfordian. Such invasions seem to be related to a general rise of the sea level, the cause of which is unknown, however. Toward the end of the Jurassic and after a second trend toward marine invasion, an extensive retreat into the geosynclinal regions occurred. This macrocycle of invasion and regression played an important part in the demarcation of the limits of the Jurassic System, especially in Europe.

In addition to epeirogenic movements, orogenic faulting and folding movements take place. It is evident from the Jurassic sediments of the Alpine geosynclines that subsidence was associated with a distension process, producing tension cracks through which basic magma (ophiolites) ascended. The compression (Alpine-type orogenesis with folding and overthrusts) was a subsequent process, which in the Alpine geosynclines was post-Jurassic, continuing into post-Cretaceous time.

Jurassic mountain building

In the North American Cordilleran geosynclines, however, compression took place after Lower Jurassic magma extrusion and at the threshold of the Middle Jurassic. This was the Nevadan orogeny, of rather uncertain date. In the California Coast Ranges, a strong geosynclinal subsidence produced 10,000 metres of Tithonic-type deposition (Franciscan formation) with associated radiolarites and basic magmas. Folding occurred later at the threshold of the Cretaceous. Similarly, in the South American Andean geosynclines, great quantities of lavas and volcanic tuffs were extruded, which are interbedded with marine and other sediments. The orogenesis here, however, was post-Jurassic. Immense igneous intrusions took place in the Caucasus (peripheral depression of the Tethys) and in the Amur and Verkhoyansk geosynclines of eastern U.S.S.R.

In Alpine-type orogenesis, subcrustal anatexis (melting and mixing) usually produces plutonic granites. In certain Asiatic mountains such plutonic rocks of Jurassic age are exposed. But among the vast Jurassic–Cretaceous, plutonic granodiorites of the Cordilleras and the Sierra Nevada only the granites of the Klamath Mountains in Oregon–California are positively known to be of Jurassic age.

Faulting and folding also occurred on pre-existing rocks outside the active geosynclines in Jurassic times. There are a number of examples: in the region of older orogenies in China; in the Yenshan Phase of Mongolia, Korea, and Japan; and in eastern Australia, in the region of the Paleozoic Tasman Geosyncline. Some geosynclinal and epicontinental sedimentation areas (*e.g.*, the Saxonian tectonics in northwest Germany) also experienced faulting and folding. Such tectonic action may be associated with igneous activity: the basalt flows of the Drachensberge in the region of the east African peripheral faults; similar flows in Queensland, New South Wales, Tasmania, and east Antarctica; tuff-filled explosion pipes of the Stormberge (Africa); volcanic rocks in Lebanon, China, eastern Siberia, and New England; and lavas and granitic intrusions in northern Nigeria.

CLIMATE

Plant growth over the entire earth and widespread coal formation during Jurassic time give evidence of a moist, warm climate that was warmer and more uniform than that of modern historic times. Seasons existed, however, as evidenced by annual growth rings of trees in the Northern and Southern hemispheres. The distribution of corals on both sides of the Tethys indicated that the Equator lay as much as 20° farther north than at present. The supplanting of mainly argillaceous rocks of the Lower Jurassic by Middle and Upper Jurassic limestones indicates increasing temperature in Europe, as also does the advance of reef corals as far as northern Scotland in the Oxfordian. A cooling trend is indicated toward the end of the Jurassic.

Oxygen-isotope determinations from belemnite shells indicate more uniform ocean temperatures than in modern times. The poles, whose Jurassic location is still problematical, seem not to have been covered with ice caps (see CLIMATIC CHANGE).

BIBLIOGRAPHY. W.J. ARKELL, *Jurassic Geology of the World* (1956), is a comprehensive textbook, treating biostratigraphic classification and correlation that presents an image of the earth surface in the Jurassic Period and includes a complete bibliography; H. HOLDER, *Jura*, in F. LOTZE (ed.), *Handbuch der stratigr. Geologie*, vol. 4 (1964), supplements Arkell's book, but focusses on the middle European and Alpine Jurassic.

Important regional monographs are presented by E.D. MCKEE *et al.*, "Paleotectonic Maps of the Jurassic System," *Misc. Geol. Investigations, U.S. Geol. Surv.* (1956); W.J. ARKELL, *The Jurassic System in Great Britain* (1933); R.W. IMLAY, "Correlation of the Jurassic Formations of North America Exclusive of Canada," *Bull. Geol. Soc. Am.*, 63:953–992 (1952); H. FREBOLD, "The Jurassic Fernie Group in the Canadian Rocky Mountains and Foothills," *Mem. Geol. Surv. Can.* 287 (1957); K. HOFFMANN and W. SCHOTT, "Oil Accumulation and the Jurassic System in Northwest Germany," *World Petrol. Congr. Proc. 4* (1955).

Some recent zoogeographical publications include: R.W. IMLAY, "Jurassic Marine Faunal Differentiation in North America," *J. Paleont.*, 39:1023–1038 (1965); H. FREBOLD, "The Jurassic Faunas of the Canadian Arctic: Middle and Upper Jurassic Ammonites," *Bull. Geol. Surv. Can. 74* (1961); W.J. ARKELL, "Jurassic Ammonites from Jebel Tuwaiq, Central Arabia," *Phil. Trans. R. Soc.*, series B, vol. 236, no. 633 (1952); T. SATO, "Etudes biostratigraphiques des Ammonites du Jurassique du Japon," *Mém. Soc. Géol. Fr., 94* n.s., vol. 41 (1962); G.R. STEVENS, "Faunal Realms in Jurassic and Cretaceous Belemnites," *Geol. Mag.*, 100:481–497 (1963); G.E.G. WESTERMANN, "The Ammonite Fauna of the Kialagvik Formation at Wide Bay, Alaska Peninsula (Bajocian)," *Bull. Am. Paleont.*, pt. 1, vol. 47 (1964), pt. 2, vol. 57 (1969); and A. HALLAM, "Tectonism and Eustasy in the Jurassic," *Earth-Sci. Rev.*, 5:45–68 (1969), a modern approach to sea-level fluctuation in the Jurassic Period.

(H.H.)

Jury

The jury is an historic Anglo-American legal institution in which a group of laymen participate in a major way in deciding cases brought to trial. Its exact characteristics and powers depend on the laws and practices of the countries, provinces, or states in which it is found, and there is considerable variation. But its basic characteristics make it a remarkable political institution. It recruits laymen at random from the widest population for the trial of a particular case and allows them to deliberate in secrecy, to reach a decision by other than majority vote, and to make it public without giving reasons. Throughout its history it has been both overpraised as a charter of liberty and overcriticized as reliance on incompetent amateurs in the administration of justice.

History **and** use. The jury's origin is lost in the past. It may have been indigenous to England or have been brought there by the Norman invaders in 1066. Originally, the jurors were neighbourhood witnesses who passed judgment based on what they themselves knew. But the breakdown of medieval society and the growth of the towns changed this; the jury was called upon to determine the facts of the case, based upon the evidence presented in court. The availability of the jury in the king's courts may have been a key factor in centralizing the nation's courts under the king and in creating the common law. By the 15th century, nonrational modes of trial such as ordeal, in which the defendant was subjected to various tortures that, if successfully endured, proved his innocence, were replaced by the jury trial, which became the established form of trial for both criminal and civil cases at common law.

Early development in England

Two forces moved the jury abroad. One was the expansion of the British Empire, which brought the jury to Asia, Africa, and the American continent. The other was the French Revolution and its aftermath, which brought it, as a symbol of popular government, to the European continent: first to France itself, then, through Napoleon, to the Rhineland, later to Belgium, most of the remaining German states, Austria-Hungary, Russia, Italy, Switzerland, Holland, and Luxembourg, although the last two abolished it immediately after Napoleon's defeat. In each of these countries, use of the jury was from the outset limited to trials of major crimes and of political crimes against the state.

Beginning in the mid-19th century, the jury was emasculated in a variety of ways: in 1850, Prussia, for example, removed treason from its jurisdiction; in 1851, the duchy of Nassau removed all political crimes; in 1923, Czechoslovakia removed treason and, one year later, libel; in 1919, Hungary suspended jury trial entirely and never restored it. Germany abandoned the jury in 1924. Both the Soviet bloc and the fascist states abolished it outright; France never restored the jury abolished during the German occupation in the 1940s, and Japan did away with its short-lived jury courts in 1943. After World War II, Austria reintroduced the jury in a weakened form.

United States as major user of jury system

Thus, there are three important points about the history and development of the jury as a legal institution: first, the effort to introduce it outside the Anglo-American legal orbit has failed; further, in England itself, its use has limited by statute to a small category of cases; and thus, the United States has emerged today as the home of the jury system for both criminal and civil cases. Some 120,000 jury trials are conducted there annually, more than 90 percent of all jury trials in the world.

Use of the jury in the United States depends on two factors: the degree to which it is available as a matter of right, and the degree to which the parties themselves choose to use it. The laws as to its availability have varied from state to state, but in 1968 in *Duncan* v. *Louisiana,* the United States Supreme Court declared that a jury trial is a constitutional right in all criminal cases in which the penalty may exceed six months. In civil cases its constitutional status is less clear, but in general jury trial is available. The practice of allowing the parties to waive a jury trial also varies widely from region to region, and as a result, the number of jury trials per year also varies widely. The annual number of criminal jury trials per 100,000 population ranges between 3 for Connecticut to 144 for Georgia.

Jury procedures. *Selection.* Historically, there were some minimum requirements of property and competence for jury service. More recently, the idea of genuine random selection from the population, to achieve a cross section of the community, has been gaining ground. Since 1969, it has been the principle of selection in the federal courts. Most jurisdictions exempt some groups from jury service: policemen, lawyers, doctors, and so on. Some jurisdictions exempt women entirely. All jurisdictions excuse jurors if the service imposes undue hardship.

The voir dire

The commitment of important decisions to a random group of laymen has been moderated, particularly in the United States, by an elaborate screening, voir dire, conducted by trial counsel at the inception of a trial. The law permits counsel to challenge prospective jurors either for cause (if there is specific likelihood of bias) or, for a limited number, "peremptorily"—that is, without having to give a reason. American trial tradition attaches a great deal of significance to the strategies of juror selection, and in celebrated cases, the lawyers' voir dire examination has extended for several weeks.

Size and unanimity. Traditionally, the jury had 12 members and was required to reach its decisions with unanimity, a striking arrangement in Anglo-American countries that make all other decisions by majority vote. Over the years some modifications have been made. Some jurisdictions prescribe or allow in minor cases a jury of six. Oregon allows 10:2 verdicts—that is, a majority of 10—in all criminal cases, except capital ones; and in 1968 Great Britain followed the Oregon example. A few Southern states in the United States allow majority verdicts in misdemeanour trials. In civil cases, many states now allow 10:2 verdicts. When the required number (12 or 10) of jurors cannot agree on a verdict (termed a hung jury in the United States), the judge declares a mistrial, which means the case, unless it is withdrawn, must be tried anew. It is somewhat remarkable that "hung" juries occur with relative infrequency even when unanimity is required. In Europe juries operate under a different principle. Unless at least two-thirds of all the jurors vote guilty, the defendant must be acquitted. The U.S. Army court-martial jury also operates under this principle.

Sentencing. Although in civil cases the jury decides both issues of liability and amount of damages, in criminal cases it has been restricted generally to the issues of guilt, with punishment left to the judge. In some Southern U.S. states, however, the jury also decides the sentence within a range that the law provides. And in all jurisdictions that have retained the death penalty, if the jury finds the defendant guilty of the capital crime, it decides, or at least expresses an opinion, as to whether the death penalty is to be imposed. In most jurisdictions decisions on guilt and sentence are rendered simultaneously, but California has introduced the so-called second trial in capital cases, which occurs after a guilty verdict. At such a "second trial" pleas and evidence are presented for and against the imposition of the death penalty in the specific case, and only then is the jury asked to determine the sentence.

Control. Trial by jury is, of course, trial by jury under the supervision of a judge. The formula for sharing power between judge and jury is complex. First, the judge decides what the jury may or may not hear under the rules of evidence. Second, if the judge finds that the evidence presented leaves no factual issue to be resolved, he may withdraw the issue from the jury and direct the jury to acquit a defendant or, in a civil trial, find for either plaintiff or defendant; he cannot, however, direct a guilty verdict in a criminal trial. Third, in some jurisdictions the judge may, and often will, summarize the evidence or even discuss its weight. Fourth, the judge instructs the jury as to the law it should apply in reaching the verdict. Finally, if the judge finds the jury's verdict to be manifestly against the weight of the evidence, he may with one exception set it aside and order a new trial. The only exception is in a criminal case in which the jury renders an acquittal; under Anglo-American law (though not under European continental law) the jury's acquittal is always final.

The jury normally renders a general verdict, that is, a yes or no answer to liability or guilt, and does not give reasons for its decision. At times, however, courts employ "special verdicts" or "special interrogatories" in which the jurors are asked to decide a series of specific factual issues that bear on the overall verdict.

The controversy over the jury. The jury has been enmeshed in a perennial debate as to its merits, a debate that has recruited some of the great names in law and political philosophy, from Montesquieu, William Blackstone, and Thomas Jefferson to present-day theorists and practitioners, and has centred on three issues. First, there is the debate about collateral aspects: there are favourable contentions that the jury provides an important civic experience, that it makes tolerable the stringency of certain decisions, that it acts as a sort of lightning rod for animosity that otherwise might centre on the more permanent judge, and that the jury is a guarantor of integrity since it is said to be more difficult to bribe 12 men than one. Against this it has been urged that jury duty disenchants the citizen, that it imposes an unfair burden, that the jury is expensive, and that it makes it difficult to do away with the often interminable delays that exist in civil litigation.

The jury's competence

Second, there is the issue of the jury's competence. It is argued that the judge, by training, discipline, experience, and superior intelligence, is better able to understand law and facts than laymen drawn from a broad range of levels of intelligence, without experience in matters of this sort and without durable official responsibility. But it is also argued that 12 heads are better than one, that the jury as a group has wisdom and strength beyond that of its individual members, that it makes up in common sense and experience what it lacks in training, and that its very inexperience is an asset because it secures a fresh perception of each trial, avoiding the stereotypes that may infect the judicial eye.

Finally, there is the question of the jury's interpretation of the law. The critics complain that the jury will not follow the law, either because it does not understand it or because it does not like it, and hence will administer justice unevenly and that the jury produces a government by men and not by rule of law, against which Anglo-American political tradition is so steadfastly set. The jury's champions offer this very flexibility as its most endearing characteristic. They see the jury as a remarkable device for ensuring that the rigidity of the general rule can be shaped to justice in a particular case, with government by the spirit of the law and not by its letter.

Performance. In a recent survey of some 7,000 jury trials, the presiding judges were requested to reveal how they would have decided without a jury; the results of the survey provided some major insights into the actual performance of the contemporary American jury. In both civil and criminal trials, judge and jury agreed in 78 percent of all verdicts. In civil cases the disagreement in the remaining cases was symmetrically split; in 19 percent of the criminal cases, however, the judge would have convicted, whereas the jury acquitted. The letter of the law confines the jury to "finding the facts," but the deviations from the judge are mostly due to the jury's subtle, and not always conscious, injecting its sense of justice into a case that might go either way. This sense of justice may be concerned with the person of the accused, with the threat of too harsh a punishment, or with the content of the criminal law rules. Thus, close study of the jury has revealed it as a highly sensitive institution, subtle and discerning, moved by factors far beyond gross sympathy for the defendant. On the whole, the system tolerates and even appreciates these deviations of the jury from the judge, even if in rare cases they reflect what the national community experiences as intolerable local prejudice.

Lay judges. The jury, though a distinctive institution, must be seen in the larger context of other forms of lay participation in the judicial process, a feature of most legal systems since the days of Athens and Rome. Today, the eastern European bloc, West Germany, Austria, and some Scandinavian countries have mixed tribunals in which lay judges sit side by side with learned judges, in varying proportions, on both criminal and civil cases in trial and appeals courts. The largest of these mixed tribunals is the so-called jury in West Germany and France, with three learned and six lay judges forming one tribunal. At the other extreme are single lay judges, such as the English magistrates and the now almost defunct justice of the peace in the United States.

There is a difference in the performance of juries deliberating alone and lay judges deliberating jointly with judges on mixed tribunals. Only juries reach verdicts that are often significantly different from what the learned judges would have done. Lay judges in mixed tribunals --even when they outnumber the learned judges--exert little influence on the guilt–innocence issue, somewhat more on the punishment. But perhaps the very presence of lay judges ensures that the learned judges never stray too far from what the community would allow.

BIBLIOGRAPHY. WILLIAM FORSYTH, ***History of Trial by Jury*** (1852), the classic treatise; ROSCOE POUND, "Jury" (England and the United States), ***Encyclopedia of the Social Sciences,*** vol. 8 (1932), a distinguished essay on the jury written prior to the era of empirical investigation; G.L. WILLIAMS, ***The Proof of Guilt,*** 3rd ed. (1963), a general discussion of the criminal trial, with an important chapter (10) on the jury; LORD DEVLIN, ***Trial by Jury*** (1956, reprinted 1966), a view of the jury by one of England's great judges; J.P. DAWSON, ***A History of Lay Judges*** (1960), a comprehensive history of lay judges, of which the jury is but one form, from antiquity to the present; HARRY KALVEN, JR., and HANS ZEISEL, ***The American Jury*** (1966), a major empirical study of the American criminal jury based on 3576 actual jury trials for which the presiding judge reported how he would have decided the case without a jury; W.R. CORNISH, ***The Jury*** (1968), a comprehensive British essay on the jury, combining traditional learning with new empirical material; R.J. SIMON, ***The Jury and the Defense of Insanity*** (1968), an experimental study of how differences in the law of insanity affect the outcome of jury trials.

(H.Ka./H.Z.)

Justinian I

The spectacular and enduring legal and architectural monuments of the great Byzantine emperor Justinian's reign and the vivid personalities of the Emperor and his contemporaries have caused his age to be recognized as a highly significant period in world history. This recognition has indeed sometimes even led to an unjust appraisal of Justinian's work, particularly his ambitious foreign policy, causing his achievements to be seen out of per spective and in isolation rather than as part of a long and continuous imperial development.

Alinari—Giraudon

Justinian I, detail of a mosaic, 6th century. In the Basilica of San Vitale, Ravenna, Italy.

Early career. Justinian I (Flavius Petrus Sabbatius Justinianus, called the Great), Byzantine emperor from 527 to 565, was a Latin-speaking Illyrian, born in 483 of peasant stock in Tauresium (a village probably south of Niš in present-day Yugoslavia). The names Petrus Sabbatius are known only from a consular diptych of 521 and were dropped after his accession to the throne. Justinianus was a Roman name that he took from his uncle, the emperor Justin I, to whom he owed his advancement. Nothing is known of his early life except that he went, while still a young man, to Constantinople, where his uncle held high military command. He received an excellent education, though it was said that he always spoke Greek with a bad accent. When Justin became emperor in 518, Justinian was a powerful influence in guiding the policy of his elderly and childless uncle, whose favourite nephew he was. He was legally adopted by Justin and held important offices. In 525 he received the title of caesar and, on April 4, 527, was made coemperor with the rank of augustus. At the same time his wife, the former actress Theodora, who exercised considerable influence over him, was crowned augusta. On Justin I's death on August 1, 527, Justinian succeeded him as sole emperor.

Influence in government and accession to the throne

Foreign policy and wars. Two important facets of Justinian's foreign policy were his continuation of the age-long struggle with Persia and his attempt to regain the Roman provinces in the West, lost to barbarian invaders.

Eastern wars. When Justinian came to the throne, his troops were fighting on the Euphrates against the armies of the Persian king Kavadh (Qobād) I. After campaigns in which the Byzantine generals, among whom Belisarius was the most distinguished, obtained considerable successes, a truce was made on the death of Kavadh in September 531. His successor, Khosrow I, finally came to terms and the Treaty of Eternal Peace was ratified in 532. The treaty was on the whole favourable to the Byzantines, who lost no territory and whose suzerainty over the key district of Lazica (Colchis, in Asia Minor) was recognized by Persia. Justinian, however, had to pay the Persians a subsidy of 11,000 pounds of gold, and in return Khosrow gave up any claim to a subvention for the defense of the Caucasus.

War broke out again in 540, when Justinian was fully occupied in Italy. Justinian had somewhat neglected the army in the East, and in 540 Khosrow moved into Mesopotamia, northern Syria, and Byzantine Armenia and systematically looted the key cities, including Antioch. In 541 he invaded Lazica in the north. Belisarius, now reappointed commander in chief in the East (*magister militum per Orientem*), launched counteroffensives in 541 and 542 before his recall to Italy. The war dragged on under other generals and was to some extent hindered by the spreading of bubonic plague. A five years' truce was made in 545 and renewed in 551 but still did not extend to Lazica, which the Persians obstinately refused to restore, and a fierce struggle continued intermittently in this mountainous region. When the truce was again renewed in 557, however, Lazica was included. Finally, a 50 years' truce was negotiated, probably at the end of 561; Byzantium agreed to pay an annual tribute of 30,000 solidi (gold coins), and the Persians renounced all claim to the small Christian kingdom of Lazica, an important bulwark against northern invaders. They did, however, keep their hold over Suania, a small Lazican dependency. Justinian had thus maintained his eastern provinces virtually intact in spite of the vigorous offensives of the Persian King, so his policy on this front can hardly be described as a failure.

Regaining of lost provinces

Western wars. In the West, Justinian's policy followed inevitably from his conception of imperial responsibility; he considered it his duty to regain provinces lost to the empire "through indolence," and he could not ignore the trials of Catholics living under the rule of Arians (Christian heretics) in Italy and in North Africa. In both these regions he had reason to believe that his help was desired and that political circumstances were favourable to his intervention. In the Vandal kingdom of North Africa, an Arian domain, Catholics had been subject to frequent persecution. There was also a disputed succession to the throne after the aged Vandal king Hilderich, who had been in alliance with Constantinople and had ceased persecution of the Catholics, was deposed in favour of Gelimer in 530. At the same time, the Vandals were threatened by the Moorish tribes of Mauretania and south Numidia. In the face of considerable opposition from his generals and ministers, Justinian launched his attack on North Africa to aid Hilderich in June 533. The fleet of about 500 vessels set out with 92 warships. Procopius, Belisarius' legal advisor (assessor), reports that there were about 15,000 regular troops in the transport ships, along with 1,000 barbarian allies, as well as the bodyguard (buccellarii) of Belisarius, who was in command of the expedition. An unopposed landing was made in August, and by the following March (534) Belisarius had mastered the kingdom and received the submission of the Vandal ruler Gelimer. Northern Africa was reorganized as part of the empire and now included Sardinia, Corsica, the Balearic Islands, and Septem (Ceuta). Justinian placed the conquered territory under a praetorian prefect of Africa with seven provincial governors. There was also a commander in chief (magister *militum* per Africam) with five duces, or military leaders (four African and one of Sardinia). Both the prefecture and the magisterium *militum* per Africam were centred in Carthage. Justinian, with his usual care for detail, issued precise orders regulating the duties and salaries of these officials. The native inhabitants and the Catholic Church were able to regain land extorted from them by the Vandals. Heretics and Jews were penalized. The reconquest was celebrated by a triumph in Constantinople granted to Belisarius, and the treasure (part of the Vandal loot) was displayed together with the Vandal prisoners, who were subsequently enrolled into imperial regiments active on the eastern frontier.

After Belisarius' departure from North Africa in 534 there was some discontent, particularly among Roman soldiers who had married the dispossessed Vandal women and also among the Arians. Mutiny in the army, Moorish attacks, and revolts—particularly during 543–548—were, however, suppressed, and good government was restored along the North African coast. In 547 Septem was attacked by the Visigoths from Spain. This was followed by a disputed succession and a revolt by the Catholics of Córdoba in Spain. In 552 Justinian sent an expedition to support Athanagild, the claimant to the throne, and to strengthen the conquered part of southeast Spain, which for purposes of administration was attached to North Africa.

In Italy, the mother province of the Roman Empire in which the elder capital city (Rome) was situated, Justinian found a situation similar to that in North Africa and particularly favourable to his ambitions. Under his immediate predecessors, Italy had been ruled by a barbarian, the Ostrogoth Theodoric, who, though virtually independent, was the nominal representative of the Byzantine emperor. He was an Arian and, though at first a tolerant and wise ruler, toward the end of his reign had begun to persecute the Catholics. He had no male heir, and on his death there was not only antagonism between Arian Goths and Catholic Italians but a rift within the ranks of the Ostrogoths, some of whom were violently anti-Byzantine. Theodoric's daughter, Amalasuntha, appealed to Justinian for help, but she was killed in 535 by the Gothic nationalist party after the death of her young son Athalaric in 534.

Re-establishment of imperial administration in Italy

Thinking that this was now his opportunity to support his fellow Catholics and to reassert direct control over the province, Justinian dispatched an army to Italy via Dalmatia and sent Belisarius with a fleet to attack Sicily, while an embassy set off to gain the support of the powerful Franks now settled in Gaul. After the defeat of the Ostrogothic king Witigis and the capture of Ravenna in 540, imperial administration was re-established in Italy under the praetorian prefect Athanasius. Rigorous financial exactions and the rapacity of the soldiers made the new regime unpopular. Many of the Ostrogoths had never submitted, and after the two short and unfortunate reigns of Hildebad and Eraric, they proclaimed Totila (Baduila) as their king in the autumn of 541. Totila proved a vigorous and able leader: in 542 he took the offensive in southern Italy and in 543 captured Naples. In 544 Belisarius was sent against him with inadequate forces. As contributing factors to the successes of Totila, Belisarius encountered discontent among the Italians at the financial scrutiny of Justinian's auditors, as well as demoralization among the troops stationed there, because of arrears of pay. City after city was captured by the Ostrogoths until only Ravenna, Otranto, and Ancona remained in Byzantine hands. Belisarius could make no headway without adequate reinforcements. which his wife Antonina's personal appeal to Justinian failed to obtain, and in 549 he was recalled to Constantinople. Meanwhile, Totila took over the administration of the country, though at the expense of alienating the great landowners. He hoped to come to terms with Justinian, but in 552 a powerful army was sent against him under the eunuch commander Narses. Totila was defeated by superior numbers and strategy and was mortally wounded at the battle of Busta Gallorum. Narses entered Rome and soon afterward defeated Ostrogothic resistance at Mount Lactarius, south of Vesuvius. Pockets of resistance, reinforced by Franks and Alemanni who had invaded Italy in 553, lingered on until 562, when the Byzantines were in control of the whole of the country.

Justinian hoped to restore the social and economic well-being of Italy by a series of measures embodied in the Pragmatic Sanction of 554. These dealt with such problems as the restoration of confiscated property, the return of slaves and *coloni* (serfs) to owners and landlords, and the collection of taxes. The country was so ravaged by war, however, that any return to normal life proved impossible during Justinian's lifetime, and only three years after his death part of the country was lost to the Lombard invaders.

On the northern frontier in the Balkans, the Roman provinces faced continual attacks from barbarian raiders. Thrace, Dacia, and Dalmatia were harried by Bulgars and Slavs (known as Sclaveni). In 550–551 the invaders even wintered in Byzantine territory, despite the efforts of the army to dislodge them. In 559 the Bulgars and Slavs were joined by the Kotrigur Huns, who got as far south as Thermopylae and eastward through Thrace to the long wall protecting Constantinople. The veteran Belisarius saved the situation by mustering the civilian population. In 561 the Avars joined the raiders but were bought off with a subsidy. These attacks from beyond the Danube did immense damage, and although fortifications and defense works were built and strengthened in the Balkans and in Greece, the newcomers were neither effectively repulsed nor assimilated by the Byzantines. The Slavs, and later the Bulgars, eventually succeeded in settling within the Roman provinces. Failure to keep them out is one of the criticisms sometimes made against Justinian. The Slavs in the Balkans and in other parts of the empire, however, did provide a source of manpower, which was particularly welcome in view of the ravages of the 6th-century bubonic plague.

The Code of Justinian

Internal policy. Legislation. Justinian's best known work was as a codifier and legislator. He greatly stimulated legal studies, and in 528 he set up a commission to produce a new code of imperial enactments or constitutions. This was published in 529, and in 530 a second commission sat to codify the Roman jurists; the work of this commission, known as the Digest, appeared in 533. At the same time a handbook for the use of law students, the Institutes, was prepared and published in 533. A second edition of the Codex Justinianus containing Justinian's own laws up to the date of issue was published in 534. His subsequent legislative work to 565 is known as the Novels (Novellae constitutiones). Much of this legal activity was supervised and inspired by Tribonian, a gifted advocate, who had by 529 become quaestor of the sacred palace, the most important judicial minister. He probably played a major part in framing these laws.

Administration. Justinian was genuinely concerned with promoting the well-being of his subjects by rooting out corruption and providing easily accessible justice. This involved adequate control over provincial governors and some administrative reorganization. At the same time it was essential to provide revenue for Justinian's various campaigns, particularly in the West. Justinian knew how to pick his servants. He had two outstanding ministers. One was John of Cappadocia from Asia Minor—outspoken, blunt, efficient, incorruptible, treating rich and poor alike, a clerk who rose to the office of praetorian prefect in 531; he was a man with many enemies, including the empress Theodora, who achieved his dismissal in 541. The other was Peter Barsymes, a Syrian clerk in the financial department of the praetorian prefecture, who became praetorian prefect of the East in 543 and was evidently a careful administrator, even though his name is blackened by Procopius.

Administrative reforms

The first important reform was the prohibition of the *suffragia*, or sale of provincial governorships, in 535, for it was clear that desire to recoup oneself for this heavy initial outlay to buy the offices accounted for much extortion suffered by the provincials. Instructions were drawn up for provincial governors, and the position of the *defensores* civitatis, the officials whose duty it was to protect the cities, was strengthened and their jurisdiction widened so that provincials did not need to have the expense of going to the governor's court. At the same time there was a reorganization of the provincial system. Changes were made in the dioceses of Asiana, Pontica, and Oriens and in Egypt, involving the abolition of the vicariates (administrative officials) and the regrouping of the provinces, thus effecting some economy. In some cases (*e.g.*, Egypt and Cappadocia, where only purely internal civil disorders might arise), civil and military authority was combined and the governor given the higher rank of *spectabilis* (notable) with final jurisdiction in cases of less than 500, later 750, solidi. But where it might be necessary for the duces to defend a strategic point against an enemy, the civil and military powers remained distinct, as in Syria.

Financial reforms

These rather piecemeal changes sometimes resulted in situations in which, despite Justinian's efforts, the provincial governor was unwilling or not sufficiently strong to enforce good government; and with the disappearance of the larger unit of the vicariate, there was nothing left but an expensive appeal to Constantinople. In order to meet these difficulties, there was a partial renewal of a different kind of vicariate after John of Cappadocia's fall. Under both John and his successor, Peter Barsymes, finance was of primary concern, since wars and subsidies to foreign powers and barbarian rulers were expensive. Regained provinces, especially in Italy, were often so devastated that they could not pay their way, and income from these provinces was, therefore, very irregular. The money was found, however, largely through the financial ability of Justinian's two ministers. Taxes were efficiently collected, accounts audited, misappropriations tracked down, public expenses pruned, and city spectacles and municipal services cut down—thus alienating both rich and poor alike in the provinces and in the capital. One important source of revenue came from trade and industry that had long been vigorously promoted in the Roman Empire, particularly with India, Southeast Asia, and China. Spices, perfumes, and raw silk were among the most important imports, silk coming either by way of Ceylon and Persia or by caravan through Central Asia. Persian hostility could cause prices to rise or even stop supplies. A feature of this period was the introduction of the silkworm from Sogdiana (Samarkand and Bukhara), so that the Byzantine market eventually achieved an independent supply. The sale of raw material was a government monopoly and Peter Barsymes, Justinian's finance minister, extended this to silk fabrics, thus creating another lucrative state monopoly.

Government attempts to root out abuses and its attacks on vested interest, whether of rich or poor, were unpopular. In 532 public discontent was voiced most dramatically in Constantinople by the Nika revolt (Nika, "Conquer" or "Win," was the cry of rival factions at the races in the hippodrome). The city parties known as the Greens and the Blues united and attacked and set fire to the city prefect's office and public buildings, as well as to part of the imperial palace and the Church of the Holy Wisdom adjoining it. Then they gathered in the hippodrome, calling for the dismissal of the city prefect and of Justinian's two ministers, John the Cappadocian and the advocate Tribonian. Justinian agreed, but the mob was by now out of control and perhaps exploited by interested parties. The next day the late emperor Anastasius' nephew Hypatius was proclaimed emperor and was supported by certain senators. The crowd and the usurper assembled in the hippodrome. At this point two factors saved Justinian. Theodora persuaded him to stand his ground, and the generals in the city, Belisarius and Mundus, mustered what troops they could and resolutely turned on the mob in the hippodrome. They gained the upper hand, helped by Narses, though only at the cost of a wholesale massacre of the rebellious citizens. Hypatius was executed. The uprising had nearly cost Justinian his throne.

Buildings. Justinian's costly building program was motivated by his sense of imperial responsibility. It is described in detail in Procopius' book, On Buildings (De aedificiis), and ranged from assisting civic authorities in special need, such as rebuilding cities devastated by earthquakes (*i.e.*, Antioch in 528), to the provision for aqueducts and bridges, or the strengthening of the massive fortifications on the extensive frontiers, the remains of which can still be seen. Monasteries, orphanages, and hostels were provided, as well as superb churches, such as Hagia Sophia in Constantinople, the daring construction of which was owed to the genius of the mathematicians and engineers Isidore of Miletus and Anthemius of Tralles, and which still stands as a breathtaking monument to Byzantine genius.

Ecclesiastical policy. To treat Justinian's ecclesiastical policy separately is in a sense misleading, since it obscures the extent to which church and state were indissolubly linked as essential aspects of a single Christian empire, the terrestrial counterpart of the heavenly polity. It

Support of orthodoxy

was therefore the duty of Justinian, as it was for later Byzantine emperors, to promote the good government of the church and to uphold orthodox teaching. This explains why so many of his laws deal in detail with religious problems, being either directed against pagans or heretics or concerned with matters of ecclesiastical and monastic discipline. Pagans, heretics, and Samaritans, for instance, were forbidden to teach any subject whatsoever, and though fully appreciative of the classical heritage, Justinian expelled pagan teachers from the once famous Academy at Athens, an action directed against paganism rather than Greek philosophy. In 542–543 Justinian took the initiative in condemning the teaching of Origen, the great theologian of Alexandria in the 3rd century, which was spreading in monastic circles in Palestine. His edict, which had the support of the episcopate and the Pope, also anathematized Origen himself.

Justinian's main doctrinal problem was the conflict between the orthodox view accepted at the Council of Chalcedon (451), that the divine and human natures coexist in Christ, and the Monophysite teaching that emphasized his divine nature. Monophysitism was strongly held in Syria and Egypt and was closely allied to growing national feelings and resentment of Byzantine rule. Justinian, whose wife, Theodora, was a strong champion of the Monophysites, did not wish to lose the eastern provinces, but, on the other hand, any concessions to them would almost certainly alienate Rome and the West. Soon after his accession he had tried to win over the Monophysite leaders, treating them with such consideration that Pope Agapetus protested. He had made a further attempt at compromise when in 544 he issued an edict condemning parts of the writings of Theodore of Mopsuestia, Theodoret of Cyrrhus, and Ibas of Edessa on the ground that these were tainted with Nestorianism (a heretical view holding that there are two separate Persons in the Incarnate Christ, the divine and the human), thus hoping to placate the Monophysites, who charged the Chalcedonians with Nestorian sympathies. This edict roused an outcry in the West but had some support from the eastern bishops. Pope Vigilius, summoned to Constantinople, knew that if he supported imperial policy he would be repudiated by the West. The second Council of Constantinople (553) finally reaffirmed the Chalcedonian position and condemned the so-called Three Chapters (this originally referred to Justinian's three heads of condemnation but was later used of the suspect writings). Justinian achieved nothing by the episode of the Three Chapters: he did not conciliate the Monophysites, he enraged Antioch by the attack on its teachers, and he aroused Rome particularly by his handling of Pope Vigilius and his attempt to determine doctrinal matters. The decrees of the council were not accepted by Vigilius' successors, and a schism thus occurred between Rome and Constantinople lasting until 610.

Toward the end of his reign, Justinian to some extent withdrew from public affairs and was occupied with theological problems. He even lapsed into heresy when, at the end of 564, he issued an edict stating that the human body of Christ was incorruptible and only seemed to suffer (the doctrine called Aphthartodocetism). This roused immediate protest and many ecclesiastics refused to subscribe to it, but the matter was dropped with the Emperor's death, at which time the throne passed to his nephew Justin II on November 14, 565. To describe Justinian's interest and activity in church affairs as caesaropapism (state in control of the church) is misleading. Justinian, like succeeding Byzantine emperors, regarded himself as the vicegerent of Christ, and the Eastern Roman Empire knew no such clear-cut distinction between church and state as developed in Latin Christendom.

Character and significance

Assessment. Two mosaic pictures of Justinian exist at Ravenna, one in San Vitale, the other in Sant' Apollinare Nuovo. In appearance he was somewhat above middle height, with fresh colour in his cheeks; he had an extraordinary power of doing without food and rest and was called "the emperor who never sleeps." Justinian was a man of large views and great ambitions, of wonderful activity of mind, tireless energy, and an unusual grasp of detail. In attempting to restore the Old Roman Empire, Justinian's work has all too often been regarded in isolation. His aims—though by reason of historical circumstances on a more grandiose scale than most rulers—were those of nearly every Byzantine emperor, even though such plans were generally doomed to failure. It should not be forgotten that Justinian renewed Byzantine rule and Hellenic influence in parts of Italy for several centuries and that, for over a half century, sound government was given to North Africa, from which came salvation for Constantinople in the person of Heraclius in 610. Justinian's legal work and the magnificent Great Church (as Hagia Sophia was called) have won him unending fame, and the literature, poetry, and philosophical achievements of his contemporaries witnessed to the outstanding quality of 6th-century civilization in the Eastern Roman Empire.

BIBLIOGRAPHY. J.B. BURY, ***History of the Later Roman Empire,*** vol. 2 (1923), a narrative survey; E. STEIN, ***Histoire du Bas-Empire,*** vol. 2 (1959), a detailed study; C. DIEHL, ***Justinien et la civilisation byzantine*** (1901), the classic study now in need of some revision; B. RUBIN, ***Das Zeitalter Iustinians,*** vol. 1 (1960); G. DOWNEY, ***Constantinople in the Age of Justinian*** (1960), a lively account of secular and ecclesiastical life; P.N. URE, ***Justinian and His Age*** (1951), a stimulating account, but also needs some revision; R. BROWNING, ***Justinian and Theodora*** (1971), a good introduction for the general reader, with lavish illustrations, useful notes on sources, and bibliography; A.H.M. JONES, ***The Later Roman Empire, 284–602,*** 3 vol. and maps (1964), a valuable and objective analysis; J.W. BARKER, ***Justinian and the Later Roman Empire*** (1960), a general introduction.

(J.M.H.)

Juvenal

Juvenal (Decimus Junius Juvenalis) was the last and most powerful of the Roman satiric poets. Since he seldom speaks directly of himself in his poems and gives no important facts about his life and character, conjecture —based on the topical and personal allusions in his Satires, three poems about him by the epigrammatist Martial, a few references by later writers, and an inscription (now lost)—must take the place of documentation.

Life. The one contemporary who ever mentions Juvenal is Martial, who claims to be his friend, calls him eloquent, and describes him as living the life of a poor dependent cadging from rich men. There are a few biographies of him, apparently composed long after his death; these may contain some nuggets of fact, but they are brief, ill-proportioned, and sometimes incredible.

From these sparse sources it can be inferred that Juvenal was born in the town of Aquinum southeast of Rome, perhaps between AD 55 and 60; that his family was well-to-do; and that he became an officer in the army as a first step to a career in the administrative service of the emperor Domitian (AD 81–96), but failed to obtain promotion and grew embittered. He wrote a satire declaring that court favourites had undue influence in the promotion of officers, and for this he was banished—possibly to the remote frontier town of Syene, now Aswān, in Egypt—and his property was confiscated. In 96, after Domitian's assassination, he returned to Rome; but, without money or a career, he was reduced to living as a "client" on the grudging charity of the rich. After some years his situation improved, for autobiographical remarks in Satire 11 show him, now elderly, living in modest comfort in Rome and possessing a farm at Tibur (now Tivoli) with servants and livestock. Although still pessimistic, the later Satires show a marked change of tone, and some touches of human kindness, as though he had found some consolation at last. Though no details of his death exist, he probably died in or after 127.

The Satires. Juvenal's 16 satiric poems deal mainly with life in Rome under the much-dreaded emperor Domitian and his more humane successors Nerva (96–98), Trajan (98–117), and Hadrian (117–138). They were published at intervals in five separate books. Book One, containing Satires 1–5, views in retrospect the horrors of Domitian's tyrannical reign and was issued between 100 and 110. (The historian Tacitus, a contemporary of Juvenal, was also embittered by the suspicion and fear of that epoch.) Book Two, the single, enormous Satire 6, contains topical references to the year 115. The third Book, with Satires 7, 8, and 9, opens with praise of an emperor—surely Hadrian, who endowed a literary institute to assist deserving authors—whose generosity makes him the sole hope of literature. There is no datable allusion in Book Four, which comprises Satires 10–12. Book Five, made up of Satires 13, 14, 15, and 16, has two clear references to the year 127.

Themes of the Satires

The Satires attack two main themes: the corruption of society in the city of Rome and the follies and brutalities of mankind. In the first Satire, Juvenal declares that vice, crime, and the misuse of wealth have reached such a peak that it is impossible *not* to write satire, but that since it is dangerous to attack powerful men in their lifetime, he will take his examples from the dead. He does not maintain this principle, for sometimes he mentions living contemporaries; but it provides a useful insurance policy against retaliation, and it implies that Rome has been evil for many generations. Male homosexuals are derided in two poems: passives in Satire 2, actives and passives together in Satire 9. In the third Satire a friend of Juvenal explains why, abandoning the humiliating life of a dependent, he is determined to live in a quiet country town and leave crowded and uncomfortable Rome, which has been ruined by Greeks and other foreign immigrants; while in the fifth Juvenal mocks another such dependent by describing the calculated insults he must endure on the rare occasions when his patron invites him to dinner. The fourth relates how Domitian summoned his cringing Cabinet to consider an absurdly petty problem: how to cook a turbot too large for any ordinary pan.

Satire 6, over 600 lines long, is a ruthless denunciation of the folly, arrogance, cruelty, and sexual depravity of Roman women. The seventh Satire depicts the poverty and wretchedness of the Roman intellectuals who cannot find decent rewards for their labours. In the eighth, Juvenal attacks the cult of hereditary nobility. One of his grandest poems is the tenth, which examines the ambitions of mankind—wealth, power, glory, long life, and personal beauty—and shows that they all lead to disappointment or danger: what mankind should pray for is "a sound mind in a sound body, and a brave heart." In Satire 11, Juvenal invites an old friend to dine quietly but comfortably and discourses on the foolishly extravagant banquets of the rich. The 12th is a quiet little poem distinguishing between true and mercenary friendship. In the 13th Juvenal offers sarcastic consolation to a man who has been defrauded of some money by a friend, telling him that such misdeeds are commonplace; while in the 14th he denounces parents who teach their children avarice. Satire 15 tells of a riot in Egypt during which a man was torn to pieces and eaten: a proof that men are crueller than animals. In the 16th Juvenal announces that he will survey the privileges of professional soldiers, an important theme; but the poem breaks off at line 60 in the middle of a sentence: the rest was lost in ancient times.

Style

Technically, Juvenal's poetry is very fine. The structure of the individual Satires is—with a few exceptions—clear and forceful. They are full of skillfully expressive effects in which the sound and rhythm mimic and enhance the sense; and they abound in trenchant phrases and memorable epigrams, many known to people who have never heard of Juvenal: "bread and circuses"; "slow rises worth, by poverty oppressed"; "who will guard the guards themselves?"; "the itch for writing"; "the greatest reverence is due to a child." Vivid, often cruelly frank, remarks appear on almost every page: after describing a rich woman's efforts to preserve her complexion with ointments, tonics, donkey's milk, and poultices, Juvenal asks, "Is that a face, or an ulcer?" He describes striking and disgusting scenes with a clarity that makes them unforgettable: we see the statues of the emperor's discarded favourite melted down to make kitchenware and chamber pots; the husband closing his disgusted eyes while his drunken wife vomits on the marble floor; the emperor Claudius (poisoned by his consort) "going to heaven" with his head trembling and his lips drooling long trains of saliva; the impotent bridegroom whimpering while a paid substitute consoles his wife. Juvenal is not a poet to be relished by soft hearts or optimists, but he has power.

Influence. His work was forgotten for a time after his death. Later it began to be read and quoted, first by the Christian propagandist Tertullian—who lived and wrote about AD 200 and was as full of passionate indignation as Juvenal—then by other Christian authors and also by pagan students of literature. A commentary on the Satires (which survives) was compiled at some time between 350 and 420, and two editions of the text were produced, based on one master copy: apparently the only copy that had been preserved till then. Thenceforward Juvenal has never ceased to be studied and admired: he has been imitated by many satirists; for instance, by Giovanni Boccaccio, Nicolas Boileau, and Lord Byron.

BIBLIOGRAPHY

Bibliographies: G. HIGHET, *Juvenal the Satirist* (1954), pp. 339–346 lists and describes 81 books and articles dealing with many aspects of Juvenal's life and art. A more recent, specialized bibliography appears in sect. 6 of M. COFFEY'S new introduction to the reissue of J.D. DUFF'S edition of Juvenal's satires (1970).

Editions: The best text of the satires is that edited by U. KNOCHE with a critical apparatus in Latin (1950); W.V. CLAUSEN'S edition (1959), is also recommended. The best English-language edition with explanatory annotations is by J.D. DUFF (1898, reissued 1970). It is expurgated, omitting satires 2 and 9 and parts of 6, as well as occasional sentences elsewhere.

Translations: The first good English verse translation was published by JOHN DRYDEN and a group of friends, *The Satires of Decimus Junius Juvenalis* (1693), and is still highly readable. Less vivacious but equally vigorous is that by WILLIAM GIFFORD, *Satires, with the Satires of Persius* (1802, reissued with revisions by J. WARRINGTON, 1954). G.G. RAMSAY'S prose version, *Juvenal and Persius* (Loeb series, 1918; rev. ed., 1961), misses much of the poetic values, but is reliable as a guide to the meaning. The best modern English verse rendering is by ROLFE HUMPHRIES, *Satires* (1958), bright, slangy, rather slapdash. SAMUEL JOHNSON adapted satire 3 in his "London" (1738) and satire 10 in "The Vanity of Human Wishes" (1749). There are translations into French, German, Italian, and Spanish.

Biography and criticism: G. Highet's book (cited above) reconstructs the poet's career, analyzes each of the satires individually, estimates their value as poetry, and surveys his influence in later literature. Various aspects of Juvenal's work have recently been examined by W.S. ANDERSON, "Juvenal and Quintilian," *Yale Classical Studies,* 17:1–93 (1961); M. COFFEY, in his new introduction (cited above); H.A. MASON, "Is Juvenal a Classic?" pp. 93–176 in J.P. SULLIVAN (ed.), *Roman Literature: Satire* (1963); E.J. KENNEY, "Juvenal: Satirist or Rhetorician?" and D. WIESEN, "Juvenal's Moral Character," both in *Latomus,* vol. 22 (1963).

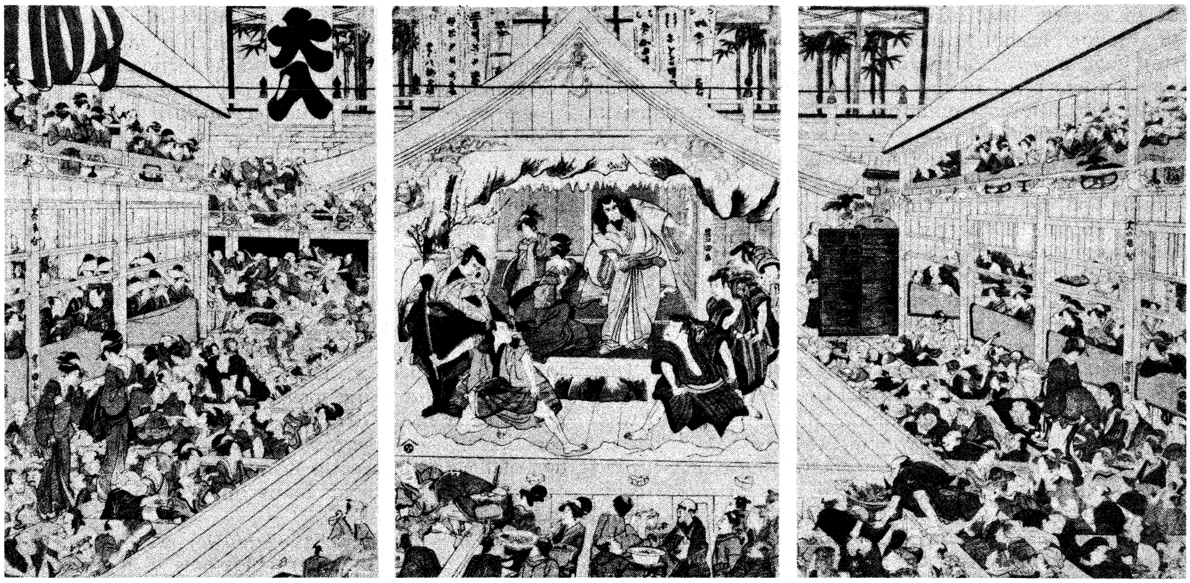
(G.Hi.)

Kabuki Theatre

For nearly 400 years Kabuki theatre has been a popular entertainment in Japan. It is a rich blending of realism and formalism, of music, dance, mime, and spectacular staging and costuming. In modern Japanese, the word is written with three characters; ***ka*** signifying "song"; *bu,* "dance"; and *ki,* "skill."

To many Westerners, Kabuki is the most perfect and complete theatrical expression. Its highly lyrical plays are regarded, with notable exceptions, less as literature than as vehicles for its actors to demonstrate their enormous range of skills in visual and vocal performance. These actors have carried the traditions of Kabuki from one generation to the next with only slight alterations. Many of them trace their ancestry and performing styles to the earliest Kabuki actors and add a "generation number" after their names to indicate their place in the long line of actors bearing the same name.

Traditionally, a constant interplay between the actors and the spectators took place in the Kabuki theatre. The actors frequently interrupted the play to address the crowd, and the latter responded with appropriate praise or clapped their hands according to a prescribed formula. They also could call out the names of their favourite star



Interior of a Kabuki theatre, *c.* 1800; coloured woodcut triptych by Utagawa Toyokuni. In the British Museum.
By courtesy of the trustees of the British Museum

actors during the performance. Since Kabuki programs ran from morning to evening and many spectators often attended for only a single play or scene, there was a rather constant coming and going in the theatre. At mealtimes food was served to the viewers. The programs incorporated themes and customs that reflected the four seasons or inserted material derived from contemporary events. Unlike most Western theatres, in which since the late 17th century a proscenium arch has separated actors and audience, the Kabuki performers constantly intruded on the audience. When two *hanamichi*, elevated passageways from the main stage to the back of the auditorium, were used, the audience was fenced in by three stages.

The Kabuki form dates from the late 16th century, when a former priestess, Okuni, achieved popularity with parodies of Buddhist prayers. She gathered a company of women, introduced bolder dancing, and began a tradition whose long history was both notorious for the sensuality of its performers and illustrious for its art. Kabuki owes much to the tastes of the Tokugawa period (1603–1867), during which the ukiyo-e school of artists used as subjects for their wood-block prints the everyday scenes and the brilliant costumes of the common people. Always able to assimilate new forms and ideas with ease, Kabuki came rapidly to reflect this interest in the commonplace as well as the sensational. As a result its thematic material is often pedestrian, but the Kabuki has invariably infected it with great aesthetic appeal.

Relations with other Japanese theatre

Kabuki is regarded as an eclectic theatre incorporating earlier traditions of Japanese drama. Bugaku, the dance ceremony of the imperial court, and the Nō theatre (*q.v.*), both of great antiquity, were long the exclusive domain of the nobility and the warrior-aristocrats known as samurai; Kabuki became the theatre of the townspeople and the farmers. Bugaku and Nō have a fragile elegance and an extreme subtlety of movement. In contrast, Kabuki is somewhat coarse and unrestrained, and its beauty is gaudy and extravagant. It appeals to the senses rather than to the intellect.

The strongest ties of Kabuki are to the Nō, its predecessor, and to *jōruri*, the puppet theatre that developed during the 17th century. Kabuki derived many of its materials from the Nb and constantly parodied this theatre. When Kabuki was banned in 1652, it re-established itself by adapting and parodying the Kybgen, the plays that provided comic interludes during the Nb performances.

Kabuki programs

Kabuki plays and playwrights. Within each of these categories, the subject matter creates such distinctions as that between the historical play (*jidaimono*) and the domestic play (sewamono). A Kabuki program generally presents them in this order, separated by one or two dance plays featuring ghosts, courtesans, and other exotic creatures. It ends with a lively dance finale (*ōgiri shosagoto*) with a large cast.

The historical plays often depicted events that were more or less contemporary but were moved back in time to satisfy censors. One of the Kabuki masterpieces, the cycle *Chūshingura* ("Treasury of Royal Retainers"; 1748), a collaboration of three leading playwrights, was based on an actual event in 1702 but was shifted to an earlier period. In this process, the *jidaimono* became an interesting mixture of fact and fiction.

The early domestic plays developed out of this custom of dramatizing recent, sensational events. They were presented as an added attraction to the jidaimono and were somewhat related to it in theme. The sewarnono became a distinct genre around the middle of the 19th century in the plays of Kawatake Mokuami, which brought about a renewed focus on the lives of everyday people and a revival of realism on the Kabuki stage. Among the sensational topics of the sewamono, which flourished in the 17th century, the love-suicide (*shinju*) theme figures prominently because of its appeal to the townspeople. The dialogue and events of these plays are more realistic than those of the historical genre. Playwrights tried at times to link the historical and domestic plays in a program by maintaining that the principal characters of the former appeared again under new names in the latter.

Although the basic purposes of Kabuki are to entertain and to allow the actors to demonstrate their skills, there is a didactic element, an ideal represented by the notion of kanzen-chdaku ("reward the virtuous and punish the wicked"). Thus the plays often present conflicts involving such religious ideas as the transitory nature of the world (from Buddhism), and the importance of duty (from Confucianism), as well as more general moral sentiments. Tragedy occurs when morality conflicts with human passions. At the same time, erotic, grotesque, and sadistic elements are often introduced to please public taste, but such scenes rarely fail to achieve a high aesthetic stature.

Structurally, the plays often comprise two or more themes in a very complex plot, but they lack the strong unifying element for which Western drama strives. Kabuki plays include a variety of intermingled episodes, which develop toward a final dramatic climax.

Written texts

The early written texts of Kabuki themselves were more scenarios than fully developed scripts. Like the scenarios of the commedia dell'arte in the West, they were essentially notes on plot that gave the performer both license and obligation to improvise upon them. Over time many of these texts were expanded and revised, as were the productions based on them, especially as single authorship gave way to collaboration. In this system, subordinate writers filled in the outline provided by the chief playwright after consultation with the theatre manager and principal actors.

Despite the collaborative nature of much Kabuki drama, several individual playwrights stand out for their writing and innovation. Chikamatsu Monzaemon (1653–1724), the earliest writer of note, has been called the Shakespeare of Japan. For the puppet theatre and the Kabuki, he wrote over 100 full-length plays, of which about 50 authenticated ones survive. Some of his famed works featured the double suicide of lovers, drawn from actual cases. Namiki Gohei (1747–1808) first clearly separated the *jidaimono* and sewamono genres. The dance dramas of Sakurada Jisuke I (1734–1806) are still popular. The complete definition of the sewamono was the work of Kawatake Mokuami (1816–93).

The most important work on playwriting is Kezairoku ("Valuable Hints on Playwriting"), written in 1801 by Namiki Shbzb II (?–1807). There are no Tokugawa writings dealing with the theoretical aspects of Kabuki. Popular introductions, dictionaries, and other miscellaneous works on Kabuki were written by novelists.

Theatre and stage design. Kabuki is commonly known as the *shibai*, a term that may mean "play" or "theatre." In the early, outdoor Kabuki, shibai referred to the grass-covered viewing area around the stage or to the broad, open space directly in front of the stage. Later, it came to be identified with the theatre structure itself. Originally, most spectators sat on carpets or mats spread on the ground; when it rained, the performance was cancelled. The more affluent occupied an elevated, roofed structure (*sajiki*), a social distinction by elevation that was retained when Kabuki went indoors.

Stage components

The stage was borrowed entirely from the Nb, but gradually acquired its own features. The entire theatre was placed under a single roof during the Genroku period (1688–1703). About 1716 the hanarnichi ("flower path"), originally a temporary path for sending flowers and gifts to the actors, became a permanent fixture for the entrance and exit of the performers. This passageway, which permitted the actors to enter the main stage from the rear of the auditorium, was first located at the centre but was later shifted to stage right. (The terms right and left as used here and in the remainder of the article refer to the actor's right and left when facing the audience.) When necessary to the production, a narrower ramp was placed on stage left. With the main stage, the three virtually encircle the audience. The *hanamichi* can serve a multitude of functions. It can represent a road, a river, or the banks of a river. On occasion, the actors may start down the left *hanamichi*, follow the passageway in the back of the auditorium, walk onto the right *hanamichi*, and finally appear on the main stage. During this time, the main stage can be prepared for the next scene.

Another special feature is the revolving stage (*mawari*butai). In 1758, the Osaka playwright Namiki Shōzō I used this device to show two simultaneous events. In 1847 two concentric stages (*janome-mawashi*) revolving around a common centre were introduced. This invention could simulate such events as two boats passing each other. Other stage devices include the rising platform (seridashi) and the trap door (*suppon*). These mechanisms were installed when the area under the stage was enlarged, and the actors could appear from below. The actors could emerge from the *suppon*, located on the hanantichi, when they played spirits, ghosts, and magicians or when they appeared in a dream sequence. In Sanmort Gosan *no* Kiri ("The Story of Ishikawa Goemon"), Goemon, the notorious thief, stands in a large temple gate that emerges from below the stage, and in *Sendaihagi* ("The Dispute in the Household of Lord Sendai"), a mansion rises through the stage floor to a point above the main stage, and the action occurs under the floor of this huge structure.

The curtains behind the apron are the draw curtain (*joshikimaku* or hikimaku) and the scene curtain (*dōgumaku.*) The brilliantly coloured draw curtain, used mainly for changing scenes, is often opened to musical accompaniment. It can serve a ceremonial function as well, as in the grand opening scene of *Chūshingura*.

Properties and lighting

Large stage properties (*ōdōgu*) include a special platform (yatai or *nijubutai*), which is 16.7 inches (42.4 centimetres) high if used for a domestic play and 25 inches (64 centimetres) or 33.4 inches (84.8 centimetres) high for an historical play. For plays adapted from the puppet theatre, a chanter's platform (choboyuka) for the musicians and chanters is placed at an angle at the left corner of the stage. In the dance plays, which feature a style of Kabuki music known as *nagauta*, singers sit directly in front of the audience on a two-tiered, carpeted platform (*hinadan*) and such dance music as kiyomoto and *tokiwazu* is performed on another platform to the right.

The role of the large property man (dddgugata) appeared early in the development of Kabuki and became hereditary. The dddgugata even supervised the construction of theatre. Hasegawa Kambei XI (1781–1841) was especially known for his many important inventions for the Kabuki stage.

The small stage properties (koddgu) are those held by the actors. A row of cherry blossoms strung from the ceiling is an *ōdōgu*, whereas the cherry branch, which can be broken, is a koddgu. Among the many other small properties is the long, stiff wire (*sashigane*), that has a butterfly or bird on the end and is manipulated by the stage assistant (kdken).

The modern Kabuki stage is illuminated by electric lights, but no dimming of lights or change in colour is allowed. Following the ancient traditions of performance, the light or darkness is indicated by the behaviour of the actors and by the symbolic opening and closing of windows. A face light (tsura-akari), a candleholder attached to a long stick, may be wielded by the *kōken* to achieve certain traditional moods. When, for example, the leading actor appears on the harzamichi or performs a *mie*, striking an attitude, the *kōken* brings the light near his face to suggest the effect of a closeup. Before the electric lamp, candles were used for lighting. In the Kyōto–Osaka area there was decorative lighting from the ceiling of the main stage, a custom still used occasionally.

Costumes, wigs, and makeup

Kabuki costumes, whether the realistic fashions of the sewamono or the stylized garments of the *jidaimono*, are based on historically authentic Japanese dress. Many standard costumes exist for various character types as well as historical periods. Improvements in costumes and their manipulation followed advances in weaving and dyeing, and the Kabuki stage and popular fashions have long had a strong influence upon one another.

An unusual feature of Kabuki is the practice of making costume changes in full view of the audience, often as a symbolic or metaphorical expression of the play's action or the characters' experiences. The two methods for such change are, first, the bukkaeri, in which the threads holding the outer garment are pulled out at the shoulders, the upper garment then falling to reveal a new and striking inner one, and, second, the hikinuki, in which both the upper and lower portions of the costume are removed with the aid of the *kōken*, who must not disturb the tempo or rhythm of the dance in progress. The former technique is often used to show sudden changes in character, as when a scoundrel reveals his true colours, or to signify responses to events, such as the angry and humiliated reaction of a Buddhist monk to his seduction by a beautiful woman. The latter technique is used in Musume *Dōjōji* ("Dōjōji Dance"), in which the basic pattern of the successive costumes remains the same while the background colour changes.

Costumes, wigs, and makeup form an integral unity for the actor, reflecting the rather clearly defined "types" that exist in Kabuki for both plays and characters. Hair styles are distinguished by sex, age, class, and occupation, and like the costumes they are realistic or stylized. The fantastic characters of the "rough plays" (aragoro) popular in the Edō (modem Tokyo) area wore highly stylized "wheel wigs" (kurumabin) of the heroes or "great hundred-day wigs" (daibyaku) of the brigands. The female impersonator (onnagata) wears wigs that are functional as well as decorative. Of these, the gattari can collapse to one side, the sabaki has forelocks that remain in place while the rest falls into disarray, and the *sō-sabaki* becomes totally dishevelled as during a murder or suicide. Ghosts and other fantastic and supernatural characters require special wigs. Persons caring for the wigs are distinguished from both wigmakers and hair stylists, and the male and female wigs have separate hair setters.

The stylized makeup for the jidaimono comprises white and red types, indicative, respectively, of the good and bad characters. The romantic rascals (iroaku) wear white. A special type, kumadori, consists of bold designs in either red or blue. The warm red represents anger, strength, or justice; the cool blue, wickedness, death, ghosts, or fiends. Comic qualities may be indicated by the red kurnadori, and animal figures may have their special kumadori.

Conventions of staging

Staging of Kabuki. Despite the ease with which it can assimilate new forms, Kabuki is a very formal theatre. It retains numerous conventions adapted from earlier forms of theatre that were performed in shrines and temples. The opening call of a Kabuki production, "*Tōzai-tōzai*" ("East–west, east–west"), was originally addressed to the audience seated in the left and right sides of the viewing area, and the tradition remains, although the seating arrangement has altered.

In Japan the left is regarded as sacred, the right as profane, a factor that is the basis of many conventions. The principal curtain is opened from right to left. Although stage right (shimote, the westerly direction) is the entrance, stage left (*kamite*, the easterly direction) is the position of honour. Characters of high rank, guests, and messengers cross to the *kamite* and climb the double-tiered platform (*nijū-butai*). The master of the house must retire to the shimote from the katnite before paying respects to his guests. Men begin moving on their left foot, women on their right. In climbing stairs, men are on the left, women the right. These traditions can be used to dramatic purpose: in *Chūshingura* a young lord, in a state of great anxiety, starts to climb a platform with his right foot before realizing and correcting his error.

The idea of the picturesque (*emen*) plays an important part in Kabuki. In each scene there must be a harmonious balance in the actors' positions and the colours of costumes and stage. Each major scene in the jidaimono culminates in a mie, the striking of an attitude by one or more actors. The *sewamono*, though it tries to preserve such a balance at the ending of a scene, is generally less formal, and the curtain closes on a more realistic picture.

In Western theatre the drama is ordinarily portrayed "as if" it were really happening, but Kabuki does not engage in this pretense. Its actors always address the audience directly, an indication of its storytelling quality. Other nonactors perform functions that would destroy the belief of a Western audience. The *kōken* may assist in costume changes and bring on or remove stage properties. Other stage assistants are the kurogo, men dressed completely in black to symbolize their invisibility, and those in traditional ceremonial attire who frequently appear in the ritualistic plays known as *shosagoto*.

Speech, acting, and dance

The dialogue of Kabuki may be extremely stylized in its rhythm and tempo and it may be almost colloquial. A blend of the two styles in a given play creates a distinctive dramatic effect. Both the jidiamorzo and the *sewamono* are mixtures of formalistic and realistic elements, but the speech patterns of the former tend toward stylization, of the latter toward the everyday. In certain kinds of play the actor can introduce himself in his role by delivering a long speech to demonstrate his skill in elocution. In the *warizerifu*, two actors may deliver lines that are seemingly unrelated to each other but that produce a cumulative dramatic effect when the audience combines them as a single story. In the watarizerifu, a number of actors may speak their lines in succession, with a finale in unison. The various male and female roles also have distinctive manners of delivery.

Geographical differences in Kabuki include the preference in the Kyōto–Osaka area for the kind of play called wagoto. In this, romantic and sentimental elements are combined with the comedic for a more realistic production. In Edō, the most stylized forms of play and of acting traditionally have been preferred. There the *mie,* the dramatic moment when the stage becomes frozen into a picture, and the *roppō*, a spectacularly exaggerated exit, were developed as part of the *aragoto* style. The aragoto and wagoto represent opposing extremes in Kabuki acting.

Kabuki dance is probably the best known feature of this theatre. Rarely is an opportunity missed to insert dancing, whether the restrained, flowing movement of the women or the exaggerated posturings of the men. The acting in Kabuki can be so stylized that it becames virtually indistinguishable from dancing. Most dances are gay and colourful, and usually they are identified with an object—flowers, a fan, a towel, a musical instrument—carried by the dancer. They may be, for example, dances of priestesses, townspeople, or lovers, or they may be slow and delicate or wildly acrobatic. The best known of Kabuki dances in the West is probably the *Musume Dōjōji* dance.

Production of Kabuki. Kabuki was performed under government supervision during the Tokugawa period. Kyōto, Osaka, and Ed6 established their own theatres, which have since become hereditary. Besides the major theatres, which were given special status, large numbers of lesser theatres were scattered throughout Japan. Temporary licenses were issued to even smaller theatres, which often performed in temples, shrines, and market places. The major theatres were under the direct charge of the town magistrate, whereas the smaller ones were administered by the magistrate responsible for religious affairs. The actors and playwrights were hired on an annual basis, and no theatre enjoyed a monopoly on talent. Thus the three largest cities each had their own performers who circulated among the major theatres. The Kyōto–Osaka area and Ed6 developed their distinctive traditions in acting, playwriting, and theatre management. The actors performing in lesser theatres were forbidden to work on a major stage, and the same rule kept outsiders from joining a major theatre.

Although in the beginning the theatres were centrally located, they were moved out into the suburbs by the government officials. The first theatres in the three largest cities were along the river banks: the dry bed of the Kamo River in Kyōto, where Okuni is alleged to have performed in 1586, is regarded as the birthplace of Kabuki. Traditionally, the government regarded Kabuki as being akin to the city's pleasure quarter (the *akubasho*, the "wicked place"). In Ed6 the original downtown theatre district was moved in 1842 to a notorious location, which continues to be the centre of theatrical activity.

In the Tokugawa period, Kabuki performances were required by law to begin at six o'clock in the morning and to end at five in the afternoon. Ticket prices were determined by the location of the viewing area; those in the sajiki were equivalent to the annual wage of a low-salaried worker. The opening day, however, was free. In this first performance of the new year, a courtesan play (*keis*eimono) was given in the Kyōto–Osaka area. Scenes of the pleasure quarter with its beautiful courtesans were a prominent feature of the jidaimono at this season. In Edō, however, a well-known local vendetta was the main theme for the new-year performance, and this custom has continued for over two centuries.

Position of actors

During the Tokugawa period, the Kabuki actors were not regarded as ordinary citizens. They were subject to moral and social stigmas sometimes assigned to performers in the West, and their living quarters were restricted by the government. Nonetheless, the top stars earned high wages and lived extravagantly. The Kabuki actors exerted enormous influence on popular fashions, painting, and writings.

Kabuki actors are extremely conscious of their status within the acting hierarchy. They often begin training as early as five, and at advanced ages Kabuki actors may portray young men or women. When actors assume the names of predecessors who belonged to distinguished acting families, they go through the ***shiimei,*** an elaborate on-stage ceremony to proclaim their new position. Such names are given to a direct descendant or to a pupil. The successor is expected to perpetuate the style of the acting family to which he belongs. The innovations of famous actors become part of an established tradition in the next generation. The function of the director has been assumed by the chief actor ***(zagashira),*** who has absolute authority.

The ***Yakusha Hydbanki*** ("Actors' Annual") was published yearly for 230 years (1656–1886). Its treatment of Genroku period is outstanding, and it remains a unique chronicle for understanding the history of this theatre. The ***Yakiishabanaslzi*** ("The Stories about Actors"; 1776) is regarded as a kind of bible by Kabuki actors; it contains accounts by actors and playwrights from the Kyōto–Osaka area who were active during the Genroku period. The autobiographies of Nakamura Nakazb I (1736–90) and Nakamura Nakazb III (1809–86) describe the life of Kabuki actors in the Tokugawa period. Since the Meiji period, many actors have written accounts of their lives.

Kabuki in the 20th century. In the Taishō period (1912–26), plays were written by non-Kabuki playwrights, including novelists and journalists, and were called ***Shin-Kabuki*** ("new Kabuki play"). The representative dramatist of the period was Okamoto Kidō (1872–1939). Today both old and new Kabuki plays are given on the same program.

At present, regular performances are held at the Kabuki Theatre (Kabuki-za)—with a capacity of 1,600—and the National Theatre, both in Tokyo. Other theatres have occasional performances. Troupes of Kabuki actors also perform outside of Tokyo. There are several such companies, but their memberships often overlap. A typical program at the Kabuki Theatre consists of a ***jidaimono,*** a ***shosagoto,*** and a ***sewamono,*** in that order; two performances are held daily. At the National Theatre, a complete *jidaimono* and a ***sewamono*** or a ***Shin-Kabuki*** is given daily. At the Kabuki Theatre, the length of an average program is about five hours; at the National Theatre it is about four. The Kabuki Theatre upholds the actors' traditional position by adopting a conservative approach, planning its program around the major stars. The National Theatre stresses the importance of the play itself, trying to maintain the historical tradition and to preserve Kabuki as a classical form. The prestige of the leading acting families is as strong as ever, and lineage is regarded as more important than talent.

Kabuki
and motion
pictures

The influence of Kabuki is evident in the theories of cinematic montage propounded by the Russian film director Sergey Eisenstein. Kabuki also influenced such later film directors as the Italians Federico Fellini and Pier Paolo Pasolini. In Japan, themes from Kabuki have been adapted for motion pictures and Western-style plays.

It is difficult to predict the future of Kabuki theatre. Many critics complain that its traditions, unlike those of the Nō theatre, have been watered down since the Meiji period through excessive assimilation of new materials. Kabuki travelled abroad before World War II but much more afterward, and the form has become internationally recognized. This contact with foreign cultures and theatrical traditions could erode its distinctive qualities. To remain a significant part of contemporary world theatre, Kabuki must preserve its traditions and at the same time keep itself open to growth and adaptation as it has in the centuries of the past.

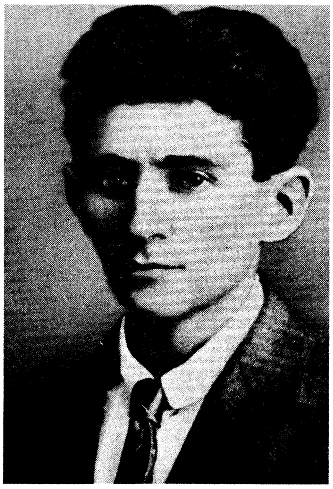
BIBLIOGRAPHY. FAUBION BOWERS, ***Japanese Theatre*** (1952), a popular, authoritative history that emphasizes Kabuki, includes three translations of Kabuki plays; EARLE ERNST, ***The Kabuki Theatre*** (1956), an outline of the Kabuki form considered as an expression of Japanese life and culture, includes a short glossary of Japanese theatre terms and a selected bibliography of works in Western languages; ADOLPHE C. SCOTT, ***The Kabuki Theatre of Japan*** (1955), in which the author brings to his study of Kabuki a relevant knowledge of the classical drama of China, includes a glossary of Japanese terms used and a list of Japanese eras with which the history of Kabuki theatre is involved.

(M.Gu.)

Kafka, Franz

Franz Kafka was a creator of visionary fiction and the most influential master of modern German prose whose work has become a symbol of 20th-century anxiety and of the sense of alienation pervasive in Western society. Judged by external standards, his life was sheltered, uneventful, and secure. The conflicts and upheavals he suffered, which he expressed in terms that revealed the despair of an age, were almost entirely internal events.

Archiv fur Kunst und Geschichte



Kafka.

Early influences. Franz Kafka, the son of Julie (*née* Lowy) and Hermann Kafka, a merchant, was born into a Jewish middle class family on July 3, 1883, in Prague. After two brothers nearest to him in age died in infancy, he became the oldest child, remaining forever conscious of his role as older brother; Ottla, the youngest of his three sisters, became the family member closest to him. Kafka identified with his maternal ancestors not because of their social superiority but in view of their spirituality, intellectual distinction, piety, rabbinical learning, eccentricity, melancholy disposition, and their delicate physical and mental constitution. He was not, however, particularly close to his mother, ***a*** simple woman devoted to her children. Subservient to her overwhelming, ill-tempered husband and his exacting business, she shared with her spouse a lack of comprehension of their son's unprofitable and possibly unhealthy dedication to the literary "recording of [his] . . . dreamlike inner life."

Kafka's
father
figure

The figure of Kafka's father overshadowed his work as well as his existence; the figure is, in fact, one of his most impressive creations. For, in his imagination, this robust, tenacious, coarse, and domineering shopkeeper and patriarch, who liked to complain and to brag about the hardships of his childhood and who worshipped nothing but material success and social advancement, belonged to a race of giants and was an awesome, admirable, but repulsive tyrant. In Kafka's most important attempt at autobiography, the "Letter to His Father" (written 1919), a letter that never reached the addressee, Kafka attributed his failure to live—to cut loose from parental ties and establish himself in marriage and fatherhood—as well as his escape into literature, to the prohibitive father figure, which instilled in him the sense of his own impotence. He felt his will had been broken by his father. Yet there is something solid and consoling in this dramatization of a classical Oedipal conflict, an experience typical of Kafka's generation and, particularly, of his literary contemporaries among the Expressionists. The conflict with the father is reflected directly in Kafka's story ***The Judgment.*** It is projected on a grander scale in Kafka's novels, which portray, in lucid, deceptively simple prose, a protagonist, K., in desperate struggle with a mysterious and overwhelming power that may persecute its victim, as in ***The Trial,*** or be sought after and begged in vain for approval, as in ***The Castle.*** Yet the roots of Kafka's anxiety and despair go deeper than his relationship to his father and family, with whom he chose to live in close and cramped proximity for the major part of his adult life. The source of Kafka's despair lies in a sense of ultimate isolation from a true community and communion with all human beings—the friends he cherished, the women he loved, the job he detested, the society he lived in—and with God or, as he put it, with true indestructible Being.

This is not to say that Kafka failed to live up to the expectations of others. He was considered a good, obedient, sensible child. He did well in German elementary school and in the Altstadter Staatsgymnasium, an exacting high school designed for the academic elite. He was respected and liked by his teachers. Inwardly, however, he rebelled against the authoritarian institution and the dehumanized humanistic curriculum, with its emphasis on rote learning and classical languages. Nonetheless, he owed to it some of his intellectual discipline, meticulous sense of verbal structure, and abiding love of German literature. Moreover, his very protest against conventional authority allied him, in turn, with the insular intellectual and literary vanguard of Prague.

The socialist and atheist

Intellectual life in Prague. Kafka's opposition to established society became apparent when, as an adolescent, he declared himself a Socialist as well as an atheist. He continued in his later life to reject capitalism as "a state of the world and of the mind" (as in his image of Amerika). But it should also be pointed out that he was inclined to reject the entire sensory world as a sinful aberration from true Being. Throughout his adult life he expressed qualified sympathies for the Socialists; attended meetings of the Czech Anarchists (before World War I); and, in his later years, showed marked interest and sympathy for a socialized Zionism. Yet even then he remained a politically unengaged and inactive spectator.

In this respect he was rather unlike many of the German Jewish intellectuals and literati of Prague (*e.g.*, Max Brod, E.E. Kisch, Franz Werfel) with whom he soon became associated. By external standards they were as isolated as he, not only from the Czech population but also, as Jews, from the privileged German minority. Indeed, by virtue of their modern sophistication, they were isolated from their own Jewish heritage. Even if his fellow students felt a glass wall separating Kafka from them, he nonetheless sought their friendship actively; later on, he was sought after and wooed by a small group of devoted friends, including Max Brod, the most intimate and solicitous of his friends, who was to survive as promoter, saviour, and interpreter of Kafka's writings and as his most influential biographer.

They became acquainted in 1902, when Kafka, resigned to the necessity of preparing for an indifferent nonliterary profession, was studying law at the University of Prague. He received his doctorate in 1906, and, after a respite of one year—spent in a perfunctory legal apprenticeship—he took up regular employment with an insurance company. The long hours and exacting requirements of the Assicurazioni Generali, however, did not permit Kafka to devote himself to writing. In 1908 he found in Prague a "single shift" job (8 AM to 2 PM, including Saturday) in the semi-nationalized Workers' Accident Insurance Institute for the Kingdom of Bohemia. There he remained until 1917, when tuberculosis forced him to take intermittent sick leaves and, finally, to retire (with a pension) in 1922, about two years before he died. In his job he was considered tireless and ambitious; he soon became the right hand of his boss, whose efficiency he admired, and was esteemed and liked by colleagues, subordinates, and superiors. Kafka himself, however, found his futile office job and the exhausting double life into which it forced him (for his nights were frequently consumed in writing) to be excruciating torture. The question whether "to earn one's living or to live one's life" proved to be insoluble.

This is not to suggest that Kafka's existence should be stylized into a mere symbol of despair. The tall, well-dressed, retiring young man who, as a student, was received into socially distinguished circles, took interest in a variety of subjects, including Alfred Weber's analysis of capitalist society and Franz Brentano's philosophy of ethics. Though Kafka's social life became increasingly restricted, he continued to frequent intellectual salons, where he met leading philosophers, mathematicians, physicists, and other men of learning; thus he became acquainted, in some measure, with science, the classics of German Idealistic philosophy, psychoanalysis, and theosophy. Unlike other German poets of Prague, such as Rainer Maria Rilke or Franz Werfel, he actively sought contact with the Czech people and frequently visited the assemblies of their political parties. He travelled to Switzerland, Paris, Berlin, Hungary, and, in 1912, to Weimar, home of Goethe, and to a centre for methods of natural cure. Consistent with his concern for a life-style in keeping with nature were his vegetarianism, his love of the countryside, his enjoyment of long hikes, and his skill and endurance in swimming and rowing.

Personal charm and complex personality

Kafka's friends have attested to the charm and fascination of his conversation. His sensitive, undoctrinaire intelligence was subservient to a precise sensibility and a uniqueiy original imagination. It was natural for him to express himself in startling metaphors or in statements that combined a complex thought refined to the point of paradox with a quietly simple and entirely unassuming manner. His contemporaries speak of his scrupulous conscientiousness, veracity, self-control, and quiet courage; of a grace even in shyness, a sense of tact and readiness to help; of his childlike integrity and ethical rigour. His subtle irony and sense of humour were rooted in the awareness of the incongruity of human affairs vis-à-vis the—unquestionable and unattainable—absolute. Brod, in particular, has been at pains to stress the positive aspects of Kafka's personality: "His spiritual bent was not in the direction of the morbidly interesting, the bizarre or grotesque." Characteristically, he took no interest in the literary Decadents or the authors of horror stories but instead sought out "the curative, health-giving, sound, firmly established, simple things."

Yet such statements apply only to one aspect of Kafka's complex and ambivalent personality. With good reason, he referred to himself as to a battlefield. Some of his contradictory tendencies and conflicting inclinations became explicit in his erotic and sexual relationships. It is odd that a man who could speak of sexual intercourse as punishment for the happiness of being together with his beloved should have relished on occasion (and quite in keeping with the customs of the time) the easy girls or the professional prostitutes of Prague. Apparently, his sexual life was sporadic, occasionally intense, and frequently interrupted by periods of indifference. In his writings sex is most frequently associated with dirt and guilt and is treated—with ascetic disgust—as an abomination. Yet Kafka also experienced the sweetness of erotic intimacy in two fleeting vacation affairs and with the women he loved: his two fiancées, Felice Bauer and Julie Wohryzek; his married mistress, Milena Jesenská Pollak (1920), a Czech woman of extraordinary vitality, élan, and intelligence; and Dora Dymant (1923–24), a young girl who lived with him in Berlin during the last year of his life. Kafka's most desperate conflict was not in relation to sex, however, but rather in relation to marriage, which he wanted yet feared in the years of his tormented courtship of Felice Bauer (1912–17) and again in his brief engagement to Julie Wohryzek (1919–20). To marry and found a family, he felt, was the greatest achievement possible to man, though he also felt himself incapable of assuming the powers and privileges his father enjoyed, and he found family life incompatible with solitary nights spent in writing, the one and only activity he was meant for. Moreover, he also felt that marriage represented the "world," which was essentially evil, though, to be sure, man should meet the world to give it its due and to suffer his way through it.

Outbreak of tuberculosis. Kafka himself thought that the outbreak of his tuberculosis in 1917 was a psychosomatic event initiated by a conspiracy between head and body for the purpose of putting an end to his insoluble and unbearable internal struggles. From this time on, he spent half of his remaining life in sanatoriums and health resorts. His disease released him from his office job and his marriage dilemma. It did not inhibit his creative development as a writer, the intensity of his vital experience, or the range of his interests. Indeed, it could be argued that his illness enabled him to pursue his most essential concerns and to promote his self-realization through a study and rediscovery of Judaism and a confrontation not merely with religious authors but ultimately with questions of faith.

Literary influences

Kafka had always been a passionate reader and was fond of reading his favorite prose authors aloud to an audience. In his early phase these included Friedrich Nietzsche, Thomas Mann, Hugo von Hofmannsthal—the latter a distinct influence on Kafka's early works. Apart from his perennial concern with Goethe, Kafka's diaries suggest his sense of affinity to a variety of literary figures; *e.g.*, to Heinrich von Kleist, Franz Grillparzer, Flaubert, Dostoevsky, Strindberg, and Knut Hamsun; his appreciation of the Bible; of Friedrich Hebbel, J.P. Hebel, Adalbert Stifter; of Dickens and Tolstoy; of some Chinese poetry; as well as of contemporaries, among whom he admired Werfel excessively. Characteristic of Kafka as a reader is his marked preference for autobiographical documents, diaries, correspondence, and memoirs. His interest in the Jewish heritage was first activated through close contact with a group of Yiddish players in 1911–12. Evidence of his increased concern with religion and with biblical, linguistic, literary, historical, and contemporary aspects of Judaism may be found in his study of the Danish philosopher Søren Kierkegaard in 1917–18; in the beginning of his intensive study of Hebrew in 1918 and, presumably, of the Jewish mystical practices and beliefs known as Hasidism; and in his reading of Yiddish authors and his sympathetic attitude toward the Zionism of Brod and Dora Dymant.

In the final year of his life, Kafka still sought to escape from his paternal family and the native city that had shaped his life and work. He settled down in Berlin, where his stay was cut short by a decisive deterioration of his health during the winter of 1924. After a brief final stay in Prague, where Dora Dymant joined him, he was moved to a clinic in Vienna, then to a sanatorium nearby at Kierling, where he was cared for by Dora Dymant and his friend and physician Robert Klopstock. There he died, after painful suffering, on June 3, 1924. His body was buried in the family tomb in the ancient Jewish cemetery of Prague.

Publications and posthumous fame. Sought out by leading publishers of the literary vanguard, Kafka reluctantly published a mere fraction of his writings during his lifetime. These publications include two sections (1909) from ***Description of a Struggle*** (1936), written in Kafka's early, somewhat florid style; ***Meditation*** (1913), a collection of short prose pieces; and works representative of Kafka's maturity as an artist, notably "The Stoker" (1913; the first chapter of ***Amerika***); ***The Judgment*** (1913), a long story written in 1912, which, in Kafka's opinion, represented a breakthrough and his coming into his own; two further long stories entitled ***Metamorphosis*** (1915) and ***In the Penal Colony*** (1919); and a collection of short prose, ***The Country Doctor*** (1919). ***A Hunger Artist*** (1924), four stories exhibiting the concision and lucidity characteristic of Kafka's late style, had been prepared by the author; the book did not appear until after his death.

Characteristically, Kafka wrote under the impact of creative spells that were both inspired and frustrated by his love affairs. His breakthrough in 1912, the year he produced ***The Judgment***, ***The Metamorphosis***, and the major portion of his first extant novel, ***Amerika***, took place when he first met Felice Bauer; and the failure of this relationship is reflected in ***The Trial*** (written in 1914). Similarly, the relationship to Milena appears to have influenced the composition of ***The Castle*** (probably written mainly in 1922), while the last stories are to be associated with Kafka's life with Dora Dymant and with his attempt to come to terms with his sickness (as in "The Burrow," 1923).

Publication of ***The Trial*** and ***The Castle***

The works published by Kafka himself include less than half of his shorter narrative prose, and, apart from the famous parable "Before the Law," scarcely any of the essential parables and aphorisms that are among his most perfect creations. These publications gave no inkling of his major, if fragmentary, novels, ***The Trial*** and ***The Castle***, nor of the fragments, notes, diaries, and extensive correspondence (including the letter to his father, the letters to Felice Bauer, to Milena, and the still unpublished letters to his sister Ottla).

At the time of his death, Kafka was appreciated only by a small literary coterie. His name and work would not have survived if Max Brod had honoured Kafka's testament—two notes requiring his friend to destroy all unpublished manuscripts and to refrain from republishing the works that had already appeared in print. Brod took the opposite course, and thus the name and work of Kafka gained worldwide posthumous fame. This development took place first during the Hitler regime, in France and the English-speaking countries—at the very time when Kafka's three sisters were deported and killed in concentration camps. After 1945 Kafka was rediscovered in Germany and Austria and began to greatly influence German literature. In the 1960s, this influence extended even to the intellectual, literary, and political life of Communist Czechoslovakia.

MAJOR WORKS

NOVELS (all published posthumously): ***Der Prozess*** (1925; ***The Trial***, 1937); ***Das Schloss*** (1926; ***The Castle***, 1930; definitive ed., 1969); ***Amerika*** (1927; Eng. trans., 1938)—all translated by Willa and Edwin Muir.

STORIES: *Betrachtung* (1913; ***Meditation*** in ***In the Penal Settlement: Tales and Short Prose Works***, 1949); ***Das Urteil*** (1916; ***The Sentence***, 1928; also trans. as ***The Judgment***, 1945); ***Die Verwandlung*** (1915; ***Metamorphosis*** in ***Metamorphosis and Other Stories***, 1961); ***In der Strafkolonie*** (1919; ***In the Penal Colony*** in ***Metamorphosis and Other Stories***, 1961); *Ein Landarzt* (1919; ***The Country Doctor***, 1940); *Beim **Bau der Chinesischen Mauer*** (published posthumously, 1931; ***The Great Wall of China*** in ***Description of a Struggle and The Great Wall of China***, 1960).

BIBLIOGRAPHY. Useful bibliographies include HARRY JARV, ***Die Kafka-Literatur*** (1961), most comprehensive; and ANGEL FLORES (ed.), ***The Kafka Problem***, pp. 447–477 (1963), with a list of works available in English. A major collection of Kafka manuscripts is kept in the Bodleian Library, Oxford, England.

Editions: The first collection of Kafka's works was the *Gesammelte* ***Schriften***, ed. by MAX BROD, 6 vol. (1935–37). The most comprehensive edition (though neither critical nor complete) is the ***Gesammelte Werke***, ed. by Brod, 8 vol. (1952–), which includes two volumes of letters, ***Briefe*** 1902–1924 (1958) and ***Briefe an Felice*** (1967).

Biography: The most important biographies are the intimate, congenial, basic (if biased) account of MAX BROD, ***Franz Kafka*** (1937; Eng. trans., 2nd ed., 1960); and the more factual and detailed study of Kafka's youth by KLAUS WAGENBACH, ***Franz Kafka: Eine Biographie seiner*** *Jugend, 1883–1912* (1958). See also Wagenbach's summary presentation of Kafka's entire career, ***Franz Kafka in Selbstzeugnissen und Bilddokumenten*** (1964), and his edition of an annotated collection of photos and documents, ***Franz Kafka, 1883–1924*** (1968), available in English.

Critical studies: HEINZ POLITZER, ***Franz Kafka, Parable and Paradox*** (1962), aesthetic and psychological analyses of texts; WILHELM EMRICH, ***Franz Kafka*** (1958; Eng. trans., 1968), philosophical; WALTER H. SOKEL, ***Franz Kafka. Tragik und Ironie*** (1964), a psychoanalytical reading, and his brief ***Franz Kafka*** (1966); RONALD D. GRAY (ed.), ***Kafka: A Collection of Critical Essays*** (1962), which includes important studies by F. Beissner, Albert Camus, Erich Heller, Martin Buber, and Edmund Wilson; ANGEL FLORES (ed.), ***The Kafka Problem*** (1963), representative selections from the literature on Kafka; HERBERT TAUBER, ***Franz Kafka: Eine*** *Deutung* ***seiner Werke*** (1941; Eng. trans., ***Franz Kafka: An Interpretation of His Works***, 1948), an early attempt at thorough interpretation; and GUNTHER ANDERS, ***Kafka*** (1951; Eng. trans., 1960), influential and challenging.

For a discussion of divergent (aesthetic, psychoanalytical, Marxist, existentialist, and theological) interpretations of Kafka, see PETER HELLER, "Kafka: The Futility of Striving," in ***Dialectics and Nihilism***, pp. 227–306 (1966).

(Pe.H)

Kalahari (Desert)

A basin-shaped, largely sand-covered region in southern Africa, the Kalahari (called the Kalahari Basin by geographers) is bounded by the Orange River to the south; the headwaters of the Zambezi to the north; the plateaus of the Transvaal, and Rhodesia to the east; and the highlands of South West Africa to the west. Its area is more than 100,000 square miles (260,000 square kilometres); the part colloquially known as the desert is in central and western Botswana, north of the Molopo River, and south of the Chobe and Okavango swamps.

THE LANDSCAPE

Frequently termed a desert, the Kalahari is more aptly described as a thirstland. With the exception of the Boteti River and related drainages, there is no permanent surface water in this sandy but well-vegetated region.

The Kalahari is a featureless, gently undulating, sand-covered plain lying 3,000 to 3,600 feet above sea level, rising to a height of some 4,000 feet on its southeastern and northwestern margins. Its lowest point of about 2,600 feet is at the confluence of the Nossob and Molopo rivers.

Relief features. The only projections above the relatively uniform surface of the Kalahari are occasional residuals of older rocks that form hills in the southeast, in the southwest near Tshabong, and in the northwest. At the southwestern extremity of the Okavango Swamps, south and southwest of Lake Ngami. a rib of ancient rock forms a low swell with a northeasterly strike. Hills some 500 feet high, also south and southwest of Lake Ngami, form the most prominent topographical features of the Kalahari.

Formation of dunes

Barchan dunes (crescent-shaped dunes, with the convex side facing the wind) on the Ntwetwe Pan in the western Makgadikgadi (formerly Makarikari) region are the only naturally formed sand dunes in the area. Following destruction of vegetation by man and stock, however, bare sand dunes have been formed in the southwestern Kalahari. Transport of sand by wind erosion occurs only where the vegetation has been destroyed. The minor amount of wind erosion is offset by local sheetflood erosion, which decreases differences in relief. A remarkably flat surface has consequently developed, but minor relief is evident in the form of low, fixed sand dunes, sandy hollows, pans, and fossil river valleys.

Near the lower parts of the Molopo and Nossob rivers, fixed dunes occur as west-northwest trending sand ridges. In the northeastern Kalahari low linear dunes extend southwestward from Rhodesia in the area north of the Makgadikgadi. All are fixed by vegetation.

South of the Boteti River, fossil river valleys form striking topographical features. These ancient drainage lines have undergone repeated cycles of aggradation and rejuvenation during past climatic regimes. Pronounced valleys are characteristic of the middle portions of their courses, but they become less clearly defined when traced both headward and into the Kalahari. where they normally degenerate into strings of shallow, grass-covered pans and flats, upon which the bush is encroaching.

Pans south of the Boteti River may be several miles long and may lie more than 50 feet below the plain. They invariably have a well-defined sand ridge along their southern and southwestern margins. Three main types of pan can be defined: grassed pans, with a dark-coloured, sandy, clay floor; ungrassed pans, on which there is a sparsely developed, halophilic (*i.e.*, thriving in a saline environment), scrubby vegetation; and saline pans, where the pan floor consists of whitish, saline, and highly alkaline clay, which does not support any plant growth higher than reddish-coloured algae. Some pans have well-defined characteristics, but all gradations may occur.

North of the Boteti River, fossil river valleys can seldom be traced. Pans of the type found in the area to the south are absent, though a different type of pan—a shallow basin-like drainage depression, with relatively high water-holding characteristics—is frequently encountered.

The Makgadikgadi depression

A major feature in the northeastern Kalahari is the Makgadikgadi depression, erroneously identified as a "salt pan" on most maps. This feature is an ancient interior drainage basin consisting of two extensive bare pan areas floored by alkaline sandy clays, with a combined floor area of over 3,000 square miles. The basin is fed mainly by waters from five ephemeral rivers flowing into the Sua, or eastern pan area, from the east and northeast. On occasion floodwaters from the Boteti River enter the southwestern part of the western, Ntwetwe, pan.

Drainage and soils. The only permanent drainage system is the Boteti River, which is fed by a seasonal overflow from the inland deltaic Okavango Swamps. High rainfall areas in Angola drain into the Okavango Swamps and eventually into the Boteti River, which runs from the southeastern margin of the swamps to flow east. It then turns south through the north-central Kalahari, flowing into Lake Xau (formerly Dow), after which it drains northeast into the Makgadikgadi area. Elsewhere, apart from seasonal accumulations in pans and valleys, water has to be obtained from wells and boreholes.

South of the Boteti River there is little or no runoff, except along certain stretches of the ancient valley systems, and within limited areas that are mainly found around pans. No rivers or streams, permanent or otherwise, are to be found. Rainfall is eventually lost by evaporation and—in the sand veld areas away from the ancient valleys and the pans—by evapotranspiration.

Two distinct ancient drainage systems can be traced in the fossil valleys, along most of which no flow now occurs. An ancient drainage divide, known as the Bakalahari-Schweile, crosses the southern part of the Kalahari. To the south of this divide, ancient valleys occur as former tributaries of the Molopo River, which originally was a tributary of the Orange River. (At present the Molopo is only ephemeral in its upper bed. In the extreme southwestern Kalahari, the dry Nossob River joins the Molopo. The Nossob flows only if exceptionally heavy rain falls in its headwater regions in South West Africa.)

North of the Bakalahari-Schwelle, the former drainage of the region was directed toward the Makgadikgadi. Northerly draining fossil river valleys can be traced from their ancient headwaters in the southeast. From the west, another ancient drainage system, the Okwa, can be traced eastward for more than 250 miles through the central Kalahari, running from the South West African border to join the northerly drainage system flowing into the Makgadikgadi. Tributaries of the Okwa are fossilized, but after exceptional rains in South West Africa the main river may flow into the Kalahari, for as much as 70 miles.

North of the Boteti River, runoff drains into shallow local depressions, or—occasionally—into an ancient valley system draining toward the Okavango.

Types of sands

Sands of the characteristic Kalahari type mantle over 90 percent of the region. They are of both eolian (wind) and fluviatile (river) origin. South of the Boteti River they are typically reddish brown in colour, with grayish-white sands prominent in low-lying areas. Grayish-white sands are generally more typical in the north, where fluviatile sands predominate. The wide distribution of these sands is broken only where pans, drainage lines, or large outcrop areas occur. In pan areas, along the fossil valley systems, along the Boteti River, and in the Makgadikgadi region, silicified sands, calcareous sandstones and grits, surface limestones, lake limestones, diatomaceous sandy limestones, marls, conglomerates, pan tufas, gravels, gritty sands, and the full lithological range of silcretes and calcretes are to be found.

Climate. The climate is of the low latitude, hot steppe type. Precipitation is local and erratic, with wide variations in any one locality. The mean annual rainfall is 16 to 18 inches in the east, 12 to 14 inches in the west, and less than 8 inches in the southwest. North of the Boteti River, the mean rises, reaching nearly 26 inches in the extreme north.

The climate is characterized by marked differences in seasonal temperatures and by large diurnal temperature ranges. Shade temperatures may be as high as 117" F in the summer—from October until March. In winter, freezing nights are common, with temperatures as low as 8° F. Such extremes are not normal in the northern Kalahari, where the humidity is relatively higher.

Vegetation and wildlife. The Kalahari possesses a well-developed cover of trees, low scrub, and tussocky grasses. The vegetation is xerophytic (adapted to dry conditions). Most trees and shrubs shed their leaves in season.

Tall savanna woodlands, which include a number of timber trees, occur in the north. Shrubby vegetation, with areas of scattered palm trees and baobabs, as well as open grassy flats, fringe the Makgadikgadi depression.

South of the Boteti River the vegetation predominantly consists of thorny woodland, with trees that seldom exceed about 25 feet in height. Parkland country with scattered tree growth occurs in certain parts. Proceeding to the south of the Kalahari, there is a progressive decrease in the number of tree species until only four kinds of trees are found. Around pans, in the ancient fossil valleys, and in patches in the sand veld plain, relatively dense, thorny, species of scrub acacia sometimes form thickets. Certain woodland and scrub areas subjected to repeated burning have degenerated into pure grassland.

Close to the ground — apart from a variety of grasses and low-growing shrubs — seasonal annuals, tuberous and bulbous plants, creepers, and vines are found.

Animal life is varied and is richer in the north than in the south. Even in the more arid southern areas, however, the variety of animal life is great. Many species exist for long periods of the year despite the absence of surface water. South of the Boteti River the principal animals found are springbok, wildebeest (gnu), hartebeest — all occasionally in vast herds; gemsbok (oryx); eland; and many smaller nongregarious species, such as duiker and steenbok. Kudu occur in denser bush areas.

Northward from the Boteti River and Makgadikgadi region, giraffe, zebra, and elephant are also found. In wetter areas the density and variety of animal life increases, and buffalo, roan and sable antelope, tsessebe (a species of South African antelope), impala, and other species are to be seen. Predators, such as lions, are ubiquitous. There are also some leopards, cheetahs, and lynxes.

Other animals common to the area include the hyena, Cape hunting dog, foxes and jackals, varieties of wild cat, warthog and bushpig, baboons, badgers, scaly anteaters, ant bears, and hares, as well as numerous rodents including porcupines, ground squirrels, springhares, gerbils, mice, and dormice. A wide variety of snakes and other reptiles occur, and there is also a wealth of bird life. (C.B.)

THE INHABITANTS

The Kalahari is inhabited by approximately 79,000 Bantu-speaking Africans, 500 Europeans, and 17,000 Bushmen.

The Bantu. The Bantu — represented by the Tswana, the Kgalagadi, and the Herero — appear to be relative newcomers to the Kalahari. In the late 18th century the Tswana spread westward from the Limpopo basin into the northern and eastern Kalahari; the Kgalagadi moved north and westward into the southern and western Kalahari; and the Herero refugees from the German colonial wars fled eastward into the western and northern Kalahari at the beginning of the 20th century.

Cattle-raising among the Bantu

Bantu in the Kalahari live in villages housing between 20 and 1,000 people. The villages are composed of mud-walled, single-roomed, thatched huts, situated near wells or boreholes. Cattle, the basis of the economy, are kept on the outskirts of villages, or at distances of up to 50 miles away. Grazing lands are state-owned, and their use is determined by local tribal government or by village councils. Wells and boreholes are owned by villages, by syndicates of cattle owners, by the state, or by private individuals. As the only regular supplies of water, away from the Okavango Swamps, are provided by bores and wells, these sources constitute the key factor controlling access to grazing in dry seasons. Cattle selectively feed upon a small range of the available vegetation and, in the absence of fences or pasture-improvement measures, soon degrade the indigenous pasture in a growing circle about the water point. Drought, disease, internal parasites, and wild animals also threaten the cattle, nearly all of which can be marketed only after an overland trek of up to 400 miles to the railhead slaughter house. Despite these drawbacks, herds continue to increase and to provide herdsmen with a living. Cattle are valued as wealth and as a measure of social status.

Goats furnish most of the meat and milk for home consumption, and nearly all households cultivate crops of maize, sorghum, and pumpkins. Due to drought, more crops fail than are successful. Wild food plants and the meat of game animals are important components of diet in the smaller and more remote villages. All villages have trading stores or are visited by hawkers who sell maize meal and other foodstuffs, clothing, ammunition, patent medicines, and other commodities.

All but the smallest villages have state-run primary schools, which are attended by the great majority of children, although few proceed to secondary education. The literacy rate is about 30 percent. State-run health clinics and hospitals in the larger villages supplement the services of herbalists and diviners.

Riding horses and donkeys are the usual means of local travel. Trucks belonging to traders or to the mine labour recruiting agency are used for longer journeys.

Europeans. Europeans first penetrated the Kalahari as travellers in the mid-19th century. The principal European settlement is in the Ghanzi District, where some 200 whites are engaged in cattle ranching. Since 1960 determined efforts at pasture management, better utilization of groundwater supplies, and scientific stock breeding have increased productivity. Similar methods are also being adopted by Bantu pastoralists.

The remainder of whites in the Kalahari are expatriates employed by the state or by traders.

Bushmen. Bushmen are short in stature. The average height of the men is five feet two inches; the women are some two inches shorter. Skin colour at birth is coppery-yellow, darkening to light brown with exposure to the sun. The facial profile is almost flat.

Life style of the Bushmen

Most Bushmen are now clients of Bantu pastoralists, hunting game and working at cattle posts in return for support. Some 4,000 Bushmen work on the ranches of Ghanzi, which are owned by either whites or Africans; the Bushmen are employed there as ranch labourers, or else are dependents of such labourers.

An estimated 6,000 Bushmen still follow their traditional pattern of hunting and gathering. Of these, the Kung (!Kung), ! xong, and G/wi tribes (the "!" and "/" represent clicks) have been intensively studied in recent years. While each group is distinct, the G/wi of the Central Kalahari Game Reserve may be considered as illustrative of the Bushman hunter-gatherer ecology.

They travel together in bands, each consisting of 5 to 16 families united by bonds of kinship and friendship. Each band has a recognized territory of 300 to 400 square miles, selected for its resources of food plants (which contribute the main part of the diet), wet-season waterholes (for drinking water during the six to eight weeks when sufficient rainwater gathers in pools), trees (for shade, shelter, firewood, and wood for making artifacts), and areas of grazing to attract and sustain herds of game animals. Subsistence is based on 34 species of edible plants, of which eight are staples in their various seasons. This diet is supplemented by the meat of antelope and other herbivorous mammals, by tortoises and other reptiles, and by the flesh and eggs of all but raptorial and scavenging birds. Plant-gathering is mostly done by the women ranging within five miles of the camp, while men hunt over a much larger area of the territory. The main hunting weapon is a light bow shooting flimsy, unfletched, poisoned arrows. The range is only 25 yards, and great skill is needed to stalk the quarry within this distance. Antelope leather provides clothing material. Women wear fore and hind skirts, as well as cloaks, which also serve as blankets and carrying-bags. Men wear breechclouts and cloaks. Sandals are sometimes worn as protection from the hot sand.

From November to some time between late June and early August, when there is sufficient food, the band lives as one community, moving camp every three or four weeks as the local supply of food plants becomes exhausted. Blighting frosts deplete the available food plants in winter (which lasts from May to September), and the band then splits into its constituent households, each retreating to a separate part of the territory. Early-fruiting plants increase the food supply just before the approach of the wet season, allowing the band to reunite at a joint campsite.

During dry seasons, shelters are little more than open windbreaks made of branches and grass. In the wet season, domed structures of branches are thatched and made rainproof. (G.Si.)

TRAVEL AND TRANSPORT

The Kalahari thirstland constituted a barrier to early travellers. Many avoided the central Kalahari, since some early attempts to cross it ended in disaster.

The advent of the internal-combustion engine opened up previously inaccessible areas. The earliest crossing by this method was made in 1925, when the central Kalahari was traversed from east to west by truck. Because of its sparse population, the Kalahari is served by infrequent roads and tracks, the majority of which are passable only by trucks or four-wheel drive vehicles.

Maintained roads connect administrative centres, major habitations, and marginal farming areas in the south, southwest, and northwest. Constructed roads now link eastern Botswana with the Okavango Swamps and with mining developments south of the Makgadikgadi.

(C.B.)

BIBLIOGRAPHY. C. BOOCOCK and O.J. VAN STRATEN, "Notes on the Geology and Hydrogeology of the Central Kalahari Region, Bechuanaland Protectorate," *Trans. Geol. Soc. S. Afr.*, 65:125–171 (1962), a review paper on the general geology and hydrology of the central Kalahari region, with a discussion of pan characteristics and other surficial features and a list of pertinent references; A.T. GROVE, "Landforms and Climatic Change in the Kalahari and Ngarniland," *Geogrl. J.*, 135:191–213 (1969), a review article on geomorphological characteristics of the Kalahari with special reference to dune structures and former climatic regimes; C.M.H. JENNINGS, *Groundwater in Botswana: Symposium on Groundwater in Southern Africa* (1969), an article summarizing groundwater development possibilities in Botswana; L.K. LUSTIG, "An Inventory of Geographical Research on Desert Environments," in *Inventory of Research on Geomorphology and Surface Hydrology of Desert Environments*, ch. 4 (1967), a report on the state of knowledge of geomorphology and surface hydrology of the arid regions, including a section on the Kalahari with a detailed list of publications; O.B. MILLER, "The Woody Plants of the Bechuanaland Protectorate," *Jl. S. Afr. Bot.*, vol. 18 (1952), a checklist of trees and shrubs, with both scientific and Setswana and other vernacular names, listing species found in various areas of Botswana, including parts of the Kalahari; P. MAHOME and J.B. WEBSTER, *A Bibliography on Bechuanaland* (1966), a bibliography of books, periodicals, articles, and pamphlets, though largely excluding government publications, with major emphasis upon literature in the social sciences, but a number of titles on indigenous languages, agriculture, and the natural sciences are included; G. PASSARGE, *Die Kalahari*, 2 vol. (1904), a classic work containing early data on most aspects of the northwestern, north-central, and parts of the eastern Kalahari; G.B. SILBERBAUER, *Bushman Survey Report: An Official Report for the Bechuanaland Government* (1965), a detailed eco-sociological study on various Bushman tribes of the Kalahari including details of the ecology of the central Kalahari; J.H. WELLINGTON, *Southern Africa: A Geographical Study*, vol. 1, *Physical Geography* (1955), a standard textbook on the geography of the southern African region, including a number of references to the geography and geomorphology of the Kalahari.

(C.B./G.Si.)

Kālidāsa

Traditions concerning his life

Kālidāsa (fl. c. 5th century AD) was an Indian poet and dramatist, the outstanding master of the Sanskrit language and probably the greatest writer of any Indian epoch. As with most classical Indian authors, little is known about his person or his historical relationships. His poems suggest but nowhere declare that he was a Brahmin (priest), liberal yet committed to the orthodox Hindu world view. His name, literally "servant of Kālī," presumes that he was a Saivite (follower of the god Śiva, whose consort was Kālī), though occasionally he eulogizes other gods, notably Viṣṇu (Vishnu). A Sinhalese tradition says that he died in Ceylon during the reign of Kumāradāsa (ascended the throne AD 517). A more persistent legend makes Kālidāsa one of the "nine gems" at the court of the fabulous King Vikramiiditya of Ujjain. Unfortunately, there are several known Vikramādityas ("sun of valour"—a common royal appellation); likewise, the nine distinguished courtiers could not have been contemporaries. It is certain only that the poet lived sometime between the reign of Agnimitra, the second Śuṅga king (c. 170 BC), the hero of one of his dramas, and the Aihoḷe inscription of AD 634, which lauds Kālidāsa. He is apparently imitated, though not named, in the Mandasor inscription of AD 473.

No single hypothesis accounts for all the discordant information and conjecture surrounding this date. An opinion accepted by many—but not all—scholars is that Kālidāsa should be associated with Candra Gupta II (reigned AD c. 380–*c.* 415).

The most convincing but most conjectural rationale for relating Kālidāsa to the brilliant Gupta dynasty is simply the character of his work, which appears as both the perfect reflection and the most thorough statement of the cultural values of that serene and sophisticated aristocracy.

Works and significance

Tradition has associated many works with the poet; criticism identifies six as genuine and one more as likely ("Ṛtusaṃhāra," the "Garland of the Seasons," perhaps a youthful work). Attempts to suggest a poetic and intellectual development across them run afoul of the impersonality that is characteristic of classical Sanskrit literature. His works are judged by the Indian tradition as realizations of literary qualities inherent in the Sanskrit language and its supporting culture. Kālidāsa has become the archetype for Sanskrit literary composition.

In drama, his *Abhijñānaśakuntala*, "Śakuntalā Recognized," is the most famous and is usually judged the best Indian literary effort of any period. Taken from an epic legend, it tells of the seduction of the nymph Śakuntalā by King Duṣyanta, his rejection of the girl and his child, and their subsequent reunion in heaven. The epic myth is important because of the child, for he is Bharata, eponymous ancestor of the Indian nation (Bhāratavarṣa, "sub-continent of Bharata"). Kālidāsa remakes the story into a love idyll whose characters represent a pristine aristocratic ideal: the girl, sentimental, selfless, alive to little but the delicacies of nature, and the King, first servant of the *dharma* (religious and social law and duties), protector of the social order, resolute hero, yet tender and suffering agonies over his lost love. The plot and characters are made believable by a change Kālidāsa has wrought in the story: Duṣyanta is not responsible for the lovers' separation; he acts only under a delusion caused by a sage's curse. As in all of Kālidāsa's works, the beauty of nature is depicted with a precise elegance of metaphor that would be difficult to match in any of the world's literatures.

The second drama, *Vikramorvaśī*, "Urvaśī Won by Valour" (possibly a pun on *vikramdditya)*, tells a legend as old as the Veda (earliest Hindu scripture), though very differently. Its theme is the love of a mortal for a divine maiden; it is well-known for the "mad scene" (Act IV) in which the king, grief-stricken, wanders through a lovely forest apostrophizing various flowers and trees as though they were his love. The scene was intended in part to be sung or danced.

The third drama, *Mālavikāgnimitra*, "Mālavikā and Agnimitra," is of a different stamp—a harem intrigue, comical and playful, but not less accomplished for lacking any high purpose. The play (unique in this respect) contains datable references, the historicity of which have been much discussed.

Kālidāsa's efforts in *kdvya* (strophic poetry) are of uniform quality and show two different subtypes, epic and lyric. Examples of the epic are the two long poems *Raghuvaṃśa*, "Dynasty of Raghu," and *Kumārasambhava*, "Birth of the War God." The first recounts the legends of the hero Riima's forebears and descendants; the second tells the picaresque story of Śiva's seduction by his consort Pārvatī, the conflagration of Kāma (the god of desire), and the birth of Kumāra (Skanda), Siva's son. These stories are mere pretext for the poet to enchain stanzas, each metrically and grammatically complete, redounding with complex and reposeful imagery. Kālidāsa's mastery of Sanskrit as a poetic medium is nowhere more marked.

A lyric poem, the "Meghadfita," "Cloud Messenger," contains, interspersed in a message from a lover to his absent beloved, a series of unexcelled and knowledgeable vignettes, describing the mountains, rivers, and forests of north India.

Modern and Western interest in Kālidāsa begins with Sir William Jones's translation of the story of Śakuntalā (1789)--one of the earliest Indic works into a Western language—which, through a German retranslation, awakened the enthusiasm of Johann Wolfgang von Goethe.

The society reflected in Kālidāsa's work is that of a courtly aristocracy sure of its dignity and power. Kālidāsa has perhaps done more than any other writer to wed the older, Brahminic religious tradition, particularly its ritual concern with Sanskrit, to the needs of a new and brilliant secular Hinduism. The fusion, which epitomizes the renaissance of the Gupta period, did not, however, survive its fragile social base; with the disorders following the collapse of the Gupta Empire, Kālidāsa became a memory of perfection that neither Sanskrit nor the Indian aristocracy would know again.

MAJOR WORKS

PLAYS: *Abhijñānaśakuntala*, major critical edition by S.K. Belvalkar (1965; Eng. trans. by M.B. Emeneau, 1962); ***Vikramorvaśī***, ed. by H.D. Velankar (1961; Eng. trans. by E.B. Cowell, 1851); *Mālavikāgnimitra*, ed. by F. Bollensen (1879; Eng. trans. by E. Gerow, 1972).

POEMS: *Raghuvaṃśa*, ed. by S.P. Pandit (1869–74; French trans. by L. Renou, 1928); ***Kumdrasambhava***, ed. by Ganapati Sastri (1913–14; Eng. trans. by R.T.H. Griffith, 2nd ed., 1878); "Meghadfita," ed. by E. Hultzsch (1911; Eng. trans. by F. and E. Edgerton, 1964); "Ṛtusaṃhāra," ed. by A. Scharpé (1958; Eng. trans. by G.R. Mathiers, 1929).

All the works of Kālidāsa (including the "Ṛtusaṃhāra") have been edited by A. Scharpé and were published in four volumes of ***Rijksuniversiteit te Gent: Werken uitgegeven door de Faculteit van de Wijsbegeerteen Letteren*** (vol. 117, 120, 122, and 134). For the "Ṛtusaṃhāra," this constitutes the only modern critical edition. All the works are available in excellent but not critical (by Western standards), editions, published by the Nirnayasagara Press, Bombay.

BIBLIOGRAPHY

General works: MORIZ WINTERNITZ, ***Geschichte der indischen Litteratur***, vol. 3 (1920); LOUIS RENOU, ***L'Inde classique***, vol. 2 (1953); A.B. KEITH, ***A History of Sanskrit Literature*** (1928), ***Sanskrit Drama*** (1924); S.N. DASGUPTA and S.K. DE, A ***History of Sanskrit Literature, Classical Period***, 2nd ed., vol. 1 (1962), the two authors differ on Kālidāsa's date, the former preferring 1st or 2nd century BC (pp. 728–739) and De defending the time of Candra Gupta II.

Articles: BHAU DAJI, "On the Sanscrit Poet, Kālidāsa," J. ***Botnbay Branch R. Asiat. Soc.***, 6:19–30 (1861), an early effort to date Kālidāsa; R.G. BHANDARKAR, "Peep into the Early History of India from the Foundation of the Maurya Dynasty to the Fall of the Imperial Gupta Dynasty," ***ibid.***, 20:356–408 (1900), proposes Candra Gupta II as the Vikrāmaditya with whom Kālidāsa is to be associated; A.F. RUDOLPH HOERNLE, "Some Problems in Ancient Indian History," *Jl.* R. ***Asiat. Soc.*** (1909), esp. pp. 108–113, equates Vikrimaditya with Yaśodharman (AD 530–540); A.B. KEITH, "Vikrāmaditya and Kālidāsa," ***ibid.***, pp. 433–437, refutes Hoernle; B. LIEBICH, "Das Datum des Kālidāsa," *Indische Forschungen*, 31:198–203 (1912), resumé of various theories, refuting Candra Gupta II date; D.R. BHANDARKAR, "Can We Fix the Date of Kalidasa More Accurately?" ***Ann. Bhandarkar Inst.***, 8:200–204 (1927), again suggests Yaśodharman as the Vikrāmaditya.

Special studies: HARI CHAND SASTRI, ***Kalidasa et l'art poétique de l'Inde*** (1917); A. HILLEBRANDT, *Kālidāsa* (1921); W. RUBEN, ***Kalidasa: Die menschliche Bedeutung seiner Werke*** (1956; Eng. trans., ***Kalidasa: The Human Meaning of His Works***, 1957); AUROBINDO GHOSE, ***Kalidasa***, 2nd ed. (1950).

See also the introductions to the various critical editions of Kālidāsa's works now appearing under the auspices of the Sāhitva Akademi, Delhi: and RABINDRANATH TAGORE, ***shipwreck*** (1906).

(Ed.G.)

Kandinsky, Wassily

Among the vanguard artists who changed the history of art in the early years of the 20th century, the Russian painter and aesthetic theorist Wassily Kandinsky occupies a special niche of his own, for he is often referred to as the creator of the first pure abstraction—the first picture that broke completely with the Western tradition of depicting recognizable objects. His absolute priority in this respect, like most historical firsts, is open to argument. But, in his paintings and his writing just before and just after World War I, he was certainly one of the most undeviating and most eloquent of European abstractionists.

By courtesy of Mme Nina Kandinsky; © A.D.A.G.P. 1972 by French Reproduction Rights, Inc.

Kandinsky.

Kandinsky was born in Moscow on December 4, 1866 (November 22 on the old Russian calendar). His mother was a Muscovite, one of his great-grandmothers a Mongolian princess, and his father a native of Kyakhta, a Siberian town near the Chinese border; the boy thus grew up with a cultural heritage that was partly European and partly, although not very specifically, Asian. His family was genteel, well-to-do, and fond of travel; while still a child he became familiar with Venice, Rome, Florence, the Caucasus, and the Crimean peninsula. At Odessa, where his parents settled in 1871, he completed his secondary schooling and became an amateur performer on the piano and the cello. He also became, along with several of his relatives, an amateur painter; and he later recalled, as a sort of first impulse toward abstraction, an adolescent conviction that each colour had a mysterious life of its own. In 1886 he began to study law and economics at the University of Moscow, but he continued to have unusual feelings about colour as he contemplated the city's vivid architecture and its collections of icons; in the latter, he once said, could be found the roots of his own art. In 1889 the university sent him on an ethnographic mission to the ***oblast*** (provincial subdivision) of Vologda, in the forested north, and he returned with a lasting interest in the often garish, nonrealistic styles of Russian folk painting. During that same year he discovered the Rembrandts in the Hermitage at St. Petersburg (Leningrad), and he furthered his visual education with a trip to Paris. He pursued his academic career and in 1893 was granted the degree equivalent of a doctorate.

Early understanding of colour

By this time, according to his reminiscences, he had lost much of his early enthusiasm for the social sciences. He felt, however, that art was "a luxury forbidden to a Russian." Eventually, after a period of teaching at the university, he accepted a post as the director of the photographic section of a Moscow printing establishment. In 1895 he had what proved to be a prophetic experience: looking at one of Claude Monet's haystack pictures, he was momentarily unable to recognize the depicted object because of the Impressionist handling of flecked light and colour. He remembered,

> I had the feeling that here the subject of the picture was in a sense the painting itself, and I wondered if one couldn't go much further along the same route. After that I looked at Russian icons with different eyes; I had eyes, that is, for what was abstract. . . .

The next year, when he was approaching his 30th birthday, he was forced to choose among his possible futures, for he was offered a professorship in jurisprudence at the University of Dorpat (later called Tartu), in Estonia, which was then undergoing Russification. In what he called a "now or never" mood, he turned down the offer and took the train for Germany with the intention of becoming a painter.

Munich period. He already had the air of authority that would contribute to his success as a teacher in later years. He was tall, large-framed, impeccably dressed, and equipped with pince-nez glasses; he had a habit of holding his head high and seeming to look down at the universe. He resembled, according to acquaintances, a mixture of diplomat, scientist, and Mongol prince. But for the moment he was simply an overage art student. and he enrolled as such in a private school at Munich run by the minor painter Anton Azbé. Two years of study under Azbé were followed by a year of work alone and then by enrollment at the Munich Academy in the class of Franz von Stuck, who had a considerable reputation as a painter of bacchanalia and lurid, heavily emphatic allegories. Among the students at the Azbé school was Alexey von Jawlensky, who was heading toward a personal kind of Expressionism after studying with the Russian Realist Ilya Repin in Moscow; among those at the Munich Academy was the Swiss-born Paul Klee, who was developing his own inimitable kind of semi-abstract fantasy.

Academic training in art

Kandinsky emerged from the academy with a diploma in 1900 and, during the next few years, achieved moderate success as a competent professional artist in touch with modern trends. Starting from a base in 19th-century Realism, he was influenced by Impressionism, by the whiplash lines and decorative effects of Art Nouveau (called Jugendstil in Germany), by the dot technique of Neo-Impressionism or Pointillism, and by the strong, unrealistic colour of central European Expressionism and French Fauvism. Often he revealed that he had not forgotten the icons of Moscow and the folk art of Vologda; sometimes he indulged in patterns of violent hues that would have delighted his Asian ancestors. He exhibited with the vanguard groups and in the big nonacademic shows that had sprung up all over Europe—with the Munich Phalanx group (of which he became president in 1902), with the Berlin Sezession group, in the Paris Salon d'Automne and Salon des Indépendants, and with the Dresden group that called itself Die Brucke (the Bridge). In 1903 in Moscow he had his first one-man show, followed the next year by two others in Poland. Between 1903 and 1908 he travelled extensively, from Holland to as far south as Tunisia and from Paris back to Russia, stopping off for stays of several months each in Kairouan (Tunisia), Rapallo (Italy), Dresden, the Parisian suburb of Sèvres, and Berlin.

In 1909 Kandinsky and the German painter Gabriele Münter, who had been his mistress since 1902, acquired a house in the small town of Murnau, in southern Bavaria. Working part of the time in Murnau and part of the time in Munich, he began the process that led to the emergence of his first strikingly personal style and finally to the historic breakthrough into purely abstract painting. Gradually, the many influences he had undergone coalesced around a question that had evidently been on his mind since at least 1895, when he had felt that Monet's picture of a haystack was communicating something that had nothing to do with haystacks. His impulse to eliminate the haystack altogether was not, it should be noted, due merely or even primarily to strictly aesthetic considerations. No one could have been less of an aesthete, less of an "art for art's sake" addict, than Kandinsky. Nor was he the sort of born painter who could enjoy the physical properties of oil and pigment without caring what they meant. In his book *Über* das Geistige in der *Kunst* (Concerning the Spiritual *in Art*). which he completed in 1910 and published late the next year, he emphasizes what he presents as a modern need for an art that springs not from "the exterior" but from an "interior necessity." In his *Rückblicke* (*Reminiscences*), published in 1913, he remarks: "I saw clearly in front of me the danger of an ornamental kind of painting. . . ." In summary, he wanted a kind of painting in which colours, lines, and shapes, freed from the distracting business of depicting recognizable objects, might evolve into a visual "language" capable—as was, for him, the abstract "language" of music—of expressing general ideas and evoking deep emotions.

Development of a theory of purely abstract painting

The project was not, of course, entirely new. Analogies between painting and music had long been common; many thinkers had attempted to codify the supposed expressiveness of colours, lines, and shapes; and more than one fairly ancient sketch might compete for the honour of being called the first abstract picture. Moreover, in these years just before World War I, Kandinsky was by no means alone in his attack on figurative art. By 1909 the Cubists were turning out intellectualized and fragmented visions of reality that completely baffled the ordinary viewer. Between 1910 and 1914 the list of pioneer abstract artists included the Russians Mikhail Larionov, Natalya Goncharova, and Kazimir Malevich; the Swiss Augusto Giacometti (uncle of the sculptor Alberto Giacometti); the Frenchmen Robert Delaunay and Fernand Léger; the Czech František Kupka; and the Dutchman Piet Mondrian. A strict examination of works and dates can show, therefore, that Kandinsky does not quite deserve to be called, as he often is, the "founder" of nonfigurative painting; at least he cannot be called the only founder. But, when this historical point is conceded, he remains a pioneer of the first importance. He had no consequential precedent for his theorizing in *Über das* Geistige in der *Kunst* and no fellow researcher before 1910 whose program was as coherent as his. Moreover, in the beginning he was alone in possessing the peculiar but probably required combination of rationalism with an emotionalism that verged on mysticism. He was a trained lawyer and social scientist and at the same time capable, according to his own testimony, of seeing colours as he listened to Wagner's opera Lohengrin. His orderly, rather academic mind was not immune to the fascination of the large, loose notions of the pre-1914 Theosophists. He had, according to his friend Klee, "his heart in his head."

Place as a pioneer of nonfigurative painting

Stylistic evolution toward a nonrepresentational art

Kandinsky's widely accepted claim to historical priority rests mainly on an untitled work dated 1910 and commonly referred to as "First Abstract Watercolour." On the basis of research done in the 1950s, however, this work can be dated somewhat later and can be regarded as a study for the 1913 "Composition VII"; and in any event it must be considered merely an incident—among many for which the evidence has not been preserved—on Kandinsky's route. In "Blue Mountain" (1908) the evolution toward nonrepresentation is already clearly under way; the forms are schematic, the colours non-naturalistic, and the general effect that of a dream landscape. In "Landscape with Steeple" (1909; Musée National d'Art Moderne, Paris) similar tendencies are evident, together with the beginning of what might be called an explosion in the composition. By 1910, "Improvisation XIV" is already, as its somewhat musical title suggests, practically abstract; with the 1911 "Encircled," there has definitely developed a kind of painting that, though not just decoration, has no discernible point of departure in the depiction of recognizable objects. After that come such major works as "With the Black Arch," "Black Lines," and "Autumn"; in such pictures, done between 1912 and 1914 in a slashing, splashing, dramatic style that anticipates the New York Abstract Expressionism of the 1950s, most art historians see the peak of the artist's achievement.

Kandinsky's theorizing and his mysticism kept pace with the evolution of his style. In 1913, in his ***Riickblicke,*** he remarks:

> Thus the domain of Art was separating itself, for me, more and more from the domain of Nature, to the point where I could experience each of them as a domain by itself. I fully realized this for the first time this year.

He adds:

> To paint is to detonate a shock of different worlds. . . . In terms of technique, each work comes into existence as the Cosmos did, by means of catastrophes that, beginning with the chaotic cries of the instruments, finish in the symphony we call the music of the spheres.

Not surprisingly, he is said to have often reached a point in his conversations beyond which his friends could not follow his thinking. But he is also said to have possessed an unexpectedly quick, deadpan sense of humour, and this report is supported by some of his occasional verse, written in a vein that recalls both a Russian fairy tale and a Dada joke. Kandinsky also wrote, in 1909, three theatrical pieces, entitled ***Der Gelber Klang*** ("The Yellow Sound"), *Der Grüner Klang* ("The Green Sound"), and *Schwartz und Weiss* ("Black and White"), and in 1911 followed them with a fourth, ***Violet.***

Formation of Der Blaue Reiter

He was an active animator of the avant-garde movement in Munich, helping to found in 1909 the Neue Kiinstlervereinigung (New Artists' Association). Following a disagreement within this group, he and the German painter Franz Marc founded in 1911 an informally organized rival group, which took the name Der Blaue Reiter (The Blue Rider), from the title of one of Kandinsky's 1903 pictures, and which stated its aims in terms that clearly reflected Kandinsky's rather Oriental philosophical stance:

> To give expression to inner impulses in every form that provokes an intimate reaction in the beholder, such is the goal which the Blaue Reiter will aspire to achieve. We seek today, behind the veil of external appearances, the hidden things that seem to us more important than the discoveries of the Impressionists.

Among the important members, in addition to Kandinsky and Marc, were Klee and the gifted young German artist August Macke. The Blaue Reiter disappeared with the outbreak of World War I, but during its short life it had a significant role in the development of modern art. In the exhibitions it staged at Munich in 1911 and 1912, it gave central Europeans an opportunity to see the work of such artists as Delaunay, Henri Rousseau, Larionov, Goncharova, Georges Braque, André Derain, and—as a painter—the composer Arnold Schoenberg. The Blaue Reiter "almanac"—a volume of essays and reproductions—featured material on modern music and on modern, primitive, popular, medieval, and Oriental visual art.

Russian interlude. When World War I was declared in 1914, Kandinsky broke off his relationship with Gabriele Münter and returned to Russia by way of Switzerland, Italy, and the Balkans. An early marriage to a cousin had been dissolved in 1910 after a long period of separation, and in 1917 he married a Moscow girl, Nina Andreevskaya, whom he had met the previous year. Although he was past 50 and his bride was many years younger, the marriage turned out to be extremely successful, and he settled down in Moscow with the intention of reintegrating himself into Russian life. His intention was encouraged by the new Soviet government, which at first showed itself anxious to win the favour and services of avant-garde artists. In 1918 he became a professor at the Moscow Academy of Fine Arts and a member of the arts section of the People's Commissariat for Public Instruction. His autobiographical ***Riickblicke*** was translated into Russian and published by the Moscow municipal authorities. In 1919 he created the Institute of Artistic Culture, became director of the Moscow Museum for Pictorial Culture, and helped to organize 22 museums across the Soviet Union. In 1920 he was made a professor at the University of Moscow and was honoured with a one-man show organized by the state. In 1921 he founded the Russian Academy of Artistic Sciences. But by then the Soviet government was veering from avant-garde art to Social Realism, and so, at the end of the year, he and his wife left Moscow for Berlin.

Influence of Constructivism

In spite of the war, the Revolution, and official duties, he had found time to paint during this Russian interlude and even to begin a quite drastic transformation of his art. Whereas, in his Munich work as late as 1914 one can still find occasional allusions to landscape, the canvases and watercolours of his Moscow years show a determination to be completely abstract. They also show, possibly because of the influence of such Russian revolutionaries as the severely geometrical Malevich and the Constructivists Naum Gabo and Antoine Pevsner, a growing tendency to abandon the earlier spontaneous, lyrical, organic style in favour of a more deliberate, rational, and constructional approach. The change is evident in such pictures as "White Line" and "Blue Segment."

Bauhaus period. By this time Kandinsky had an international reputation as a painter. He had always, however, been interested in teaching, first as a lecturer in law and economics just after getting his university degree, then as the master of a painting school he had organized in Munich, and more recently as a professor at the University of Moscow. He seems not to have hesitated, therefore, when early in 1922 he was offered a teaching post at Weimar in the already famous Bauhaus school of architecture and applied art. At first his duties were a little remote from his personal activity, for the Bauhaus was not concerned with the formation of "painters" in the traditional sense of the word. He lectured on the elements of form, gave a course in colour, and directed the mural workshop. Not until 1925, when the school moved from Weimar to Dessau, did he have a class in "free," nonapplied painting. In spite of the somewhat routine nature of his work, however, he appeared to have found life at the Bauhaus rewarding and very pleasant. Among the professors were his old friend Klee, the American-born Cubist Lyonel Feininger, the German abstractionist Josef Albers, the Hungarian Constructivist László Moholy-Nagy, the Hungarian architect and furniture designer Marcel Breuer, and the German architect Walter Gropius, who was the school's director until 1928. The climate was one of research and craftsmanship combined with a certain amount of aesthetic puritanism; it was classical, to use the term rather loosely, by comparison with the warm romanticism of his pre-1914 days in Munich.

Kandinsky responded to this climate by continuing to evolve in the general direction of geometrical abstraction, but with a dynamism and a taste for detail-crowded pictorial space that recall his earlier sweeping-gesture technique. Although circles and triangles became his dominant forms, they were combined with arrows, spikes, and warped grids in a way that defied the Bauhaus notion of bare Euclidean purity. The colours sometimes suggested Russian folk art and sometimes something vaguely Aztec. Typical works are the transitional "In the Black Circle," "One Center," "Yellow-Red-Blue," "Three Sounds," "On the Points," and "Affirmed Pink." Although this period of the artist's career has always had its fervent admirers, a probable majority of art historians has tended to have reservations concerning the frequent lack of ease and spontaneity—the look, despite the energy displayed, of something contrived to illustrate a theory.

Treatises on theory

That Kandinsky was keenly interested in theory during these years is evident from his publication in 1926 of his second important treatise, ***Punkt und Linie zu Fläche*** (Point and Line to Plane). In his first treatise, ***Concerning the Spiritual in Art,*** he had emphasized in particular the supposed expressiveness of colours, comparing yellow, for example, to the aggressive, allegedly earthly sound of a trumpet and comparing blue to the allegedly heavenly sound of the pipe organ. Now, in the same spirit, he analyzed the supposed effects of the abstract elements of drawing, interpreting a horizontal line, for example, as cold and a vertical line as hot. He continued to have a partly mystical belief in the possibility of an abstract visual language; in 1931 he wrote:

> The impact of an acute angle of a triangle on a circle produces an effect no less powerful than the finger of God touching the finger of Adam in Michelangelo; and if fingers are not just anatomical or physiological, but something more, a triangle or circle is something more than geometry.

Paris period. Although he had been a German citizen since 1928, he emigrated to Paris when, in 1933, the Gestapo forced the Bauhaus to close. The last, and one of the finest, of his German pictures is the sober "Development in Brown"; its title probably alludes to the Nazi brown-shirted storm troopers, who regarded his abstract art as "degenerate." He lived for the remaining 11 years of his life in an apartment in the Parisian suburb of Neuilly-sur-Seine, becoming a naturalized French citizen in 1939.

"Concrete" style of late paintings

During this final period his painting, which he began to prefer to call "concrete" rather than "abstract," became to some extent a synthesis of the organic manner of the Munich period and the geometrical manner of the Bauhaus period. The visual language he had been aiming at since at least 1910 turned into collections of signs that look like almost-decipherable messages written in pictographs and hieroglyphs; many of the signs resemble aquatic larvae, and now and then there is a figurative hand or a lunar human face. Typical works are "Violet Dominant," "Dominant Curve," "Fifteen," "Moderation," and "Tempered Elan." The production of such works was accompanied by the writing of essays in which the artist stressed the alleged failure of modern scientific positivism and the need to perceive what he termed "the symbolic character of physical substances."

Kandinsky died at Neuilly-sur-Seine on December 13, 1944. His influence on 20th-century art, often filtered through the work of more accessible painters, was profound.

MAJOR WORKS

"The Sluice" (1901; Neue Pinakothek, Munich); "The Blue Rider" (1903; Ernst Biihrle Collection, Ziirich); "Beach Baskets in Holland" (1904; Neue Pinakothek); "The Night" (1907; Neue Pinakothek); "Street in Murnau with Women" (1908; Musée National d'Art Moderne, Paris); "Blue Mountain" (1908; Solomon R. Guggenheim Museum, New York); "First Abstract Watercolour" (1910; Nina Kandinsky Collection, Neuilly-sur-Seine, France); "Encircled" (1911; Nina Kandinsky Collection); "All Saints Picture 11" (1911; Neue Pinakothek); "With the Black Arch" (1912; Nina Kandinsky Collection); "Black Lines" (1913; Solomon R. Guggenheim Museum); "Improvisation No. 30" (1913; Art Institute of Chicago); "Composition VII" (1913; State Tretyakov Gallery, Moscow); "Picture with Three Spots" (1914; Solomon R. Guggenheim Museum); "Autumn" (1914; Solomon R. Guggenheim Museum); "White Line" (1920; Nina Kandinsky Collection); "Blue Segment" (1921; Solomon R. Guggenheim Museum); "In the Black Circle" (1923; Fondation Maeght, Saint-Paul, France); "One Center" (1924; Solomon R. Guggenheim Museum); "Yellow-Red-Blue" (1925; Nina Kandinsky Collection); "Several Circles" (1926; Solomon R. Guggenheim Museum); "Three Sounds" (1926; Solomon R. Guggenheim Museum); "On the Points" (1928; Nina Kandinsky Collection); "Scarcely (Kaum)" (1930; Boymans-von-Beuningen Museum, Rotterdam); "Affirmed Pink" (1932; Solomon R. Guggenheim Museum); "Development in Brown" (1933; Musée National d'Art Moderne, Paris); "Violet Dominant" (1934; Fondation Maeght, Saint-Paul, France); "Succession" (1935; Phillips Collection, Washington); "Dominant Curve" (1936; Solomon R. Guggenheim Museum); "Tensions Relaxed" (1937; W. Gidwitz Collection, Chicago); "Sweet Trifles" (1937; Nina Kandinsky Collection); "Fifteen" (1938; Nina Kandinsky Collection); "Composition X" (1939; Nina Kandinsky Collection); "Moderation" (1940; Solomon R. Guggenheim Museum); "Various Actions" (1941; Solomon R. Guggenheim Museum); "Division-Unity" (1943; Nina Kandinsky Collection); "Circle and Square" (1943; Nina Kandinsky Collection); "Seven" (1943; Max Bill Collection, Zurich); "White Balancing Act" (1944; Professor W. Loffler Collection, Zurich); "Tempered Élan" (1944; Nina Kandinsky Collection).

BIBLIOGRAPHY. MAX BILL *et al.*, *Wassily Kandinsky* (1951), is a collection of essays by Bill and other critics, with a useful chronology by NINA KANDINSKY. WILL GROHMANN, *Wassily Kandinsky: Life and Work* (1959), is a sympathetic biography by a friend of the artist. PAUL OVERY, *Kandinsky: The Language of the Eye* (1969), offers the best analysis of the painter's not always clear intentions.

(R.McMu.)

K'ang-hsi

K'ang-hsi (Pin-yin romanization, Kang-xi; given name, Hsiian-yeh; temple name, Sheng-tsu; posthumous name, Jen Huang-ti) was the fourth Ch'ing emperor but only the second to reign in China. Usually counted among the ablest monarchs ever to govern the vast Chinese empire, K'ang-hsi reigned for 61 years, laying the foundation for a long period of dynastic stability and prosperity that became commonly known as the K'ang-hsi–Ch'ien-lung age.

By courtesy of the Metropolitan Museum of Art, New York, Rogers Fund, 1942

K'ang-hsi, painting on silk by an unknown Chinese artist, 19th century. In the Metropolitan Museum of Art, New York.

Early life. The third son of Emperor Shun-chih, K'ang-hsi was born on May 4, 1654, in the Forbidden City of Peking, to Empress Hsiao-k'ang, daughter of Tulai, a famous Manchu general from the prestigious T'ung clan. (The Manchus were a people of Tungusic stock living in Manchuria who conquered China and established the Ch'ing dynasty in 1644.) Upon Shun-chih's sudden death from smallpox at age 23, in February 1661, K'ang-hsi was immediately raised to the Imperial throne over his five brothers, who had been born to mothers lower in birth than his; and the era's name was changed starting the next Chinese New Year, from Shun-chih (Manchu: Ijishūn Dasan), or Obedient Rule, to K'ang-hsi (Elhe Taifin), or Peaceful Harmony.

Elevation to the Imperial throne

Because the new emperor was not yet quite seven years old, his government was first administered by Sonin, Suksaha, Ebilun, and Oboi—four conservative Manchu courtiers from the preceding reign. One of the first political acts of the four Imperial advisers was to replace the so-called Thirteen Offices (Shih san ya men) with a Nei-wu-fu (Dorgi yamun), or Office of Household. The Thirteen Offices, all organized solely by Chinese eunuchs, had been the abomination of the Manchus ever since they were introduced by the late emperor, to handle affairs of the Imperial household, patterned after an elaborate model that had existed under the preceding dynasty—the Chinese Ming. Now the private sector of the Emperor's life would be run by his personal Manchu bond servants who staffed the newly created Office of Household. Thus, the Manchu Ch'ing successfully prevented court eunuchs from meddling with politics, in sharp contrast to many other dynasties, the Ming in particular, that had recurrently let eunuchs gain access to actual power, often with most disastrous results.

In 1667, advised by Sonin and other ministers, K'ang-hsi began attending to affairs of state at the age of 13, as his father had before him. He ruled only in name, however; the real power was still firmly in the hands of the four advisers. Sonin soon died, and Oboi became a virtual dictator, putting Suksaha to death for an alleged crime and cowing Ebilun into submission. Finally, in 1669, Oboi and Ebilun were eliminated by K'ang-hsi, who must have enlisted the help of his grandmother, the Grand Empress Dowager, and of Hsiung Tz'u-li, his Chinese tutor. The actual arrest of Oboi was made in the audience room by young wrestlers, who jumped upon the powerful minister from their hiding place behind the throne. With this coup the 15-year-old Emperor proved to the public that he was their real master.

Potential rivals in South China

Acquisition of actual power. Once in power, K'ang-hsi was confronted by the grave problem of what to do with three vassal kings in South China. The three kings—Wu San-kuei of Yunnan, Shang K'o-hsi of Kwangtung, and Keng Chi-mao (after his death succeeded by his son Keng Ching-chung) of Fukien—were among the Chinese warlords who, with their powerful firearms, had been welcomed into the Manchu camp even before the Manchu conquest of China in 1644. When Shun-chih had entered Peking in that year, the rest of China was still in the hands of the remaining Ming forces or of roving bands of robbers. Having made a major contribution toward subduing them, the three warlords had been created kings and had stayed in South China with their private armies. It was inevitable that the three vassal kings, with their virtual immunity, should become a menace to the Peking government.

A chance to improve the situation came in 1673, when Shang K'o-hsi offered to give up the command of his army and retire to Manchuria, his birthplace. The offer was so promptly accepted by K'ang-hsi that the other two kings were forced to make the same offer, if only for courtesy's sake. Now the chief issue at the Imperial council in Peking became whether or not to challenge the formidable military strength of Wu San-kuei, the very person responsible for the 1644 Manchu take-over of the capital. The young Emperor again showed his resoluteness by deciding in favour of trying to deprive Wu of his army, arguing that the three kings were sure to eventually rebel against Peking and that it would be better to forestall them by taking advantage of this opportunity. A shocked Wu immediately went to war against the Manchus, starting the so-called Revolt of the Three Feudatories. Initial reverses suffered by the Imperial forces prompted King Burni of the Chakhar Mongols—the supreme royal tribe until the Manchu conquest of Inner Mongolia in 1635—to revolt also against the Ch'ing and sent ripples of political unease among other East Asian countries. K'ang-hsi's youthful energy and genius in military strategy finally triumphed over the senility of Wu San-kuei, who never even attempted to march on Peking but died soon after styling himself emperor. The Ch'ing army entered the city of Kunming, in Yunnan, in 1681; the war was over, and the dynasty was saved.

After eliminating the three kings, K'ang-hsi turned his attention to the Cheng regime on Taiwan. Originally from Fukien, the Cheng family had been for generations a sea power that monopolized trade across the vast expanse of the China seas; Gen. Cheng Ch'eng-kung (the Koxinga), who refused to submit to the Manchus, had moved his headquarters to Taiwan, which he took from the Dutch in 1662, and his descendants had continued resistance to the Ch'ing from there. Lack of naval power prevented the Ch'ing from mounting an effective attack on Taiwan. Their only strategy, adopted in 1661, though to little avail, was the forced relocation of the inhabitants of coastal areas deeper into the continent, so that the Cheng might be isolated. In 1683 an internal strife in the Cheng family gave K'ang-hsi a chance to order his troops across the Taiwan Straits. The Cheng surrendered, and Taiwan was incorporated into the province of Fukien.

Enemies in the north

With China securely under his power, K'ang-hsi next turned to face his enemies in the North. The Russians in Siberia, who had reached the Amur Valley in the mid-17th century, had been expelled from their fortresses of Albazin and Nerchinsk by the Ch'ing army before K'ang-hsi's reign. But the Russians restored the two fortresses and were busily building many more in that region, and K'ang-hsi prepared to deal them a blow. In 1685 Ch'ing forces attacked Albazin and captured it in a few days. As soon as they withdrew, however, the Russians manned the fortress again. Infuriated, K'ang-hsi ordered another expedition to Albazin the following year and began a protracted siege. Concurrent diplomatic negotiations between K'ang-hsi and Tsar Peter the Great of Russia led to the signing of the Treaty of Nerchinsk (1689). Drafted under the pressure of a superior number of Ch'ing troops sent into Nerchinsk by K'ang-hsi, the treaty drew the Sino-Russian borderline along the Gorbitsa, an outer tributary of the Amur, and the Stanovoy Range, thereby leaving the Amur Valley and Manchuria, homeland of the dynasty, in the hands of the Ch'ing.

Next K'ang-hsi brought Outer Mongolia under his power. Dga'-ldan Boshogtu Khan of the Dzungar Oyrats, a nomadic people who lived to the west of Outer Mongolia and to the north of the T'ien Shan, was an ambitious ruler who had conquered east Turkistan and then invaded the territories of the Outer Mongolian Khalkhas. The Khalkha Mongols fled in great numbers to Inner Mongolia, seeking protection under the Ch'ing. In 1691 K'ang-hsi met with representatives of the Khalkha tribes at Doloon Nuur (later called To-lun Hsein) in Inner Mongolia, where he received their formal pledge of allegiance. In 1696 he embarked on a daring and extremely dangerous venture—a military expedition to Outer Mongolia across the scorched desert of Gobi. In personal command of the middle corps, K'ang-hsi managed to overcome hunger and thirst in the hostile terrain and annihilated the Dzungars at Juun Modu, east of the present-day Ulaanbaatar. Dga'-ldan committed suicide the following year at his hideout in the Altai Mountains. With the return of the Khalkhas to their homeland, Outer Mongolia became an integral part of the Ch'ing empire.

Two decades of peace between the Ch'ing and the Dzungars ensued, until the latter invaded Tibet and took Lhasa in 1717. Mindful of the spiritual sway of Tibet's Dalai Lama over the Mongols, K'ang-hsi sent his army into Tibet and expelled the Dzungars from Lhasa in 1720, thus incorporating the country into his empire. Hoping to check the Dzungar power, K'ang-hsi sent, in 1712, an embassy to the Torguts, or Volga Kalmyks, who had migrated to southern Russia in the earlier half of the 17th century. When the Manchu envoys, who travelled the length of Siberia back and forth by its myriad waterways, returned to Peking three years later, one of them, Tulishen, wrote a detailed account of the journey under the title of *I yü lu* (Record of *Strange Regions*).

Personal attention to detail

Administration of the empire. K'ang-hsi was an accomplished military leader who was endowed with exceptional physical strength and with skill in archery; he poured his inexhaustible energy into his daily administrative duties. Under the traditional Imperial system of China, nothing in the empire was too small to come under the personal scrutiny of the Emperor. K'ang-hsi read all the reports and memorandums presented to him, meticulously correcting even the smallest scribal errors, and he often boasted that he routinely took care of all the documents, even in wartime, when 300–400 arrived daily.

The Huang Ho was one of the subjects that commanded K'ang-hsi's attention. Long neglected, the river repeatedly flooded the land near where it joined the Huai Ho, causing great damage to northern Kiangsu. In 1677

K'ang-hsi appointed Chin Fu superintendent of riparian works; in 1683 Chin finished embanking and dredging to stabilize the flow of the river. At the same time, the Grand Canal, the important arterial waterway that connected the Huang Ho with the lower Yangtze River, was repaired to allow the smooth flow of large quantities of rice needed in the north from the rich granaries of the south. To inspect the results of the works and to acquaint himself with cultural and economic leaders of the wealthy south, K'ang-hsi travelled to the lower reaches of the Yangtze six times between 1684 and 1707, financing each journey with his own private funds. In his private life, K'ang-hsi was frugal in court expenditures, waited upon only by a small number of court ladies and eunuchs. He never raised taxes, even in wartime; in fact, taxes were reduced or exempted many times during his reign; in three years starting in 1711, all provinces received a tax relief that totalled more than 30,000,000 taels. In 1711 K'ang-hsi declared that from then on the number of tax-paying adults should be perpetually fixed at the present level for head-tax purposes, permanently exempting the balance of future population growth from taxation.

After the conquest of Taiwan, K'ang-hsi lifted restrictions on coastal trade and opened four ports, including Canton, to foreign ships. Foreigners brought silver to China to purchase such Chinese products as tea, silk, and chinawares. Such activities and internal peace stimulated a tremendous industrial growth, particularly in the lower reaches of the Yangtze.

Fondness for learning

K'ang-hsi was very fond of learning. His avidity for study steadily increased with his age, to a degree that even when ill from overwork he did not stop reading books. In 1677 he opened a small study hall called the Nan shu-fang in the Forbidden City, where he engaged himself in lively discussions on philosophical and historical topics with the leading scholars of his time. His inclination toward the scholar Chu Hsi's philosophy and arduous emulation of its Confucianist ideals were a most effective means for the Manchu Ch'ing to gain confidence of the Chinese majority.

Besides the traditional civil service examinations for recruiting Chinese officials educated in Confucian disciplines, K'ang-hsi opened in 1678 a special channel through which persons with exceptional talents in learning and writing were admitted into his service on recommendation. The 50 men who thus won appointments at the Hanlin Academy, the famous scholar Chu I-tsun among them, worked on compilation of the *Ming-shih*, an official history of the Ming dynasty. Other great books commissioned by K'ang-hsi included the dictionary of Chinese characters, *K'ang-hsi tzu-tien*, listing about 42,000 characters (1716); the rhyming dictionary of Chinese compounds, *P'ei wen yün-fu* (1711); and the subject-matter encyclopaedia, *Yuan-chien lei-han* (1710). Another great encyclopaedia, the *Ku-chin t'u-shu chi-ch'eng*, which was to consist of 10,000 chapters, was also started in K'ang-hsi's reign.

In the field of popular education, K'ang-hsi issued, in 1669, an Imperial precept consisting of 16 articles, a revision and expansion from the six articles issued by Shun-chih, in which he detailed practicalities of village life. The precept, expanded further by his son and successor, Yung-cheng, offered moral guidelines to the Chinese peasantry for almost two and a half centuries.

Always eager to absorb new knowledge and technologies from Europe, K'ang-hsi employed many Jesuit missionaries. He learned geometry from Ferdinand Verbiest, who became deputy director of the Imperial Observatory and compiled the official calendar of the empire. Verbiest was also responsible for the production of cannons that proved effective against the three rebellious kings and the Dzungars. Jean-François Gerbillon and Joachim Bouvet taught K'ang-hsi mathematics. K'ang-hsi ordered Pierre Jartoux, Jean-Baptiste Régis, and others to compile an accurate atlas of the empire; after long and laborious trigonometric surveys that covered every corner of the empire, starting in 1708, the atlas *Huang yü ch'üan lan t'u* was completed in 1717. The famous *Nouvel Atlas de la Chine, de la Tartarie chinoise et du Thibet* of Jean-Baptiste Bourguignon d'Anville is a French version of this original. European painting also fascinated K'ang-hsi. Gio Ghirardini, an Italian layman brought by Bouvet, and Giuseppe Castiglione were the Emperor's favourite court artists, who influenced Chinese painting with their European-style perspective drawing and other activities.

The Rites Controversy

Those cultural contributions by the Jesuits endeared Catholicism to K'ang-hsi, who gave official permission for its propagation in 1692 and later gave French missionaries a residence within the Imperial City and built a church for them in Peking in gratitude for curing him of malaria. His sympathy attracted to China more missionaries from such orders as the Dominicans, the Franciscans, and the Augustinians. In contrast to the Jesuits, who were lenient with such traditional Chinese rites as ancestor worship and state cult of Confucius and Heaven, those newcomers condemned the traditions as superstitions incompatible with Christian faith. The Rites Controversy raged on until 1704, when Pope Clement XI issued a decree forbidding Chinese Catholics to take part in such rites. Angered by this interference in what he considered his exclusive domain, K'ang-hsi ordered the Portuguese to arrest Carlo Tommaso Maillard de Tournon, an apostolic legate to China carrying the papal decree; and in 1706 he expelled missionaries who would not adhere to the Jesuit line. In 1720 Carlo Ambrogio Mezzabarba, another legate, was ordered back home by the Emperor.

K'ang-hsi had three empresses—one a granddaughter of Sonin of the Hesheri clan, another a daughter of Ebilun of the Niohuru clan, and the third a granddaughter of Tulai of the T'ung clan—in addition to many concubines; they bore him 35 sons in all. He nominated the second son, Yin-jeng, crown prince in 1675, at the age of little more than a year and a half; this was against the Manchu tradition of giving all sons equal rights of succession, and it resulted in vicious fights among K'ang-hsi's sons. The hapless Yin-jeng was deposed in 1708, restored in 1709, and again deposed in 1712, this time permanently. Terribly hurt by the experience, K'ang-hsi and his successors never again tried to nominate a crown prince.

At the Chinese New Year of 1722, K'ang-hsi celebrated his long and prosperous reign by inviting many elders to a great banquet at the court. That winter he fell ill while staying at the Imperial villa of Ch'ang ch'un yüan, in the northwestern outskirts of Peking, and he died on December 20. The next year he was buried at Ma lan yü, to the northeast of Peking, in a mausoleum called the Ching ling. His throne was taken over by his fourth son, Yin-chen (Yung-cheng).

BIBLIOGRAPHY. A.W. HUMMEL (ed.), *Eminent Chinese of the Ch'ing Period, 1644–1912*, 2 vol. (1944–45), a biographical dictionary with useful entries on K'ang-hsi and his contemporaries; J.D. SPENCE, *Ts'ao Yin and the K'ang-hsi Emperor: Bondservant and Master* (1966), nature of K'ang-hsi's reign described through his relationship with a bondservant in the Imperial household; J. BOUVET, *Portrait historique de l'Empereur de la Chine* (1697), K'ang-hsi as seen through the eyes of a Jesuit missionary in his service.

(N.Ka.)

K'ang Yu-wei

K'ang Yu-wei (in Pin-yin romanization Kang Yu-wei), an original thinker, influential writer, and the leader of the Chinese Reform Movement, which reached its zenith in 1898, was a key figure in the intellectual development of modern China.

Born on March 19, 1858, he came from a scholarly gentry family in the district of Nan-hai in Kwangtung Province. His teacher imbued him with the Confucian ideal of service to society, and his study of Buddhism impressed him with its spirit of compassion. He rebelled against convention, Neo-Confucian authoritarianism, and the demands of the civil service examination system. After reading about the outside world, he came to admire Western civilization. In the 1880s he began to conceive some of his basic ideas: ideas of historical progress, social equality, a world government, and the nature of the universe.

K'ang Yu-wei, 1905.
By courtesy of the Library of Congress, Washington, D.C,

Efforts at reform

K'ang's first venture in social reform was in 1883, when he tried to abolish in his village the custom of foot-binding imposed on women. The decay of the Ch'ing Empire (1644–1911) prompted K'ang and other concerned Chinese to urge fundamental institutional reforms. After his plans for the salvation of China—submitted in 1888 to the Ch'ing court—were ignored, K'ang set out to convert the educated class to his views and to arouse the people from their lethargy. In 1890 he opened a school in Canton to teach new learning. Assisted by his students, among whom was Liang Ch'i-ch'ao, who collaborated in his reform movement, he wrote *The Forged Classics* (1891), which reveals that the Confucian classics held sacrosanct as bases of the state cult had been tampered with in the Han period (206 BC–AD 220). This book was followed by *Confucius as a Reformer* (1897), which expounded K'ang's belief that Confucius was concerned with contemporary problems and stood for change and that the progress of mankind was inevitable. His interpretation of Confucian teachings and researches on ancient texts later inspired modern scholarship in the reappraisal of China's past, although critics have charged that he invoked Confucius to further his reform aims and was undermining the established way of life.

When China was defeated by Japan in 1895, K'ang mobilized hundreds of provincial graduates then in Peking to protest against the humiliating peace terms and to petition for far-reaching reforms to strengthen the empire. To arouse the people to the dangers confronting China, he and his associates published newspapers and founded the Society for the Study of National Strengthening, the archetype of political parties in modern China. The society was suppressed in 1896.

In 1898, when foreign powers threatened to partition China, K'ang and his followers suggested an alliance with Britain and Japan to check Russia's advance and insisted that only institutional reforms could save China. He urged the clearing of channels for the expression of public opinion, the convocation of assemblies, and even the acceptance of popular sovereignty and the separation of state powers, and he organized the Society to Preserve the Nation to marshal support. Finally, he prevailed upon the emperor Kuang-hsii to launch the reform program. Among the many measures that were promulgated were the streamlining of the government, strengthening of the armed forces, new standards in the civil service examination system, the development of commerce and industry, the promotion of local self-government, and the opening of Peking University and modern schools.

Years of exile

The reform measures were annulled, however, when the dowager empress Tz'u-hsi (1835–1908) reasserted control. The Emperor was placed in confinement, six of the reform leaders, including K'ang's brother, were executed, and scores were arrested. K'ang and Liang Ch'i-ch'ao escaped to Japan. Unable to persuade the Japanese and British governments to intervene for the Emperor, K'ang founded the China Reform Association (popularly known as the Save the Emperor Association and in 1907 renamed the Constitutional Party) in Victoria, British Columbia, to carry on his plans.

After the failure of the revolts instigated by the reformers in 1900 in Anhwei and Hupeh provinces to restore the Emperor, K'ang resumed his writing in exile. His most significant work completed at this time was *The Great Commonwealth*, in which he envisaged a utopian world attainable through successive stages of human development, a world where the barriers of race, religion, state, class, sex, and family would be removed and where there would be an egalitarian, communal society under a universal government.

K'ang emerged from his retreat in 1903. To help the overseas Chinese and to unite them in a common effort, he and his colleagues founded an international business firm and established schools and newspapers. These activities, conducted in the United States, Mexico, Japan, and Southeast Asia, brought them into sharp competition with the Chinese revolutionists.

During his exile, K'ang travelled extensively, circling the globe four times. His stay in Europe and his study of Western history moved him to shun the violence and destructiveness of revolution as means of political change, and he proposed as an alternative course the promotion of science, technology, and industry to rebuild China.

Return to China

After his return in 1914 to a weak and troubled China, he was soon involved in the campaign to thwart the monarchical scheme of the Chinese statesman Yüan Shih k'ai (1859–1916). In 1917, in line with his idea of a constitutional monarchy to bridge the transition to a truly democratic republic, he participated in the abortive restoration of the Ch'ing ruler. In the years that followed, animated by the fear of a divided country, he opposed the South China government of the revolutionary leader Sun Uat-sen (1866–1925). He called for the preservation of the best of China's heritage and the establishment of a reformed Confucian Church to provide the people with spiritual guidance. Partisan writers have criticized him for holding to these views. In his later years, he renewed his philosophic reflections, completing his last book, *The Heavens*, in which he blends astronomy with his own metaphysical musing, a year before his death at Tsingtao on March 31, 1927.

Besides prolific writings on the Chinese classics, politics, and economics, K'ang also left travel accounts and an anthology of his poems; he was also a famous calligrapher.

BIBLIOGRAPHY. JUNG-PANG LO, *K'ang Yu-wei: A Biography and a Symposium* (1967), the only book-length biography of K'ang Yu-wei available in a Western language—contains a translation of the *Chronological Autobiography* of K'ang, which narrates the events of his life to the end of 1898, and a discussion of his career and thought from 1898 to 1927; LAURENCE G. THOMPSON (ed. and trans.), *Ta T'ung Shu: (The) One-World Philosophy of K'ang yu-wei* (1958), a translation of K'ang's work (completed in 1902) in which he sets forth his vision of an attainable utopian world —includes a brief biographical introduction; KUNG-CHUAN HSIAO, "In and Out of Utopia: K'ang Yu-wei's Social Thought," *Chung Chi Journal*, 7:1–18 (1967), 7:101–149 (1968), and 8:1–52 (1968), the best analyses of K'ang's ideas and their impact on modern China published in a Western language.

(J.-p.L.)

Kansas

Location and general character

Lying amid the westward-rising landscapes of the Great Plains of the North American continent, Kansas in 1861 became the 34th state of the United States. In that year the capital was located in Topeka by popular vote, outpolling nearby Lawrence by some 2,700 ballots. The state's 82,264 square miles (213,063 square kilometres) are bounded by Nebraska on the north, Missouri on the east, Oklahoma on the south, and Colorado on the west. The state's name was derived from that of the Kansa, or Kaw, Indians.

Whatever the Kansas slogan, "Midway U.S.A.," lacks in inspiration, it makes up for in aptness. The geographical centre of the 48 coterminous states of the nation is marked by a limestone shaft and a flag located in a pasture near Lebanon, Kansas, close to the Nebraska border. Some 40 miles (65 kilometres) to the south is the magnetic, or geodetic, centre of the terrestrial mass of North America.

The more than 2,300,000 Kansans recorded in the 1980 census live in what was once looked upon as the agricultural heartland of the nation, After 1952, however, industry contributed more to the economy than the vast wheat fields and cattle ranches of the plains. Wichita, the state's largest city, is known locally as the "Air Capital of the World" because it produces more small aircraft than any other city. (For information on related topics, see the articles UNITED STATES OF AMERICA; UNITED STATES, HISTORY OF THE; NORTH AMERICA; and GREAT PLAINS.)

THE HISTORY OF KANSAS

Indians, explorers, and settlers. Archaeological exploration has uncovered evidence of Indian cultures that existed in Kansas for many centuries before the Europeans settled on the land. From about 1200 to 1500 there had been a thriving agricultural society in the area of the Republican and Blue rivers.

The first known European explorers were Spaniards under Francisco Vázquez de Coronado, who in 1541 rode northward from Mexico seeking the gold of the legendary Seven Cities of Cibola. Juan de Padilla, a priest with the expedition, founded the first mission in the territory, possibly north of present-day Wichita. The territory was claimed for France in 1682 by Robert Cavelier, sieur de La Salle. During the 18th century French fur traders had a flourishing exchange with the Indians in what is now the northeastern part of the state.

The region passed to the United States as a part of the Louisiana Purchase in 1803. The explorer Zebulon Montgomery Pike passed through Kansas in 1806 and described it as the "Great American Desert"—a false image that still persists, to the dismay of the state's boosters. Kansas was thoroughly explored during the following decades, but westward-bound settlers and miners passed through it without staying.

From 1830 to 1854 Kansas was in an area designated as Indian Territory, actually an area in which to relocate Indian tribes who occupied lands wanted by whites. The Kansas–Nebraska Act of 1854, however, created two territories and opened both to settlement. It also provided that residents could determine whether their future states would be free or slave. The rush began, and Kansas became a major breeding ground for the U.S. Civil War as North and South each attempted to send the most settlers into the new territory. Most early settlement was near the eastern border, and free staters were harassed constantly by "Border Ruffians" from Missouri. One notable incident was the sacking of Lawrence by Southern guerrillas in 1856. The abolitionist John Brown, with his sons and a few other men, retaliated by dragging five of their proslavery neighbours from their homes and killing them. Proslavery forces attempting to avenge this massacre were captured by Brown, who became a hero to the Northern sympathizers. Hundreds of such incidents won the territory the name "Bleeding Kansas."

Statehood. Kansas entered the Union as a free state in 1861. Before and after the Civil War, sporadic fighting occurred between the settlers and the Indians. In 1867 a peace treaty was signed in which the Indians agreed to sell their land; in return, the United States agreed to build homes for them in what is now Oklahoma and to provide money, food, and clothing. The U.S. Congress did not honour the treaty, and when the Indians returned they found their land occupied by white settlers. Further sporadic battles continued until the last Indian raid, in 1878.

Early settlers in wooded eastern Kansas lived in log cabins, but in the west they had only dugouts or houses made of buffalo-grass sod. The unpredictable weather, the recurring Indian raids, the droughts and dust storms, and the grasshopper invasions discouraged many early settlers. One of the heroes of this era was William Mathewson, known as the original Buffalo Bill, who hunted buffalo for the settlers all of one winter without pay, providing meat by the wagonloads. The coming of the railroads in the late 1860s and the 1870s made first one village and then another into boisterous cow towns. Texas cattlemen drove herds northward to Caldwell, Wichita, Dodge City, Ellsworth, Newton, and Abilene to reach the railhead. Although this development brought prosperity to Kansas and created a persistent image, the cow-town era lasted less than a decade.

The most important event in Kansas' agricultural history may have been the arrival of the Mennonites in 1874, who brought with them trunks full of hand-selected grains of Turkey Red wheat. This excellent strain provided the basis of the abundant crops that became an important part of the Kansas economy. Many of the Mennonites' descendants remain as prosperous farmers.

Expansion of agriculture

By about 1890 most of the land was occupied, and Kansans settled into a peaceful life dominated by agriculture. World War I produced a great demand for food, and more and more virgin prairie was plowed and put into production, which led to temporary prosperity but contributed directly to the terrible dust storms that devastated the state in the 1930s. World War II contributed to Kansas' growing eminence in aircraft and brought many people from Oklahoma and Arkansas to work in Wichita's aircraft plants.

THE NATURAL AND HUMAN LANDSCAPE

The natural environment. ***Surface features.*** Kansas has been erroneously characterized as a featureless plain, but its topography, while rarely spectacular, is varied. The land rises slowly but steadily from 700 feet (200 metres) above sea level in the southeast to 4,039 feet near the Colorado border. The far western section consists of high plains with few natural trees and appears to be flat and endless. Actually these seemingly flat plains are creased with shallow gullies, called draws, the product of millennia of erosion. In this part of Kansas are some of the state's most spectacular geological formations. Castle Rock, south of Quinter, consists of chalk spires rising high above the level plains. Monument Rocks, a few miles to the west, resemble sphinxes. Near Jetmore is Horse Thief Canyon, a miniature of the Grand Canyon.

Millions of years ago, much of Kansas was the floor of an inland sea. The land was built up by the deposit of soil and vegetable matter from streams feeding the sea. This residual soil is some of the most fertile in the world, and in it prehistoric fossils of great importance have been found.

Under irrigation, southwestern Kansas in recent years has produced truck crops and sugar beets. Northeastern Kansas, once covered by the glacier that crept over most of the northern United States, is hilly and timbered, with many creeks and springs. The southeast, lying near the foothills of the Ozark Mountains, is rough and covered in parts with scrub oak. In south central Kansas, near Medicine Lodge, are the Gypsum Hills, which resemble the mesas of the Southwest and are named for the gypsum found in them. In east central Kansas, the Flint Hills stretch from north to south; gentle, rolling, largely treeless, and covered with bluestem grass, they provide one of the world's best natural grazing regions. The principal rivers are the Kansas and the Arkansas. Tributaries of the Kansas are the Blue, the Republican, the Solomon, the Saline, and the Smoky Hill, all in northern Kansas. The Arkansas flows into the state from Colorado and winds through southwestern and south central Kansas, continuing through Oklahoma and Arkansas to the Mississippi. Tributaries of the Arkansas are the Cimarron, the Verdigris, Neosho (Grand), and the Marais des Cygnes.

Climate. The climate of Kansas is temperate but continental, with great extremes between summer and winter temperatures but few long periods of extreme hot or cold. The annual average temperature is 55° F (13° C). The growing season ranges from mid-April to mid-September. Normal annual rainfall ranges from less than 20 inches

(500 millimetres) in the west to more than 40 inches in the southeast.

Vegetation and animal life. Buffalo grass is native in the west and central areas of the state, bluestem around the Flint Hills, and bluegrass in the east. Wild flowers of many kinds are to be found in all parts of the state, and sunflowers grow in profusion (Kansas is popularly known as the Sunflower State). The cottonwood grows throughout Kansas, while in the northeast there are many oak, walnut, and maple trees, as well as cedar and elm. Western Kansas abounds in quail, prairie chicken, and pheasant. Deer, once almost extinct, were protected by law for many years and have multiplied to the degree that hunting is again allowed in season. The buffalo that once proliferated across the plains are to be found only in parks and zoos.

Rural and urban life

Human imprints. ***Patterns of settlement.*** Most western Kansas farms or ranches are large, covering not less than one section (a square mile, or 640 acres [260 hectares]) of land, though a farmer's holdings may not always be contiguous. Eastern Kansas began with small farms, some of no more than 40 acres; but these have grown in size. A Kansas law forbids other than family corporations for farming purposes. Most of the small towns are modern and well-kept, with paved streets and full utilities. Many of the small cities, especially in the west, present unexpected cultural and commercial resources, perhaps because they often lie far apart and draw from large trade territories. In the east the cities are older, closer together, and generally less progressive, though most of them are attractive, with broad, well-shaded residential streets and adequate downtown shopping facilities.

Wichita, the largest city, is characterized by the state's largest buildings, biggest industries, and most venturesome businesses. In Topeka, where state government once was the largest industry, more people now are employed in services. Kansas City, Kansas, merges with its larger neighbour, Kansas City, Missouri, and contains a significant part of the industrial complex of that region. Leavenworth, the state's oldest city, is built around institutions, including an army post at Ft. Leavenworth, a federal prison, a state penitentiary, and a veterans' hospital. Lawrence, home of the state's largest university, depends heavily on the school for its economy, though the city has been working aggressively and successfully for industry since the 1960s. Most of the other cities depend primarily on farm trade and agriculturally related business.

Regional spirit. Kansas suffered during most of its history from two kinds of regionalism: one that pits rural against city dwellers; the other, the east against the west. The two are related in some ways, for none of the state's principal cities is in the western half of the state. More thinly populated than the east, western Kansas has always feared and fought eastern domination, while the east often has ignored the west. Wichita, one of Kansas' three metropolitan areas, contains approximately one-seventh of the state's population. The Kansas City–Topeka area of northeastern Kansas, containing two metropolitan areas, is less populous but still is the centre of much industry. Rivalry has existed between these two urban areas, but it is diminishing gradually as the cities gain more representation in the state legislature. People from the rural areas, mostly farmers, ranchers, and owners of small businesses, as well as residents from the smaller towns, have tended to distrust the cities, often bringing about an impasse in the state legislature.

THE PEOPLE OF KANSAS

Migrations past and present

Ethnic groups. Kansas' early settlers were principally antislavery New Englanders of Anglo-Saxon stock. After the Civil War and with the building of the railroads, many central Europeans were attracted by the promise of jobs laying track and of free land when the jobs were finished. Small communities populated by citizens of predominantly Russian, Bohemian, German, or Scandinavian ancestry still dot the state. The original languages have largely disappeared, though here and there church services are still conducted in German or Swedish, and a few communities hold festivals each year at which the old folkways, foods, and languages are featured. During World War II there was an influx of servicemen and aircraft workers, many of whom remained. The state is largely Protestant, though there are some small Roman Catholic communities. Virtually every sect is represented in the state, including such rare ones as the Amish, Dunkard Brethren, Mennonite, and Eastern Orthodox.

Contemporary demography. Because of insufficient employment opportunities, Kansas loses a considerable number of its young people to other states, but this loss is almost balanced by in-migration. The birth rate, however, produces a slight natural increase in population in most years. The most conspicuous demographic trend is from the farms to the cities. In 1980 about 67 percent of the population was living in urban areas and 33 percent in rural areas. As further technological advances in farming are made and individual landholdings increase in size, this trend undoubtedly will continue.

THE STATE'S ECONOMY

Components. Both agriculture and manufacturing contribute significantly to Kansas' economy—the former contributing many of the raw materials for the latter. The production of the some 77,000 farms and ranches places Kansas first among the U.S. states in wheat, first in sorghum grains, and fourth in wild hay; in beef, it ranks fifth; in hogs, eighth. Some 4,300 manufacturing and processing plants produce everything from airplanes to zinc castings. Wichita is the world's largest producer of camping gear; it also manufactures heating and air-conditioning equipment, snowmobiles, and a variety of other products. In addition to ranking first in the world in production of small private aircraft, Wichita also is an important manufacturing centre for military aircraft. Other plants in the state turn out baby foods, pet foods, prefabricated houses, mobile homes, greeting cards, tires, paint, and dishwashers. Grain milling, though not as important as it once was, still is a major industry, and the meatpacking and processing industry is growing. Garvey Elevator, near Wichita, is the world's largest single-unit grain elevator.

Kansas has large mineral resources, a good labour force, a healthy retail trade, ample electrical power, plenty of water, and a central location. Kansas is among the 20 top mineral-producing states. It is rich in oil and gas, helium, portland cement, stone, clay and clay products, sand, salt, gravel, zinc, bituminous coal, and lead. There is an almost unlimited supply of chalk. Many plants produce gasoline, oil, grease, and paving materials from the oil pumped from Kansas wells.

Despite large oil production, Kansas has imposed only a small severance tax (laid at the time of severance, or extraction from the ground) on oil or gas for the purposes of conservation and pollution prevention. It has a tax rate on manufacturing plants that compares favourably with that of other states. Most Kansas cities issue revenue bonds to encourage new industry. The state has a right-to-work law that forbids compulsory unionism.

Two long-established army posts have contributed significantly to the state's economy. Ft. Leavenworth, with its renowned Command and General Staff College, dates from 1827. The fort was a major outpost in the early Indian wars and during the Civil War and has offered the most sophisticated training to field-grade officers for many years. Ft. Riley, near Junction City, was established in 1853 and was also an Indian outpost. In the 20th century it has been an important infantry training centre. During World War II air force bases were established in Topeka and Wichita.

Transportation. Kansas has an excellent system of railroads for east–west transport but, except in the east, has less adequate north–south lines. The same may be said of its highways, with the best ones generally carrying east–west traffic. An exception is the state's single toll road, the Kansas Turnpike, which runs southwesterly from Kansas City to the Oklahoma line south of Wichita. Although Kansas has more than 200 airports and is served by eight airlines, the only major airport with transcontinental service is in Wichita.

ADMINISTRATION AND SOCIAL CONDITIONS

Government and politics. Kansans elect a governor, lieutenant governor, attorney general, and secretary of state, while other state officers are appointed. Its legislature comprises 125 representatives and 40 senators, the former elected for two-year terms, the latter for four-year terms. The legislature holds a session in odd-numbered years and meets for a briefer period in even-numbered years. Each of the 105 counties elects three commissioners, a county attorney, a treasurer, and several other officers. Judges of the 29 judicial districts are elected, but the seven justices of the Supreme Court are appointed by the governor, subject to senate approval, from a panel presented by a Supreme Court nominating commission. The justices must subject themselves to the approval of the voters, but none has ever received a vote of "no confidence." If that were to happen, the justice would be replaced by appointment.

Innovations in government and law

The first legislative council in the United States was inaugurated in Kansas in the 1930s. It was an interim body designed to work between legislative sessions at analyzing and drafting laws. Several other states later adopted legislative councils. The 1969 legislature provided for prefiling of bills between sessions, a change that persuaded the legislature that the council was no longer necessary. It was replaced in the 1971 session by the Legislative Coordinating Council, made up of the leadership of both houses.

In 1933 Kansas enacted a "cash basis law," which requires that no state money be expended until it has been raised and appropriated by the legislature. Bonds have been issued only for capital improvements, such as college dormitories and highways, in which case they are retired by user fees.

Kansas once was known as the most Republican state in the nation, but it now has a sizable Democratic minority, a growing independent vote, and small Conservative and Prohibition parties. The first legislature, in 1861, gave women the right to vote in school elections. In 1887, women's suffrage was extended to city and bond elections, and in that year the world's first woman mayor was elected in Argonia. The state constitution of 1861 granted women equal rights to own property and to have control of children. Universal suffrage was granted in Kansas in 1912.

The Farmers' Alliance and the Populist Party both had their origin in Kansas, and for a few years in the 1890s they played an important part in the politics of the Midwest. Kansas pioneered the direct primary, and a Kansas senator introduced the resolution in the U.S. Congress that put direct election of U.S. senators into the federal Constitution.

Kansas was the first state to adopt the constitutional prohibition of alcoholic beverages. The prohibitory amendment was added to the state constitution in 1880 and was not repealed until 1948. Liquor is sold only in package stores operated by licensed individuals under strict state control. An attempt to amend the constitution to permit sale of liquor by the drink failed by a narrow margin in 1970. Drinkers have little difficulty slaking their thirst, however, in a multitude of "private clubs" that operate under state license.

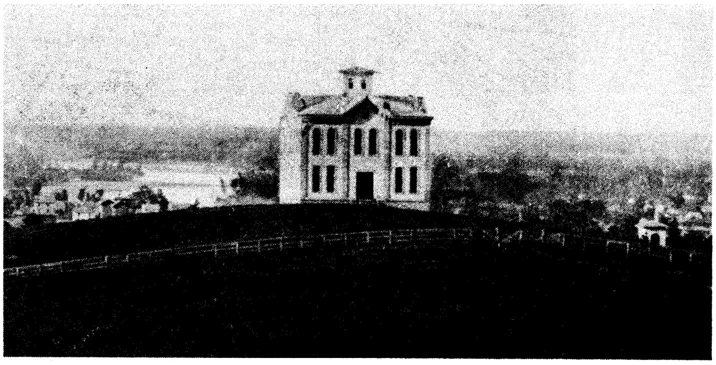
The social milieu. ***Education.*** A landmark civil rights case of the 20th century, ***Brown*** v. ***Board of Education,*** originated in Topeka in 1951, when the clergyman father of a nine-year-old black girl led her to the door of an all-white school. She was denied enrollment, and the decision that was handed down by the U.S. Supreme Court in 1954—basically stating that segregated, even "separate but equal," education is inherently unequal and must be eliminated with all due speed—became the basis for most of the civil rights decisions that have been applied to schools since that time.

In the mid-1960s Kansas abolished its office of state superintendent of public instruction and substituted the Department of Education headed by a commissioner and an elected state board of education. More than 400,000 pupils attend public schools in the some 300 school districts throughout the state. A number of two-year junior colleges are operated by the communities in which they are located.

Universities and colleges

Kansas has six state universities with a total enrollment of approximately 64,000. Fort Hays State University, Kansas State University of Pittsburg, and Emporia Kansas State University offer liberal arts degrees but specialize in training teachers. The University of Kansas is located in Lawrence, Kansas State University in Manhattan, and Wichita State University in Wichita. Kansas State, recognized as having one of the country's leading agricultural colleges, was the first land-grant college in the United States. The state's medical school is part of the University of Kansas College of Health Sciences and Hospital, with its campus at Kansas City. In 1971 the School of Medicine established a second campus at Wichita to expand its clinical teaching facilities. There are two law schools, one at the University of Kansas, the other at Washburn University of Topeka, a municipal school. In addition, there are 17 private four-year colleges in Kansas, all church-affiliated and all offering liberal arts degrees. They have a total enrollment of about 11,000 students.

By courtesy of the Library of Congress Washington DC

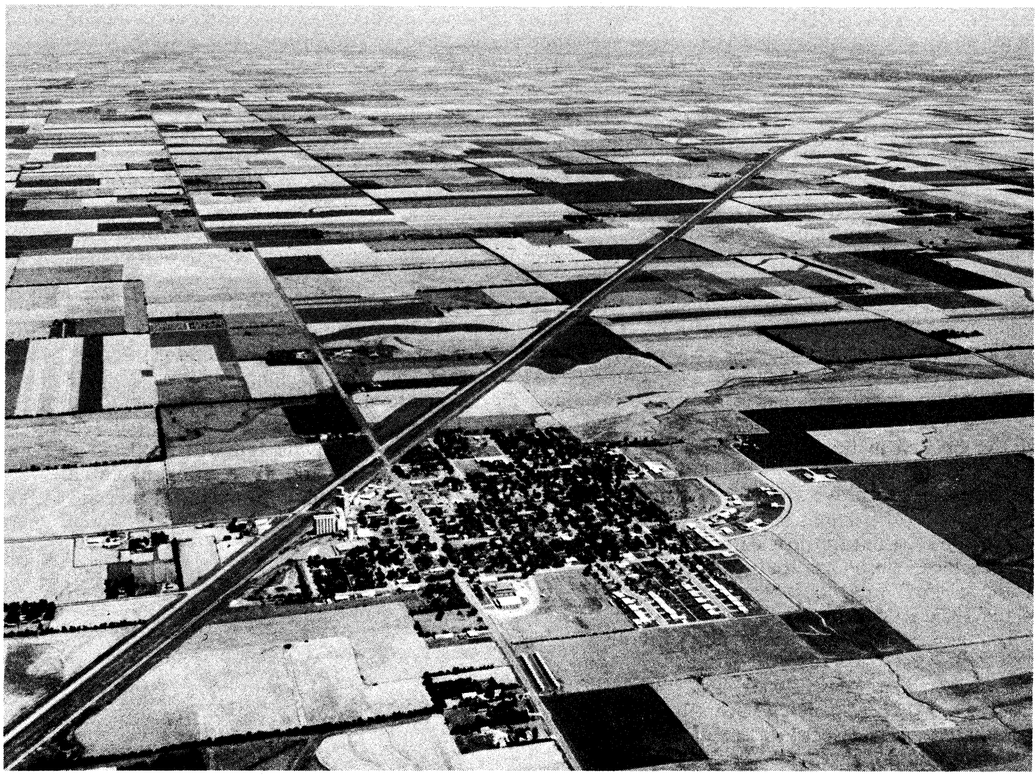


University of Kansas at Lawrence, photographed in 1867 by Alexander Gardner.

Health and welfare. The State Board of Health, consisting of appointive members, supervises the activities of one of Kansas' largest administrative departments, including doctors, dentists, sanitary engineers, hospital administrators, and veterinarians. Kansas operates three mental hospitals; they have been rated among the best in the nation and work in close conjunction with the renowned Menninger Foundation of Topeka. The health education division supplies information films on preventable diseases to more than 500,000 Kansans each year and distributes an equal number of pamphlets on subjects related to health. The board of health also has divisions of geriatrics and chronic diseases, hospital facilities, and services in public health nursing, nutrition, maternal and child health, mental hygiene, sanitation, and tuberculosis control.

Since 1862 Kansas has had some form of public assistance for the needy. The state welfare department, which is overseen by an appointed board, offers both financial assistance and special education to those requiring it. Special divisions care for crippled children, dependent children, the aged, and the blind. Vocational and rehabilitational services also are provided for the handicapped. There are schools for the mentally retarded at Parsons and at Winfield, and a children's home has operated at Atchison since 1855. There is a Kansas treatment centre for mentally disturbed children at Topeka and a reform school for girls at Beloit.

Standards of living. Kansas businesses and industries employ a civilian work force of more than 1,200,000 people. A state minimum wage was established in 1977. In 1970 Kansas had a per capita personal income just below the national average, but by 1980 it was well above the national average. The cost-of-living index for urban consumers was about equal to the national average in the late 1970s.



Flat farmland near Inman

' Grant Heilman

The state has a fair-housing law and a civil rights commission that hears grievances and attempts to mediate them.

CULTURAL LIFE AND INSTITUTIONS

The arts. The citizens of Kansas resent the suggestion that they live in a cultural desert, but the assertion is at least partially true. Most of the larger cities have amateur theatre groups, while Topeka and Wichita support symphony orchestras. The numerous colleges and universities in the state provide a concentration of art and music in many small communities that otherwise would have no comparable activities. In the sparsely populated areas of western Kansas, however, a large number of the small communities have few cultural institutions except a public library. Wichita, however, has two art galleries and a cultural and civic centre with two theatres, an exhibition hall, and a convention hall. Civic music series are sponsored in most of the larger cities of Kansas, but most professional road companies appear in theatrical productions only in Wichita. The extreme eastern areas of Kansas mainly rely upon Kansas City, Missouri, for cultural attractions. In the mid-1960s the Kansas Cultural Arts Commission was formed; funded by the state, it seeks to encourage the development of the arts, often providing seed money for communities or organizations that have the desire to develop cultural events. The University of Kansas has an outstanding museum of natural history and an art museum.

Traditional commemorations

Folk culture. In addition to an art gallery, the small community of Lindsborg has a biennial folk festival, the Svensk Hyllnings Fest, which honours the Swedish pioneers who settled the town. It features Swedish costumes, traditional food, folk dances, and displays of the arts and crafts of local artisans. Wilson has a Czech festival each year. Examples of bizarre sculpture are found in Lucas, where a self-taught artist, working in wet concrete, sculpted his own idea of the Garden of Eden and other biblical stories.

Communications media. Kansas has approximately 250 weekly and 50 daily newspapers. Among the latter, the largest are the *Wichita Eagle* and the *Wichita Beacon,* the *Topeka Daily Capital,* the *Hutchinson News,* and the *Salina Journal,* but there are many influential smaller dailies, including the *Emporia Gazette.* The state has radio stations in almost every town and television stations in all the major cities, including public television stations in Wichita and Topeka.

PROSPECTS

Kansas is adjusting to a decreasing demand for one of its major agricultural crops, wheat. It is making serious efforts to attract new industry and is developing its recreational and cultural facilities, hoping to make itself more appealing to the people additional industry would bring to the state. It is a state of pleasant towns and good living conditions, but it still has major problems. Kansas needs to streamline its governmental structure at every level to reduce duplication and cut unnecessary demands upon the taxpayers. Its tax structure, a patchwork that has grown to meet the demand for revenue over more than 100 years, needs a thorough overhaul, while service industries must be developed to employ farm labourers displaced by changing technologies and employment patterns. Legislative sessions have addressed themselves to these problems, and it seems probable that at least limited solutions can be reached.

BIBLIOGRAPHY. Sources of general information about the state are *The Kansas Directory* (biennial), containing a brief history, statistics, and information about state government, officials, and points of interest; and THE UNIVERSITY OF KANSAS, *Kansas Statistical Abstract* (annual). LOUISE BARRY, *The Beginning of the West: Annals of the Kansas Gateway to the American West, 1540–1854* (1972), is an excellent account of the earliest history to its opening for settlement. WILLIAM E. UNRAU, *The Kansa Indians: A History of the Wind People, 1673–1873* (1971), is a good account of the tribe from which Kansas got its name. One of the best histories of the turbulent days from the opening of the Kansas Territory to the U.S. Civil War is ERIC CORDER, *Prelude to Civil War: Kansas–Missouri, 1854–61* (1970). The colourful history of the trail that brought thousands of people to and through Kansas is outlined, with suggestions for additional reading, in JACK D. RITTENHOUSE, *The Santa Fe Trail: A Historical Bibliography* (1971). The story of the cattle trails and legendary cow towns is well recounted in ROBERT R. DYKSTRA, *The Cattle Towns* (1968). One of the most difficult periods in modern Kansas history, the drought and dust-bowl days, is admirably covered in FRANCIS W. SCHRUBEN, *Kansas in Turmoil, 1930–1936* (1969). One of the most readable brief histories of the state

from its earliest days to recent times is BLISS ISELY and WALTER M. RICHARDS, *The Kansas Story* (1961). Another that contains both history, good illustrations, and information about the geography and geology of the state is MOSE J. WHITSON and BARNEY SLAWSON, *Kansas: Its Geography, History and Government* (1968). A good account of the present condition and habits of Indians in Kansas is RUTH LANDES, *Prairie Potawatomi: Tradition and Ritual in the Twentieth Century* (1970). WILLIAM F. ZORNOW, *Kansas: A History of the Jayhawk State* (1957), is among the best of the modern overviews of the entire development of Kansas from earliest times to the recent era. For an interesting picture of the legends, early customs, hardships, and adventures of the pioneers, SAMUEL J. SACKETT and WILLIAM E. KOCH, *Kansas Folklore* (1961), is recommended.

(C.G.P.)

Kansu

The province of Kansu (Gan-su in Pin-yin romanization), administratively a part of the Northwest region, reaches into the geographic centre of the People's Republic of China. It is bordered by the Mongolian People's Republic to the north; by the Chinese provinces of Tsinghai and Szechwan to the south; by the Ningsia Hui Autonomous Region, the province of Shensi, and the Inner Mongolian Autonomous Region to the east; and by the Sinkiang Uighur Autonomous Region to the west. A vital strategic pivot, linking China proper with the vast territory in the extreme west, the narrow corridor of Kansu has served for several centuries as a passageway between the Upper Huang Ho (Yellow River) area and Chinese Turkistan.

Territorial changes

Kansu covers 238,100 square miles (616,600 square kilometres) of land, almost 5 percent of the total territory in China. The area within its jurisdiction has undergone several changes during the past two decades. In 1954 Kansu annexed the province of Ningsia. In 1956 the A-la-shan and Edzen Gol (in Chinese, O-chi-na Ho) region in northwestern Kansu was detached and incorporated into the Inner Mongolian Autonomous Region. In 1958, however, the affixed Ningsia Province was separated from Kansu to become the Ningsia Hui Autonomous Region, and in 1970 the A-la-shan and Edzen Gol region of the Inner Mongolian Autonomous Region was returned to Kansu again, leaving the territory of Kansu almost unchanged when compared with its 1950 area.

In 1970 the population of Kansu was estimated at about 13,000,000.

The capital of Kansu is Lan-chou, on the south bank of the Huang Ho; its population was 750,000 in 1960 and was expected to double in the 1970s. Endowed with rich mineral resources, traditionally backward Kansu is building itself into a vital industrial base to support the exploitation of the province of Tsinghai to the south and the autonomous region of Sinkiang to the far west. For associated physical features, see GOBI (DESERT); HUANG HO (RIVER).

History. Kansu became a part of Chinese territory during the Ch'in dynasty (221–206 BC), when Chinese power began to extend up to the Kansu Corridor and into the region of modern Ningsia and Tsinghai. In ancient times, all traffic between China proper and the far west was funnelled through the Kansu Corridor. Along the ancient Silk Road that began at Ch'ang-an (Sian) and continued through the corridor, camel caravans carried the tea, silk, and porcelain of China to bazaars in the Near East and even to the markets of Byzantium and Rome. In the train of these caravans such travellers as the Buddhist missionary Kumārajīva and the Venetian merchant Marco Polo entered China.

The name of Kansu first came into existence in the Yüan dynasty (1279–1368), when, under Mongol rule, it comprised the two districts of Kan-chou and Su-chou. In the Ch'ing dynasty (1644–1911), Kansu covered the later provinces of Kansu, Ningsia, a part of Tsinghai, and a part of Sinkiang. The area was under the administration of a governor general of Shensi-Kansu, who was stationed at Lan-chou and had authority over both provinces. One of the most prominent governors general was Tso Tsung-t'ang (1812–85), who, after 1878, brought a half century of peace to Kansu. A hero in the suppression of the Taiping Rebellion, Tso also helped the Ch'ing court to put down the Muslim Rebellion in Kansu, which lasted for 16 years (1862–78) and affected the lives of more than 10,000,000 people.

Before Tso assumed the governorship, Kansu was an area without law and order. The Muslims in Kansu were in open rebellion, committing murder, arson, and numerous other crimes. After having effectively destroyed their strongholds, Tso extended Chinese educational and civil-service systems into the conquered districts for the benefit of Muslims and non-Muslims alike. As a result, the violence subsided and peace prevailed.

Kansu remained a province of China during the period of the Chinese Republic (1911–49). The territory, however, shrank substantially when Sinkiang, Tsinghai, and Ningsia became independent provinces. The discovery of the Yii-men oil field in 1937 opened brighter economic prospects for the area. Communist influence in Kansu began in 1935 after the Chinese Red Army withdrew from southeast China to Shensi, and a Communist-controlled Shensi-Kansu-Ningsia border government was established in the late 1930s. Kansu continued to be a province after the establishment of the People's Republic in Peking in 1949.

THE LANDSCAPE

Relief. Plateaus are the dominant physical features of Kansu. Along the southern border, the lofty Nan Shan and Ch'i-lien Shan-mo ranges separate Kansu from Tsinghai. These ranges have an average height of 12,900 feet. Near Lan-chou in central Kansu, the Huang Ho Valley opens out, and excellent agricultural land is available. North of the Wushaling (Dark Sand Peak), 120 miles northwest of Lan-chou, there is a stretch of interior drainage where the land is relatively flat and glacier-fed streams, including the Hei Ho (Black River), disappear into the desert; this is the area previously referred to as the Kansu Corridor. The higher mountains nearby are covered with forests and their lower slopes are green with grass, but the floor of the corridor itself is monotonously flat and barren yellow earth. Geologically, Tertiary formations (from 2,500,000 to 65,000,000 years old) appear in a number of basins in Kansu, with strata generally composed of red clays, conglomerates, red sandstones, and gypsum.

The earthquake zone

The topographical features of Kansu are relatively uncomplicated in the west and northwest, in contrast to the southeast, where the land has suffered local dislocations from earthquakes. In the northwest there are very few mountains but rather a hilly terrain that merges into the Gobi Desert to the east. The average altitude is about 3,000 feet. The eastern part of Kansu is the principal centre of earthquakes in China. From the 6th century AD to the present, major earthquakes have taken place on an average of once every 65 years, while minor quakes occurred at least once every ten years. One of the greatest disasters of modern times occurred in 1920, when a violent earthquake, centred in eastern Kansu, caused great landslides. The death toll was estimated at 246,000; many cities and towns totally vanished.

Climate. The climate in Kansu undergoes sharp fluctuations of temperature in summer (June to August) and winter (December to February), with uneven and unpredictable precipitation throughout the year. In the west, the average January temperature in Chiu-ch'üan, for instance, is 18° F (—8" C), and in Tun-huang, 200 miles west of Chiu-ch'iian, is 19" F (—7" C). The temperature in July in Chiu-ch'iian is 70° F (21° C), and in Tun-huang it is 81° F (27° C). Annual temperature variations for most parts of Kansu are more than 54° F (12° C).

Rainfall is meagre throughout most of Kansu. As one goes further inland, the precipitation becomes increasingly less frequent. In the western part of the province annual rainfall ranges from two inches at Tun-huang to three inches at Chiu-ch'iian. Irrigation depends mainly on melting snow from the Ch'i-lien Shan-mo.

The southeastern part, something of an exception to the general pattern, receives a relatively abundant rainfall. In P'ing-liang, 170 miles east of Lan-chou, rainfall reaches 20 inches. Summer is usually the period of maximum precipitation, when 50 to 70 percent of the annual rainfall occurs. Near Lan-chou, because of the high rate of evaporation, the farmers spread layers of pebbles over the cultivated surface to check evaporation and to keep the soil moist.

Vegetation. Wheat is the chief crop throughout the province; barley, millet, beans, and sweet potatoes are also important food crops. Kaoliang (Chinese sorghum), buckwheat, corn (maize), oats, and rice are grown to a lesser extent. Cotton, wool, and tobacco are produced as cash crops. Kansu is famous for its water-pipe tobacco, which is raised near Lan-chou and farther west, around Chang-yeh. Although vegetation is rather limited in the mountain area, primeval forests still exist in the high mountains of Liu-p'an Shan and Ho-lan Shan in the eastern part of Kansu. On the floor of the Kansu Corridor, willows and poplars grow along the roads and ditches. Lan-chou and its vicinity are known for fruit growing: pears, peaches, apricots, dates, apples, palms, and watermelons are all cultivated there.

Animal life. Wild animals include marmots (burrowing rodents), deer, and foxes. Domestic livestock is comprised of cattle, goats, and sheep (including the Ho-si sheep that produces fine wool). About 40,000 bactrian (two-humped) camels are reared in the Kansu Corridor.

PEOPLE AND POPULATION

In 1946 the population of Kansu was estimated at about 6,500,000 and that of Ningsia at about 700,000. The 1953 census reported a total of 12,900,000 for the two areas—an increase of about 5,700,000 people, largely due to immigration of the Chinese population from the coastal areas. Average density was 48 people per square mile. In 1970, as mentioned, the population was estimated at about 13,000,000.

Population distribution

The population is concentrated in the Lan-chou Basin, in the fertile valley plains of the south and central sections where irrigation is possible, and in the dry terrace land of the Liu-p'an Shan. Population density around Lan-chou is as high as 200 per square mile. In western Kansu, population is intensively concentrated in a number of small, isolated oases scattered along the base of the high snow-capped ranges.

Ethnic groups. The Chinese, numbering more than 10,000,000, comprise the main racial group in Kansu. Other ethnic groups are the Monguors (Mongols), the Turks (Salars and Sarigh-Uighurs), and the Tibetans. There are more than 300,000 Mongols to the west of Lan-chou, and there are 400,000 Tibetans scattered over an area enclosed by the rivers P'ingfan, Ta-t'ung Ho, and Huang Ho.

Religious groups. The Chinese majority tends to follow the same traditional religious practices, such as Buddhism, generally observed elsewhere in China. The most important minority group in Kansu is formed by the Muslims (Hui-hui), living mostly in the north and west, who number about 1,500,000. In Asia, in general, Muslims are referred to as Dungans (Tungans), but in Kansu they are called Hui-hui or Hui. Some are of Arab, Turkish, or Mongol origin, whereas a few are converted Chinese. The Muslims include believers in both the Sunnī and Shi'ah traditions; the Sunnis follow the more traditional Islāmic interpretation, but the Shi'ahs do not recognize hereditary religious leadership. Tibetans and Mongols follow Tibetan Buddhism. Almost every Tibetan family has at least one son in a Buddhist monastery.

Linguistic groups. Most of the ethnic groups, including the Tibetan minority, speak Chinese as a second language. The Monguors, however, whose language differs completely from either Western or Eastern Mongolian, rarely speak a second language. Chinese Muslims use both Chinese and Arabic scripts, although Arabic is usually used only for religious purposes.

Patterns of rural and urban settlement. The Chinese and the Muslims are essentially agriculturists, although some engage in trade and industry. The Mongols are pastoralists or are seminomadic. In 1957 the agricultural population represented about 80 percent of the total, industrial population (including those engaged in handicrafts) about 3 percent, and those engaged in trade about 3 percent.

Important urban areas are centred on Lan-chou, which had a population of 750,000 in 1960. The largest city in eastern Kansu is P'ing-liang, which had a population of more than 100,000 in 1957. A major centre in western Kansu is Chiu-ch'iian, which had a population of 50,000 in 1957.

ADMINISTRATION AND SOCIAL CONDITIONS

Administration. The provincial government has its headquarters in Lan-chou. Two municipalities are under the direct supervision of the provincial government—Lan-chou itself and Chia-yii-kuan, at the western terminus of the Great Wall (which runs from northwest to southeast through the province). Intermediate administrative divisions include eight special districts (*chuan-chu'ü*)—Ting-hsi, P'ing-liang, Ch'ing-yang, T'ien-shui, Wu-tu, Wu-wei, Chang-yeh, and Chiu-ch'iian—and two autonomous wards or districts (*chou*)—the Lin-hsia Hui Autonomous District, inhabited by Muslims, and the South Kansu Tibetan Autonomous District, inhabited by Tibetans. After the transfer of part of Pa-yen-Nao-erh League (*meng*) from Inner Mongolia certain additional districts were added, but their composition is not yet fully known.

On the third level of administration before the 1970 administrative changes the province was grouped into 66 counties (hsien), six autonomous counties (*tzu-chih* hsien), two banners (ch'i), and three municipalities (*shih*) of T'ien-shui, Yii-men, and Lin-hsia.

The Muslim villages

Village life. Village life among the Chinese inhabitants is generally similar to that elsewhere in North China. There usually are from 20 to 30 families in a village. In Muslim villages, however, the religious–communal life-style is distinctly different. There is a small public building that serves as a mosque, where children gather regularly to receive religious instruction and to learn the alphabet and phonetics. Muslim villages are, by comparison, more organized and possess more community spirit than is usual in the Chinese villages. Hitherto the two peoples have been mutually segregated.

Tibetan villages, in many apsects, are similar to Chinese villages. Those Tibetans who are sedentary, however, have no clearly defined clan organization, and their family ties are much looser than among the Chinese.

Village dwellings are the conventional types of mud huts. Some people live in caves—which may be elaborate, with fine furnishings, or simply scooped out of the porous yellow earth cliffs. Brick structures predominate in cities and towns.

The eating habits of the people are slightly different from those of the Chinese in other parts of the nation. Coarse food and wheat flour, rather than rice, are consumed.

Education. The educational standard is comparatively lower than elsewhere in North China. In 1946 there were about 200 secondary schools, 800 primary schools, and 6,500 junior primary schools, with a total enrollment of almost 500,000 students. Since 1950 educational facilities have been greatly expanded. In 1950 there were more than ten universities and colleges, mostly located in Lan-chou, including Lan-chou University, the Northwest Normal College, and the Northwest Institute for Minorities Special colleges providing training for railway work, the petroleum industry, animal husbandry, and veterinary medicine are also established in Lan-chou.

Health and sanitation. By Western standards. the area is backward in health and sanitation. The most common diseases are the fecal-borne intestinal diseases spread through the use of human waste as fertilizer. The shortage of water supplies and the lack of modern doctors, nurses, and pharmacists constitute a serious problem.

Welfare. Welfare work in this area is more concerned with the victims of natural disasters than with the poor in general. Frequent earthquakes and severe droughts require the government to assume responsibility for relief. In the Muslim community, a part of the public welfare is organized by the Muslims themselves; Muslim officials collect obligatory charity for this purpose.

Since 1949 the Communist government has made general progress in Kansu with its welfare program for workers and peasants. New residential areas, for instance, have been built in Chiu-ch'iian for families of workers in the Yii-men oil fields. Medical clinics have been established in remote areas, where most people previously relied on local herb doctors.

The economy. Traditionally, Kansu has been an area of poverty. The region is frequently subject to earthquakes, droughts, and famines. On the average, calamities destroyed 20 percent or more of the crop each year during the period from 1904 to 1929.

Agricultural resources

Resources. Although it is predominately an agricultural area, and despite the fact that the per capita landholding is much larger than the national mean, output of food grain is insufficient to feed the population. The extent of cultivation in different areas depends upon the elevation, the steepness of the slope of the land, and the dryness of the climate. High elevation has a greater precipitation and is therefore more favourable for farming. Terracing is prevalent and is practiced on about 18 percent of all of the cultivated land. Much of the hill land is cultivated by the use of a modified form of contour plowing. Because the slopes of the fields are so steep, however, and the fields so extensive, erosion is a serious problem, and some of the land has been abandoned. Agriculture in this area depends on the improvement of irrigation.

Kansu is rich in minerals of great value. The oil field of Yii-men, which is located in the northwest part of Kansu, has a reserve of 729,000,000 barrels. A new oil reserve found in Ya-er-hsia (Duckling Gorge) near Yü-men is estimated at 670,000,000 tons. Coal reserves are estimated at 2,400,000,000 tons. A large deposit of iron ore has been discovered in the Nan Shan area in western Kansu and is tentatively estimated to amount to at least 100,000,000 tons.

Resource exploitation. Since 1950 strenuous efforts have been made to develop Kansu into an industrial base for northwest China. During the First Five-Year Plan (1953–57) the government emphasized efforts to build Lan-chou into a centre for heavy industry. Lan-chou won fame for its petroleum production and now has dozens of large modern industrial enterprises. The Lan-chou Petroleum Machinery Plant and the Lan-chou Petroleum and Chemical Machinery Plant were both constructed with Soviet aid and are the two key enterprises manufacturing modern petroleum drilling and refining equipment for China's fast-growing petroleum industry. The Lan-chou Refining Plant, one of the largest in China, handles more than 16,790,000 barrels of petroleum a year. The Lan-chou Locomotive Plant, the largest in the northwest, is the regional centre for locomotive repairs and also manufactures internal-combustion-engine locomotives for operation in desert areas.

Since 1960, efforts have also been made to build Lan-chou into a base for nuclear industry. A gaseous diffusion plant, a major installation of the nuclear industry, was constructed and put into operation in 1963 near the city. It is capable of separating the fissionable uranium-235 from uranium-238 and is one of the major nuclear plants in China.

From 1950 to 1960, extensive geological surveys for minerals and oil were conducted in Kansu. The oil field in Yii-men was then greatly expanded, with the output of petroleum reaching 12,400,000 barrels by 1967. Yii-men has its own refinery with an annual capacity of 6,570,000 barrels and is linked by pipeline with the Lan-chou refinery.

The chief coal mine is located at A-kan-chen, almost 20 miles south of Lan-chou. It produces 1,000,000 tons of coal a year. Total coal production for the province was estimated in 1966 at 7,000,000 tons. A large iron and steel complex is under construction at Chiu-ch'iian with a designed capacity of 2,600,000 tons of pig iron and 1,800,000 tons of steel.

Hydro-electric projects

An ambitious project is under way to harness the waters of the Huang Ho by establishing two huge hydroelectric power stations on its upper course. One of them, the Liu-chia Hsia (Liu-chia Gorge) hydroelectric station above Lan-chou, now under construction, will produce 1,000,000 kilowatts of electricity. It will not only supply ampler and cheaper electric power to the new industrial enterprises around Lan-chou but will also free the city of Lan-chou from the menance of floods caused by the Huang Ho and bring large tracts of land under irrigation.

In agriculture, attempts have been made to increase agricultural output by transforming the vast areas of wasteland along the Kansu Corridor into cotton fields. Of more than 1,600,000 acres of wasteland in this area, more than one-third is suitable for cotton.

Transportation and communication. The major barrier to development in this area has been the absence of transportation facilities. Before 1952, only the Lunghai Railway connected Kansu with the coastal area; in that year a 215-mile extension between Lan-chou and T'ien-shui to the southeast was completed. In addition, a new railway extends westward from Lan-chou via Yii-men to Urumchi, the capital of Sinkiang, spanning a distance of about 1,730 miles. The 600-mile-long Lan-chou-to-Pao-t'ou railway, linking Kansu with the Inner Mongolian Autonomous Region, was completed in 1958. Survey work for a new railway between Lan-chou and Tsinghai began in 1955. The line, completed in 1959, runs for 930 miles, connecting Lan-chou with the rich mineral area of the Tsaidam Basin in northwest Tsinghai.

The highway system has also been greatly expanded. Highways radiate from Lan-chou toward Tsinghai, Sinkiang, Inner Mongolia, Shensi, and Szechwan. Domestic aviation services connect Lan-chou with Sian, Peking, Urumchi, Hsi-ning in Tsinghai, and Chungking in Szechwan. Telephone and telegram services are in operation throughout the province.

Navigation on the Huang Ho is limited to the section between Lan-chou and Chung-wei because of considerable silting and the river's seasonal flow.

Cultural life. Kansu represents a colourful mixture of races, customs, and cultures. The land abounds with mosques, lamaseries (monasteries of lamas), and Chinese temples.

Religious festivals

The cultural milieu. Communal life in Chinese villages is marked by religious observances, particularly rituals connected with ancestor worship; seasonal celebrations, such as the New Year, the Dragon Boat Festival, and the Moon Festival; and customs relating to birth, marriage, funerals, and burials. All of these activities are similar to those of the Chinese throughout the nation. Village theatricals provide another type of communal activity.

Most of the Monguors and Tibetans have abandoned their nomadic way of life and have settled down. They live in brick and mud houses resembling their former tents (yurts). Tibetans insist on simultaneous group actions within the village. Every year, when the first day of spring planting is determined by the horoscope, for instance, all the villagers go to the fields in their best clothing. The fields are then plowed simultaneously, and the seeds are sown at the same time in each field. During the course of the growing season, all the villagers periodically parade through the fields carrying their holy books on their heads.

Cultural traditions. The Muslims are faithful followers of their religion and strictly observe the fast of Ramaḍān, during which they eat nothing between sunrise and sunset. Before darkness falls, pious, bearded men say their prayers in public, and one or two of the elders may preach on points of theology, quoting the Qur'ān in oddly mutilated Arabic. At nightfall a communal feast is eaten; the community fires blaze all night, and people call and shout to one another.

Chinese Muslims also observe the practice of making a pilgrimage to Mecca. The *ḥājīs*, those who have completed the pilgrimage, are highly respected in the community. The number of pilgrimages has, however, decreased considerably since 1949.

The western part of Kansu has long been a region renowned for ancient and classic artistic works. Numerous stone caves in Tun-huang have many kinds of religious paintings on their walls, dating from the T'ang dynasty (AD 618–907). In Wu-wei, in western Kansu, large numbers of writings on bamboo slips have been found on the sites of the old frontier garrisons of the Han Empire (206 BC–AD 220). In 1964, a coherent bamboo text comprising a large part of one of the classic works on ritual (the *I Li*) came to light in western Kansu. In Tun-huang, within a Buddhist cave-temple, a library was discovered that had been immured there in the year 1035. It consisted of voluminous rolls of texts, including many valuable paintings and Buddhist classics.

The mass media are government controlled. The most influential newspaper is the provincial party organ, the *Lan-chou Jih-pao* ("Lanchow Daily"). Radio services are provided by the Kansu People's Radio station in Lan-chou.

BIBLIOGRAPHY. A Regional Handbook on Northwest China (1956) provides basic materials on Kansu during the pre-Communist era and some information on the early years of the Communist rule. CHANG KUEI-SHENG, "Geographical Bases for Industrial Development in Northwestern China," *Econ. Geogr.*, 39:341–350 (1963), gives data for the new discoveries of mineral resources in this area. LIU EN-LAN, "The Ho-Si Corridor," *ibid.*, 28:51–56 (1952), describes the author's personal impressions of this vital gateway. Two historical accounts of the Muslim Rebellion in Kansu are CHU WEN-DJANG, *The Moslem Rebellion in Northwest China, 1862–1878* (1966); and W.L. BALES, *Tso Tsung-t'ang: Soldier and Statesman of Old China* (1957). GEORGE B. CRESSEY, *Land of the 500 Million* (1955); and T.R. TREGEAR, *An Economic Geography of China* (1970), are general references on the geographic aspects. The official grand plan for the transformation of Kansu into a new industrial base is included in T.H. TENG, "Report on the Multi-Purpose Plan for Permanently Controlling the Yellow River and Exploiting Its Water Resources," *People's China*, supplement (September 1, 1955).

(C.y.C.)

Kant, Immanuel

Immanuel Kant was the foremost thinker of the Enlightenment and one of the great philosophers of all time, in whom were subsumed new trends that had begun with the Rationalism (stressing reason) of René Descartes and the Empiricism (stressing experience) of Francis Bacon. He inaugurated a new era in the development of philosophical thought. His comprehensive and systematic work in theory of knowledge, ethics, and aesthetics greatly influenced all subsequent philosophy, especially the various German schools of Kantianism and Idealism (*qq.v.*).

Kant, pencil portrait by Hans Veit Schnoor von Carolsfeld (1764–1841). In the Kupferstichkabinett, Dresden, East Germany.

Marburg—Art Reference Bureau

Pietist rearing and schooling

Background and early life. Kant was born on April 22, 1724, at Konigsberg in East Prussia (since 1946 a part of the Soviet Union) and lived in that remote province for his entire life. His father, a saddler, was a descendant of a Scottish immigrant; his mother, an uneducated German woman, was remarkable for her character and natural intelligence. Both parents were devoted followers of the Pietist branch of the Lutheran Church, which taught that religion belongs to the inner life expressed in simplicity and obedience to the moral law. The influence of their pastor made it possible for Kant—the fourth of 11 children, but the eldest surviving child—to obtain an education.

At the age of eight Kant entered the Pietist school that his pastor directed. This was a Latin school, and it was presumably during the eight-and-a-half years he was there that Kant acquired his life-long love for the Latin classics, especially for the naturalistic poet Lucretius. In 1740 he enrolled in the University of Konigsberg as a theological student. But, although he attended courses in theology and even preached on a few occasions, he was principally attracted to mathematics and physics. Aided by a young professor who had studied Christian Wolff, a systematizer of Rationalist philosophy, and who was also an enthusiast for the science of Sir Isaac Newton, Kant began reading the work of the English physicist and, in 1744, started his first book, dealing with a problem concerning kinetic forces. Though by that time he had decided to pursue an academic career, the death of his father in 1746 and his failure to obtain the post of undertutor in one of the schools attached to the university compelled him to withdraw and seek a means of supporting himself.

Tutor and Privatdozent. He found employment as a family tutor and, during the nine years that he gave to it, worked for three different families. With them he was introduced to the influential society of the city, acquired social grace, and made his farthest travels from his native city—some 60 miles away to the town of Arnsdorf. In 1755, aided by the kindness of a friend, he was able to complete his degree at the university and take up the position of *Privatdozent*, or lecturer.

The three early dissertations

Three dissertations that he presented on obtaining this post indicate the interest and direction of his thought at this time. In one, *De Igne* (On Fire), he argued that bodies operate on one another through the medium of a uniformly diffused elastic and subtle matter that is the underlying substance of both heat and light. His first teaching was in mathematics and physics, and he was never to lose his interest in scientific developments. That it was more than an amateur interest is shown by his publication within the next few years of several scientific works dealing with the different races of men, the nature of winds, the causes of earthquakes, and the general theory of the heavens.

At this period Newtonian physics was important for Kant as much for its philosophical implications as for its scientific content. A second dissertation, the *Monodologia physica* (1756), contrasted the Newtonian methods of thinking with those employed in the philosophy then prevailing in German universities. This was the philosophy of Gottfried Wilhelm Leibniz, a universal scholar, as systematized and popularized by Wolff and by Alexander Gottlieb Baumgarten, author of a widely used text, the *Metaphysica* (1739). Leibniz' works as they are now known were not fully available to these writers; and the Leibnizian philosophy that they presented was extravagantly Rationalistic, abstract, and cut-and-dried. Yet it remained a powerful force, and the main efforts of independent thinkers in Germany at the time were devoted to examining its ideas.

In a third dissertation, *Principiorum Primorum Cognitionis Metaphysicae Nova Dilucidato* (1755), on the first principles of metaphysics, Kant analyzed especially the principle of sufficient reason, which, in Wolff's formulation, asserts that for everything there is a sufficient reason why it should be rather than not be. Although critical, Kant was cautious and still a long way from challenging the assumptions of Leibnizian metaphysics.

During his 15 years as a *Privatdozent,* Kant's fame as a teacher and writer steadily increased. Soon he was lecturing on many subjects besides physics and mathematics—including logic, metaphysics, and moral philosophy. He even lectured on fireworks and fortifications and every summer for 30 years gave a popular course on physical geography. He enjoyed great success as a lecturer; his style, which differed markedly from that of his books, was humorous and vivid, enlivened by many examples from his reading in English and French literature, and in travel and geography, science and philosophy.

Although he twice failed to obtain a professorship at Konigsberg, he refused to accept offers that would have taken him elsewhere—including the professorship of poetry at Berlin that would have brought greater prestige. He preferred the peace and quiet of his native city in which to develop and mature his own philosophy.

Critic of Leibnizian Rationalism. During the 1760s he became increasingly critical of Leibnizianism. According to one of his students, Kant was then attacking Leibniz, Wolff, and Baumgarten, was a declared follower of Newton, and expressed great admiration for the moral philosophy of the Romanticist Jean-Jacques Rousseau.

The attack on Leibniz

His principal work of this period was *Untersuchung iiber die Deutlichkeit der Grundsätze der natiirlichen Theologie und der Moral* (1764; "An Inquiry into the Distinctness of the Fundamental Principles of Natural Theology and Morals"). In this work he attacked the claim of Leibnizian philosophy that philosophy should model itself on mathematics and aim at constructing a chain of demonstrated truths based on self-evident premises. Kant argued that mathematics proceeds from definitions that are arbitrary, by means of operations that are clearly and sharply defined, upon concepts that can be exhibited in concrete form. In contrast with this method, he argued that philosophy must begin with concepts that are already given, "though confusedly or insufficiently determined," so that philosophers cannot begin with definitions without thereby shutting themselves up within a circle of words. Philosophy cannot, like mathematics, proceed synthetically; it must analyze and clarify. The importance of the moral order, which he had learned from Rousseau, reinforced the conviction received from his study of Newton that a synthetic philosophy is empty and false.

Besides attacking the methods of the Leibnizians, he also began criticizing their leading ideas. In an essay *Versuch, den Begriff der negativen Grössen in die Weltweisheit einzuführen* (1763), he argued that physical opposition as encountered in things cannot be reduced to logical contradiction, in which the same predicate is both affirmed and denied, and, hence, that it is pointless to reduce causality to the logical relation of antecedent and consequent. In an essay of the same year, *Der einzig mogliche Beweisgrund zu einer Demonstration des Daseyns Gottes,* he sharply criticized the Leibnizian concept of Being by charging that the so-called ontological argument, which would prove the existence of God by logic alone, is fallacious because it confuses existential with attributive statements: existence, he declared, is not a predicate of attribution. Moreover, with regard to the nature of space, Kant sided with Newton in his confrontation with Leibniz. Leibniz' view that space is "an order of co-existences" and that spatial differences can be stated in conceptual terms, he concluded to be untenable.

Investigation of the world of spirits

Some indication of a possible alternative of Kant's own to the Leibnizian position can be gathered from his curious *Traume eines Geistersehers erläutert durch Träume der Metaphysik* (1766). This work is an examination of the whole notion of a world of spirits, in the context of an inquiry into the spiritualist claims of Emanuel Swedenborg, a scientist and biblical scholar. Kant's position at first seems to have been completely skeptical, and the influence of the Scottish Skeptic David Hume is more apparent here than in any previous work; it was Hume, he later claimed, who first awoke him from his dogmatic slumbers. Yet Kant was not so much arguing that the notion of a world of spirits is illusory as insisting that men have no insight into the nature of such a world, a conclusion that has devastating implications for metaphysics as the Leibnizians conceived it. Metaphysicians can dream as well as spiritualists, but this is not to say that their dreams are necessarily empty; there are already hints that moral experience can give content to the ideal of an "intelligible world." Rousseau thus acted upon Kant here as a counterinfluence to Hume.

Early years of the professorship at Konigsberg. Finally, in 1770, after serving for 15 years as a *Privatdozent,* Kant was appointed to the chair of logic and metaphysics, a position in which he remained active until a few years before his death. In this period—usually called his critical period, because in it he wrote his great *Critiques*—he published an astounding series of original works on a wide variety of topics, in which he elaborated and expounded his philosophy.

"Inaugural Dissertation" of 1770

The *Inaugural Dissertation* of 1770 that he delivered on assuming his new position already contained many of the important elements of his mature philosophy. As indicated in its title, *De Mundi Sensibilis atque Intelligibilis Forma et Principiis: Dissertatio,* the implicit dualism of the *Traume* is made explicit; and it is made so on the basis of a wholly un-Leibnizian interpretation of the distinction between sense and understanding. Sense is not, as Leibniz had supposed, a confused form of thinking but a source of knowledge in its own right, although the objects so known are still only "appearances"—the term that Leibniz also used. They are appearances because all sensing is conditioned by the presence, in sensibility, of the forms of time and space, which are not objective characteristics or frameworks of things but "pure intuitions." But though all knowledge of things sensible is thus of phenomena, it does not follow that nothing is known of things as they are in themselves. Certainly, man has no intuition, or direct insight, into an intelligible world; but the presence in him of certain "pure intellectual concepts, such as those of possibility, existence, necessity, substance, cause, enables him to have some descriptive knowledge of it. By means of these concepts he can arrive at an exemplar that provides him with "the common measure of all other things as far as real." This exemplar gives man an idea of perfection for both the theoretical and practical orders: in the first, it is that of the Supreme Being, God; in the latter, that of moral perfection.

After the *Dissertation,* Kant published virtually nothing for 11 years. Yet, in submitting the *Dissertation* to a friend at the time of its publication, he wrote:

> About a year since I attained that concept which I do not fear ever to be obliged to alter, though I may have to widen it, and by which all sorts of metaphysical questions can be tested in accordance with entirely safe and easy criteria, and a sure decision reached as to whether they are soluble or insoluble.

Period of the three "Critiques." In 1781 the *Critik der reinen Vernunft (Critique of Pure Reason)* was published, followed for the next nine years by great and original works that in a short time brought a revolution in philosophical thought and established the new direction in which it was to go in the years to come.

The Critique of Pure Reason. The *Critique of Pure Reason* was the result of some 10 years of thinking and meditation. Yet, even so, Kant published the first edition only reluctantly after many postponements; for although convinced of the truth of its doctrine, he was uncertain and doubtful about its exposition. His misgivings proved well-founded, and Kant complained that interpreters and critics of the work were badly misunderstanding it. To correct these wrong interpretations of his thought he wrote the *Prolegomena zu einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können* (1783) and brought out a second and revised edition of the first "critique" in 1787. Controversy still continues regarding the merits of the two editions: readers with a preference for an Idealistic interpretation usually prefer the first edition, whereas those with a Realistic view adhere to the second. But with regard to difficulty and ease of reading and understanding, it is generally agreed that there is little to choose between them. Anyone on first opening either book finds it overwhelmingly difficult and impenetrably obscure.

The cause for this difficulty can be traced in part to the works that Kant took as his models for philosophical writing. He was the first great modern philosopher to spend all of his time and efforts as a university professor of the subject. Regulations required that in all lecturing a certain set of books be used, with the result that all of Kant's teaching in philosophy had been based on such handbooks as those of Wolff and Baumgarten, which abounded in technical jargon, artificial and schematic divisions, and great claims to completeness. Following their example, Kant accordingly provided a highly artificial, rigid, and by no means immediately illuminating scaffolding for all three of his *Critiques.*

The *Critique* of *Pure Reason,* after an introduction, is divided into two parts, of very different lengths: A "Transcendental Doctrine of Elements," running to almost 400 pages in a typical edition, followed by a "Transcendental Doctrine of Method," which reaches scarcely 80 pages. The ". . . Elements" deals with the sources of human knowledge, whereas the ". . . Method" draws up a methodology for the use of "pure reason" and its a priori ideas. Both are "transcendental," in that they are presumed to analyze the roots of all knowledge and the conditions of all possible experience. The "Elements" is divided, in turn, into a "Transcendental Aesthetic," a "Transcendental Analytic," and a "Transcendental Dialectic."

The simplest way of describing the contents of the *Critique* is to say that it is a treatise about metaphysics: it seeks to show the impossibility of one sort of metaphysics and to lay the foundations for another. The Leibnizian metaphysics, the object of his attack, is criticized for assuming that the human mind can arrive, by pure thought, at truths about entities, which, by their very nature, can never be objects of experience, such as God, human freedom, and immortality. Kant maintained, however, that the mind has no such power and that the vaunted metaphysics is thus a sham.

The synthetic a priori in mathematics and physics

As Kant saw it, the problem of metaphysics, as indeed of any science, is to explain how, on the one hand, its principles can be necessary and universal (such being a condition for any knowledge that is scientific) and yet, on the other hand, involve also a knowledge of the real and so provide the investigator with the possibility of more knowledge than is analytically contained in what he already knows; *i.e.*, than is implicit in the meaning alone. To meet these two conditions, Kant maintained, knowledge must rest on judgments that are a priori, for it is only as they are separate from the contingencies of experience that they could be necessary and yet also synthetic; *i.e.*, so that the predicate term contains something more than is analytically contained in the subject. Thus, for example, the proposition that all bodies are extended is not synthetic but analytic because the notion of extension is contained in the very notion of body; whereas the proposition that all bodies are heavy is synthetic because weight supposes, in addition to the notion of body, that of bodies in relation to one another. Hence, the basic problem, as Kant formulated it, is to determine "How [*i.e.*, under what conditions] are synthetic a priori judgments possible?"

This problem arises, according to Kant, in three fields, viz., in mathematics, physics, and metaphysics; and the three main divisions of the first part of the *Critique* deal respectively with these. In the "Transcendental Aesthetic," Kant argued that mathematics necessarily deals with space and time and then claimed that these are both a priori forms of human sensibility that condition whatever is apprehended through the senses. In the "Transcendental Analytic," the most crucial as well as the most difficult part of the book, he maintained that physics is a priori and synthetic because in its ordering of experience it uses concepts of a special sort. These concepts—"categories," he called them—are not so much read out of experience as read into it and, hence, are a priori, or pure, as opposed to empirical. But they differ from empirical concepts in something more than their origin: their whole role in knowledge is different; for, whereas empirical concepts serve to correlate particular experiences and so to bring out in a detailed way how experience is ordered, the categories have the function of prescribing the general form that this detailed order must take. They belong, as it were, to the very framework of knowledge. But although they are indispensable for objective knowledge, the sole knowledge that they can give is of objects of possible experience; they yield valid and real knowledge only when they are ordering what is given through sense in space and time.

The synthetic a priori in metaphysics

In the "Transcendental Dialectic" Kant turned to consideration of a priori synthetic judgments in metaphysics. Here, he claimed, the situation is just the reverse from what it was in mathematics and physics. Metaphysics cuts itself off from sense experience in attempting to go beyond it and, for this very reason, fails to attain a single true a priori synthetic judgment. To justify this claim, Kant analyzed the use that metaphysics makes of the concept of the unconditioned. Reason, according to Kant, seeks for the unconditioned or absolute in three distinct spheres: (1) in philosophical psychology it seeks for an absolute subject of knowledge; (2) in the sphere of cosmology, it seeks for an absolute beginning of things in time, for an absolute limit to them in space, and for an absolute limit to their divisibility; and (3) in the sphere of theology, it seeks for an absolute condition for all things. In each case, Kant claimed to show that the attempt is doomed to failure by leading to an antinomy in which equally good reasons can be given for both the affirmative and the negative position. The metaphysical "sciences" of rational psychology, rational cosmology, and natural theology, familiar to Kant from the text of Baumgarten, on which he had to comment in his lectures, thus turn out to be without foundation.

With this work, Kant proudly asserted that he had accomplished a Copernican revolution in philosophy. Just as the founder of modern astronomy, Nicolaus Copernicus, had explained the apparent movements of the stars by ascribing them partly to the movement of the observers, so Kant had accounted for the application of the mind's a priori principles to objects by showing that the objects conform to the mind: in knowing, it is not the mind that conforms to things but things that conform to the mind.

The Critique of *Practical Reason.* Because of his insistence on the need for an empirical component in knowledge and his antipathy to speculative metaphysics, Kant is sometimes presented as a Positivist before his time; and his attack upon metaphysics was held by many in his own day to bring both religion and morality down with it. Such, however, was certainly far from Kant's intention. Not only did he propose to put metaphysics "on the sure path of science," he was prepared also to say that he "inevitably" believed in the existence of God and in a future life. It is also true that his original conception of his critical philosophy anticipated the preparation of a critique of moral philosophy. The *Critik der practischen Vernunft* (1788; *Critique of Practical Reason),* the result of this intention, is the standard source book for his ethical doctrines. The earlier *Grundlegung zur Metaphysik der Sitten (1785)* is a shorter and, despite its title, more readily comprehensible treatment of the same general topic. Both differ from *Die Metaphysik der Sitten* (1797) in that they deal with pure ethics and try to elucidate basic principles; whereas the later work is concerned with applying what they establish in the concrete, a process that involved the consideration of virtues and vices and the foundations of law and politics.

Similarity between his ethics and epistemology

There are many points of similarity between Kant's ethics and his epistemology, or theory of knowledge. He used the same scaffolding for both—a "Doctrine of Elements," including an "Analytic" and a "Dialectic," followed by a "Methodology"; but the second *Critique* is far shorter and much less complicated. Just as the distinction between sense and intelligence was fundamental for the former, so is that between the inclinations and moral reason for the latter. And just as the nature of the human cognitive situation was elucidated in the first ***Critique*** by reference to the hypothetical notion of an intuitive understanding, so is that of the human moral situation clarified by reference to the notion of a "holy will." For a will of this kind there would be no distinction between reason and inclination; a being possessed of a holy will would always act as it ought. It would not, however, have the concepts of duty and moral obligation, which enter only when reason and desire find themselves opposed. In the case of human beings, the opposition is continuous, for man is at the same time both flesh and spirit; it is here that the influence of Kant's religious background is most prominent. Hence, the moral life is a continuing struggle in which morality appears to the potential delinquent in the form of a law that demands to be obeyed for its own sake—a law, however, the commands of which are not issued by some alien authority but represent the voice of reason, which the moral subject can recognize as his own.

In the "Dialectic," Kant took up again the ideas of God, freedom, and immortality. Dismissed in the first ***Critique*** as objects that men can never know because they transcend human sense experience, he now argued that they are essential postulates for the moral life. Though not reachable in metaphysics, they are absolutely essential for moral philosophy.

Kant is often described as an ethical Rationalist, and the description is not wholly inappropriate. He never espoused, however, the radical Rationalism of some of his contemporaries nor of more recent philosophers for whom reason is held to have direct insight into a world of values or the power to intuit the rightness of this or that moral principle. Thus, practical, like theoretical, reason was for him formal rather than material—a framework of formative principles rather than a content of actual rules. This is why he put such stress on his first formulation of the categorical imperative: "Act only on that maxim through which you can at the same time will that it should become a universal law." Lacking any insight into the moral realm. men can only ask themselves whether what they are proposing to do has the formal character of law—the character, namely, of being the same for all persons similarly circumstanced.

Critiques of aesthetics and natural teleology

The "Critique of Judgment." The ***Critik der Urtheilskraft*** (1790)—one of the most original and instructive of all of Kant's writings—was not foreseen in his original conception of the critical philosophy. Thus it is perhaps best regarded as a series of appendixes to the other two ***Critiques.*** The work falls into two main parts, called respectively "Critique of Aesthetic Judgment" and "Critique of Teleological Judgment." In the first of these, after an introduction in which he discussed "logical purposiveness," he analyzed the notion of "aesthetic purposiveness" in judgments that ascribe beauty to something. Such a judgment, according to him, unlike a mere expression of taste, lays claim to general validity; yet it cannot be said to be cognitive because it rests on feeling, not on argument. The explanation lies in the fact that, when a person contemplates an object and finds it beautiful, there is a certain harmony between his imagination and his understanding, of which he is aware from the immediate delight that he takes in the object. Imagination grasps the object and yet is not restricted to any definite concept; whereas a person imputes the delight that he feels to others because it springs from the free play of his cognitive faculties, which are the same in all men.

In the second part, Kant turned to consider teleology in nature as it is posed by the existence in organic bodies of things of which the parts are reciprocally means and ends to each other. In dealing with these bodies, one cannot be content with merely mechanical principles. Yet if mechanism is abandoned and the notion of a purpose or end of nature is taken literally, this seems to imply that the things to which it applies must be the work of some supernatural designer; but this would mean a passing from the sensible to the suprasensible, a step proved in the first ***Critique*** to be impossible. Kant answered this objection by admitting that teleological language cannot be avoided in taking account of natural phenomena; but it must be understood as meaning only that organisms must be thought of "as if" they were the product of design, and that is by no means the same as saying that they are deliberately produced.

Last years. The critical philosophy was soon being taught in every important German-speaking university, and young men flocked to Konigsberg as a shrine of philosophy. In some cases, the Prussian government even undertook the expense of their support. Kant came to be consulted as an oracle on all kinds of questions, including such subjects as the lawfulness of vaccination. Such homage did not interrupt Kant's regular habits. Scarcely five feet tall, with a deformed chest, and suffering from weak health, he maintained throughout his life a severe regimen. It was arranged with such regularity that people set their clocks according to his daily walk along the street named for him, "The Philosopher's Walk." Until old age prevented him, he is said to have missed this regular appearance only on the occasion when Rousseau's ***Bmile*** so engrossed him that for several days he stayed at home.

Works on religion and the ban on them

With the publication of the third ***Critique,*** Kant's main philosophical work was done. From 1790 his health began to decline seriously. He still had many literary projects but found it impossible to write more than a few hours a day. The writings that he then completed consist partly of an elaboration of subjects not previously treated in any detail, partly of replies to criticisms and to the clarification of misunderstandings. With the publication in 1793 of his work ***Die Religion innerhalb der Grenzen der blossen Vernunft,*** Kant became involved in a dispute with Prussian authorities on the right to express religious opinions. The book was found to be altogether too Rationalistic for orthodox taste; he was charged with misusing his philosophy to the "distortion and depreciation of many leading and fundamental doctrines of sacred Scripture and Christianity" and was required by the government not to lecture or write anything further on religious subjects. Kant agreed but privately interpreted the ban as a personal promise to the King, from which he felt himself to be released on the latter's death in 1797. At any rate, he returned to the forbidden subject in his last major essay, ***Der Streit der Fakultaten*** (1798; "The Conflict of the Faculties").

The large work at which he laboured until his death—the fragments of which fill the two final volumes of the great Berlin edition of his works—was evidently intended to be a major contribution to his critical philosophy. What remains, however, is not so much an unfinished work as a series of notes for a work that was never written. Its original title was ***Ubergang von den metaphysische Anfangsgriinde der Naturwissenschaft zur Physik*** ("Transition from the Metaphysical Foundations of Natural Science to Physics"), and it may have been his intention to carry further the argument advanced in the ***Metaphysische Anfangsgrunde der Naturwissenschaft*** (1786) by showing that it is possible to construct a priori not merely the general outline of a science of nature but a good many of its details as well. But judging from the extant fragments, however numerous they are, it remains conjectural whether its completion would have constituted a major addition to his philosophy and its reputation.

After a gradual decline that was painful to his friends as well as to himself, Kant died in Konigsberg, February 12, 1804. His last words were "Es ist gut" ("It is good"). His tomb in the cathedral was inscribed with the words (in German) "The starry heavens above me and the moral law within me," the two things that he declared in the conclusion of the second ***Critique*** "fill the mind with ever new and increasing admiration and awe, the oftener and the more steadily we reflect on."

MAJOR WORKS

PRE-CRITICAL WRITINGS: ***Gedanken von der wahren Schätzung der lebendigen Krafte und Beurtheilung der Beweise derer sich Herr von Leibniz und anderer Mechaniker in dieser Streitsache bedient haben (1746); Allgemeine Naturgeschichte und Theorie des Himmels (1755; Kant's Cosmog-***

ony . . . , 1900 and 1968; *Universal Natural History and Theories of the Heavens,* 1969); *Principiorum Primorum Cognitionis Metaphysicae Nova Dilucidatio* (1755; Eng. trans. by F.E. England in *Kant's Conception of God,* 1929); *Metaphysicae cum geometria iunctae usus in philosophia naturali, cuius specimen I. continet Monadologiam physicam* (1756); *Versuch einiger Betrachtungen über den Optimismus* (1759); *Die falsche Spitzfindigkeit der vier syllogistischen Figurerewiesen* (1762; trans. in *Kant's Introduction to Logic and His Essay on the Mistaken Subtilty of the Four Figures,* 1963); *Der einzig mögliche Beweisgrund zu einer Demonstration des Daseyns Gottes* (1763; *Enquiry into the Proofs for the Existence of God,* 1836); *Versuch, den Begriff der negativen Grössen in die Weltweisheit einzuführen* (1763; *An Attempt to Introduce the Conception of Negative Quantities into Philosophy,* 1911); *Untersuchung über die Deutlichkeit der Grundsätze der natürlichen Theologie und der Moral* (1764); *Beobachtungen über das Gefühl des Schönen und Erhabenen* (1764, 1766, 1771; *Observations on the Feeling of the Beautiful and Sublime,* 1960); *Träume eines Geistersehers erläutert durch Träume der Metaphysik* (1766; *Dreams of a Spirit-Seer, Illustrated by Dreams of Metaphysics,* 1900; *Dreams of a Spirit Seer, and Other Related Writings,* 1969); *De Mundi Sensibilis atque Intelligibilis Forma et Principiis: Dissertatio* (1770; *Kant's Inaugural Dissertation and Early Writings on Space,* 1929); *Von den Verschiedenen Racen der Menschen* (1775).

CRITICAL AND POST-CRITICAL WRITINGS: *Critik der reinen Vernunft* (1781; rev. ed., *Kritik der reinen Vernunft,* 1787; *Critique of Pure Reason,* 1929, 1950); *Prolegomena zur einer jeden künftigen Metaphysik die als Wissenschaft wird auftreten können* (1783; *Prolegomena to Any Future Metaphysics,* 1951); *Grundlegung zur Metaphysik der Sitten* (1785; *The Fundamental Principles of the Metaphysic of Ethics,* 1938; *The Moral Law; or, Kant's Groundwork of the Metaphysic of Morals,* 1948; *Foundations of the Metaphysics of Morals,* 1969); *Metaphysische Anfangsgründe der Naturwissenschaft* (1786; *Metaphysical Foundations of Natural Science,* 1970); *Critik der practischen Vernunft* (1788; *Critique of Practical Reason,* 1949); *Critik der Urtheilskraft* (1790, 2nd ed. 1793; *Kant's Kritik of Judgment,* 1892, reprinted as *Kant's Critique . . . ,* 1914; new version, *Critique . . . ,* vol. 1, *Kant's Critique of Aesthetic Judgment* and vol. 2, *Critique of Teleological Judgment,* 1911–28, republished 1952); *Über eine Entdeckung, nach der alle neu Critik der reinen Vernunft durch eine ältere entbehrlich gemacht werden soll* (1790; 2nd ed., 1791); *Die Religion innerhalb der Grenzen der blossen Vernunft* (1793; 2nd ed., 4 pt., 1794; *Religion Within the Boundary of Pure Reason,* 1838; *Religion Within the Limits of Reason Alone,* 2nd ed., 1960); *Zum ewigen Frieden* (1795; 2nd ed., 1796; *Perpetual Peace,* 1796, many later editions; 1915 ed. reprinted 1972); *Die Metaphysik der Sitten* (1797; 2nd ed., 1798–1803; *The Metaphysic of Morals,* 2 vol., 1799 and 1965; *The Metaphysic of Ethics,* 1836), comprising *Metaphysische Anfangsgründe der Rechtslehre* (*The Philosophy of Law,* 1887) and *Metaphysische Anfangsgründe der Tugenlehre* (*The Doctrine of Virtue,* 1964); *Der Streit der Facultäten* (1798); *Von der Macht des Gemüths durch den blossen Vorsatz seiner krankhaften Gefühle Meister zu seyn* (1798; *Kant on the Art of Preventing Diseases,* 1806); *Anthropologie in pragmatischer Hinsicht abgefasst* (1798; improved ed., 1800; *The Classification of Mental Disorders,* 1964); *Immanuel Kants Physische Geographie,* 3 vol. in 6 pt., 1801–04); *I. Kants Logik: Ein Handbuch zu Vorlesungen* (1800; *Logic,* 1819); *Immanuel Kant über Pädagogik* (1803; *Kant on Education,* 1899; *The Educational Theory of Immanuel Kant,* 1904; *Education,* 1960); *Welches sind die wirklichen Fortschritte, die Metaphysik seit Leibnizens und Wolf's Zeiten in Deutschland gemacht hat?* (1804).

BIBLIOGRAPHY

Biography: The main sources for Kant's life are three memoirs published in 1804: LUDWIG ERNEST VON BOROWSKI, *Darstellung des Lebens und Charakters Immanuel Kants;* REINHOLD B. JACHMANN, *Immanuel Kant geschildert in Briefen an einen Freund;* and CHRISTOPH WASIANSKI, *Immanuel Kant in seinen letzten Lebensjahren* (the basis of THOMAS DE QUINCEY'S "The Last Days of Kant" included in his *Works*). See also J.H.W. STUCKENBERG, *The Life of Immanuel Kant* (1882); FRIEDRICH PAULSEN, *Immanuel Kant: Sein Leben und seine Lehre,* 2nd–3rd ed. (1899; Eng. trans., *Immanuel Kant: His Life and Doctrine,* 1902); ERNST CASSIRER, *Kants Leben und Lehre* (1921); and KARL VORLANDER, *Immanuel Kants Leben* (1921) and *Immanuel Kant: Der Mann und das Werk,* 2 vol. (1924).

Editions: The standard edition of Kant's works is that of the Berlin Academy, 1st ed., 9 vol. (1902–23; vol. 10–14, 1900-02, 1911), 4th ed., much augmented, in progress in the 1970s, which contains Kant's correspondence and literary remains as well as his published writings. There are also modern collected editions by ERNST CASSIRER, 11 vol. (1912–22) and by KARL VORLANDER, 10 vol. (1920–29). The most convenient edition of the *Critique of Pure Reason* is that by RAYMOND SCHMIDT (1926).

Aids to study: RUDOLF EISLER, *Kant-lexicon* (1930); HEINRICH RATKE, *Systematisches Handlexikon zu Kants Kritik der reinen Vernunft* (1929, reprinted 1965).

General works: The best introduction is STEPHAN KORNER, *Kant* (1955). See also EDWARD CAIRD, *The Critical Philosophy of Kant,* 2 vol. (1889, reprinted 1969); KUNO FISCHER, *Kants Leben und die Grundlagen seiner Lehre* (1860); ALOIS RIEHL, *Der philosophische Kriticismus . . . ,* 3rd ed., 3 vol. (1924–26); BRUNO BAUCH, *Immanuel Kant,* 4th ed. (1933), in German; HEINRICH RICKERT, *Kant als Philosoph der modernen Kultur* (1924); MAX WUNDT, *Kant als Metaphysiker* (1924); MARTIN HEIDEGGER, *Kant und das Problem der Metaphysik* (1929; 3rd ed., 1965; Eng. trans., *Kant and the Problem of Metaphysics,* 1962); HERMAN J. DE VLEESCHAUWER, *La Déduction transcendentale dans l'oeuvre de Kant,* 3 vol. (1934–37), and *L'Évolution de la pensée kantienne* (1939); PANTALEO CARABELLESE, *Il problema della filosofia in Kant* (1938); GOTTFRIED MARTIN, *Immanuel Kant* (1951; 4th ed., 1969; Eng. trans., *Kant's Metaphysics and Theory of Science,* 1955); HEINZ HEIMSOETH, *Studien zur Philosophie Immanuel Kants* (1956); RICHARD KRONER, *Kant's Weltanschauung* (Eng. trans. 1956); FRIEDRICH DELEKAT, *Immanuel Kant,* 3rd ed. (1969), in German.

Precritical writings: MARIANO CAMPO, *La genesi del criticismo kantiano* (1953); GIORGIO TONELLI, *Elementi metodologici e metafisici in Kant dal 1745 al 1768* (1959).

The "Critique of Pure Reason": NORMAN KEMP SMITH, *A Commentary to Kant's "Critique of Pure Reason,"* 2nd ed. rev. (1923, reprinted 1962); H.J. PATON, *Kant's Metaphysic of Experience,* 2 vol. (1936); ALFRED C. EWING, *A Short Commentary on Kant's Critique of Pure Reason* (1938, reprinted 1967); THOMAS D. WELDON, *Kant's Critique of Pure Reason,* 2nd ed. (1958); HERMANN COHEN, *Kants Theorie der Erfahrung,* 4th ed. (1925), and *Kommentar zu Immanuel Kants Kritik der reinen Vernunft,* 4th ed. (1925); HANS VAIHINGER, *Kommentar zu Kants Kritik der reinen Vernunft,* 2nd ed., 2 vol. (1922); HEINZ HEIMSOETH, *Transzendentale Dialektik,* 3 vol. (1966–69). See also GRAHAM BIRD, *Kant's Theory of Knowledge* (1962); PETER F. STRAWSON, *The Bounds of Sense* (1966).

Ethical writings: PAUL A. SCHILPP, *Kant's Pre-Critical Ethics,* 2nd ed. (1960); H.J. PATON, *The Categorical Imperative* 1947); L.W. BECK, *A Commentary on Kant's Critique of Practical Reason* (1960); A.E. TEALE, *Kantian Ethics* (1951); WILLIAM D. ROSS, *Kant's Ethical Theory* (1954); HERMANN COHEN, *Kants Begründung der Ethik,* 2nd ed. (1910); MAX SCHELER, *Der Formalismus in der Ethik und die materiale Wertethik,* 5th ed. (1966). See also PAUL MENZER'S ed. of *Eine Vorlesung Kants über Ethik* (1924; Eng. trans., *Lectures on Ethics by Immanuel Kant,* 1930); MORRIS STOCKHAMMER, *Kants Zurechnungsidee und Freiheitsantinomie* (1961).

The "Critique of Judgment": HEINRICH W. CASSIRER, *A Commentary on Kant's Critique of Judgment* (1938, reprinted 1970); ALFRED BAEUMLER, *Kants Kritik der Urteilskraft* (1923).

Particular problems: (Science): ERICH ADICKES, *Kant als Naturforscher,* 2 vol. (1924–25); JULES VUILLEMIN, *Physique et métaphysique kantiennes* (1955). *(Ontology):* CHRISTOPHER B. GARNETT, *The Kantian Philosophy of Space* (1939, reprinted 1965); MARTIN HEIDEGGER, *Kants These über das Sein* (1963) and *Die Frage nach dem Ding: Zu Kants Lehre von den transzendentalen Grundsätzen* (1962). *(Philosophy of history):* KLAUS WEYAND, *Kants Geschichtsphilosophie: Ihre Entwicklung und ihr Verhältnis zur Aufklärung* (1963). *(Political philosophy):* GEORGES VLACHOS, *La Pensée politique de Kant* (1962). *(Religion):* C.C.J. WEBB, *Kant's Philosophy of Religion* (1926); JOSEF BOHATEC, *Die Religionsphilosophie Kants . . .* (1938, reprinted 1966). *(Comparative studies):* JOHANNES B. LOTZ (ed.), *Kant und die Scholastik heute* (1955); KARL JASPERS, *Die grossen Philosophen,* vol. 1 (1957; abridged Eng. trans., *The Great Philosophers,* ed. by HANNAH ARENDT, vol. 1, 1966).

Opus postumum: ERICH ADICKES, *Kants Opus Postumum* (1920); GERHARD LEHMANN, *Kants Nachlasswerk und die Kritik der Urteilskraft* (1939).

Journals: Many important works of Kantian scholarship have been published in the periodical *Kantstudien.*

(O.A.B.)

Kantianism

As a philosophical designation, Kantianism can signify either the system of thought contained in the writings of the epoch-making 18th-century philosopher Immanuel Kant or those later philosophies that arose from the study of Kant's writings and drew their inspiration from his principles. Only the latter is the concern of this article.

NATURE AND TYPES OF KANTIANISM

The Kantian movement comprises a loose assemblage of rather diverse philosophies that share Kant's concern with exploring the nature, and especially the limits, of human knowledge in the hope of raising philosophy to the level of a science in some sense similar to mathematics and physics. Participating in the critical spirit and method of Kant, these philosophies are thus opposed to dogmatism, to expansive speculative naturalism (such as that of Baruch Spinoza, the Jewish Rationalist), and, usually, to irrationalism. The various submovements of Kantianism are characterized by their sharing of certain "family resemblances"; *i.e.*, by the preoccupation of each with its own selection of concerns from among the many developments of Kant's philosophy: a concern, for example, with the nature of empirical knowledge; with the way in which the mind imposes its own categorial structure upon experience, and, in particular, with the nature of the structure that renders man's knowledge and moral action possible, a structure considered to be a priori (logically independent of experience); with the status of the *Ding-an-sich* ("thing-in-itself"), that more ultimate reality that presumably lurks behind man's apprehension of an object; or with the relationship between knowledge and morality. **A** brief exposition of Kant's philosophical system may be found in KANT, IMMANUEL.

A system such as the critical philosophy of Kant freely lends itself to reconstructions of its synthesis according to whatever preferences the private philosophical inclinations of the reader may impose or suggest. Kant's system was a syncretism, or union, of British Empiricism (as in John Locke, George Berkeley, and David Hume) that stressed the role of experience in the rise of knowledge; of the scientific methodology of Isaac Newton; and of the metaphysical apriorism (or Rationalism) of Christian Wolff, who systematized the philosophy of Gottfried Leibniz, with its emphasis on mind. Thus it constituted a synthesis of elements very different in origin and nature, which tempted the student to read his own presuppositions into it.

Varieties of Kantianism

The critical philosophy has been subjected to a variety of approaches and methods of interpretation. These can be reduced to three fundamental types: those that conceive of the critical philosophy as an epistemology or a pure theory of (scientific) knowledge and methodology; those that conceive of it as a critical theory of metaphysics or the nature of Being (ultimate reality); and those that conceive of it as a theory of normative or valuational reflection parallel to that of ethics (in the field of action). Each of these types—known, respectively, as epistemological, metaphysical, and axiological Kantianism—can, in turn, be subdivided into several secondary approaches. Historically, epistemological Kantianism included such different attitudes as empirical Kantianism, rooted either in physiological or psychological inquiries; the logistic Kantianism of the Marburg school, which stressed essences and the use of logic; and the realistic Kantianism of the Austrian Alois Riehl. Metaphysical Kantianism developed from the transcendental Idealism of German Romanticism to Realism, a course followed by many speculative thinkers; who—like nearly all contemporary Kantians—saw in the critical philosophy the foundations of an essentially inductive metaphysics, in accordance with the results of the modern sciences. Finally, axiological Kantianism—concerned with value theory—branched, first, into an axiological approach (properly so-called), which interpreted the methods of all three of Kant's *Critiques* (*i.e., Critique of Pure Reason, Critique of Practical Reason,* and *Critique of Judgment)* as normative disciplines of thought; and, second, into an eclectic or relativistic Kantianism, which regarded the critical philosophy as a system of thought dependent upon social, cultural, and historical conditions.

The chief representatives of these submovements are identified in the historical sections below.

It is essential to distinguish clearly between two periods within the Kantian movement: first, the period from 1790 to 1831 (the death of Hegel); and, second, the period from 1860 to the present—separated by a time when an antiphilosophical Positivism, a type of thought that supplanted metaphysics with science, was predominant. The first period began with the thorough study and emendation of Kant's chief theoretical work, *Kritik der reinen Vernunft* (2nd ed., 1787; the *Critique of Pure Reason,* 1929); but it soon became intermingled with the romantic tendencies in German Idealism. The second period, called specifically Neo-Kantianism, was, first of all, a conscious reappraisal, in whole or in part, of the theoretical *Critique*, but also, as a total system, a reaction against Positivism. Earlier Neo-Kantianism reduced philosophy to the theory of knowledge and scientific methodology; systematic Neo-Kantianism, arising at the beginning of the 20th century, expressed itself in attempts at building metaphysical structures.

EARLY KANTIANISM: 17961835

Kant himself

According to Immanuel Kant (1724–1804), his major work, the *Critique of Pure Reason*, comprised a treatise on methodology, a preliminary investigation prerequisite to the study of science, which placed the Newtonian method (induction, inference, and generalization) over against that of Descartes and Wolff (deduction from intuitions asserted to be self-evident). The result was a critique of metaphysics, the value of which lay not in science but in a realm of being accessible only to the pure intellect. In exploring this "noumenal" realm, as he called it, Kant placed his *Critique* in a positive role. Recalling the revolution that occurred in astronomy when Nicolaus Copernicus discerned, in the apparent motions of the planets, reflections of the earth's own motion, Kant inaugurated a Copernican revolution in philosophy, which claimed that the subject doing the knowing constitutes, to a considerable extent, the object; *i.e.*, that knowledge is in part constituted by a priori or transcendental factors (contributed by the mind itself), which the mind imposes upon the data of experience. Far from being a description of an external reality, knowledge is, to Kant, the product of the knowing subject. When the data are those of sense experience, the transcendental (mental) apparatus constitutes man's experience or his science, or makes it to be such. These transcendental elements are of three different orders: at the lowest level are the forms of space and time (technically called intuitions); above these are the categories and principles of man's intelligence (among them substance, causality, and necessity); and at the uppermost level of abstraction the ideas of reason—the transcendental "I," the world as a whole, and God. It is by virtue of the encounter between the forms of man's sensory intuition (space and time) and his perceptions that phenomena are formed. The forms arise from the subject himself; the perceptions, however—or the data of experience—have reference, ultimately, to things-in-themselves, which nevertheless remain unknowable, inasmuch as, in order to be known at all, it is necessary for things to appear clothed, as it were, in the forms of man's intuition and, thenceforth, to present themselves as phenomena and not as noumena. The thing-in-itself, accordingly, indicates the limit and not the object of knowledge.

Early criticism

These theses of Kant provoked criticism among the followers of Christian Wolff, the Leibnizian Rationalist, and doubts among the disciples of Kant, which, as they further developed into systems, marked the first period of Kantianism. Inasmuch as these disciples took the *Critique of Pure Reason* to be a *preface* to the study of the pure reason or of the transcendental system and not the system itself, they saw in this interpretation an explanation for the ambiguities to which the *Critique* (as they felt) was subject. Their doubts revolved around two points: first, Kant had erroneously distinguished three kinds of a priori knowledge, coordinate with the three aforementioned levels or faculties of the mind; and second, Kant had accepted the thing-in-itself as constitutive of knowledge. Regarding the first point, they claimed that Kant had accepted the three faculties and their respective transcendental characteristics without investigation, in which case this structure should be viewed, in accordance with the preliminary character of the *Critique,* as a triple manifestation of a single fundamental faculty. For this reason the distinction between the levels of intuition and understanding (or between the receptivity and spontaneity of the mind) had to be rejected—for the three transcendentals—space and time, the categories, and the ideas of reason—were not existents but were only functions of thought. Finally, these disciples argued that the existence of a single transcendental subject, the Ego, would render the thing-in-itself superfluous and even pernicious for the scientific treatment of epistemology.

This function of human thought (the transcendental subject), which serves as the absolute source of the a priori, was variously designated by different early Kantian thinkers: for the German Realist Karl L. Reinhold, it constituted the faculty of representation; for the Lithuanian Idealist Salomon Maimon, it was a mental capacity for constructing objects; for the Idealist Jakob S. Beck, a protégé of Kant's, it was the act of synthesis; for the empirical critic of Kantianism G.E. Schulze, it was experience in the sense intended by Hume—a volley of discrete sense impressions; for the theory of knowledge of the outstanding ethical Idealist Johann G. Fichte, it was the original positing of the Ego and the non-Ego—which meant, in turn, in the case of the aesthetic Idealist F.W.J. von Schelling, "the absolute self," and in the case of G.W.F. Hegel "the *Geist* or absolute Spirit," and finally, in the case of the pessimistic Romanticist Arthur Schopenhauer, "the absolute Will." In each case (excepting Schulze) the interpretation of the thing-in-itself in a realistic metaphysical sense was rejected in favour of various degrees of transcendental Idealism. Removed from the main current of Kantianism was the empirically oriented thinker Jakob Friedrich Fries (the one figure in this group who was not an Idealist in the true sense), who interpreted the a priori in terms of psychological faculties and elements.

Having earlier renounced these apostates on a large scale, Kant, at the end of his life, prepared a new exposition of the transcendental philosophy (the second part of his *Opus Postumum),* which showed that he was ready tacitly to accede to the criticisms of his adversaries.

NEO-KANTIANISM: SINCE 1860

General features of Neo-Kantianism

Nineteenth-century Neo-Kantianism. The rejection of all of philosophy by Positivism had the anomalous effect of, itself, evoking an awakening of Kantianism, for many thinkers wished to give to Positivism itself a philosophical foundation that, while respecting the phenomenological attitude, would yet be hostile to the metaphysics of Positivism, which was usually a tacit, but inconsequent, Materialism. It was justifiably held that Kant could provide such a foundation because of his opposition to metaphysics and his limitation of scientific knowledge to the sphere of phenomena. The complexity of the critical philosophy was such that the theoretical criticism could be approached in diverse ways and that, through the facts themselves, diverse interpretations of the *Critique of Pure Reason* could be obtained. In the order of their origin (though not of their worth or importance), there thus arose currents of Kantianism that were empiricist, logicist, realist, metaphysical, axiological, and psychological—of which the most important have survived into the 20th century.

The return to Kant was determined by the historical fresco of the incomparable historian of philosophy Kuno Fischer entitled *Kants Leben und die Grundlagen seiner Lehre* (1860; "Kant's Life and the Foundations of his Teaching"), which replaced the earlier work of the semi-Kantian Ernst Reinhold, son of the more notable Jena scholar (published 1828–30), and especially that of the outstanding historian of philosophy Johann Eduard Erdmann (published 1834–53). In 1865 the order: "Zuriick nach Kant!" ("Back to Kant!") reverberated through the celebrated work of the young epistemologist Otto Liebmann, *Kant und die Epigonen* ("Kant and his Followers"), which was destined to extricate their spirits from the Positivistic morass and, at the same time, to divert the Germans from romantic Idealism.

Schools of Neo-Kantianism

Epistemological Neo-Kantianism. The empiricist, logistic, and realistic schools can be classed as epistemological.

Empiricist Neo-Kantianism was represented by the erudite pioneering physicist and physiologist Hermann L.F. von Helmholtz and, in part, by F.A. Lange, author of a famous study of Materialism. Helmholtz found support in Kant for his claim, first, that, although perception can *represent* an external thing, it usually does so in a way far removed from an actual description of its properties; second, that space and time comprise an empirical framework created for thought by the perceiving subject; and, third, that causality is an a priori law allowing the philosopher to infer a reality that is absolutely unknowable. Smilarly, Lange reduced science to the phenomenal level and repudiated the thing-in-itself.

Logistic Neo-Kantianism, as represented in the most well-known and flourishing school of Kantianism, that at Marburg, originated with Hermann Cohen, successor of Lange at Marburg, who, in a book on Kant (1871), argued that the transcendental subject is not to be regarded as a psychic being but as a logical function of thought that constructs both the form *and* the content of knowledge. Nothing is *gegeben* ("given"), he urged; all is *aufgegeben* ("propounded," like a riddle) to thought—as when, in the infinitesimal calculus, the analyst generates motion by imagining thin slices of space and time and adding up their areas. Hence experience is a perfect construction of man's logical spirit. The example of Cohen inspired many other authors, among them Cohen's colleague at Marburg Paul Natorp, who, in his work on the logical foundations of the exact sciences, integrated even psychology into the Marburgian transcendentalism; and Ernst Cassirer, best known for stressing the symbolizing capacities of man, who, in his memorable work *Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit* (1906–20; *The Problem of Knowledge: Philosophy, Science, and History since Hegel,* 1966), transposed this same logisticism into a form that illumines the history of modern philosophy.

Realistic Neo-Kantianism, the third manifestation of epistemological Neo-Kantianism, was represented in the Realism of the scientific monist Alois Riehl and of his disciple Richard Honigswald. In a work on the significance of the critical philosophy for the positive sciences (published 1876–87), Riehl held, in direct opposition to the Marburgian logisticism, that the thing-in-itself participates positively in the constitution of knowledge inasmuch as all perception includes a reference to things outside the subject.

Metaphysical Neo-Kantianism. Ten years after the appearance of the aforementioned ground-breaking book *Kant und die Epigonen,* its author, Otto Liebmann, introduced the new metaphysical approach in his book on the analysis of reality (1876), which came near to the Kantianism of Marburg. The Romanticist Johannes Volkelt, in turn, took up the theme of a critical metaphysics and expressed his persisting aspirations toward the Absolute in the claim that, beyond the certainties of man's own subjective consciousness, there exists a new kind of certainty in a transsubjective realm. Subjectivity is, thus, inevitably transcended, just as the sciences are surmounted when they presuppose a metaphysics. The influential spiritual moralist Friedrich Paulsen defended the claim that Kant had always behaved as a metaphysician, even in the *Critique of Pure Reason,* in spite of the epistemological restrictions that he imposed upon himself—a claim that made an impact that was felt throughout the following century.

Axiological Neo-Kantianism. Inasmuch as the two principal representatives of the axiological interpretation both taught at Heidelberg, this branch is also known as the Southwest German or Baden school. Its initiator was Wilhelm Windelband, esteemed for his "problems" approach to the history of philosophy. The scholar who systematized this position was his successor Heinrich Rickert, who had come from the tradition of Kuno Fischer. Drawing a parallel between the constraints that logic exerts upon thought and those that the sense of ought exerts upon ethical action, these thinkers argued that, while man's *action* must answer to an absolute value (the Good), his *thought* must answer to a regulative value (the True), which imposes upon him the duty of conforming to it. The *Critique of Pure Reason,* they held, elaborates this rule—which is not an entity but an imperative, or absolute, charge to act. Rickert regarded the critical endeavour as having been too narrow, since it was suited merely to physics. Actually, he charged, it should be the foundation for all of the sciences of the spirit. The distinctive characteristic of this school thus consisted in reintegrating German Idealism (as in Fichte and Hegel) into a rather personal Kantianism. Consequently, it succeeded in annexing more than one area of semi-Kantian thought: *e.g.,* "the philosophy of the spiritual sciences" of Wilhelm Dilthey, who held that intellectual life cannot be explained by means of naturalistic causality but only through historical understanding (*Verstehen*); "the life-philosophy" of the social philosopher Georg Simmel, who deviated from an earlier naturalistic relativism to the espousal of objective values; "the philosophy of value" of the experimental psychologist Hugo Munsterberg, author of one of the earliest systems of values; the "semi-Hegelianism" of Richard Kroner, a philosopher of culture and religion; and the general works of Bruno Bauch, Liebmann's successor at Jena. All of these philosophers were more or less related to axiological Neo-Kantianism.

Psychological Neo-Kantianism. An initial attempt to interpret Kantian transcendentalism in psychological terms was made by the Friesian Empiricist Jürgen Bona Meyer in his *Kants Psychologie* (1870). Later, a more important contribution in this field was made by the Göttingen philosopher of ethics and law Leonard Nelson and published in the *Abhandlungen der Fries'schen Schule* (1904 ff; "Acts of the Friesian School"). Even this title suggests an intimate agreement with the Kantianism of Fries's new critique of reason (1807); and Nelson, indeed, is regarded as the founder of the Neo-Friesian school. At a time when other Kantian schools were concerned with the transcendental analysis of objective or outer knowledge, Nelson held that, in the analysis of the subjective or inner self, the transcendental equipment of the mind—the a priori—is directly revealed. It thus fell to psychology to lay bare this equipment, which belongs in itself to the metaphysical order. It was upon this basis that the Marburg theologian Rudolf Otto, in his book *Das Heilige* (1917; *The Idea of the Holy,* 1958), ventured a type of religious phenomenology that has proved very successful.

Kantian philology

A discipline known as the Kant *Philologie,* concerned with the history, development, and works of Kant, has pre-empted a considerable portion of philosophical historiography since 1860. These studies began with the immense commentary on the *Critique of Pure Reason* produced in 1881–92 by Hans Vaihinger, known for his philosophy of the "As If" (which stresses man's reliance on pragmatic fictions), and with the founding of the new journal *Katttstudien* (1896) and the Kant-Gesellschaft ("Kantian Society," 1904)—both of which are still extant. The most conspicuous result of this philological movement, however, was undeniably the monumental edition, in 23 volumes, of all of Kant's available works prepared (1900 ff) by the Academy of Sciences at Berlin under the editorship of the champion of humanistic studies, Wilhelm Dilthey. These volumes include: Sect. 1, Works; Sect. 2, Correspondence; Sect. 3, The "Nachlass." Since the transfer of this task to the University of Münster, Sect. 4, Kant's Lectures, has been undertaken. Those on logic and metaphysics (vols. 24–25) have already been splendidly edited by Gerhard Lehmann.

Contemporary Neo-Kantianism. The recent development of Neo-Kantianism, except for innumerable historical studies, is very one-sided: no longer considered as exclusively epistemological, it merely prolongs the metaphysical school. Moreover, a large portion of the present Kant research is covered by the so-called *Problems of Kantianism* (see below). Important studies have been made on the development of Kant's philosophical thought, on Kant as a metaphysician, on his ontology and teachings on science, and on his transcendental deduction.

NON-GERMAN KANTIANISM

The Kantian awakening, in no wise limited to Germany, extended throughout Western philosophy. Its principal initiators were as follows: France was the first to open to its influence, beginning with the eclectic thinker Victor Cousin, who had studied German authors and made several trips to Germany. The relativistic personalist Charles Renouvier then defended a rather personal critical philosophy, which exerted an enduring influence through its impact upon the extreme Idealist Octave Hamelin of the Sorbonne; upon the metaphysician and cofounder of French neospiritualism Jules Lachelier; and upon his pupil, the philosopher of science Émile Boutroux.

English Kantianism

The English-speaking countries, on the other hand, have not seemed disposed to assimilate the critical philosophy as they did Hegelian Idealism. Except for the Scottish religious absolutist Edward Caird (*The Critical Philosophy of Immanuel Kant,* 1889), who was chiefly an Hegelian, there was in Britain at the close of the 19th century only another Scot, the critical Realist Robert Adamson, who was a Kantian. After him, however, can be cited the commentary, published in 1918, of Norman Kemp Smith, producer of the standard English translation of Kant's first *Critique,* and more recently, the remarkable exposition by the Oxford Kantian Herbert J. Paton, *Kant's Metaphysic of Experience* (2 volumes, 1936). Finally, Kantian methods can be discerned today in the later work of the prominent Oxford "ordinary language" philosopher, Peter F. Strawson, entitled *Individuals: An Essay in Descriptive Metaphysics* (1959). Kantianism became known in the United States toward 1840 primarily through the New England transcendentalist and poet Ralph Waldo Emerson—who was not, however, a Kantian himself. The physicist and logician Charles Sanders Peirce owes his Pragmatism largely to Kant's role as a counterweight against Hegelianism. The former southern California philosopher William H. Werkmeister represents a type of Neo-Kantianism inspired by the Marburg school (*The Basis and Structure of Knowledge,* 1948).

Italian scholars, on the other hand, have been vigorously engaged in Kantian studies since the initiative was taken by Alfonso Testa. The chief Neo-Kantian in Italy, however, was the Realist Carlo Cantoni, who took an anti-Positivist stance. Later, in the period from 1900 to 1918, Kantianism was represented by the extreme Realism of the theist Francesco Orestano. A school of Kantian philology has formed at Turin around the erudite Christian Idealist Augusto Guzzo and his journal *Filosofia.* More independent in spirit is the work of the critical ontologist Pantaleo Carabellese, Giovanni Gentile's successor at Rome.

ASSESSMENT OF KANTIANISM

At the present time there does not exist, either in Germany or elsewhere, a purely Kantian philosopher; but all acknowledge the obligation to come to grips with him. Within the four great currents of contemporary thought, however—*i.e.,* in Phenomenology, in the traditionalistic metaphysics, in Existentialism, and in the Positivistic Empiricism of the Vienna Circle and of Analytical philosophy—the predominant attitude toward Kant is negative.

Problems of Kantianism. As far as epistemology is concerned, the critical philosophy constitutes a theory of science that agrees with current trends; for science must have a base that is empirical though also real. On the other hand, the transcendental or a priori is implicated; and severe complications ensue whenever the question is posed whether a type of apprehension can be acquired apart from experience that conveys, however, some new and genuine knowledge — whether, in short, synthetic a priori judgments can be made. Significantly, the founder of Phenomenology, Edmund Husserl, came back to the fold of Kantian transcendentalism after previously opposing it bitterly. As against the Kantian position, Empiricism entirely rejects the possibility (and even the meaning) of the synthetic a priori and, robbed thereby of everything traditionally regarded as the subject matter of philosophy, directs its philosophical inquiries principally to the study of language. The foremost recent analyst of language, however, the pioneering philosopher Ludwig Wittgenstein, imposed upon philosophy the obligation to limit reason (or the transcendental element in knowledge) — a semi-Kantian position, which he nonetheless later renounced. As for Existentialism, one of recent Germany's foremost philosophers, Martin Heidegger, has presented in his ***Kant und das Problem der Metaphysik*** (1929; Eng. trans., ***Kant and the Problem of Metaphysics***, 1962) a highly personalized interpretation. A student of Cohen at Marburg, the metaphysician Nicolai Hartmann, became the harbinger of the Realistic approach, elaborating in his analysis of the metaphysics of knowledge (1921) an ontological relation that he discerned to obtain between two forms of being: between thought and reality. Accordingly, the principles of thought correspond, in his view, to those of reality — a position at odds with Kant (even when he is interpreted as a Realist). Moreover, Hartmann treated the problems of mathematics — so urgent in current philosophy — in a manner again completely opposed to Kant; in particular, he questioned the validity of Kant's a priori intuition (or positing) of the spatio-temporal framework in terms of which man thinks about the world, challenging Kant at this point not merely to accommodate the non-Euclidean geometries (with curved space) that afforded a Realist alternative to the a priori but above all to reflect the distinctly logistic position regarding the foundations of mathematics to which he adhered. Discussion of the status of the thing-in-itself in man's knowledge of the real remained on the philosophical agenda both during and after Hartmann's time, but invoked the same indecision as it always had. At a time when Hartmann was accepting the thing-in-itself almost naïvely, Empiricism (in all its forms) rejected it categorically and attempted to construe the real in terms merely of what Kant had called phenomena. In the realm of ethics, Phenomenologists and Existentialists were dissatisfied with the purely formal character of Kant's ethics—*i.e.*, with its lack of specificity — and substituted a "material" ethic, of concrete duties, which was no less absolute than that of Kant. For their part, Empiricists were only interested in the analysis of expressions of moral judgment, which they reduced to imperative statements that are emotive and aimed at winning adherents.

Objections to Kantianism. It must be acknowledged that Kant has furnished many of the most significant themes that are found in the currents of contemporary philosophy, even in the forms that they still assume today. Yet, as compared with the state of affairs that existed from 1860 to 1918, Kantianism has suffered an impressive decline — though a slight recovery seems to have occurred during the third quarter of the 20th century.

Scientism, formalism, subjectivism

What were the reasons for this decline? In general, since World War I the reduction of philosophy to the philosophy of science has no longer been accepted, though contemporary Positivistic Empiricism has offered hardly any objection to it. The philosophy of science comprises, in fact, only ***one*** problem area, not the entire assemblage of philosophical problems. From this a second objection arises: Kantianism in general is too formalistic to satisfy man's actual inquisitiveness, which inclines more and more toward concrete concerns. Kantianism restricts itself to examining the a priori forms of thought and cares little for its diverse contents. Were this objection pertinent only to the exact sciences, it would not be serious — for these sciences attend to their own applications; but the objection becomes very grave for the field of ethics. For this reason, the objection against Kant's formalism has been raised most passionately against his ethical treatise, the ***Critique*** of ***Practical Reason—as*** by Hartmann, by the Phenomenologist Max Scheler, and by others. This transcendental formalism immediately encounters the further objection of subjectivism — in spite of efforts (from the side of logic) to evade it—*i.e.*, it is blamed for obstructing man's apprehension of the real universality of his Ego, of the thinking subject, and for inexorably impelling the scholar to the view that man's knowledge is merely the product of subjective construction. This subjectivistic transcendentalism, by its intrinsic logic, denies man access to the external world. Not only does it debar him from the world of things-in-themselves but it also prevents him from granting objective reality to phenomena as such, inasmuch as the transcendental source is here viewed as playing a constructive role with respect to experience and the phenomenon.

These three major objections, which stand out in the midst of many criticisms of minor details, recur constantly in the Kantian literature of the past quarter of a century. The result of these objections, as far as the evaluation of the critical philosophy is concerned, is that it is repudiated in its entirety — without, however, being thereby considered barred by limitation. Kant thus remains, in spite of everything, an inexhaustible source of problems and ideas, comparable in this respect to Plato and Aristotle, with whom he forms the great triad of Western philosophical thought.

BIBLIOGRAPHY. Though the literature on Kant himself comprises innumerable titles, that on Kantianism is relatively scanty. The only work that contains the complete history of Kantianism is ***Friedrich Ueberwegs Grundriss der Geschichte der Philosophie,*** 13th ed. (1953), volumes 3:606–620 and 4:1–128 for the first period and pp. 410–483 for the second. For the first period, there is an abundant literature. Of particular interest for the present purposes are: J.E. ERDMANN, ***Versuch einer wissenschaftlichen Darstellung der Geschichte der neueren Philosophie,*** 2nd ed., vol. 3–5 (1931); ERNST CASSIRER, ***Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit***, 2nd ed., vol. 3 (1923); G. LEHMANN, "Kant im Spatidealismus und die Anfänge der neukantischen Bewegung," in ***Zeitschrift für philosophische Forschung,*** 17:438–456 (1963); and two works by H.J. DE VLEESCHAUWER, ***La Deduction transcendentale dans I'oeuvre de Knnt,*** vol. 3 (1937), and ***L'Evolution de la pensée kantienne*** (1939; Eng. trans., ***The Development of Kantian Thought,*** 1962). For the second period, MARIANO CAMPO has begun the history in his ***Schizzo storico della esegesi e critica Kantiana,*** vol. 1 (1959); and summaries have been written by L.W. BECK in the ***Encyclopedia of Philosophy,*** 5:468–473 (1967); and by HERMANN NOACK in his ***Die Philosophie Westeuropas,*** pp. 143–196 (1962). See also the ***Enciclopedia Filosofica,*** new ed., vol. 3, col. 1225, and vol. 4, col. 953 (1969); the article by ERNST CASSIRER on the same subject in the ***Encyclopaedia Britannica,*** 14th ed., pp. 215–216 (1929); W. RITZEL, ***Studien zum Wandel der Kantauflassung*** (1952); HENRI DUSSORT, ***L'École de Marbourg*** (1963); and HEINRICH RICKERT, ***Die Heidelberger Tradition und Kants Kritizismus*** (1934). Some recent works on Kant are reviewed by H.J. DE VLEESCHAUWER in the ***Review of Metaphysics,*** vol. *2* (1957).

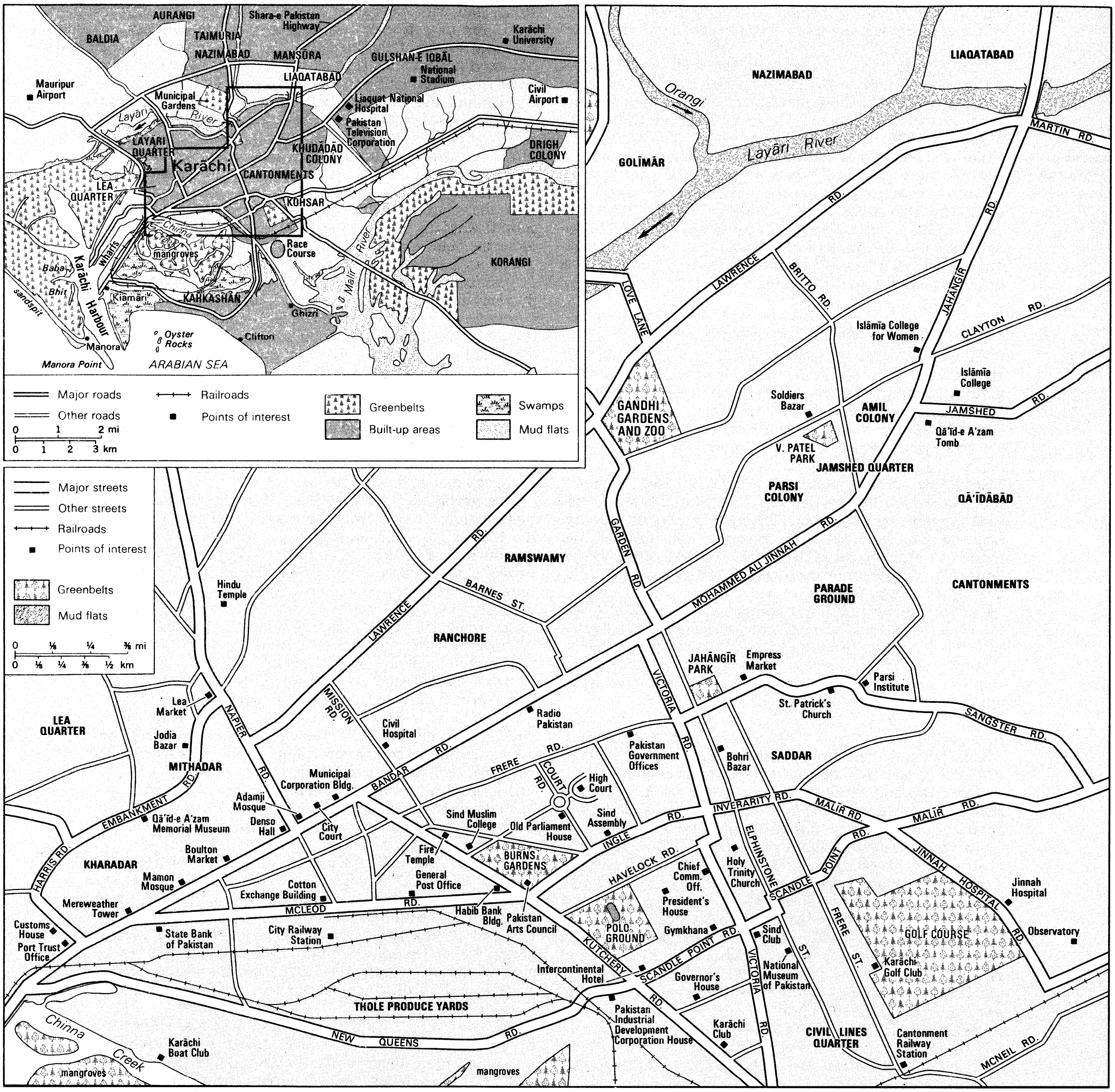
(H.J.deV.)

Karāchi

Karfichi, the principal seaport and the largest city in Pakistan, is located on the coast of the Arabian Sea immediately northwest of the Indus River Delta. It is the capital of the province of Sind as well as the headquarters of the district and division of Karfichi. It is also a major commercial and industrial centre. The city proper covers an area of 228 square miles (591 square kilometres), while the metropolitan area of Greater Karachi spreads out over an area of 560 square miles. The population is about 3,400,000 (metropolitan area).

The city has been variously called Caranjee, Crochey, Krotchey, Currachee, and Kurrachee. All its names are believed to be derived from the Sindhi name of the original settlement that initially stood on the spot — Kalachi-Jo-goth (meaning the village of Kalachi — the headman of the tribe).

The impetus to Karāchi's development originally came from its role as the port serving the Indus River Valley and the Punjab region of British India. The development



The city of KarHchi and (inset) its metropolitan area.

of air travel subsequently increased Karāchi's importance. It is also the port serving the landlocked country of Afghanistan. It is a crowded, bustling, optimistic city whose continued development and preeminence seem assured. (For an associated physical feature, see INDUS RIVER.)

HISTORY

Origins of the port

Karāchi was a small fishing village when a group of traders moved there in the early 18th century from the decaying port of Kharak Bandar nearby. It expanded rapidly, and was already of significance when it was captured in 1839 by the British, who annexed it in 1842, together with the province of Sind. It then became an army headquarters for the British, and also began to develop from a fishing village into the principal port for the Indus River region.

In 1843 a river-steamer service was introduced between Karachi and Multān, about 500 miles up the Indus. Port facilities were improved from 1854 onward. In 1861 a railway was built from Karāchi to Kotri, 90 miles upstream on the right bank of the Indus, opposite Hyderābād. In 1864 direct telegraph communications were established with London and with the interior. With the opening of the Suez Canal in 1869, the importance of Karachi grew, and it became a full-fledged seaport. By 1873 it possessed an efficient and well-managed harbour.

Karachi was connected directly with the hinterland when the railway line was extended from Kotri in 1878 to join the Delhi–Punjab railway system at Multān. In 1886, the Karachi Port Trust was established as the port authority, and between 1888 and 1910 the East Wharf—186,000 feet in length—was constructed. When the Punjab emerged as the granary of India in the 1890s, KarHchi became the region's principal outlet. By 1914, it had become the largest grain exporting port of the British Empire.

After World War I, manufacturing and service industries were installed. By 1924 an aerodrome had been built, and Karachi subsequently became the main airport of entry to India. The city became the provincial capital of Sind in 1936.

Growth after 1947

With the creation of Pakistan in 1947, Karachi not only became the capital and premier port of the new country but a nerve centre for industry, business, and administration. Although Rāwalpindi became the interim capital in 1959, some governmental agencies, including the Public Service Commission, as well as "skeleton" staffs of various ministries, remained in Karachi, which continued to play a role as a multifunctional city serving the entire country.

THE CONTEMPORARY CITY

The setting. *The city site.* Karachi Harbour, on the shores of which the city is situated, is a safe and beautiful natural harbour. It is protected from storms by Kiamāri Island, Manora Island, and Oyster Rocks, which together block the greater part of the harbour entrance in the west.

A low-lying coastal strip runs along the shore of the harbour. Inland, the ground rises gently to the north and east to form a large plain, from 5 to 120 feet above sea level, on which the city is built. The Malir River, a seasonal stream, passes through the eastern part of the city, and the Layāri River, also seasonal, runs through the most densely populated northern section. Some ridges and isolated hills occur in the north and east; Mango Pīr, the highest, is 585 feet high.

The 560 square miles that comprised the Federal Capital Area of Pakistan in 1948 are considered, for all practical purposes, to form the Karāchi metropolitan area. Almost half of the area is occupied by the city and its suburbs, and the surrounding 332 square miles consist of agricultural land and government wasteland.

Climate. Karachi has pleasant weather for the greater part of the year. May and June are the hottest months, when the mean maximum temperature is about 93° F (34" C). Spells of enervating weather occasionally prevail in May and October, during which the temperature shoots up to 105" F (41" C). The coolest months are January and February, during which the mean minimum temperature remains about 56° F (13° C). A biting north wind occasionally blows in these months and the temperature may drop to 40° F (4° C). The relative humidity varies from 58 percent in October, the driest month, to 82 percent in August, the wettest month. The average rainfall is eight inches; most of the rain falls during a total of nine or ten days in June, July, and August.

The city faces some pollution problems. High humidity does not permit evaporation of stagnant water in some places, while fumes from factories and automobiles contribute to air pollution, in spite of land and sea breezes.

Vegetation and animal life. The natural vegetation is scanty. Seaweed rises in tangles, and mangroves grow along some of the shores. Coarse grass, cactus, and castor plants occur on the plains and hills, and date and coconut palms grow in the river valleys.

The common wild animals are wolves, chinkaras (a type of gazelle), hog deers, jackals, wild cats, and hares. Domestic animals include sheep, goats, horses, and cows. Local birds include geese, ducks, snipe (game birds related to the woodcock), cranes, flamingos, and ibis (wading birds related to the heron). There are various types of snakes, particularly cobras, kraits, vipers, and python.

Layout and traditional regions. The most striking aspect of Karachi's layout is the west-to-east parallel alignment of the four arterial roads—Lawrence Road, Bandar Road, Frere Road, and McLeod Road. Beginning at Mereweather Tower in the vicinity of the port, these roads run through the centre of the city. Several roads, such as Napier Road, Kutchery Road, and Garden Road, cut across these arteries from north to south.

The old town

The old town lies near the port, to the north of Bandar Road, and with extensions stretching along the material roads for over a mile; unplanned, it is reminiscent of medieval towns of the Near East or Europe. East of the old town are such districts as the Drigh Cantonment, the Civil Lines (residential areas for senior civil service officers), and the Saddar Bazar. This area is planned on a checkerboard pattern, and shows European characteristics. Beyond this stretch several radial roads, along which recent growth has taken the form of neighbourhood units; each unit is laid out with straight, broad roads connected by smaller streets.

On the map, Karachi resembles a huge spider, with central Karachi forming the body, and the developments along the outer roads forming extended legs.

Land use. The land-use pattern of the city is complex. In the central area, the preponderance of residential property tends to form a matrix within which all other functions are set. There is, however, a marked concentration of commercial buildings at the western ends of Bandar Road and McLeod Road. Wholesale businesses are located in the old town, retail businesses along Bandar Road and in Saddar Bazar, the government offices on Frere Road, near Saddar. The outer areas are dominated by dormitory suburbs and planned industrial areas, interspersed with a scattering of cantonments (military quarters), agricultural tracts, salt works, airports, railway stations, and marshalling yards.

Transportation. *Roads.* The Karachi to Peshāwar highway links the city with the interior of Pakistan, while the Karachi to Zāhedān highway connects it with Iran and other Middle Eastern countries. There are 400 miles of roads in the central area of the city and a far greater road mileage (approximately 1,600 miles) in the outer area. Express roads radiate from the city centre, while feeder roads connect the express roads with local streets.

Railways. Karachi is the terminus of Pakistan's railway system, which mainly serves to transport goods between Karachi and the interior; about a dozen freight trains arrive daily, and about the same number leave. There are also a number of passenger trains, as well as a circular railway that skirts the city on the north and the east, catering to commuter traffic and the transport of goods between the port and the industrial areas.

Air transport. Karachi Airport is a major international traffic centre servicing flights between Europe, the Far East, and Australia; it is equipped with a runway for jet aircraft, and other facilities.

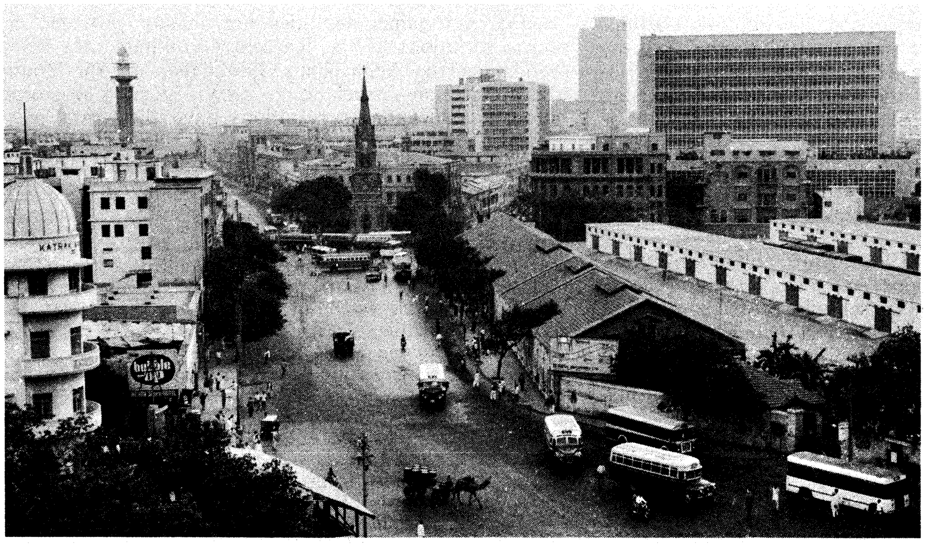
The port. The port of Karāchi is one of the busiest east of Suez. About 1,700 ships visit it annually. There are 19 general cargo berths—15 at the East Wharf and 4 at the West Wharf—with berth lengths ranging from 400 to 600 feet and depths from 30 to 34 feet. Three additional berths are under construction at West Wharf. There is also a separate wharf with 20 berths for shallow-draft craft, as well as oil-tanker berths, dry-dock facilities, and a shipyard.

Demography. In 1961 the population of Karachi numbered about 1,900,000, of which almost 1,100,000 were males and more than 800,000 were females. More than half of the inhabitants were migrants who came to the city after 1947 from India, from other cities in Pakistan, or from other countries.

The migrant population

The average population density in the city proper is 300 persons to the acre, although densities reach from 500 to 700 persons to the acre in some wards. In outlying areas the density is between 50 and 100 persons to the acre.

No ethnic group predominates in the city. Cultural and social activities essentially revolve around religion. Of the total population in 1961, 97 percent were Muslims, almost 2 percent were Christians, and 1 percent Hindus; there are also small Parsi, Buddhist, and Jain minorities. Most of the Muslims derive from Indo-Pakistani stock, except for "Makranis" and "Shiddies," who are also descended from Negro elements, and who originated during the era of the slave trade in the days before British rule, when Karachi was an important slave-trading centre. Some of the members of the Christian minority are of Indo-Pakistani origin, while others are descended from Portuguese or other European groups.



Mereweather Tower on McLeod Road in downtown Karlchi.
Frederic Ohringer — Nancy *Palmer* Agency

Housing. In the mid-1960s only 55 percent of the population possessed their own houses; the remainder lived in rented property. On the average, a family of four or five persons occupied 1.4 rooms. The city proper has old and decayed buildings, occupied by members of the middle and lower income groups. Further from the city centre are new bungalows occupied by richer persons; the outer zone is occupied by workers.

Building types. Karachi has a variety of types of building. The central area contains apartment bungalows, barracks, and multistoried buildings; the outer areas are characterized by bungalows, blocks of flats, and quarters (streets of small houses). Buildings of the British period were constructed with stone in western styles of architecture; other stone buildings in the central city show a blending of Eastern and Western styles, and have towels, domes, pillars, arches, hanging balconies, and rectangular courtyards. Buildings in the outer areas are built of cement blocks, and with few exceptions they show no uniformity in design. Some follow contemporary American design, while others incorporate features of traditional Muslim architecture.

Manufacturing

The economy. In the mid-1960s Karachi had more than 3,000 factories employing almost 113,000 workers, without including about 70 establishments employing 40,000 workers in Lāndhi and other suburbs. Textiles and footwear were the principal items manufactured, followed by such items as metal products, food and beverages, paper and printing, wood and furniture, machinery, chemicals and petroleum, leather and rubber, and electrical goods. Karāchi is also an important centre for handicrafts and cottage industries that produce handloomed cloth, lace, carpets, articles make of brass and bell metal (an alloy of copper and tin), pottery, leather goods, and gold and silver embroidery.

Karāchi handles the entire seaborne trade of Pakistan and, as mentioned, of landlocked Afghanistan. Between 1970 and 1971 total imports accounted for 6,279,000 tons of goods, and exports for 3,157,000 tons.

There are 29 large banks that have branches throughout Pakistan; these include the State Bank of Pakistan, the Habib Bank Ltd., the National Bank of Pakistan, the United Bank Ltd., the Industrial Development Bank of Pakistan, and the Agricultural Development Bank of Pakistan. The city is also the centre of about two dozen insurance companies, which play an important role in the economic development of the country by investing large sums in power development, housing programs, joint-stock companies, government loan securities, and savings certificates.

Karachi has a stock exchange that handles an overwhelming proportion of transactions in government securities and in the shares of most of the important industrial and financial institutions.

The twin bases of the city's prosperity are manufacturing and trade; together they employ 44 percent of the total labour force. Services and professions employ about 42 percent of the labour force, transport 9 percent, and agriculture 1 percent.

Political and governmental institutions. *Administration.* The city is administered by five institutions, the heads of which are appointed by the government. The Karachi Municipal Corporation, constituted in 1852, performs a large number of civic functions affecting more than three-fourths of the population of Greater Karachi. The Korangi-Lāndhi and Drigh-Malir municipal committees were established in 1966 and 1970, respectively, to provide civic facilities to the suburban areas developed after 1947. The Karāchi Cantonment Board, established in 1852, is the administrative body for the areas where the military are quartered. The Karachi Port Trust was constituted in 1886 to administer the affairs of the port, and is entrusted with the development and maintenance of the harbour.

Sources of water supply

Public utilities. The three main sources of the city's water supply are Lake Hāleji, 55 miles away, which is fed by the Indus River; wells that have been sunk in the dry bed of the Malir River, 18 miles away; and Lake Kalri, 60 miles away, which is also fed by the Indus waters. The city uses about 100,000,000 gallons of water a day, with Lake Kalri supplying about 70 percent of the total. Although the city's water mains have a total length of about 280 miles, some of the outer areas, such as Lāndhi, Malir, New Karachi, and Mauripur, still experience an acute water shortage.

The Karachi Electric Supply Corporation is responsible for electricity services. It has four power stations, all located in the city, with a total generating capacity of about 393,000 kilowatts; these stations use natural gas, diesel oil, or both. A nuclear power station has been built at Paradise Point, but in 1972 had not yet begun to function.

Karāchi Municipal Corporation maintains a fleet of vehicles for refuse collection, night soil removal, dog catching, and antimalarial and antifly operations. Sweepers are employed to clean the streets. Sewage is disposed of by two underground drainage systems, and there are two sewage treatment plants, one serving the city proper and the other the outlying areas.

Health and safety. In ***1964*** Karachi proper had 21 general hospitals, as well as five hospitals specializing in tuberculosis, skin diseases, leprosy, and epidemic diseases. There are also child-welfare centres and dispensaries, in addition to three more general hospitals in the suburbs.

The city has about two hospital beds for every 1,000 people.

There are several well-equipped fire-fighting stations; separate fire brigade units are attached to the Railway Department, which is a part of Pakistan's railway administration. In addition, the Port Trust and Pakistan International Airlines (PIA) have services that can be used in emergencies.

The police are administered by the Sind provincial administration; the inspector general of police is assisted by a force of more than 1,100. The city is divided into 40 police districts.

Education. In the early 1960s, Karachi had more than 900 schools, of which about 700 were primary schools and about 200 were secondary schools. More than half of all these are privately run, the rest being run by the government. Among schools established by different religious communities are Karachi Grammar School, St. Joseph's Convent School, and St. Patricks High School, all of which are Christian; a school for the Parsi community; and Sind Madrassa, a Muslim school.

The University of Karāchi

The University of Karachi, the primary institution of higher education, was established in 1951. It has 22 graduate departments in arts and sciences, as well as a graduate school of business administration. Courses in a variety of subjects, including commerce and law, are provided by about 75 colleges affiliated to the university. In addition, there is a medical college, as well as two engineering colleges, a polytechnic institute, a college of home economics, and two teacher-training colleges.

Cultural life. *Cultural institutions.* Pakistan Arts Council is the primary cultural institution in the city; it organizes various cultural functions including art exhibitions, and offers training in music. The Ghanshyam Art Centre and the Bulbul Academy promote Pakistani dancing and other cultural activities.

Karachi does not have well-established theatre, but amateur dramas and variety shows are frequently staged in Katrak Hall. Motion pictures are more popular, and there are more than 50 cinemas.

Karachi has a small museum containing relics of the early Indus Valley civilization and examples of the Greco-Buddhist art of Gandhāra (a region of ancient India in what is now northwestern Pakistan); it also possesses some ethnological collections representing life in different regions of Pakistan.

The library of the University of Karachi is the city's largest, but there are also ten public libraries containing books of a popular nature. Material of a more scholarly nature is to be found in the British Council Library, the American Information Centre Library, and the Liaqat Memorial Library. The departmental libraries of the State Bank of Pakistan, the Pakistan Institute of Development Economics, and the National Archives contain collections of books on economics and on national matters.

The media. Most of the national news agencies have their head offices in Karachi, and some of the larger newspapers are published in the city. Among the latter are *Dawn* and *The Morning News* (both published in English), and *Jang* and *Āghāz* (both in Urdu). There are also three Gujarati and two Sindhi newspapers, as well as several evening papers, biweeklies, periodicals, and journals published in various languages.

Television services were established in 1966, and programs are broadcast for four hours each evening. The Karāchi Radio Station, established in 1948, is equipped with medium and shortwave transmitters; programs include news, music, features, talks, and discussions. External broadcasts are received from the United Kingdom, South and East Africa, Afghanistan, and Middle Eastern and Southeast Asian countries.

Recreation. There is a general shortage of open spaces and parks in Karachi. Of the 19 existing parks, Gandhi Gardens and Burns Gardens are the most famous. There are a number of fine swimming and fishing beaches, such as Paradise Point, Hawks Bay, Sandspit, Manora, and Clifton.

The Karachi Zoo

The Karāchi Zoo is located in the Gandhi Gardens, and contains a varied collection of animals, birds, and reptiles. A 270-acre botanical garden is to be established along University Road.

Sports and games facilities are mostly provided by such associations as the Karachi Gymkhana, the Parsi Gymkhana, the Agha Khan Gymkhana, and the Young Men's Christian Association (YMCA). Various organizations and educational institutions have their own playgrounds. The largest sports area is the National Stadium, which contains playgrounds for cricket, hockey, football (soccer), and tennis. There are also boating, yachting, and flying clubs.

BIBLIOGRAPHY. NAZIR AHMAD, *Survey of Shelterless Persons in Karachi 1959* (1959), is a detailed report of the problems faced by the city due to the influx of refugees from India following the establishment of Pakistan. KARACHI DEVELOPMENT AUTHORITY, *The Greater Karachi Resettlement Housing Programme* (1961), analyzes the housing requirements, activities, and future needs in metropolitan Karāchi. A.F. BAILLIE, *Kurrachee: Past, Present and Future* (1890), gives Karachi's history and its growth in the early years of British rule. R.F. BURTON, *Scinde; or, the Unhappy Valley*, 2 vol. (1851), and *Sind Revisited*, 2 vol. (1877), are classic surveys of Sind under British rule. The *Census of Pakistan, 1961*, vol. 3 and the *District Census Report of Karachi, 1961*, provide official demographic data. H. FELDMAN, *Karachi Through a Hundred Years* (1960), written to commemorate the centenary of the Karachi Chamber of Commerce and Industry, gives details of the role played by the Chamber in the economic development of Karachi from 1860 to 1960. Karachi Development Authority MP Reports have been written for the preparation of the Master Plan of Karāchi and cover all aspects of modern city development. GUSTAV RANIS, *Industrial Efficiency and Economic Growth: A Case Study of Karachi* (1961), is an analysis of economic development; and IMTIAZUDDIN HUSAIN, MOHAMMED AFZAL, and AMJAD ALI BAHADUR RIZVI, *The Social Characteristics of the People of Karachi* (1965), a study of the socio-economic characteristics of the people.

(Z.A.K.)

Karakoram Range

A mountain system of Central Asia, the Karakoram Range (Turkish: "Black Rock Mountains") extends some 300 miles from the easternmost part of Afghanistan to the southeast. The borders of the Soviet Union, China, Pakistan, Afghanistan, and India all converge within the system.

Physiography. The Karakorams consist of a group of parallel ranges with several spurs. Only the central part is a monolithic range. The width of the system is about 150 miles; the length is increased from 300 to 500 miles, if the easternmost extension—the Ch'iang-ch'en-mo Shan (called Chāng Chenmo in Jammu and Kashmir) and Pantong ranges of the Tibetan Highlands—is included. The Karakoram Range is one of the highest mountain systems in the world; its average height is around 20,000 feet, and four peaks exceed 26,000 feet, the highest being K2 (Godwin Austen, Chogori, Dapsang) at 28,250 feet (8,611 metres). K2, the second highest peak in the world, was first climbed by an Italian expedition in 1954; the remaining high peaks were climbed in 1956–58.

Glaciation

The topography is characterized by craggy peaks and steep slopes. The southern slope is long and very steep, the northern slope steep and short. Cliffs and taluses (great accumulations of large, fallen rocks) occupy a vast area. In the intermontane valleys, rocky inclines occur widely. Transverse valleys usually have the appearance of narrow, deep, steep ravines. Because of their great height, the Karakorams are characterized by heavy glaciation, the total glaciated area amounting to 6,900 square miles (17,800 square kilometres). Glaciers occur on both slopes, but glaciation is more developed on the southern, more humid slope. The snow line on the southern slope of the Karakorams begins at an altitude of 15,400 feet; glaciers begin at 9,440 feet. On the northern slope the figures are 19,400 feet and 11,580 feet, respectively. Often, glaciers combine to form complex glacial systems occupying not only valleys but also the watersheds. Seasonal thawing of the glaciers give rise to serious floods on the southern slopes. Traces of ancient glaciation are evident at altitudes of 8,500 to 9,500 feet.

The Karakorams serve as a watershed for the basins of the Indus and Tarim rivers. The formation of river channels, for the most part, occurs in the high-altitude zone, the melted waters of seasonal and perpetual snows and glaciers being principal feeders of the rivers. Groundwaters accumulate in the rocky taluses and contribute to a more even flow throughout the year. During winter, huge layers of ice are formed.

Geology. Structurally, the Karakorams originated from folding in the Cenozoic Era (up to 65,000,000 years ago). Granites, gneisses, crystallized slates, and phyllites dominate the geological composition. To the south and north, the crystal centre of the Karakorams is edged by a region of limestones and micaceous slates of the Paleozoic and partly of the Mesozoic eras (from 190,000,000 to 570,000,000 years old). To the south the sedimentary rock is sometimes cut by intrusions of granite. Certain areas expose slate at the surface, which yields more rapidly to weathering.

At the end of the Mesozoic Era (65,000,000 years ago), the region of Karakorams was characterized by great structural changes, and the Karakorams emerged as the result of intensive, geologically recent upheavals. Today there is considerable seismic activity in the region, individual earthquakes exceeding nine points on the ten-point Richter scale. Hot springs are found in several areas.

There is also evidence of stannic-tungsten on the northern slopes of the Karakorams and of alluvial gold on the southern slopes.

Climate. The climate of the Karakoram Range is for the most part semi-arid and sharply continental. The southern slopes are exposed to the humidifying influence of the monsoons coming in from the Indian Ocean, but the northern slopes are extremely dry. In the lower and central part of the slopes, rain and snow is precipitated in small quantities; average annual precipitation does not exceed four inches. At altitudes of more than 16,000 feet precipitation always takes a solid form, but, even lower down, in June, snow is not infrequent. At altitudes of around 18,700 feet, the average temperature during the warmest month is lower than 32" F (0° C), and, at altitudes between 12,800 and 18,700 feet, the temperature is lower than 50" F (10" C). Rarified air, intensive solar radiation, strong winds, and great ranges of temperature are characteristic climatic features of the region.

High-altitude vegetation patterns

Vegetation and animal life. The high-altitude vegetation of the northern and southern slopes of the Karakorams is varied. On the northern slope, at altitudes of 7,900 to 9,200 feet on the rocky desert soil, a complex of plant combinations of such species as flowering plants of the genus *Kalidium,* and horsetail (genus *Equisetum)* has developed. Vast expanses, completely devoid of vegetation, are not infrequent. Only at the source of the Yarkand River (Tarim Basin) and its tributaries, up to altitudes of 10,000 feet are there individual thickets of brushwood (mainly barberry) and poplars. In the central part of the northern slope, at altitudes of 8,500–10,200 feet, a desert-steppe landscape is developed, with sparse thickets of coarse grasses and wintergreen. At altitudes of 10,500–11,500 feet, a mountain steppe predominates, and, in places that are most humid and well sheltered from winds, there are prairie steppes. Still higher up are found high-altitude expanses of wintergreen, wormwood, and meadows of desert-like plants. Sparse combinations of wintergreen and prickly herbs of the genus *Acanthus* are located on the coarse soils near the arable-land zone. On the moister southern slopes, more extensive and varied vegetation is found. Valleys up to 10,000–11,500 feet support forests of pine, Himalayan cedar, and, near streams, willows and poplars. Higher up, high-altitude steppes rather like typical alpine meadows predominate.

Notable animals of the region include the snow leopard, wild yak, and Tibetan antelope; in the southern foothills wild asses are also found. There are a great number of pikas and marmots. Among birds, the Pallas sand grouse, Tibetan capercaillie, partridge, ibis white dove, and red brambling are characteristic.

Human settlement. Being centred in an area of very high ranges covered with immense glaciers, the Karakorams are extremely inaccessible. Mountain passes are situated at altitudes of about 16,000 feet and are open only five or six months of the year. Sections along the banks of rivers and lakes are utilized as pasture, and in places, on the southern slopes, agriculture is developed, with two harvests in years when water is plentiful. Apricot orchards are extensive.

Exceptionally severe natural conditions in the Karakorams make life hard for man. An area of almost 80,000 square miles supports a population of only a few tens of thousands of people, mainly Tibetans, who live in villages at altitudes of up to 14,800 feet. Most of the Tibetans are farmers who grow barley, oats, and millet and who breed cattle. Individual groups of Tibetans lead a nomadic or seminomadic way of life and are occupied with the breeding of yaks, *dso* (a hybrid of the yak and common cow), sheep, and goats; they also do some hunting. Other peoples include the Baltis and Ladkhis, who, like the Tibetans, are Tibetan Buddhist. Also there are the Burishki people, who speak a language that occupies a singular place in linguistic classification. The Burishki are Muslims and are settled on farms,

BIBLIOGRAPHY. Personal accounts of exploration in the Karakorams include OSCAR ECKENSTEIN, *The Karakorams and Kashmir: An Account of a Journey* (1896); FANNY BULLOCK and WILLIAM HUNTER WORKMAN, *Ice-Bound Heights of the Mustagh: An Account of Two Seasons of Pioneer Exploration in the Baltistan Himalaya* (1908), and *Two Summers in the Ice-Wilds of Eastern Karakorum* (1917); *Explorations in the Eastern Kara-Koram and the Upper Yarkand Valley: Narrative Report of the Survey of India Detachment with the De Filippi Scientific Expedition* (1922); ARDITO DESIO, *La conquista del K2, seconda cima del mondo* (1954; Eng. trans., *Ascent of K2, Second Highest Peak in the World,* 1955; *U.S.* title, *Victory Over K2, Second Highest Peak in the World, 1956);* and the AMERICAN ALPINE CLUB, *K2, the Savage Mountain* (1954). See also the reports of the Italian expeditions to the Karakorams published by the Club Alpino Italiano.

(G.D.B.)

Kara-Kum Desert

Kara-Kum (Peski Karakumy in Russian, in the transliteration system of the Akademiya Nauk; Russianized Turkmen *karakumy,* "black sands"; *i.e.,* with vegetation) is the name given to a great sandy desert comprising about 60 percent of the area of the Turkmen Soviet Socialist Republic (*q.v.*) of the U.S.S.R. Another, smaller desert in the Soviet republic of Kazakhstan near the Aral Sea is called the Priaralskiye Karakumy.

The Turkmen Kara-Kum. The Turkmen Kara-Kum is approximately 100,000 square miles (300,000 square kilometres) in area. It is bordered on the north by the Sary-Kamysh Depression (Sarykamyshskaya Kotlovina), on the northeast and east by the Amu Darya (River) Valley, and on the southeast by the Karabil (Vozvyshennost) Highlands and Badkhyz semidesert. In the south and southwest the desert runs along the foot of the Kopet-Dag Mountains (Khrebet), and in the west and northwest it borders the course of the ancient valley of the Uzboy River. It is divided into three parts: the elevated northern Zaunguzskiye Karakumy; the central, low-lying Central Kara-Kum (Tsentralny Nizmennye Karakumy); and the southeastern Kara-Kum, through which runs a chain of salt-marshes. Across the border of the Zaunguzskiye and Central Kara-Kum passes the Unguz chain of saline, isolated, wind-eroded hollows.

The relief of the Turkmen Kara-Kum is quite sharp and clear-cut and reflects its origin and historical development. The surface of the Zaunguzskiye Karakumy has been eroded by violent winds. The plain of the Central Kara-Kum runs from the Amu Darya to the Caspian Sea along the same incline as the river. The height of wind-accumulated, half-overgrown, sand ridges ranges from five to 100 feet, depending on age and wind velocity. Somewhat less than 10 percent of the area consists of crescent-shaped dunes (*barkhany*), some of them 250 feet or more in height. There are numerous hollows of clay deposits, called *takyr,* and saline land formed by the evaporation of subsoil waters.

Climate of the Kara-Kum

The climate of the Turkmen Kara-Kum is continental, with long, hot summers and unpredictable but warm winters. The average temperature in July in the north and along the shore of the Caspian Sea ranges from 79° to 82° F (26" to 28° C), and in the central part of the Central Kara-Kum from 86° to 90° F (30" to 34° C). In January, average temperatures are 25" F (−4° C) in the north and 39° F (+4° C) in the south, but temperatures may fluctuate from as low as −4° F (−20" C) to +97° F (+36° C) within a 24-hour period. The average annual rainfall varies from 2.75 inches (70 millimetres) in the north to six inches (150 millimetres) in the south. Precipitation occurs mainly in autumn and early spring, more than half of it between November and April. There is little snow. The prevailing northeasterly and northwesterly winds are mild.

The vegetation is quite varied, consisting mainly of grass, small shrubs, bushes, and trees. The humid early spring permits the widespread growth of ephemeral plants—the main animal fodder—while in the barchan dune areas the typical vegetation consists of cereals, the wormwood shrub, and trees of the species *Ammodendron conollyi*. The most common bushes are species of *Astragalus*, *Calligonum*, and the salt-tree (*Salsola* richteri). In regions of deep underground water, the white haloxylon is the most typical plant, but in regions where water is nearer to the surface, the black haloxylon occurs. The vegetation of the Turkmen Kara-Kum can be used as hay in winter by camels, sheep, and goats.

Animals are not numerous, but they are of many kinds. The insects include ants, termites, ticks, beetles, darkling beetles, dung beetles, and spiders. Various species of lizards, snakes: and turtles also occur. The most common birds are skylarks, haloxylon (sacksal) jays, wagtails, and desert sparrows. Among the rodents are sand eels, gophers, and jerboas. The tolai hare, the hedgehog, barchan cat, and corsac fox and also the dzheran (a gazelle) usually live near takyr soils.

The population is composed of Turkmen, among whom some tribal distinctions have been preserved. Formerly nomadic, they settled into agricultural pursuits or fishing on the shores of the Caspian. Collective and state farms have developed permanent settlements, with gas and electricity. Cattle-raising brigades care for the socialized livestock. The development of oil, gas, and other industries has brought new settlements, populated by diverse nationalities. From 1939 to 1970 the population doubled.

Irrigation and economic development

Irrigation has made the desert suitable for the raising of livestock on a large scale, especially astrakhan sheep. The Karakum Canal (Karakumsky Kanal) has brought water to Southeastern Kara-Kum, the southern border of the Central Kara-Kum, and along the foot of the Kopet-Dag Mountains, where fine-fibred cotton is now grown in the oasis areas. The proposed extension of the canal to the shore of the Caspian Sea will permit the cultivation of the subtropical arid regions.

Intensive economic development after World War II brought an industrial revolution to the Kara-Kum. Factories, gas lines, railroads, and highways have changed the face of the region. Hydroelectric stations are under construction at Takhiyatas and Tyuyamuyun.

The Priaralskiye Kara-Kum. 'The Priaralskiye Kara-Kum in Kazakhstan is northeast of the Aral Sea and south of the Turgay Lowlands (Turgayskaya Dolina). Its altitude ranges from 174 to 410 feet above sea level. Its white and yellow sand contains only quartz and a little magnetite. Winding regions of half-exposed sand alternate with broad, flat-bottomed depressions and areas of grass. Here and there shallow reserves of fresh underground water occur, enabling the sands to be used as year-round pasture for sheep, goats, cows, horses, and Bactrian camels. (B.A.F.)

Kara Sea

A marginal sea of the Arctic Ocean, the Kara Sea (Karskoye More in Russian; spelled Karskoje More in the transliteration system of the Akademiya Nauk) is located between the coast of the West Siberian Plain and the Novaya Zemlya islands on the west, Franz Joseph Land on the northwest, and the Severnaya Zemlya islands on the east. In the north it is connected with the Arctic Basin, in the west with the Barents Sea, and in the east with the Laptev Sea. It has an area of about 340,000 square miles (880,000 square kilometres). The average depth is 417 feet (127 metres), the maximum depth 2,034 feet (620 metres), and the volume of water 27,000 cubic miles (113,000 cubic kilometres). The greatest extent of the sea from southwest to northeast is about 900 miles; the greatest width in the northern part is about 500 miles. Several deep inlets cut into the low, gently sloping bank of the mainland. The largest rivers flowing into the sea are the Yenisey, Ob, Pyasina, and Kara, from which the sea derives its name. The rivers have a total volume of about 310 cubic miles a year, 80 percent of which is in the summer.

Numerous islands dot the sea, totalling about 4,000 square miles in area. They are concentrated primarily in the northern part and include the Minina Islands, the Nordenshelda Islands (more than 70 islands), and others. The islands are diverse in origin and landscape. Some are mountainous, and the mountains on the islands of Vaygach and Novaya Zemlya seem to be a continuation of the Urals. In the mountains of Severnity Island, Novaya Zemlya, Franz Josef Land, and Severnaya Zemlya lie iceberg-calving glaciers. Their shores are steep and broken by fjords. Other islands—Ushakova and Shmidta, for example—are completely covered by icy domes. There are many low-lying, sandy islands.

Physical characteristics. The Kara Sea lies on the Siberian Shelf; thus, about 40 percent of it is less than 160 feet deep, and only 2 percent is more than 1,600 feet deep. The shelf is cut in the north by two wide, deep-sea troughs—the Svyatory Anny along the eastern shore of Franz Joseph Land, with a depth of 2,034 feet, and the parallel Voronina Trough, some 180 miles to the east, with a depth of 1,475 feet. Along the eastern shores of Novaya Zemlya stretches the Novozemelskaya Trough, 650–1,300 feet deep.

Geologically, the Kara Sea is one of the youngest. It formed as a result of deglaciation during the last Ice Age, traces of glacial ice being found under a thin layer of sediment—brown, gray, and blue silts in troughs and deepwater hollows and sandy silts on underwater heights and shallows. Rocky soils are found in the northeast part of the sea. There are also sandbars and sandy mainland shores.

Salinity

Salinity varies greatly: near the mouths of the Ob and the Yenisey it is 10–12 parts of salt per 1,000 parts of water; at Cape Zhelaniya it rises to 30; and at Franz Joseph Land it reaches 33. In the southwestern areas, salinity of the surface waters fluctuates from 20 to 25 parts per thousand, rising toward the southern straits to 30–31. The upper layer of water is very fresh as a result of river inflows and summer thawing. This freshwater layer could reach a thickness of 63 inches (160 centimetres)—the greatest of the whole globe.

Currents move in two slow counterclockwise rotations in the southwestern and northeastern parts of the sea. The tides are primarily twice a day. Tidal influence is from the Barents Sea and the Arctic Basin, the high–low tide range reaching as much as 2.5 feet.

Climate. The climate is severely Arctic. The polar night (period during which the sun does not rise above the horizon) lasts three to four months of the year. Air temperatures below 32" F (0° C) prevail in the north nine to ten months a year and in the south seven to eight months. The average temperature in January is from −18" to −4° F (−28" to −20" C), and the minimum is −51° F (−46" C). In July, averages are from 30° to 43° F (−1° to 6° C), with a maximum of 61° F (16" C). In the winter there are frequent gales, and snowstorms and, in the summer, snow, snow squalls, and fogs. For most of the year the sea is covered with ice. In unfavourable years, summer ice occupies almost the entire sea; in other years, considerable expanses of open water are found.

The water masses of the Kara Sea are extremely cold

Water temperatures

and stratified. Water masses with a temperature below 29.3" F (−1.5" C) predominate, and only in the troughs penetrating from the Arctic Basin are found warm Atlantic waters, at depths of 500–650 feet and at temperatures up to 36.5" F (2.5" C). In the winter the water temperature averages 29.1" F (−1.6" C); in the summer it reaches 43" F (6° C) in the southwestern part of the sea and 36" F (2" C) in the north.

Economic activity. Fish having commercial significance include cod, salmon, and sturgeon. Mammals include the seal, sea hare, white whale, walrus, and polar bear.

The Kara Sea is on the Soviet Northern Sea route. The main port is Dikson. Seagoing vessels also call at the towns of Igarka and Dudinka on the Yenisey. Timber, building materials, furs, and foodstuffs are the most important cargoes. Since the 1960s the importance of the Kara Sea route has increased in connection with the discovery in the Ob–Yenisey region of large deposits of oil and natural gas. (Y.G.N./A.O.S.)

Kármán, Theodore von

One of the great research engineers of the 20th century, Theodore von KBrmBn pioneered the use of mathematics and the basic sciences in aeronautics and astronautics, as well as in other fields of technology in which, until then, progress had been achieved mainly empirically. He made a lasting impact on the world as a scientist, teacher, research organizer, and promoter of international scientific cooperation.

By courtesy of the California Institute of Technology. Pasadena

von Kármán.

The third of five children, he was born to Maurice and Helene von KBrmBn in Budapest, Hungary, on May 11, 1881. Probably the best known of his ancestors is Judah Low ben Bezalel, the "Exalted" Rabbi of Prague, a friend of the astronomer Tycho Brahe, and himself a famous 16th-century scholar, who was credited with creating an automaton called the Golem of Prague. Von Kármán's father, a professor at the University of Budapest and Commissioner of the Ministry of Education, reformed the secondary school system of the country and founded the Minta (Model) Gymnasium, which his son attended, as did the atomic physicists George de Hevesy and Leo Szilard. Von KBrmBn showed a natural mathematical facility at an early age and was well on his way to becoming a child prodigy when his father, fearing that he would become a mathematical freak, guided him toward engineering.

On completing his undergraduate studies in 1902 at the Royal Polytechnic University in Budapest, he decided to pursue his engineering career in the academic world, which would enable him to fulfill his wide scientific interests and to practice the art of teaching, which his father had inspired in him. In later years, he was delighted when engineers to whom he had imparted his scientific attitude and methodological approach acknowledged him as their teacher.

Between 1903 and 1906 he served on the faculty of the Polytechnic University and as consultant to the principal Hungarian engine manufacturer. The research he conducted on the strength of materials prepared the way for important later contributions to the design of aircraft structures. He was awarded a two-year fellowship to the University of Gottingen, Germany, to obtain a doctor's degree, but before completing it he went to the University of Paris. There, after an all-night party, a friend suggested that, instead of going to sleep, they watch the French aviation pioneer Henri Farman fly his machine. Farman successfully completed a two-kilometre (1¼-mile) course, unknowingly providing the inspiration for the young man who was to become a founder of the aeronautical and astronautical sciences.

Shortly thereafter, Ludwig Prandtl, a pioneer of modern fluid mechanics, invited von Kármán to return to Göttingen as his assistant on dirigible research and to complete his degree. The environment at the University was admirably suited to develop von Kármán's talents. He responded, in particular, to the school of the eminent mathematician Felix Klein, which stressed the fullest use of mathematics and of the basic sciences in engineering to increase technological efficiency. In 1911, he made an analysis of the alternating double row of vortices behind a bluff body (one having a broad, flattened front) in a fluid stream, now famous as Kármán's Vortex Street. The use of his analysis to explain the collapse, during high winds, of the Tacoma Narrows Bridge in the state of Washington, in 1940, is one of the most striking examples of its value.

Kármán's Vortex Street

In 1912, after a short stay at the College of Mining Engineering in Hungary, he became director of the Aeronautical Institute at Aachen (Aix-la-Chapelle), Germany, at the age of 31, remaining until 1930. In World War I he was called into military service and while at the Military Aircraft Factory at Fischamend in Austria led the development of the first helicopter tethered to the ground that was able to maintain hovering flight. After the war, as his international reputation grew, so did that of the institute. Students came from many countries, attracted by the intellectual and social atmosphere he had created. To help re-establish contacts and friendships broken by the war, he was instrumental in calling an international congress on aerodynamics and hydrodynamics at Innsbruck, Austria, in 1922. This meeting resulted in the formation of the International Applied Mechanics Congress Committee, which continues to organize quadrennial congresses, and gave birth, in 1946, to the International Union of Theoretical and Applied Mechanics, with von Kármán as honorary president.

Von KBrmBn never married. His mother and his sister, Josephine, lived with him from 1923 onward in The Netherlands near Aachen and later in Pasadena, California. His sister was his manager and hostess until her death in 1951 in America. Brother and sister were devoted to each other, and her death plunged von KBrmBn into deep depression for several months, during which he was unable to work.

He began travelling widely in the 1920s as a lecturer and consultant to industry. After his first visit to the United States in 1926, he was invited in 1930 to assume the direction of the Guggenheim Aeronautical Laboratory at the California Institute of Technology (GALCIT) and of the Guggenheim Airship Institute at Akron, Ohio. His love for Aachen caused him to hesitate, but the darkening shadow of German Nazism decided him to accept. He never regretted his decision. When Pres. John F. Kennedy presented to him the first National Medal of Science in 1963, he "pledged his brain as long as it lasted" to the country of which he had become a citizen in 1936.

Shortly after his arrival at the California Institute of Technology, his laboratory became again a mecca of the world of the aeronautical sciences. Two years later he became a founder of the U.S. Institute of Aeronautical Sciences, consultant to various U.S. industries and to the government. His personal scientific work continued unabated with important contributions to fluid mechanics, turbulence theory, supersonic flight, mathematics in engineering, aircraft structures, and wind erosion of soil.

His openmindedness was well demonstrated by his involvement in the development of astronautics. In 1936, in spite of the general disbelief in academic circles in the possibilities of rocket propulsion and its applications, he supported the interest of a group of his students in the subject. Within two years the U.S. Army Air Corps sponsored a project at his laboratory on the use of rockets to provide superperformance for conventional aircraft—especially to reduce their distance of takeoff from the ground and from naval aircraft carriers. In 1940, von Kármán, together with Frank J. Malina, showed for the first time since the invention of the black powder rocket in China around the 10th century that it was possible to design a stable, long-duration, solid-propellant rocket engine. Shortly thereafter, the prototype of the famed jet-assisted takeoff (JATO) rocket was constructed. This became the prototype for rocket engines used in present day long-range missiles, such as the Polaris, Minuteman, and Poseidon of the United States armed forces. In 1941, von Kármán participated in the founding of the Aerojet General Corporation, the first American manufacturer of liquid- and solid-propellant rocket engines. In 1944 he became the cofounder of the present NASA Jet Propulsion Laboratory at the California Institute of Technology when it undertook America's first governmental long-range missile and space exploration research program for the U.S. Ordnance Department.

Invention of the rocket engine

When he took leave from the Institute in 1944 to establish in Washington, D.C., the Air Corps Scientific Advisory Group for Gen. Henry H. Arnold, commander of the U.S. Army air forces in World War II, von Kármán could look back on his participation in a number of major contributions to achievements in rocket technology: America's first assisted take-off of aircraft with solid and liquid propellant rockets, flight of an aircraft with rocket propulsion alone, and development of spontaneously igniting liquid propellants of the kind that were to be used in the Apollo Command and Lunar Excursion modules some 25 years later.

His dedication to international scientific cooperation led him in 1947 to propose to the United Nations the establishment of an international research centre for fluid and soil mechanics in the Middle East, which, though unfulfilled, contributed to the development by UNESCO of the Arid Zone Research Project in 1950. He conceived the idea of cooperation among aeronautical engineers of the member nations of the North Atlantic Treaty Organization (NATO) and, in 1951, obtained approval to launch the Advisory Group for Aeronautical Research and Development (AGARD), of which he was chairman until his death. In 1956 his efforts brought into being the International Council of the Aeronautical Sciences (ICAS) and, in 1960, the International Academy of Astronautics. One of the outstanding activities of the academy under his presidency was its sponsorship, in 1962, in Paris, of the First International Symposium on the Basic Environmental Problems of Man in Space, at which for the first time scientists from the United States and the Soviet Union, as well as other countries, exchanged information in this field. Between 1960 and 1963 he led NATO-sponsored studies on the interaction of science and technology.

During his lifetime, laboratories were named after von Kármán at the California Institute of Technology, the Arnold Engineering Development Center of the U.S. Air Force at Tullahoma, Tennessee, and the NATO institute for fluid dynamics at Sint-Genesius-Rode, Belgium. A crater on the Moon has carried his name since 1970.

An appreciation of von Kármán's personality must also take account of his nonscientific talents. He was much interested in poetry and literature and could always supply a story appropriate to any occasion. When the atmosphere became charged with tension in a scientific meeting, he was able to restore balance by drawing on his collection of anecdotes. He had a fantastic capacity for work and left behind him wherever he went a trail of bits of paper covered with calculations. He was an optimist and believed in the future, despite the difficulties in the world at the time of his departure, as he put it, for "the other side of space": he died in Aachen, West Germany, on May 6, 1963.

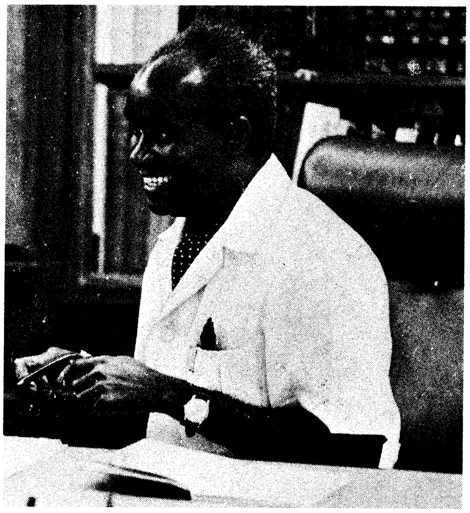
BIBLIOGRAPHY. THEODORE VON KARMAN with LEE EDSON, *The Wind and Beyond* (1967), an autobiography for the general reader; *Aerodynamics: Selected Topics in the Light of Their Historical Development* (1954), suitable for nonspecialist readers; *Collected Works of Theodore von Kármán*, 4 vol. (1956); F.L. WATTENDORF and F.J. MALINA, "Theodore von Kármán, 1881–1963," *Astronautica Acta*, 10:81 (1964); S. GOLDSTEIN, "Theodore von Kármán, 1881–1963," *Biogr. Mem. Fellows R. Soc.*, 12:335–365 (1966); D.S. HALACY, JR., *Father of Supersonic Flight: Theodor von Kármán* (1965), a popular text for the layman.

(F.J.Ma.)

Kaunda, Kenneth

Kenneth David Kaunda, president of Zambia (formerly Northern Rhodesia) and the leading figure in Zambia's independence movement, was born in 1924 in Chinsali, a small town in the north of the country. His father, who was from Nyasaland (now Malawi), eventually became a schoolteacher; his mother, also a teacher, was the first African woman to teach in colonial Zambia. Both taught among the Bemba tribe in northern Zambia, where young Kaunda received his early education, completing secondary school in the early 1940s. Like the majority of Africans in colonial Zambia who achieved some measure of modern middle class status, he also began to teach, first in colonial Zambia, in the middle 1940s in Tanganyika (now Tanzania), and returning to Zambia in 1949.

Camera Press—Publix



Kaunda.

In that year he became interpreter and adviser on African affairs to Sir Stewart Gore-Browne, a liberal white settler and a member of the Northern Rhodesian Legislative Council. Kaunda acquired knowledge of the colonial machinery as well as political skills, both of which served him well when later that same year he joined the African National Congress (ANC), the first major anticolonial organization in Northern Rhodesia. The ANC had its origin in the Federation of African Welfare Societies, a group that had sought to extend African political influence; the group organized into a federation of welfare societies in 1946 and two years later transformed itself into an explicit political agency, the ANC. In the early 1950s Kenneth Kaunda became the ANC's secretary general, functioning as its chief organizing officer, a role that brought him into close contact with the movement's rank and file. Thus, when the leadership of the ANC clashed over strategy in 1958–59, Kaunda carried a major part of the ANC operating structure into a new organization, the Zambia African National Congress.

Role in the African National Congress

Kaunda became president of the new organization and skillfully used it to forge a militant policy against the British plan for a federation of the three central African colonies——Southern Rhodesia, Northern Rhodesia, and Nyasaland. African leaders opposed and feared any such federation because it would tend to place ultimate power in the hands of a white minority of settlers. Kaunda employed the Zambia congress as an instrument for executing what he called "positive nonviolent action," a form of civil disobedience against the federation policy. His campaign had two major results: first, the British government modified the federation policy and eventually agreed to discard it; second, the imprisonment of Kaunda and other militant leaders elevated them to the status of national heroes in the eyes of the masses. Thus, from 1960 on, the nationwide support of Zambia's independence movement was secured, as was too the dominant status of Kenneth Kaunda in that movement.

Struggle for independence

Kaunda was released from prison by the colonial government on January 8, 1960. At the end of that month he was elected president of the United National Independence Party (UNIP), which had been formed in October 1959 by Mainza Chona, a militant who was disenchanted with the older African National Congress. The party enjoyed a spectacular growth, claiming 300,000 members by June 1960. This strength, however, did not prevent severe harassment by the government and especially by the local chiefs who owed their status to their colonial rulers. Yet the government also was quick to recognize the legitimacy of the UNIP leaders among the Zambian populace. Thus in December 1960 it invited Kaunda and several other UNIP leaders to participate in discussions on the status of the three colonies at the Federal Review Conference in London. Early in the following year the British government announced that formal decolonization of Zambia would commence.

The procedures through which Zambia moved toward independence were complicated, however, by the presence of the European settler community of some 77,000 and also an Asian community of 11,000. Both groups sought to delay decolonization, hoping to strengthen their positions. The British government proved indecisive in its handling of this conflict. The leadership of Africans by Kaunda and the UNIP was, however, largely able to allay the deeply felt grievances among Zambians, although there was an outbreak of widespread anti-white rioting in urban areas in 1961. The first major elections leading to final decolonization were held in October 1962. The constitutional proposals upon which the election was based provided the European settlers with a disproportionate share of the votes. Yet the two major African parties—the UNIP and ANC—gained a majority of the votes. The UNIP was the winner, gaining 15 of the 37 seats in the new Legislative Council.

The UNIP success was attributed overwhelmingly to the leadership of Kaunda. He had been astute both in allaying European fears that an African regime would unfairly disregard their interests and at quelling anarchistic militancy among wide sections of the country's African population. It was this same skill that enabled Kaunda to negotiate further constitutional advances, and in 1964 Zambia was granted independence. Like other African leaders, Kenneth Kaunda faced many complex post-independence problems, especially the issue of tribalism. He succeeded in continuing to negotiate on this issue, saving Zambia the trauma of tribal civil war. Nonetheless political violence began to occur, and this, together with pressures on his borders, led him in 1973 to impose single-party rule on his country. In 1976, with a declining economy at home and civil war in Angola, just to the east, he assumed emergency powers.

BIBLIOGRAPHY. For further information on Kaunda, see RICHARD S. HALL, ***The High Price of Principles: Kaunda and the White South*** (1969), a reliable description of Kaunda's political relations with the white-ruled regimes in southern Africa; DAVID C. MULFORD, ***Zambia: The Politics of Independence, 1957–1964*** (1967), the best account of Zambian politics in the years of decolonization; and ***The Northern Rhodesia General Election, 1962*** (1964), the only published work on the Zambian election; ROBERT I. ROTBERG, ***The Rise of Nationalism in Central Africa: The Making of Malawi and Zambia, 1873–1964*** (1965), a study of the emergence of modern anti-colonial nationalism in the British territories of Nyasaland and Northern Rhodesia, noted for its attention to the role of African voluntary organizations, called Native Welfare Associations, in this development; and RONALD SEGAL, ***Political Africa: A Who's Who of Personalities and Parties*** (1961), a valuable compendium of biographical data, party histories, and concise descriptions of key political events.

(M.Ki.)

Kazakh Soviet Socialist Republic

Stretching across the rolling tablelands of the heart of the Eurasian landmass, the Kazakh Soviet Socialist Republic (or Kazakhstan) has been since December 5, 1936, one of the 15 constituent republics of the Soviet Union. It is bounded by the People's Republic of China on the east; by the Kirgiz and Uzbek S.S.R.'s on the south; by the Caspian Sea and a small portion of the Turkmen S.S.R. on the west; and by the Russian S.F.S.R. on the north. It covers 1,049,200 square miles (2,717,300 square kilometres), an area larger than that of Algeria. It stretches about 1,200 miles (1,900 kilometres) from east to west and 800 miles (1,300 kilometres) from north to south. By the mid-1970s, it was the home of 14,170,000 persons. The capital is at Alma-Ata.

The ancient nomadic way of life in this part of the world has been changed greatly during the Soviet period. Industry, particularly the extractive industries, now plays the major role in the economy of Kazakhstan, the republic functioning as an important supplier of raw materials for the entire Soviet Union. Agriculture, however, continues to be of significance. The transformation was accompanied by an influx of settlers—the Kazakhs themselves now form only slightly more than a third of the population. Urbanization, too, is increasing, but many traditional customs have been preserved, side by side with such incursions of modernity as the Soviet cosmodrome, or space-launching centre, Baykonyr, near Leninsk.

For related information, see RUSSIA AND THE SOVIET UNION, HISTORY OF; INNER ASIA. HISTORY OF; TURKISTAN, HISTORY OF; CENTRAL ASIAN PEOPLES, ARTS OF; ALTAIC LANGUAGES; RUSSIAN STEPPE; BALKHASH, LAKE; CASPIAN SEA; and ARAL SEA.

THE LANDSCAPE

Topography. Although it exhibits a variety of relief forms, Kazakhstan is essentially a tableland. Lowlands account for more than a third of the total area, hilly plains and plateaus for nearly half, and mountainous regions make up barely 20 percent. The highest point is Khan-Tengri Peak, on the border between Kazakhstan and Kirgiziya. The lowest point in the whole Soviet Union is found in the Karagiye salt depression of the east Caspian region.

The Caspian Depression

The western and southwestern portions of the republic are dominated by the Caspian Depression (Prikaspiyskaya Nizmennost), south of which lies the Ustyurt Plateau and, on the Mangyshlak Peninsula, the Karatau and Aktau mountains. Farther east, the Ural Plateau and its extension, the Mugodzhar Hills, separate the Caspian Depression from the extensive Turan Plain, while the Aral Sea lies to the south. The rivers that lose themselves in this region have deposited vast amounts of sand, forming the Greater and Lesser Barsuki deserts in the north, the Aral region of the Kara-Kum Desert in the centre, and the sandy Kyzylkum desert in the south. Most of the deserts retain some vegetative cover, fed by subterranean groundwater.

Central Kazakhstan is an undulating upland, having depressions filled by salt lakes, the water of which has largely evaporated. In the north, the mountains reach about 5,000 feet (1,500 metres), and there are similar high areas in the west (the Ulutau Mountains) and in the east (the Chingiz-Tau Range).

In the east and southeast, high mountain massifs are furrowed by valleys. The Altai mountain complex to the east sends three ridges—the Listvyaga, Kholzun, and Tigirek—into the republic, and, farther south, the Tarbagatay Range is a similar offshoot of the Narym-Kolbin complex.

Another range, the Dzhungarian Alatau, penetrates the republic to the south of the depression containing Lake Balkhash. The Tien Shan rises along the southern frontier with the Kirgiz S.S.R., and, to their northwest, the much lower Chu-Ili Mountains and the Karatau Range are penetrated by the Muyunkum Desert.

Drainage patterns. The republic has more than 7,000 streams and rivers, most of them part of the inland drainage systems of the Caspian and Aral seas and lakes Balkhash and Tengiz. The major exceptions are the Irtysh, Ishim, and Tobol, which flow across Kazakh territory and ultimately drain to the Arctic. Many of the other rivers cease to flow in summer or wander through salty lagoons. In addition to the Irtysh in the northeast, the major rivers include the Ural in the west and the Syrdarya in the south.

The Irtysh River

Every year the Irtysh discharges some 988,000,000,000 cubic feet (28,000,000,000 cubic metres) of water into the vast West Siberian catchment area, and the Irtysh-Karaganda Canal will ultimately divert 8 percent of this mass of water into central Kazakhstan. The only major tributary of the Irtysh is the Bukhtarma. The Ural winds its way along a wide, flood-prone valley dotted with lakes; the Syrdarya, rising in the Tien Shan, is vital to Kazakh agriculture, several dams having been built to prevent floods and aid irrigation. Its major tributary, the Arys, irrigates more than 500,000 acres (200,000 hectares).

The rivers of the east and southeast are swollen by melting snow and by glaciers, of which there are more than 1,500, covering more than 800 square miles. Kazakhstan also contains about 48,000 small lakes, most of which have variable water levels, and some of which dry up periodically. The Caspian Sea is the largest inland body of water in the world, and Kazakhstan has 1,450 miles of its coastline. The other large bodies of water are the Aral Sea and Lake Balkhash, followed by Zaysan, Alakol, Tengiz, and Seletytengiz.

Climate and plant and animal life. The climate is sharply continental, especially in the plains and valleys, with hot summers alternating with equally extreme winters. Temperatures fluctuate remarkably: the average January temperature in northern and central regions ranges from 3" to –2° F (–16" to –19° C), and in the south from 23° to 29° F (–5" to –1.4° C), while the average July temperatures range from 68" F (20" C) in the north to 84" F (29" C) in the south. Absolute temperatures have been recorded at –49" F (–45" C) and 113" F (45" C), with desert-sand temperatures occasionally reaching 158" F (70" C). Precipitation ranges from eight to 12 inches (200 to 300 millimetres) in the north and centre to 16 to 20 inches (400 to 500 millimetres) in the southern mountain valleys, although it rises to twice the latter figures in the high ranges and falls to one-fifth in the deserts.

Strong, dry winds are common in the northwest and in the centre of the country, and mountain valleys (*e.g.*, the Dzungarian Gate) are often hit by hurricane-like winds.

The rich, black chernozem soils of the north account for about 7 percent of the total area; farther south they are replaced by fertile chestnut-brown soils and steppe areas (about 26 percent), and these give way to infertile alkaline soils (35 percent) consisting of sands and saline clays; there are also saline solonchak soils. Wind-blown loess soils are found in the Dzhambul and Chimkent regions. Kazakhstan has 74,000,000 acres (30,000,000 hectares) of arable land, with cultivation most highly developed on the chestnut-brown soils of the northern steppes. Other cultivated areas fringe the mountains on the south and east, and, where irrigation and reclamation are feasible, extend along river valleys into the deserts.

Plant life

Steppes and deserts dominate the landscape, the vegetation of the latter including wormwood and tamarisk. Wormwood is also found in the drier steppes, along with feather grass. The wooded area is very small (about 3 percent of Kazakh territory).

The fauna of Kazakhstan includes 155 mammal species, including antelope, elk, and, in the mountains, wolf, bear, and snow leopard. Ermine and sable have commercial importance. The rivers and lakes contain many fish species, including carp, perch, pike, and white salmon, and sprats and trout have become acclimatized. The fishing industry also makes use of the sturgeon, herring, and roach of the Caspian.

The human imprint. Kazakhstan's varied historical heritage has given rise to distinct regional patterns of settlement. Large villages, centres of collective and state farms (kolkhozy and sovkhozy, respectively) characterize the northern steppes, forming green oases separated by wheat fields, often situated along a stream or on a lake shore. The more arid steppes, semideserts, and deserts also contain large villages, housing the state farms of sheep breeders, while *chabany*, or herdsmen, live in temporary settlements made up of yurts, tents made of felt. The foothills are fringed by a string of village settlements, clustering along highways and surrounded, in the north, by fields of wheat and sugar beets and, in the south, by orchards, vineyards, and fields of melons.

Urban settlements are of two types: older communities such as Alma-Ata, Semipalatinsk, Petropavlovsk, and Uralsk, which have one- and two-storied houses and whole districts of modem multistoried buildings; and entirely new, planned towns such as Karaganda, Ust-Kamenogorsk, and Rudny, with wide, straight thoroughfares, tail buildings, and fringing industrial areas.

THE PEOPLE

The Kazakh S.S.R. in 1975 was made up of 19 regions, including 82 cities and 183 semi-urban settlements, with a total population estimated at 14,170,000.

The Kazakh people—Muslims, who speak a Turkic language but are Mongol in physical type—were never the exclusive inhabitants of Kazakhstan, and at the 1970 census they made up only about 33 percent of the total population; Russians constituted 42 percent, Ukrainians just over 7, and Germans (deported from the Soviet west to Central Asia in 1941) nearly 7. The remainder consisted of small percentages of Tatars, Uzbeks, Belorussians, Uighurs, Dungans, Koreans, and others. At the census of 1970, slightly more than half of the total population (but only about 26% of Kazakhs) lived in urban areas.

Immigration

Around 400,000 Russians arrived in Kazakhstan before 1897. Some Uighur and Dungan settlers (both Muslim) also appeared there in the last third of the 19th century (after Russia conquered Kazakhstan) and settled in the south and in the Alma-Ata and Taldy-Kurgan regions, to the east. Some Uzbeks and Kirgiz also inhabited the Chimkent and Dzhambul regions in the south. The main immigration, however, took place in the first third of the 20th century. More than a million had come to the area by 1916 and remained. Large numbers of Russians, Ukrainians, Belorussians, Mordvins, Germane, Bulgarians, Poles, Jews, and Tatars, most of them non-Muslim, moved in, first from the tsarist and then from the Soviet west. Koreans were transported by Joseph Stalin's orders from the Soviet Far East to Central Asia. Immigration by additional settlers occurred mainly between 1954 and 1956, as a result both of industrialization and of the Virgin and Idle Lands program.

The population is very unevenly distributed, with highest densities in the developed agricultural regions of the north and southeast. Since World War II, as a consequence of rapid industrialization, there has been a great movement to the urban areas. This process occasioned an urban housing problem, which was attacked by the development of a building industry, using industrialized methods, and an intensive regional planning program (see below Housing, *health*, and recreation).

A notable feature of the urbanization process has been the fact that it involved an influx of people from other republics rather than a movement (of young people, for example) from the countryside. Major factors in this respect have been the virtual monopolizing of urban and industrial employment by Russians and other outsiders and, to some extent, the modernization of agriculture, which demands skilled, well-educated manpower. During the mid-1950s, there was even an exodus from the towns to the Virgin Lands areas.

Like its neighbouring Central Asian republics, Kazakhstan has one of the highest rates of natural increase in the Soviet Union. This approximates 16.7 per 1,000 persons, with a very high birth rate (23.2 per 1,000) offset by a low mortality rate (6.5 per 1,000). In their republic, Kazakhs alone increased by 52 percent between 1959 and 1970.

THE ECONOMY

During the Soviet period, the economy of Kazakhstan has become characterized by industry and mechanized agriculture, and the republic acts as a major supplier of raw materials for other parts of the Soviet Union.

Resources. The development of the economy has been determined largely by the availability of natural resources. More than 90 different minerals have been discovered in the republic. Copper, in central Kazakhstan and the Aktyubinsk ***oblast;*** lead, zinc, and silver in the Rudny Altai, the Dzhungarian Alatau, and the spurs of the Karatau; tungsten and tin in the Kolbin Ridge and the southern Altai; chromite, nickel, and cobalt in the Mugodzhar Hills; titanium, manganese, and antimony in the central regions; vanadium in the south; and gold in the north and east are the most important.

Mineral wealth

Iron ore is worked in the Karaganda and Kustanay ***oblasti,*** and an extensive iron-ore basin has been prospected in the east; phosphorite deposits lie beneath the Karatau Mountains, and borates and other salts are worked in the deposits at Inderborsky. The leading coalfields lie in the Karaganda, Turgay, Ekibastuz, and Maykuben basins, while the Mangyshlak Peninsula and the Caspian Depression are promising regions for oil and gas. The republic also contains other reserves, ranging from asbestos to molybdenum, while its building materials include marble and limestone.

Fuel and electrical power resources are estimated at about 114,500,000,000,000 tons of conventional fuels and a yearly potential of 177,300,000,000,000 kilowatt-hours of electricity generated by waterpower.

Industry. Industry is now the leading branch of the Kazakh economy, though it engages less than one-tenth of the indigenous Kazakhs, most industrial manpower being supplied by Russians and other Slavs. Cast iron, steel, rolled metal, cement, and mineral fertilizer are manufactured, as well as consumer goods. There are steelworks in Temirtau and Karaganda. Kazakhstan's nonferrous industry, concentrated in the east, holds first place in the Soviet Union for lead production and second for copper.

Fuel production is on the increase; the main coalfields are still located in the Karaganda Basin, although production at Ekibastuz is expanding. The discovery of petroleum deposits in the Mangyshlak Peninsula stimulated oil production from the late 1960s onward.

The building industry is vital, and prefabricated construction has become as important as traditional methods. Manufacturing was stimulated during World War II, and Kazakhstan's plants now range from the automatic press factory at Chimkent to the X-ray equipment centre at Aktyubinsk.

Meat-packing plants are widespread; creameries are found chiefly in the north and east, and sugar refineries in the Alma-Ata, Dzhambul, and Taldy-Kurgan ***oblasti.*** Grain milling, brewing, and wine making are also important, as are fruit- and vegetable-canning plants. Newer industries include a synthetic-fibre plant at Karaganda and pharmaceutical manufacturing at Chimkent.

Agriculture. Kazakhstan has 460,000,000 acres (190,000,000 hectares) of land suitable for agriculture, and about three-quarters of this is used for pasture. About 20 percent is arable land, and the rest is made up of hayfields and relatively small portions of fallow land and land under orchards and vineyards. Agriculture still occupies almost half the population, and the republic in 1975 had about 1,800 state and about 410 collective farms, most of the former in the central and northern regions, and the latter in the south and east. Grain-growing state farms represent more than a third of these farms, and another third are occupied in sheep raising. By the mid-1970s, the republic contained some 7,945,600 head of cattle and additional large numbers of sheep, goats, pigs, and poultry. In 1974, 1,127,000 tons of meat, 4,633,000 tons of milk, more than 2,630,000,000 eggs, and some 116,000 tons of wool were produced.

Arable farming, not introduced till the 1880s, is now very important, with grain crops (20,350,000 tons in 1974) alone accounting for almost two-thirds of Kazakhstan's gross agricultural output. The republic is also a leading Soviet producer of fodder and industrial crops, potatoes, vegetables, and melons. The principal grain-producing areas are the rolling plains of the north, which produce up to a third of all wheat grown in the Soviet Union. In the south, fruits, vegetables, melons, sugar beets, rice, and grapes predominate.

Granary regions

Transportation. The great distances, scattered industrial centres, and developing extractive industries of Kazakhstan mean that transportation plays a critical role. By the early 1970s, the total amount of freight hauled approached 159,000,000,000 ton-miles (232,000,000,000 ton-kilometres) annually. Forty percent of this total was intra-republic traffic, though much more freight is shipped out of Kazakhstan than is brought in. Coal, iron ore, and grain, as well as metal, machinery, and livestock products, are exported, and timber, oil products, machinery, and other equipment are imported.

The railways take almost 85 percent of the total freight turnover, and the track length, at about 8,749 miles (14,080 kilometres), approximates to a density of 11.9 miles per 1,000 square miles (7.4 kilometres per 1,000 square kilometres), with the densest network in the north. Kazakhstan is crossed by long sections of the Trans-Siberian, South Siberian, and Turkistan–Siberian trunk lines, among others. while the Orenburg–Tashkent line serves the west and south. Electrification of the railroads is proceeding.

The total length of the republic's highways is about 69,200 miles, almost half hard-surfaced, and freight hauled on the roads approached 14,929,000,000 ton-miles (21,765,000,000 ton-kilometres) annually by the mid-1970s. The road network includes the 500-mile "Eastern Ring," reaching many areas of the Rudny Altai.

River transport accounts for less than 1 percent of the freight turnover. The Irtysh is the most important waterway, using tugboats and barges.

The bulk of passenger traffic is by air, and air transport provides a vital link between Kazakhstan and the rest of the Soviet Union. There is also an air network connecting regional centres and minor settlements, including individual collective and state farms.

Kazakhstan also possesses an extensive network of pipelines, including the Guryev–Orsk and the Uzen–Zhetybay–Shevchenko oil pipelines. Important sections of Soviet trunk oil pipelines also cross the republic.

ADMINISTRATION AND SOCIAL CONDITIONS

Constitution and government. According to the constitution of the Kazakh Soviet Socialist Republic, adopted on March 24, 1937, all power in the republic belongs to working people, with a socialist system, based on state ownership of the means of production, forming its economic foundation. The highest legislative body, which possesses no real power in the Soviet political system, is the Supreme Soviet of the Kazakh S.S.R., selected (no opposing candidates compete with a single list) every four years and meeting for a short period semi-annually. Its members select a Presidium, composed of a chairman, three deputy chairmen, a secretary, and 15 members, to function between the widely spaced sessions of the Supreme Soviet.

The socialist state

Local government operates through ***oblast,*** district, city, village, and ***aul*** (Kazakh village) soviets, selected for two-year terms.

The republican Supreme Soviet also generates the Council of Ministers, which coordinates and supervises the work of republican ministries and agencies. The chairman of the Supreme Soviet of the Kazakh S.S.R. serves also as a vice chairman of the union-wide Supreme Soviet, which meets in Moscow, and the chairman of the republican Council of Ministers also serves in the Council of Ministers of the Soviet Union.

Political life in the republic is largely determined by the Communist Party authorities in Moscow, acting via the more than 610,000-member Communist Party of Kazakhstan, a republican branch (there is no Kazakh Communist Party as such) of the CPSU (Communist Party of the Soviet Union). Substantially less than half of that membership in 1972 was Kazakh. The Komsomol (Young Communist League) of Kazakhstan has more than 1,600,000 members, including many Slavs and other non-Kazakhs.

The trade unions, with a membership approaching 5,500,000, do not, as in the West, have the right to strike, to picket, or otherwise energetically to protect labour's interests in relations with management.

Education. All forms of education are provided tuition-free, the institutions being supported mainly by turnover and business taxes affecting most employees and enterprises in the republic. A high percentage of literacy —96.9 in 1959 and 99.7 in 1970—is claimed for everyone in Kazakhstan between nine and 49 years of age. Only 2.5 percent of Kazakhs over 60 (391,450 persons), however, had completed any formal schooling beyond grade school.

Multilingual instruction

By the mid-1970s almost 3,400,000 children attended the republic's 10,000 schools of all types annually, while children of preschool age attended about 5,550 kindergartens and nurseries. Languages of instruction are mainly Russian and Kazakh but include minority tongues. In sparsely populated areas, there are boarding-type secondary schools for pupils whose homes are far away. Evening secondary schools for young workers and farmers are widely attended. Vocational education is also available at numerous specialized secondary and technical schools.

Kazakhstan has 148 students in higher education per 10,000 population. Major educational establishments are the Kazakh S.M. Kirov State University, the Abay Teachers College, and polytechnical, agricultural, veterinary, and other institutions in Alma-Ata; and the medical and teachers' institutes in Karaganda. Other regional centres have at least one institution of higher learning. Many Kazakhs study in other Soviet cities. Kazakhs are poorly represented in graduate schools; in 1970, when they constituted 4.35 percent of the total Soviet population, they constituted only 1.7 percent of the total number of graduate students.

In 1970 half the 630,000 high school and college graduates in the Kazakh S.S.R. were women, though among Kazakhs themselves women holding college degrees represented but 2 percent, compared with 4.4 percent of men. At the same time, Russian men and women in the republic holding college degrees were near parity, with 4.4 percent of males and 4 percent of females over the age of nine recorded as having graduated from college.

Scientific research is under the auspices of the Academy of Sciences of the Kazakh S.S.R., founded in 1945. The work of Kazakh scholars on the history of the Kazakh language and literature, on the history and ethnography of the Kazakh people, and in industrial and agricultural research has been widely acclaimed.

Housing, health, and recreation. Urbanization and population growth have been accompanied by a housing boom. Living conditions in heavily industrialized regions such as Karaganda, however, have been said by visitors to be largely substandard, with considerable attendant ethnic friction resulting from overcrowding and competition for adequate living space. As industry has developed, anti-pollution measures have been introduced.

Health standards are high. By 1973 the republic had 35,200 physicians and about 122,500 qualified nurses and other auxiliary personnel. Medical establishments are represented by 1,900 outpatient clinics and hospitals with 169,600 beds, sustained by a wide network of auxiliary centres. Five medical institutions of higher learning had 18,500 students in 1970, training physicians of all specialties. In addition, the republic has 25 medical secondary schools with a student body of about 23,000.

Special attention is given to the care of mothers and children, with about 1,100 infant and maternal care centres in 1973 and more than 16,000 beds in hospitals for expectant mothers.

Sanatoriums of different types are scattered through the republic, the best known being those for treatment of respiratory diseases with koumiss (fermented mare's milk), which are often situated in settings of woods and lakes. Hospitals for rheumatics have been built around mineral springs, and mud cures are also available.

Cultural life and institutions. Cultural traditions are deeply rooted in the distinctive ways of life found in Kazakhstan. This heritage is evident in clothes, food, and interior decoration. The national costume style is best preserved in women's clothes: long, wide dress with a stand-up collar and bloomers gathered at the ankle, supplemented by a kerchief or cotton headdress with a face slit, bracelets, and earrings. Young girls wear both European and such national clothes as wide, white gowns, brightly coloured velvet waistcoats, and beads. Elderly men usually wear wide, white shirts, wide trousers, and woollen or cotton robes. Kazakh homes are decorated with carpets and *q̣oshma*s, felt rugs in gay colours. Kazakh ornamentation (stucco work and wall facing) is widely used in building.

Kazakh literature

Kazakh folklore has roots deep in the past. Abay Ibragim Kunanbayev, a late 19th-century humanist, poet, man of letters, and translator, gave a strong start to the development of modern Kazakh written literature and literary language. Most important to the development of Kazakh as a national written language in the first decades of the 20th century was the work of Aqmet (Ahmed) Baytursin-uli, reformist politician, author, intellectual leader, and editor of the influential Kazakh newspaper Qazaq, published intermittently between 1913 and 1918 in Orenburg and Troitsk. The *aqïn* (bard) Jambil Jabayev gained major recognition from the Soviet authorities as a folk poet.

One of the first Soviet Kazakh playwrights, the academician Mukhtar Auezov, also wrote a long novel, Abay, the first part of which won a Stalin Prize in 1949; on completion, it was awarded the Lenin Prize for literature in 1959. Abay, translated into English and many other languages, is an encyclopaedic but nevertheless somewhat psychological representation of Kazakh steppe life in the late 19th century; its main characters are based upon real historical counterparts. This tradition is actively continued by Kazakh writers and poets as well as through the observance of such traditional Kazakh rites as the recital of folk lyrics and epic poems, the performances of popular comedians, and the competitions between *aqïn*s, any of which events may turn into grand theatrical performances. There is also a large number of modern theatres of all types, including Uighur and Korean musical show houses, as well. as opera, ballet, and a puppet theatre. The Kazakh film studio, founded in 1944, has released many feature films and a great number of documentaries, popular science films. and newsreels.

There are many art schools, and Alma-Ata has a national picture gallery; museums and libraries are numerous, Amateur art, drama, musical groups, and dance ensembles abound.

By 1973, 385 newspapers (146 in the Kazakh language) and 168 magazines (29 in Kazakh) were being published in the Kazakh S.S.R. In the republic, annual book publishing in Kazakh per person rose from 2.4 copies around 1959 to 2.5 copies about 1970, a very small increase in comparison to Kazakh population growth (more than 50 percent) during the same period. Kazakhstan also has a multilingual radio centre, in Alma-Ata, and local broadcasting stations in the regional centres. Many collective farms, state farms, and large plants have their own relay systems. Television relay systems enable the viewers to watch programs from Moscow and Alma-Ata and, in the south, from Uzbekistan and Kirgiziya.

Every year many thousands of families find accommodation at rest homes, tourist and skiing camps, and the Young Pioneer summer camps. Many people also take part in physical culture and sports, including horseback riding, and the network of sports facilities includes a world-famous skating rink at Medeo, near Alma-Ata.

PROSPECTS

The Kazakh industrial profile shows a major emphasis on ferrous and nonferrous metallurgy, the coal, oil, and chemical industries, power generation, and manufacturing. Mid-1970s production under the Ninth Five-Year Plan (1971–75), following the introduction of many new plants, was originally expected to reach 100,000,000 tons of coal, 33,000,000 tons of oil, 5,500,000 tons of pig iron, and about 8,000,000 tons of cement annually; the revised plan for 1975 indicated that only in coal production were original goals likely to be met. About half the total capital investment went into the light and food industries, and accelerated housing construction also reflected a greater emphasis on social welfare generally. In agriculture, the plan anticipated an increase in the average total yearly production of grain (to 26,400,000 tons), and particular attention was also given to sugar beets, vegetables, and melons and to horticulture and viticulture in southern *oblasti*, with emphasis on new irrigation systems.

Cattle and, especially, sheep breeding were to be aided by expansion of pastures, again relying heavily on irrigation. Kazakhstan planned to produce about 1,100.000 tons of meat, 120,000 tons of wool, and 4,400,000 tons of milk annually by 1975, and, except for wool, met these goals by 1974. The most intensive economic development was to take place in the northern, western, and southern *oblasti*. Population increase and a planned rise in the labour force was expected to produce a string of new settlements.

With the rising higher and secondary educational level of Kazakhs, in addition to the increase in the urban segment of the Kazakh population, new and higher expectations became general among this dynamic nationality. The anticipated growth in the Kazakh S.S.K. population resulting from further Slavic immigration appears likely to increase tension over securing employment in white-collar or skilled, desirable (higher paying) jobs as more and more educated and trained Kazakhs from their rapidly growing population insistently enter the industrial and management employment market.

BIBLIOGRAPHY. For further information on Kazakh life, see THOMAS G. WINNER, *The Oral Art and Literature of the Kazakhs of Russian Central Asia* (1958); EDWARD ALLWORTH (ed.), *Central Asia: A Century of Russian Rule* (1967); GEORGE J. DEMKO, *The Russian Colonization of Kazakhstan, 1896–1916* (1969); EDWARD ALLWORTH (ed.), *The Nationality Question in Soviet Central Asia* (1973); *Canadian Slavonic Papers (Russian and Soviet Central Asia)*, vol. 18, no. 2 and 3 (1975).

(V.F.Ko./E.Al.)

Kean, Edmund

An actor whose megalomania was the source of his histrionic genius as well as the source of his human tragedy, Edmund Kean ranks with David Garrick and Sir Henry Irving as one of the three great English tragic actors of all times.

Though no official record of his birth exists, it has been well established that he was born in London, probably on March 17, 1789, the bastard son of Ann Carey, who described herself as an itinerant actress and street hawker, and Edmund Kean, a mentally unbalanced youth who committed suicide at the age of 22. The story of Kean's upbringing is overladen with legend, much of it the product of his own later fantasies, but during his formative years he was in the charge of Charlotte Tidswell, mistress of his father's eldest brother Moses. Miss Tidswell, then a small-part member of the Drury Lane Theatre Company, had been born to a higher station in life and had accepted her position at Drury Lane as the cast-off mistress of Charles Howard, the 11th duke of Norfolk. Extremely ambitious for her adopted child, she gave him both an early stage training and the rudiments of a general education. Her efforts to provide a disciplined home background were defeated, however, by his willfulness and vagrancy, and for much of his childhood he lived as a waif and stray. At times he was with his mother, who exploited his precocious talents for singing, dancing, reciting, and acrobatics. For a while he was adopted by a wealthy middle class couple, from whom he fled when he discovered that they considered him their social inferior. This behaviour was typical of the boy Kean, who displayed in miniature the character of the man he was later to become. He was as sensitive to slight as he was obsessively anxious for applause, and as self-destructive as he was determined to fulfill his destiny.

Kean, pencil drawing by Samuel Cousins, 1814. In the National Portrait Gallery, London.
By courtesy of the National Portrait Gallery, London

At the age of 15 he was his own master and set out to conquer the stage, the only world he knew. Joining one Samuel Jerrold's company at Sheerness, Kent, for 15 shillings a week he engaged to "play the whole round of tragedy, comedy, opera, farce, interlude and pantomime." The ensuing ten-year struggle was especially hard for him to endure, not only because of the privations of a strolling player's existence but also because it prolonged the agony of his frustrated ambition.

In 1808 Kean married Mary Chambers, a fellow member of his theatrical company, a woman of lower middle class background who had little understanding of his ferocious ego. She bore him two sons, whom he called Howard and Charles, presumably in recognition of Miss Tidswell's relationship with the Duke of Norfolk and his own latent fantasy that he was its offspring. The elder son, his favourite, died a few days after he had at last received the offer of a London engagement.

Style of acting

Kean's long apprenticeship left scars, particularly an addiction to alcohol, which he had come to rely on as a substitute for recognition. But the experience of adversity may well have been essential to his artistic achievement. By the standards of the time, he was unsuited to the great tragic roles. The style then in vogue was artificial, declamatory, and statuesque, and its leading exponent, John Philip Kemble, was an actor of classic good looks, imposing figure, and vocal eloquence. Though Kean had handsome features, notably unusually expressive eyes, he was small, with a voice forceful and commanding rather than melodious. He could never have hoped to compete with Kemble on Kemble's terms, so he had to school himself to become an innovator as well as a virtuoso. On January 26, 1814, when, as Shylock, he made his Drury Lane debut, the measure of his triumph was not to outshine Kemble but to outmode him.

Kean reintroduced a so-called naturalistic style of acting; by rejecting static poses in favour of agility of movement, and beautiful recitation in favour of unleashed passion, he gave the illusion of *being*, rather than *playing*, a character. There was nothing improvised about his performances, however. Technically they were carefully planned, and it was said of his portrayal of Othello that, with its unvarying tones and semitones, rests and breaks, forte and piano, crescendo and diminuendo, it might have been read from a musical score. He permitted himself melodramatic tricks now long regarded as unnatural; his range, moreover, was limited. The poet Samuel Taylor Coleridge said that Kean revealed Shakespeare by "flashes of lightning," but he failed in parts which left no opportunity for the thunder to roll as well as for the lightning to flash. As a grim archvillain in Philip Massinger's *New Way to Pay Old Debts,* Kean was so convincing as a rapacious extortioner that he was reputed to have sent the poet Lord Byron into convulsions; as Romeo he was almost laughably unpersuasive. Nevertheless, he had a profound and lasting influence on the art of acting. In magnetic power and domination of the stage he has, possibly, never been equalled.

Reputation and character

Though Kean remained a passionately admired actor, as a public figure he became increasingly unpopular. Haunted by his fear of losing his position as head of the British stage, he was betrayed into displays of jealousy against potential rivals. At the same time, his fame and fortune (he earned, on the average, $10,000 a year) were insufficient to satisfy his ambitions. Rebelling against his low social status as an actor, he surrounded himself with tavern riffraff and founded quasi-political secret societies dedicated to the "damnation of all lords and gentlemen." The climax came in 1825, when he was successfully sued for adultery with the wife of a city alderman. This provided the pretext for a virulent press campaign, in which he was subjected to hostile demonstrations in England and during his second, and last, tour of the United States. Though he courageously withstood this persecution and eventually quelled it, Kean never recovered from the blow it dealt to his pride. The last eight years of his life were a story of slow suicide by drink and other excesses.

At Covent Garden on March 25, 1833, playing Othello to his son Charles's Iago, he collapsed during the performance—his last. On May 15, he died at his house at Richmond, Surrey, leaving his son only his name. The name proved to be a valuable asset, however, for Charles John Kean (1811–68), who established a reputation as the pioneer of representational realism and who in this sense is considered the forerunner of Sir Henry Irving.

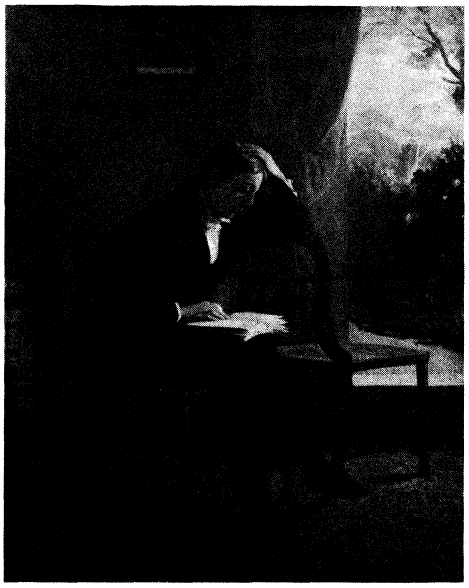
BIBLIOGRAPHY. Early accounts of Kean's life, though inaccurate, are of interest for his contemporary reputation. See especially "BARRY CORNWALL" (B.W. PROCTER), Life of *Edmund Kean* (1835). H.N. HILLEBRAND, *Edmund Kean* (1933), provides the first scholarly examination of the evidence on Kean's parentage, birth, and upbringing. For a fuller documentation, see G.W. PLAYFAIR, *Kean,* 2nd ed. (1950); and the *Letters* of *Mr. and Mrs. Charles Kean Relating to Their American Tours,* ed. by W.G.B. CARSON (1945).

(G.W.P.)

Keats, John

One of the most immediately attractive of English poets and one of the greatest of the 19th century, John Keats devoted his short life to the perfection of a poetry marked by vivid imagery, great sensuous appeal, and an attempt to express a philosophy through classical legend.

Youth. John Keats was born in London on October 31, 1795. His maternal grandfather, John Jennings, was a prosperous man who leased a livery stable and inn and employed Thomas Keats, a west-country man of uncertain origin, who married Jennings' daughter Frances. In 1802, Keats's father took over management of the business; in April 1804 he died in a riding accident. His widow remarried almost immediately. The death of Keats's grandfather, with litigation over his will and the breakup of his mother's second marriage, resulted in the Keats children living with their widowed grandmother at Edmonton, Middlesex. The mother apparently did not return to them for some years.



Keats, oil painting by Joseph Severn. 1821. In the National Portrait Gallery, London.

By courtesy of the National Portrait Gallery, London

Keats and his two surviving brothers attended a school at Enfield, two miles away. This liberal and humane establishment became his real home, and he formed a lasting friendship with the headmaster's son, Charles Cowden Clarke. At school, Keats was noted as being "a creature of passion," a fighter, and "not literary," but in 1809, possibly prompted by the return of his mother, he began to read voraciously and won several academic prizes. In March 1810 his mother died of tuberculosis, after Keats had spent his holidays nursing her. Keats's grandmother appointed two guardians and trustees to administer her estate for the benefit of her grandchildren. Shortly afterward Keats left school and was apprenticed to the family doctor, though he still paid frequent visits to Charles Cowden Clarke. He showed no dislike of his chosen career, but he quarrelled with the doctor. His association with Clarke led him, about this time, to compose his first poem, an "Imitation of Spenser." Other poems followed, all derived from popular models, many composed under the influence of a small circle of poetical amateurs.

Schooling

On October 1, 1815, Keats became a student at Guy's Hospital in the Borough of Southwark, London. In four weeks he was promoted to dresser, the equivalent of a junior house surgeon, but poetry had now become a passion with him, and he even threatened suicide if he did not succeed in it. His reading of serious contemporary poets, such as Leigh Hunt and William Wordsworth, led him away from his old circle to find a style of his own. In May 1816 his sonnet "O Solitude" was printed in Hunt's *Examiner.* In July 1816 Keats was examined and made a licentiate of the Society of Apothecaries.

Associations with Hunt. After a summer holiday at Margate with his brother Tom, poetry became his dominant interest, and, although he attended surgical lectures with a view to obtaining membership of the Royal College of Surgeons, he did not complete the course. He now lodged in Dean Street, just south of London Bridge, and, after an evening spent with Clarke, reading a translation of Homer by George Chapman, he returned there to write his first mature sonnet, "On First Looking Into Chapman's Homer." In October 1816, Clarke introduced Keats to Leigh Hunt, who had already printed one of his poems. He was instantly received into Hunt's circle, where he met the historical painter Benjamin Robert Haydon, the young poet John Hamilton Reynolds, and, a little later, Shelley. His literary pursuits could no longer be combined with further medical studies, and Keats himself increasingly felt doubts about his fitness for surgery. In November he moved to lodge with his brothers, and, as he was now 21, he informed his surviving guardian that he intended to abandon medicine for poetry. By this change of plans Keats forfeited his guardian's sympathy, but suppositions that the latter subsequently mishandled or even misappropriated Keat's legacy from his grandmother have no foundation.

First sonnet

On December 1, 1816, Leigh Hunt's article "Young Poets" appeared in *The Examiner,* praising Keats, together with Reynolds and Shelley, and quoting the whole of Keats's sonnet on Chapman's Homer. According to a fellow medical student, "This sealed his fate." He prepared a book of poems for publication, including his latest, "Sleep and Poetry," which reflected his intention to be a poet and nothing else. He bad little money and no employment, but his attractive personality made him welcome everywhere. Strikingly handsome, with reddish-gold hair and classical features, Keats had an alert expression and upright bearing that made him look more than his actual height, which was only five feet. His animation, burning thought, and deeply ingrained sense of humour are best seen in the remarkable letters that he began to write at this time, which still provide the chief authority for his life and his poetic development. His volume of *Poems,* dedicated to Leigh Hunt, was published in March 1817 but was practically unnoticed by the public. Nevertheless, another publisher engaged him to write his first long poem, *Endymion,* planned, according to one account, in friendly competition with Shelley. The poem is based on a subject that had already entered his poems, the Greek legend of the love of the moon goddess, Cynthia, for a mortal, Endymion.

First major work. Keats left London in April 1817 to begin the poem in the isolation of the countryside. The first of four 1,000-line books, into which the poem was divided, was written at Carisbrooke (Isle of Wight), Margate, Canterbury, and Hastings. He returned to write Book II in Hampstead, and there he became intimate with two neighbours, the antiquary and critic Charles Wentworth Dilke and the literary dilettante Charles Brown. Book III was written in September at Oxford, where he stayed with another new friend, Benjamin Bailey, a theological student with whom he had religious and philosophic discussions. Book IV was written partly in London and partly at Box Hill, near Dorking, and was completed in the last days of November. While writing it, Keats formulated many ideas on philosophy and art, mostly expressed in letters to Bailey. An illness mentioned in these letters may have been venereal.

In the winter of 1817–18, while leading an intensely active social life, Keats met William Wordsworth. who was visiting London. Though hardly wavering in his admiration for Wordsworth's poetry, he was disappointed by the egotism of the older poet. He also became disillusioned by Hunt and felt himself in danger of being attacked by critics whom Hunt had offended. During the early months of 1818, Keats prepared *Endymion* for the press, wrote much minor poetry, and attended the lectures by the critic William Hazlitt on the English poets. He began to plan another poem, though it was not written for another year, on a Greek legend, the supplanting of the Titan Hyperion by the sun god Apollo.

Adversity. Personal and family difficulties now began to cloud his life. His money was exhausted; one brother, Tom, was suffering from the family disease, tuberculosis, while the other, George, was planning to marry and emigrate to America. In March 1818, Keats relieved George as companion to Tom, who was wintering for his health at Teignmouth, Devonshire. There he wrote the narrative poem "Isabella; or, The Pot of Basil," based on a story from the *Decameron* of Boccaccio and suggested by Hazlitt. Tom's worsening health and his own tendency toward introspection led Keats to expand, in prose and verse, the profound series of letters he was already writing to J.H. Reynolds.

In these, he rejected the doctrine of his letters to Bailey —"O for a Life of Sensations rather than of Thoughts!" —and insisted on the need for thought, especially as a preparation for his own epic *Hyperion.* Returning to London with Tom early in May, shortly after the publication of *Endymion,* Keats accepted an invitation from Charles Brown to accompany him on a walking tour among the mountains of Westmorland, Cumberland, and Scotland. His intention was to enlarge his vision for the new poem, and he was financed by £500 from his brother George, who left for America with his bride at the end of June. Keats and Brown, who had accompanied George to Liverpool, then proceeded on the tour, passing through the Lake District and the Western Highlands, with a brief visit to Ireland. Contracting a feverish cold on a strenuous detour to the Scottish islands of Mull and Iona, Keats was obliged to leave Brown at Inverness and return by sea to London. There, in mid-August, he found Tom dangerously ill; he devoted himself to nursing his brother through his last illness.

Illness

In September, three reviews of *Endymion* had appeared, in *Blackwood's Edinburgh Magazine,* the *Quarterly Review,* and the *British Critic.* All attacked him for his association with Leigh Hunt; but, whereas the *Quarterly* attempted some serious though grudging and biassed appraisal, the review in *Blackwood's* was irresponsible and flippant, and that in the *Critic* was mere vulgar abuse. Treating Keats throughout as an ignorant apothecary's apprentice and quoting derisively from *Poems* as well as *Endymion, Blackwood's* article advised, "It is a better and a wiser thing to be a starved apothecary than a starved poet; so back to the shop, Mr. John."

Mature work. Keats met these reviews with a calm assertion of his talents, "I think I shall be among the English Poets after my death," and proceeded to write the first two books of *Hyperion* during his brother's last months. In the same period, he was attracted by three women, Jane Cox, a cousin of J.H. Reynolds; Isabella Jones, an enigmatic lady he had met at Hastings in 1817; end his final love, Frances (Fanny) Brawne, who lived with her widowed mother in Hampstead. The exact date of his meeting with Fanny Brawne is controversial; but from Tom's death, when he came to lodge with Brown in Hampstead, he began to see her frequently. Her recollection of Christmas 1818 as a particularly happy time may indicate some understanding, though not an engagement, between them at that date, but Keats continued to see Isabella Jones, who suggested to him the subject of a new narrative poem, "The Eve of St. Agnes." Written in the second half of January 1819 on a visit with Brown to Dilke's relatives in Sussex and Hampshire, the poem was full of rich, vivid, and dramatic detail, coloured certainly by Tom's death and possibly by his unresolved relationship with Fanny Brawne. The unfinished "Eve of St. Mark" followed on his return in February, but during the next several weeks he became increasingly disturbed and despondent and abandoned *Hyperion* halfway through Book III.

Early in April, Fanny and her mother came to live next door, and Keats entered into a phase of intense creative activity. In April and May, he wrote the sonnet "On a Dream," the lyric "La Belle Dame sans Merci," and his most distinctive poetical achievements, the great odes "To Psyche," "On Melancholy," "To a Nightingale," and "On a Grecian Urn," together with the less important "On Indolence." The odes are not united formally but share a common theme, examined in various aspects, of the transitoriness of youth, beauty, and life.

Odes

A long-awaited letter from George announced financial difficulties in America, and Keats determined to raise money by a popular narrative poem "Lamia," and a play, *Otho the Great,* the latter constructed in collaboration with Brown. For this purpose, he left London to spend the summer at Shanklin (Isle of Wight) and Winchester, where he completed both works, though the absence from Fanny Brawne caused friction between them. He also began to remodel *Hyperion* as a more personal vision, *The Fall of Hyperion: A Dream,* and late in September he wrote the serene "To Autumn." In October he returned to London, planning to earn money by journalism and to live away from Fanny, near Dilke, who had moved to Westminster. A visit to Hampstead broke his resolve; he took up lodgings again with Brown and became officially engaged to Fanny. Tuberculosis, caught while nursing Tom, threatened, and worry over an indefinitely long engagement and over George's financial troubles marred their happiness. His hopes for the play failed, and he worked without much conviction on a topical satirical poem, "The Cap and Bells."

Last years. In January 1820, George appeared in London, obtained his portion of Tom's estate, and left, taking most of John's share too, as a repayment of the £500 he had lent his brother in the summer of 1818. Though Fanny and many of Keats's friends thought George at fault, the settlement seems to have been financially correct. On February 3 Keats had a lung hemorrhage, that, as he himself realized, heralded his impending death. He struggled on through spring and summer, torn by unreasoning jealousy over Fanny and despair over his work. In fact, his volume ***Lamia, Isabella, The Eve of St. Agnes, and Other Poems,*** which appeared at the beginning of July 1820, was well received. After further hemorrhages, his doctor advised Keats to go to Italy as a last hope. Financed by his publishers and accompanied by a young artist, Joseph Severn, Keats sailed on September 17. On his 25th birthday he landed at Naples and journeyed to Rome. After a few weeks' respite, massive hemorrhages showed there was no hope. Keats wanted to commit suicide but was prevented by Severn. By February 1821 he grew calmer, and on February 23 he died in Severn's arms. At his request (though with additional phrases invented by his friends), the words "Here lies one whose name was writ in water" were engraved on his tombstone. Shelley's memorial poem ***Adonais,*** Byron's gibes, and Leigh Hunt's reminiscences created the false legend that Keats's death was in some way caused by the critics' lack of understanding of his work.

Assessment

Keats's poetry, at its best, conveys a direct, almost pictorial impression of the object or situation he describes. This follows from his own doctrine of the chameleon nature of the poet, who takes on the character of every experience without allowing his own identity to interfere with the communication of sensations. His letters reveal an astonishingly mature outlook on life and art; they display, with few lapses, a personality all the more lovable because its balance was achieved by victory over an inward violence of temperament. His impact on 19th-century poetry was considerable and continues to be so.

MAJOR WORKS

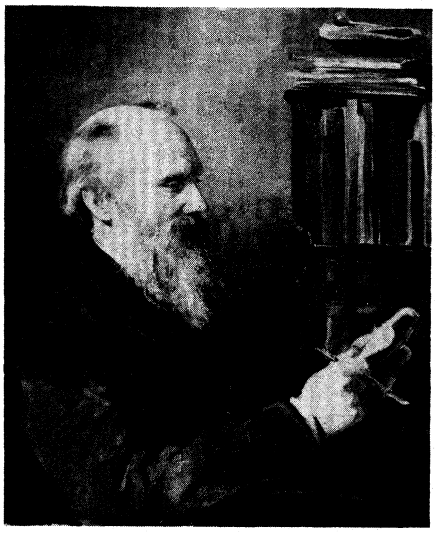
Poems (1817); ***Endymion: A Poetic Romance*** (1818); ***Lamia, Isabella, The Eve of St. Agnes, and Other Poems*** (1820). Some individual poems in the above volumes were first published in periodicals, for example, the sonnets "O Solitude" and "On First Looking into Chapman's Homer" both appeared in ***The Examiner*** in 1816; "Ode to a Nightingale" and "Ode on a Grecian Urn" in ***Annals of the Fine Arts*** in 1819; and "La Belle Dame sans Merci" in ***The Indicator*** in 1820. Other poems of Keats were not collected or published until after his death.

BIBLIOGRAPHY. The three major critical biographies are WALTER JACKSON BATE, ***John Keats*** (1963); ROBERT GITTINGS, ***John Keats*** (1968); and AILEEN WARD, ***John Keats: The Making of a Poet*** (1963). There is no entirely satisfactory edition of the texts of Keats's poems. For general use, the Oxford Standard Authors edition, ***Poetical Works*** (1939; definitive ed., 1958), ed, by H.W. GARROD, is adequate. The ***Letters of John Keats, 1814–1821*** (1957), are edited complete in two volumes by HYDER E. ROLLINS. The most comprehensive selection is ROBERT GITTINGS (ed.), ***Letters of John Keats*** (1970). Special studies are C.L. FINNEY, ***The Evolution of Keats's Poetry,*** 2 vol. (1936, reprinted 1963), a detailed account of the origins of poems; and M.R. RIDLEY, ***Keats' Craftsmanship*** (1933, reprinted 1963), an examination of Keats's working methods. Valuable articles are D.G. JAMES, "Purgatory Blind," in ***Romantic Comedy*** (1948, reprinted 1963); F.R. LEAVIS, "Keats," in ***Revaluation*** (1936, reprinted 1964); and LIONEL TRILLING, "The Poet As Hero: Keats in His Letters," in ***The Opposing Self*** (1955).

(R.Gi.)

Kelvin, William Thomson, Lord

William Thomson, who became Lord Kelvin in 1892 in recognition of his work in engineering and physics, was foremost among the small group of British scientists who helped to lay the foundations of modern physics. His contributions to science included a major role in the development of the law of conservation of energy; the absolute temperature scale (in degrees Kelvin); the dynamical theory of heat; the mathematical analysis of magnetism and electricity, including the basic ideas for the electromagnetic theory of light; the geophysical determination of the age of the Earth; and fundamental work in hydrodynamics. His inventions for use on submarine cables and his theoretical work on submarine telegraphy aided Britain in capturing a pre-eminent place in world communication during the 19th century. He had, in general, a profound influence on the scientific thought of his day.



Kelvin, oil painting by Elizabeth King, 1886–87. In the National Portrait Gallery, London.
By courtesy of the National Portrait Gallery, London

Influences on his career

Thomson was born on June 26, 1824, in Belfast, the fourth child in a family of seven. His mother died when he was six years old. William Thomson's father, who was a textbook writer, taught mathematics first in Belfast and later at the University of Glasgow; he taught young Thomson the most recent mathematics, much of which had not yet become a part of the British university curriculum. An unusually close relationship between a dominant father and a submissive son served to develop the son's extraordinary mind.

His family having moved to Glasgow in 1832, Thomson entered the University there at the age of ten. At the age of 15 he won a University gold medal for "An Essay on the Figure of the Earth," which exhibited a high order of mathematical ability and opened a field for his later research. While at Glasgow, Thomson read the book by the French mathematician and physicist Joseph Fourier, ***Analytical Theory of Heat,*** which applied abstract mathematical techniques to the study of the flow of heat through any object. Although the book was well-known in France, very few British scientists had studied or appreciated the wide generality of Fourier's approach. Thomson's first two published articles, at age 16 and 17, were a defense of Fourier's work. Thomson first promoted the idea that Fourier's mathematics could be applied to many energy questions in physics—whether fluids in motion or electricity flowing through a submarine cable.

Thomson entered the University of Cambridge in 1841 and took his B.A. in 1845 with high honours. In 1845 he was given a copy of George Green's publication ***An Essay on the Application of Mathematical Analysis to the Theories of Electricity and Magnetism.*** That essay, with Fourier's book, were the components from which Thomson created his pioneering synthesis. After finishing at Cambridge, Thomson went to Paris, where he worked in the laboratory of the physicist and chemist Victor Regnault to gain practical experimental knowledge to supplement his theoretical education.

The chair of natural philosophy at Glasgow fell vacant in 1846. Thomson's father then mounted a carefully planned and energetic campaign to have his son named to the position, and at the age of 22 William was, in fact, unanimously elected to it. Despite blandishments from Cambridge, Thomson remained at Glasgow for the rest of his career.

In 1847–49 Thomson and the British mathematician and physicist George Gabriel Stokes collaborated in a series of articles on hydrodynamic principles, which Thomson applied to electrical and atomic theory. In 1848 Thomson described his absolute temperature scale, now given in degrees Kelvin, based on the theory of heat of Sadi Carnot, the French physicist. Although at first Thomson rejected ideas that heat could be converted into mechanical energy, in 1851 he proposed a statement that effectively synthesized the ideas of Carnot and James Prescott Joule. Embodied in his paper, "On the Dynamical Theory of Heat,"—a major mathematical treatise—his statement is one version of the second law of thermodynamics. In the 1850s he collaborated with Joule in experiments on the thermal effects of gases and liquids.

The trans-atlantic cable

Although mild by nature, Thomson became involved in the controversial questions dealing with the projected cable under the Atlantic through an inquiry from G.G. Stokes, a lifelong correspondent on scientific matters, who asked him in 1854 for a theoretical explanation of the apparent delay in an electric current passing through a very long cable. In his reply Thomson referred to his early paper "On the Uniform Motion of Heat in Homogeneous Solid Bodies, and its Connexion with the Mathematical Theory of Electricity" (1842).

The publication of Thomson's reply to Stokes prompted a rebuttal by a man helping to plan the cable, E.O.W. Whitehouse. Whitehouse claimed that practical experience refuted Thomson's theoretical findings. Despite their disagreement, Thomson participated, as chief consultant, in the hazardous early cable-laying expeditions. The advice of Whitehouse, who was the company's chief electrician, prevailed with the directors of the Atlantic Telegraph Company.

In 1858 Thomson patented his telegraph receiver for use on the cable. The device, the mirror galvanometer, along with his later modification called the siphon recorder, was the instrument used on most of the worldwide network of submarine cables. Eventually, the Atlantic Telegraph Company directors fired Whitehouse and adopted Thomson's suggestions for the design of the cable and for his mirror galvanometer. In 1866 Thomson was knighted by Queen Victoria for his work on the cable. He became rich and well-known through his cable patents and consulting.

Thomson had broad interests in science, including electricity, magnetism, thermodynamics, and hydrodynamics. He was interested in many geophysical questions concerning tides, the shape of the Earth, atmospheric electricity, thermal studies of the ground, rotation of the Earth on its axis, and geomagnetism. Believing that all science is related, he was easily drawn into the controversy over Darwin's theory of evolution.

Investigations of the age of the Earth

Thomson challenged the views on geological and biological change of the early uniformitarians, including Darwin, who claimed that the Earth and its life had evolved over an incalculable number of years, during which the forces of nature always operated as at present. On the basis of thermodynamic theory and Fourier's studies, Thomson estimated in 1862 that more than 1,000,000 years ago the Sun's heat and the temperature of the Earth must have been considerably greater than at present. These conditions had produced violent storms and floods and an entirely different type of vegetation. His published views of 1868 particularly angered Darwin's supporters. Thomas Henry Huxley used the occasion of the Anniversary Address of the President of the Geological Society of London in 1869 to reply to Thomson. Although Thomson's speculations as to the age of the Earth and the Sun were inaccurate, he did succeed in pressing his contention that biological and geological theory had to conform to the well-established theories of physics.

Thomson's intellectual dynamism and his physical vigour enabled him to pursue, with extraordinary success, the two careers of engineer and scientist. He travelled extensively in Europe and the United States chiefly for pleasure, although in 1884 he delivered a series of lectures at Johns Hopkins University. With the wealth resulting from his cable projects, he bought a sailing yacht, the "Lalla Rookh," of 126 tons. From his interest in the sea came a number of patents: a compass used by the British admiralty; a calculating machine for measuring tides in a harbour and for calculating tide tables for any hour, past or future; and sounding equipment. He manufactured these items himself. The textbook, Treatise on Natural Philosophy (1867), he co-authored with P.G. Tait helped to shape the thinking of future physicists.

Lord Kelvin was said to be entitled to more letters after his name than any other man in Britain. He received honorary degrees from major universities throughout the world and was lauded by engineering societies and scientific organizations. He published more than 600 papers and was granted dozens of patents. He was elected a Fellow of the Royal Society in 1851 and served as its president from 1890 to 1895. Lord Kelvin died on December 17, 1907, at Netherhall, near Largs, Scotland, and was buried in Westminster Abbey.

BIBLIOGRAPHY. S.P. THOMPSON, ***The Life of William Thomson, Baron Kelvin of Largs,*** 2 vol. (1910), written by one who knew and admired Thomson, has a complete bibliography of his published works, a complete list of his patents, and a list of his honours and awards. HAROLD I. SHARLIN, ***Kelvin: A Biography of William Thomson*** (in prep.), provides a new interpretation of the sources of originality of an extraordinary scientist-engineer that blends the scientist's life with his work. ANDREW GRAY, ***Lord Kelvin: An Account of His Scientific Life and Work*** (1908), is an admiring but not thorough account. ELIZABETH THOMSON KING, ***Lord Kelvin's Early Home*** (1909), are reminiscences of his early family life by Thomson's oldest sister.

(H.I.S.)

Kennedy Family

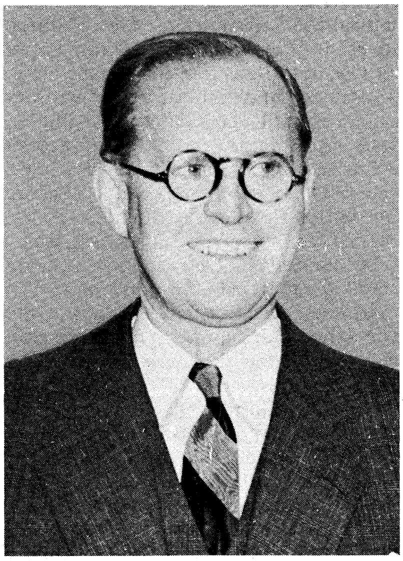
The Kennedys of Massachusetts are unique in United States politics. With the exception of the Adamses and the Roosevelts, no family has played a more prominent role in the nation's public life. As much through legend as through deeds, the Kennedys provided a recognizable theme in the confusing events in the 1960s. By the decade's end, all but one of the Kennedy men were dead, leaving behind an extraordinary record of accomplishment, drama, and tragedy.

Joseph P. Kennedy. Joseph P. Kennedy (1888–1969), was the son of a Bay State politician and grandson of an Irish immigrant. In 1912 young Kennedy graduated from Harvard. Two years later he married Rose Fitzgerald, the daughter of John F. ("Honey Fitz") Fitzgerald, mayor of Boston. Rose bore her husband nine children and taught them love, compassion, and serenity. From their father they acquired a fierce competitive drive and a yearning for distinction.

Acquisition of the Kennedy fortune

A freckled, vigorous man whose horn-rimmed spectacles gave him a slightly owlish look, Joseph Kennedy was a bank president at age 25 and a millionaire at 30. He became a shipbuilder, a motion-picture tycoon, and a heavy contributor to the Democratic Party. In five years (1929–34) as manager of the stock division of Hayden,

Keystone



Joseph P. Kennedy, 1937.

The three Kennedy brothers who carried on the family's political tradition: (left to right) John, 1981; Robert, 1987; and Edward, 1968.
(Left) Wide World Photos. (centre and right) Pictorial Parade

Stone and Company, an investment-banking firm, he mastered the art of stock-exchange manipulation. Striking out on his own, he became a brilliant plunger in the bull market of the 1920s. He retired early in 1929, having acquired enough capital to establish a million-dollar trust fund for each of his children. Then as chairman of the Securities and Exchange Commission (1934–35) under Pres. Franklin D. Roosevelt, Kennedy outlawed the very speculative practices that had made him rich.

He became chairman of the United States Maritime Commission and, in 1937, the first Irish-American to serve as the U.S. ambassador to Great Britain. In November 1940 he resigned; he was convinced that Britain was doomed to Nazi conquest and believed America's only hope lay in isolationism.

By then, Ambassador Kennedy's family was complete—four sons: Joseph P., Jr. (born 1915), John F. (born 1917), Robert F. (born 1925), Edward M. (born 1932); and five daughters: Rosemary (born 1918), Kathleen (born 1920), Eunice (born 1921), Patricia (born 1924), and Jean (born 1928). Prodded by their parents, the children were competitive and cohesive. They vied with one another in sailboat races, touch football games, tennis matches, and rough and tumble fights. No quarter was asked or given. The important thing was winning. Often they would emerge bruised and bleeding. The girls were given no preferential treatment. If they ran home sobbing, their mother would comfort them, but if their father saw their tears, he sent them back into the fray.

The ambassador was not trying to breed professional athletes. He had made enough money for them all. He expected his sons to prepare for public life and his daughters to be ready for marriage to distinguished men who would become "naturalized" Kennedys. The children were encouraged to read ***The New York Times*** at an early age. Small talk was not allowed at the dinner table. Instead, the family discussed national issues, sometimes with consequences not seen till years later. Joseph, Jr., for example, became an isolationist; John an ardent advocate of United States participation in world affairs; and Robert, perhaps because of the age gap, became shy—an affliction he battled throughout his life.

The family experienced its first misfortune when Rosemary was found to be mentally defective. Then tragedy befell Kathleen; in 1944 she married an Englishman, William Cavendish, marquess of Hartington. That same year Hartington was killed during the Allied invasion of Normandy. Four years later she herself died in a European plane crash.

Passing of the family's political standard

John F. Kennedy. Joseph Kennedy had expected his eldest son to be the first Kennedy to run for public office, but young Joe was killed too, in 1944, during the war. Thus, the family's political standard passed to the next oldest son, John, who had planned to pursue an academic or journalistic career. John's first book, ***Why England Slept,*** based on his senior thesis at Harvard, had become a best seller in 1940, the year he was graduated cum laude. Now, keeping the flame of his father's pride burning became a matter of duty. As he explained prophetically in the 1950s, when he had become a senator:

> Just as I went into politics because Joe died, if anything happened to me tomorrow, my brother Bobby would run for my seat in the Senate. And if Bobby died, Teddy would take over for him.

John Kennedy himself barely escaped death in battle. Commanding a navy motor torpedo (PT) boat, he was gravely injured when a Japanese destroyer sunk it in the Solomon Islands. Marooned far behind enemy lines, he led his men back to safety. He was awarded the Navy and U.S. Marine Corps medal for heroism and returned to active command at his own request. But his back, which had bothered him since his teens, never really healed. Despite operations in 1944, 1954, and 1955, he was in pain for much of the rest of his life. None of the Kennedys, however, including him, ever thought that John might choose a career less strenuous than politics. He was expected to run for office and to win.

Congressman. He did win; indeed, he never lost an election. His first opportunity came in 1946, when he ran for Congress. Though still weak from his wartime injuries, he campaigned aggressively, bypassing the Democratic organization in the Massachusetts 11th congressional district and depending, instead, upon his family, college friends, and fellow navy officers. In the Democratic primary he received nearly double the vote of his nearest opponent; in the November election he overwhelmed the Republican candidate. He was only 29.

John Kennedy served three terms in the House of Representatives (1947–53) as a bread-and-butter liberal. He advocated better working conditions, more public housing, higher wages, lower prices, cheaper rents, and more Social Security for the aged. In foreign policy he was an early supporter of Cold War policies. He backed the Truman Doctrine and the Marshall Plan but was sharply critical of the Truman administration's record in Asia. He accused the State Department of trying to force Chiang Kai-shek into a coalition with Mao Tse-tung. "What our young men had saved," he told the House on January 25, 1949, "our diplomats and our President have frittered away."

Senator. His congressional district in Boston was a safe seat. Had he remained there, eventually he could have become a power in the House, perhaps even its speaker. As a Kennedy, however, he was expected to raise his sights. In 1952 he ran for the U.S. Senate against the popular incumbent, Henry Cabot Lodge, Jr. John's mother and sisters held "Kennedy teas" across the state. Thousands of volunteers flocked to help, and 27-year-old Robert managed his brother's campaign. That fall the Republican presidential candidate, General Dwight D. Eisenhower, carried Massachusetts by 208,000 votes; but Kennedy defeated Lodge by 70,000 votes. Less than a year later, on September 12, 1953, John enhanced his electoral appeal by marrying Jacqueline Lee Bouvier.

John's election to the Senate

Senator Kennedy ordered that the door to his office be kept always open, and he quickly won a reputation for responsiveness to requests from constituents, except on certain occasions when the national interest was at stake. In 1954 he was the only New England senator to approve an extension of President Eisenhower's reciprocal-trade powers; and he vigorously backed the opening of the St. Lawrence Seaway, despite the fact that over a period of 20 years no Massachusetts senator or congressman had ever voted for it.

He was not as sensitive as liberal Democrats wished, however, to the demagogic excesses of Sen. Joseph R. McCarthy of Wisconsin, who in the early 1950s conducted witch-hunting campaigns against government workers accused of being Communists. His father, Joseph Kennedy, liked McCarthy; he contributed to his campaign and even entertained him in the family's compound at Hyannis Port on Cape Cod. John disapproved of McCarthy, but as he once observed, "Half my people in Massachusetts look on McCarthy as a hero." Yet on the Senate vote over condemnation of McCarthy's conduct (1954), Kennedy expected to vote against him. He prepared a speech explaining why, but he was absent on the day of the vote. Later, at a National Press Club Gridiron dinner, costumed reporters sang, "Where were you, John, where were you, John, when the Senate censured Joe?" Actually, John had been in a hospital, in critical condition after back surgery. For six months afterward he lay strapped to a board in his father's house in Palm Beach, Florida. It was during this period that he wrote *Profiles in Courage* (1956), his account of eight great American political leaders who had defied popular opinion in matters of conscience. In 1957 the book was awarded a Pulitzer Prize.

Back in the Senate, Kennedy led a fight against a proposal to abolish the electoral college, crusaded for labour reform, and became increasingly committed to civil rights legislation. As a member of the Senate Committee on Foreign Relations in the late 1950s, he advocated extensive foreign aid to the emerging nations in Africa and Asia, and he stunned Washington by calling upon France to grant Algerian independence.

Movement to the political left

During these years his political outlook was moving leftward. Possibly because of their father's dynamic personality, the sons of Joseph Kennedy matured slowly. Gradually John's stature among Democrats across the nation grew, until he had inherited the legions that had once been pledged to Gov. Adlai E. Stevenson of Illinois, the two-time presidential candidate who by appealing to idealism had transformed the Democratic Party and made Kennedy's rise possible. "A whole new generation," Edward Kennedy later said of Stevenson, "was drawn to take an interest in public affairs when he came on the scene. They were led by him, taught by him, and inspired by him." According to the historian Arthur M. Schlesinger, Jr., John Kennedy was emerging as "the heir and executor of the Stevenson revolution."

Presidential candidate. John had nearly become Stevenson's vice presidential running mate in 1956. The handsome, young New Englander's near victory and his speech of concession on television brought him into 40,000,000 American homes. Overnight he had become one of the most famous political figures in the country. Already his campaign for the 1960 nomination had begun. One newspaperman called him "young man in a hurry." Kennedy felt that he had to redouble his efforts because of the widespread conviction that no Roman Catholic candidate could be elected president. He made his 1958 race for re-election a test of his popularity in Massachusetts. "Senator Kennedy is on the move," James Reston wrote in *The New York Times*. "He is swinging for the fences now." His margin of victory was 874,608 votes—the largest ever in Massachusetts politics and the greatest of any senatorial candidate that year.

A steady stream of speeches and periodical articles followed. Photographs of him and his wife appeared on magazine covers. To transport him and his staff around the country, his father bought a 40-passenger Convair aircraft. His brothers Robert ("Bobby" or "Bob") and Edward ("Teddy" or "Ted") pitched in. After graduating from Harvard University (1948) and from the University of Virginia Law School (1951), Bob had embarked on a brilliant career as a Justice Department attorney and counsellor for congressional committee. John called young Ted "the best natural politician in the family." After a shaky start at Harvard—he was suspended after a friend had taken an examination for him—Ted had been graduated with honours (1956) and from Virginia Law School (1959). Both were astute campaigners.

In January 1960 John F. Kennedy formally announced his presidential candidacy. His chief rivals were Senators Hubert H. Humphrey of Minnesota and Lyndon B. Johnson of Texas. Kennedy knocked Humphrey out of the campaign and dealt the religious taboo against Catholics a blow by winning the primary in Protestant West Virginia. He tackled the Catholic issue again, by avowing his belief in the separation of church and state in a televised speech before a group of Protestant ministers in Houston. Nominated on the first ballot, he balanced the Democratic ticket by choosing Johnson as his running mate. In his acceptance speech Kennedy declared, "We stand on the edge of a New Frontier." Thereafter the phrase New Frontier was to become attached to his programs.

Presidential nomination

Another phrase, the Kennedy style, encapsulated the candidate's emerging identity. It was glamorous and elitist, an amalgam of his father's wealth, John Kennedy's charisma, Jacqueline Kennedy's glamour and sophistication, the charm of their children and relatives, and the erudition of the Harvard advisers who surrounded him.

President. Kennedy won the general election, narrowly defeating the Republican candidate, Vice Pres. Richard M. Nixon, by a margin of 118,550 out of the total of 68,335,642 votes cast. Nixon had defended the Eisenhower record; Kennedy, whose slogan had been "Let's get this country moving again," had deplored unemployment, the sluggish economy, what he called a missile gap, and the Communist government in Havana. A major factor in the campaign was a unique series of four televised debates between the two men; an estimated 85,000,000–120,000,000 Americans watched one or more debates. Kennedy, who had been the lesser known candidate, was impressive for his poise, firmness, and grasp of issues.

Concern with foreign affairs

John Fitzgerald Kennedy was the youngest man and the first Catholic ever elected to the presidency. His administration lasted 1,037 days. From the onset he was concerned with foreign affairs. In his memorable inaugural address he called upon Americans "to bear the burden of a long twilight struggle . . . against the common enemies of man: tyranny, poverty, disease, and war itself." He declared:

> In the long history of the world, only a few generations have been granted the role of defending freedom in its hour of maximum danger. I do not shrink from this responsibility—I welcome it. . . . The energy, the faith, the devotion which we bring to this endeavor will light our country and all who serve it—and the glow from that fire can truly light the world. And so, my fellow Americans: ask not what your country can do for you—ask what you can do for your country.

The administration's first brush with foreign affairs was a disaster. In the last year of the Eisenhower presidency, the Central Intelligence Agency had equipped and trained a brigade of anti-Communist Cuban exiles for an invasion of their homeland. The Joint Chiefs of Staff unanimously advised the new president that this force, once ashore, would spark a general uprising against Fidel Castro. But the operation was a fiasco; every man on the beachhead was either killed or captured. Kennedy assumed "sole responsibility" for the setback. Privately he told his father that he would never again accept a Joint Chiefs recommendation without first challenging it.

The Soviet premier, Nikita Khrushchev, thought he had taken the young president's measure when the two leaders met in Vienna in June 1961. Khrushchev ordered a wall built between East and West Berlin and threatened to sign a separate peace treaty with East Germany. The President activated National Guard and reserve units, and Khrushchev backed down on his separate peace threat. In

October 1962 a buildup of Soviet missiles was discovered in Cuba. Kennedy demanded that the missiles be dismantled; he ordered a "quarantine" of Cuba—in effect, a blockade. For 13 days nuclear war seemed near; then the Soviet premier announced that the offensive weapons would be withdrawn. Ten months later Kennedy scored his greatest foreign triumph when Khrushchev and Prime Minister Harold Macmillan of Great Britain joined him in signing a Nuclear Test-Ban Treaty.

Relations with Congress

Because of his slender victory in 1960, Kennedy approached Congress warily, and with good reason; Congress was largely indifferent to his legislative program. It approved his Alliance for Progress (Alianza) in Latin America and his Peace Corps, which won the enthusiastic endorsement of thousands of college students. But his two most cherished projects, massive income tax cuts and a sweeping civil rights measure, were not passed until after his death.

Nevertheless, he was an immensely popular president, at home and abroad. At times he seemed to be everywhere at once, encouraging better physical fitness, improving the morale of government workers, bringing brilliant advisers to the White House, beautifying Washington, D.C. His wife joined him as an advocate for U.S. culture. Their two young children, Caroline Bouvier and John F., Jr., were familiar throughout the country.

Joseph Kennedy had been invalided in Hyannis Port by a stroke, but the other Kennedys were in and out of Washington, D.C. Robert Kennedy, as John's attorney general, was the second most powerful man in the country. He advised the President on all matters of foreign and domestic policy, national security, and political affairs. A strong civil rights supporter, he was widely admired by the country's blacks and Mexican-Americans and by poor people everywhere.

In 1962 Edward Kennedy was elected to the President's former Senate seat in Massachusetts. Eunice's husband, R. Sargent Shriver, Jr., became director of the Peace Corps. Jean's husband, Stephen Smith, was preparing to manage the Democratic Party's 1964 presidential campaign. Patricia had married Peter Lawford, an English-born actor, and Lawford served the family as an unofficial envoy to the entertainment world. Almost all Americans knew who Rose, Jackie, Bobby, and Teddy were, and most could identify Bobby's wife as Ethel and Teddy's wife as Joan.

Assassination. President Kennedy believed that his Republican opponent in 1964 would be Sen. Barry Goldwater of Arizona. He was convinced that he could bury Goldwater under an avalanche of votes, thus receiving a mandate for major legislative reforms. One obstacle to his plan was a feud in Vice President Johnson's home state of Texas between Gov. John B. Connally, Jr., and Sen. Ralph Yarborough, both Democrats. To present a show of unity, the President decided to tour the state with both men. On Friday, November 22, 1963, he and Mrs. Kennedy were in a motorcade riding slowly through downtown Dallas in an open limousine. At 12:30 PM CST a sniper opened fire.

The death of John F. Kennedy

Two rifle bullets struck the President, at the base of his neck and in the head. He was dead upon arrival at Parkland Memorial Hospital. Governor Connally, though gravely wounded, recovered. Vice President Johnson took the oath as president at 2:38 PM CST. Lee Harvey Oswald, a 24-year-old Dallas citizen, was accused of the slaying. Two days later Oswald was shot to death by Jack Ruby, a local nightclub owner, in the basement of the Dallas police station. A presidential commission appointed to investigate the assassination and headed by the chief justice of the United States, Earl Warren, later found that neither the sniper nor his killer "was part of any conspiracy, domestic or foreign, to assassinate President Kennedy," but that Oswald had acted alone.

John Kennedy was dead, but the Kennedy mystique was still very much alive. Jacqueline Kennedy and her two children moved from the White House to a Georgetown home. Continuing crowds of the worshipful and curious made peace there impossible, however, and in the summer of 1964 she moved to New York City. Pursuit continued until October 20, 1968, when she married Aristotle Onassis, a wealthy Greek shipping magnate. The Associated Press said that the marriage "broke the spell of almost complete adulation of a woman who had become virtually a legend in her own time." She found a haven on her second husband's island home of Skorpios in the Ionian Sea.

Robert F. Kennedy. Robert Kennedy continued to serve as attorney general until he resigned in September 1964. The months after his brother's death were a desperate time for him. He was stooped by grief and spent long periods staring out windows or walking in the Virginia woods. He had presided over the Department of Justice for 44 months. He had emerged as a statesman of the law, improving the lot of many. Learning on May 20, 1961, that a hostile mob threatened the civil rights leader Martin Luther King, Jr., and 1,200 of his supporters in Montgomery, Alabama, Attorney General Kennedy sent 400 federal marshals to protect them. In subsequent racial crises he used long telephone sessions to work out the strategies of peace officers in the South. He also led a tough and imaginative drive against organized crime. One of his proudest achievements was assembling the evidence that convicted the head of the Teamster Union, James R. Hoffa. On Robert Kennedy's departure from the Department of Justice ***The*** New ***York Times,*** which had criticized his appointment three years earlier, said editorially, "He named excellent men to most key posts, put new vigor into protecting civil rights through administrative action, played a pivotal role in shaping the most comprehensive civil rights law in this country. . . . Mr. Kennedy has done much to elevate the standard."

In November 1964 he was easily elected U.S. senator from New York. Within two years Robert had established himself as a major political figure in his own right. He became the chief spokesman for liberal Democrats and a critic of President Johnson's Vietnam policy. On March 16, 1968, he announced his candidacy for the presidency. By June 4 he had won five out of six presidential primaries, including one that day in California. Shortly after midnight on June 5 he spoke to his followers in Los Angeles' Ambassador Hotel. As he left through a kitchen hallway he was fatally wounded by an Arab immigrant, Sirhan Bishara Sirhan. Robert Kennedy was buried near his brother at Arlington National Cemetery.

Edward M. **Kennedy.** Edward Kennedy remained in the Senate. He was unable to campaign actively for reelection in 1964—his back had been broken in an airplane crash—but was swept back into office by a landslide vote. Early in 1969 he was elected majority whip in the Senate and had become front-runner for the 1972 Democratic presidential nomination until July 18, 1969. On that night he drove his car off an unmarked bridge on Chappaquiddick Island, near Martha's Vineyard, Massachusetts. His companion, 28-year-old Mary Jo Kopechne, was drowned, and Edward was found guilty of leaving the scene of the accident. Kennedy was reelected to the Senate in 1970 but announced that he would not seek the presidency in 1972.

During the 1970s Kennedy continued to be a prominent spokesman for the policies that had come to be associated with the family name, and he became an advocate for many liberal causes, including a system of national health insurance. Supporting the nomination of Jimmy Carter for the presidency, Kennedy in 1976 won reelection to a third full term in the Senate. From 1979 to 1981 he was chairman of the Senate Judiciary Committee. In 1980 Kennedy entered the race against Carter for the Democratic presidential nomination. He won only 10 primaries and withdrew from the race during the convention.

The death of Joseph P. Kennedy

On November 18, 1969, Joseph P. Kennedy died at Hyannis Port, age 81. Edward Kennedy read the eulogy, which had been written in 1967 by his brother Robert:

> He has called on the best that was in us. There was no such thing as half-trying. Whether it was running a race or catching a football, competing in school—we were to try. And we were to try harder than anyone else. We might not be the best, and none of us were, but we were to make the effort to be the best. "After you have done the best you can," he used to say, "the hell with it."

BIBLIOGRAPHY. JOE MCCARTHY, *The Remarkable Kennedys* (1960), a lively account of the Kennedy family up to the presidential years; ROSE FITZGERALD KENNEDY, *Times to Remember* (1974), memoirs of the President's mother; JAMES MACGREGOR BURNS, *John Kennedy: A Political Profile* (1960), an account of John Kennedy up to his campaign for the presidency; THEODORE H. WHITE, *The Making of the President* (1960), on John Kennedy's successful presidential campaign of 1960: THEODORE C. SORENSEN, *Kennedy* (1965), a detailed history of John F. Kennedy's 11 years as senator and president, by his chief aide; ARTHUR M. SCHLESINGER, JR., *A Thousand Days: John F. Kennedy in the White House* (1965), the Kennedy Presidency as seen by the historian designated by John Kennedy to keep this record, and *Robert Kennedy and His Times* (1978), an overview of the subject's life written 10 years after his death; WILLIAM MANCHESTER, *The Death of a President: November 20–November 25, 1963* (1967), an account of John Kennedy's last days, death, and funeral and the transition to a new administration—written at the request of Mrs. John F. Kennedy; EDWIN GUTHMAN, *We Band of Brothers* (1971), an account of Robert Kennedy's entire career, by a friend who witnessed it from the beginning; WILLIAM J. VANDEN HEUVEL and MILTON S. GWIRTZMAN, *On His Own: Robert F. Kennedy, 1964–1968* (1970), Robert F. Kennedy from November 1963 to June 1968, by two of his associates; WILLIAM H. HONAN, *Ted Kennedy: Profile of a Survivor* (1972); THEO LIPPMAN, JR., *Senator Ted Kennedy* (1976); and MAX LERNER, *Ted and the Kennedy Legend* (1980), accounts of the political career of Edward Kennedy.

(W.Man.)

Kentucky

The Commonwealth of Kentucky is a south central state of the United States. Long the home of various Indian tribes, Kentucky was settled by Daniel Boone and other frontiersmen in 1769. By 1792, when it was admitted as the 15th state of the Union—the first west of the Appalachian Mountains—Kentucky had drawn nearly 75,000 settlers.

Location and size

Rivers define Kentucky's boundaries except on the south, where it shares a border with Tennessee along a nearly straight line of about 425 miles (685 kilometres), and on its mountainous southeastern border with Virginia. The Tug and Big Sandy rivers separate it from West Virginia on the east and northeast. From the point where the Big Sandy empties into the Ohio River, the pyramid-shaped northern boundary cuts a jagged line across the country, following the Ohio and meeting the states of Ohio, Indiana, and Illinois to the north. Where the Ohio flows into the Mississippi, the short western edge of the state is separated by the Mississippi from Missouri. The 40,395 square miles (104,623 square kilometres) encompassed by these boundaries are occupied by more than 3,660,000 people, about one-half of whom are rural residents. The capital, Frankfort, lies between the two major cities, Louisville, which lies on the Ohio River, and Lexington.

Kentucky brings to mind images of coal mines, of the bourbon whiskey that was developed there, of white-suited colonels and their ladies sipping mint juleps on summertime verandas, of mountaineers and moonshiners, of horse breeding and the Kentucky Derby. In actuality, Kentucky has a curious mixture of poverty and wealth, ugliness and beauty. From the 1950s through the 1970s, at least 940 lives were lost in Kentucky's coal mines, and strip-mining has continued to leave countless hillsides to erode after denuding them of their forest cover. Yet the seemingly endless landscape of white-railed horse pens and paddocks, characteristic of the rolling Bluegrass region around Lexington, symbolizes an unhurried and genteel way of life that looks more to Kentucky's ties with the pre-Civil War South than to its position in the industrial frenzy of the nation. Kentucky has always existed in the middle: as a state looking back and ahead, as a crossroads for westward expansion, and as a split personality during the Civil War. It was the birthplace both of Abraham Lincoln, 16th president of the United States, and of Jefferson Davis, president of the Confederate States during that strife. (For information on related topics, see the articles UNITED STATES; UNITED STATES, HISTORY OF THE; APPALACHIAN MOUNTAINS; and NORTH AMERICA.)

THE HISTORY OF KENTUCKY

Exploration and settlement. Before the arrival of Europeans, the Kentucky region was a hunting ground and battlefield for such Indian tribes as the Shawnee from the north and the Cherokee from the south. Even earlier agricultural and hunting peoples left burial mounds and other traces. French and Spanish explorers must have seen Kentucky from the rivers of the Mississippi basin, and traders entered the region from the eastern colonies during the early 18th century. During the 1750s and 1760s Indian resistance and rough terrain hindered successful exploration of the region. In 1769, however, Daniel Boone penetrated to the central plateau region, or Bluegrass country.

Indian raids and immigration

Settlement was rapid during the 1770s, though the prophecies of an angry Cherokee chieftain, Dragging-Canoe—that Boone and other whites would find Kentucky "a dark and bloody land"—were in large part fulfilled. British officers spurred the Indians during the Revolution, notably in raids on Boonesboro in 1777 and 1778 and at a bloody ambush at Blue Licks in 1782, and settlers encountered numerous other sieges, scalpings, and skirmishes. Following the war immigrants poured down the rivers and travelled the Wilderness Road from Cumberland Gap. The settlers founded towns and before long began to call for separation of the judicial district of Kentucky from Virginia. Statehood conventions at Danville in the 1780s were somewhat ruffled by the "Spanish Conspiracy" of James Wilkinson and others to ally the region with Spain, but they led ultimately to admission into the Union on June 1, 1792, and to the organization of state government, which took place in a Lexington tavern.

Statehood and crises. Events leading to a second state constitution in 1800 revealed an internal division that has continued to characterize Kentucky. Farmers, who floated their grain, hides, and other products on flatboats down the Mississippi to Spanish-held New Orleans, allied themselves with other antislavery forces to oppose slaveholders and businessmen. The federal Alien and Sedition Acts of 1798, passed in an attempt to control criticism of the government, were vigorously opposed. One of the leading spokesmen for the opposition was the young politician Henry Clay, who was to stamp his personality on the state and national scenes as the "great compromiser."

Kentucky took a lead in the War of 1812, much of which was fought in the adjacent Northwest Territory against combined British and Indian forces. Following the war a land boom, with attendant speculation and inflation, and the chartering of 40 independent banks that flooded the state with paper money led to financial disaster during the national economic panic of 1819. Fierce controversy over relief to debtors split Clay's Whigs and Andrew Jackson's Democrats. Signs of progress from 1820 to 1850, however, included the building of a canal at Louisville, the chartering of railroads, and increased manufacturing. The slavery question was uppermost, however, until the Civil War. The few large slaveholders were located mainly in the plantation agriculture of the Bluegrass and Pennyrile sections, but by 1833, when the legislature forbade importation of slaves for resale, the state was already one-quarter black. Until the Civil War, proslavery forces maintained an iron control of government and prevented any constitutional change that endangered their property.

Civil War and its aftermath. During the war Kentucky was a state divided. Officially, it had sought to avoid war by continuing Clay's tradition of compromise. Some 90,000 soldiers fought for the Union armies and about 40,000 for the Confederacy, though after the war popular sentiment became strongly pro-South. Following the defeat of the Confederate general Braxton Bragg at Perryville, in 1862, the only action in the state consisted of widespread guerrilla warfare.

The opening of new rail lines into the eastern coal country and the introduction of a tobacco economy stirred the state in the last decades of the 19th century. The Ku Klux Klan evoked fears and hatred, but the freed blacks were given the vote and settled as tenant farmers or urban workers. As elsewhere, however, blacks were not to become first-class citizens. Lexington and the Ohio

River cities—Louisville, Owensboro, Paducah, and Covington—grew rapidly, and the state was intensely involved in the populist agrarian politics of the period. Warfare between tobacco growers and tobacco trusts brought on an era of barn burning and similar attempts to keep tobacco prices up.

Continued diversification marked the 20th century, though the Great Depression and strikes by the United Mine Workers of America brought serious problems and open strife in many sectors. The state followed the national trends toward the loss of rural population to industrial centres, both inside and outside the state. In 1966 the devastations of strip-mining were partially ameliorated by a law requiring restoration and reforestation of the landscape.

THE LANDSCAPE

Climate. Kentucky enjoys a temperate climate, plentiful rainfall, and distinctive soils, which combine to create variety in vegetation, animal life, and landscape. The state's mean annual temperature is between 55° and 60° F (13° and 16° C). The growing season lasts from 176 to 197 days a year. Mean annual rainfall for the entire state is about 45 inches (1,143 millimetres), evenly distributed throughout the year. The greatest differences occur between the southern areas, which average as much as 48 inches annually, and the northeast, which may receive only 40 inches. Prevailing winds are from the south and southwest, although winter's chill frequently arrives on north and northwest winds.

Vegetation and animal life. Kentucky was part of the great hardwood forest region covering the nation from the Allegheny Mountains to the western prairies. Three-fourths of the state was once covered with magnificent stands of yellow poplar, oak, chestnut, sycamore, and walnut. By the close of the 19th century, however, all but a fraction of these virgin forests had been felled. Trees, shrubs, and plants of many kinds still flourish in all parts of the state, ranging from the native hardwoods and pines on the eastern slopes to the picturesque bald cypresses in the western river marshes, and the maples, beeches, and magnolia found throughout the state, Rhododendron, laurel, dogwood, redbud, and trillium are prominent among the dozens of flowering plants in the Kentucky mountains.

Birds and animals of Kentucky include those native to the Deep South as well as those of southern Canada. Of the numerous hoofed animals that once roamed Kentucky—bison, elk, moose, and deer—only deer remain. Wolves, beavers, and panthers have likewise disappeared. Among the many small animals found in the state are rabbits, squirrels, foxes, raccoons, opossums, woodchucks, and—in the numerous caves—bats and rodents. The northwestern corner of Kentucky, where the Green River flows into the Ohio, is the site of one of the world's great migratory bird routes. More than 200 species of birds frequent this area, while close to 300 species have been found in the state as a whole. The marshes of the southwestern Kentucky–Tennessee border provide breeding places for such waterfowl as the American egret, great blue heron, and double-crested cormorant. A few wild turkeys remain as a reminder of pioneer days. The swift mountain streams, wide rivers, and man-made lakes of Kentucky provide habitats for more than 100 species of fish. The muskellunge, the largest member of the pike family, and commonly considered a Great Lakes fish, is found especially in the Barren and Green rivers.

Bird life

Soils. The landscape of Kentucky is as diverse as the life it supports, extending from the wrinkled outcroppings of early Paleozoic rocks to the 2,300-mile shoreline of Kentucky Lake. Rich alluvial deposits lie along the rivers, while the rest of the state's soil derives from the long and gradual breakdown and decay of underlying rock.

Physiographic regions. Kentucky has six major physiographic regions: Mountain, Knobs, Bluegrass, Pennyrile, Western Coal Field, and Purchase.

Mountain region. More than 10,000 square miles of the easternmost part of Kentucky lie in the Mountain region, a sloping plateau of the Cumberland and Pine mountain ranges. It is a scenic land of narrow valleys, steep pinnacles, and transverse ridges. The state reaches its highest point at Big Black Mountain, 4,150 feet (1,265 metres) in altitude. An area of deep gorges, natural rock arches, and small valley farms, eastern Kentucky is drained by three major rivers and their tributaries: the Big Sandy, Cumberland, and Kentucky rivers. Natural passages through these mazes of mountains are sometimes provided by winding gaps, such as historic Cumberland Gap, or water gaps, which include the picturesque Breaks of Sandy.

Physical diversity

The great eastern coalfields of Kentucky lie in the mountains and, though the region has been a major coal-producing area throughout the 20th century, there are more than 13,000,000,000 tons of coal still buried in the eastern hills. Mining has created the state's paradox of wealth and poverty, and the dilemma of whether to exploit natural resources or preserve the ecology.

The Knobs. A long, narrow region shaped like an irregular horseshoe with both ends touching the Ohio River, the Knobs embraces the Bluegrass country on its inner side, the Mountain area on the east, and the Pennyrile on the west. Its landscape is one of cone-shaped or rounded hills and ancient escarpments. The weathered shale soil is not rich and is easily eroded, making it better adapted to forest growth than to cultivation. Canebrakes grew along some of the lower ground before European settlement and attracted large herds of buffalo and deer. A major portion of the Daniel Boone National Forest lies in the eastern Knobs.

The Bluegrass. There is a folk saying that when east Kentuckians die they want to go to Lexington, the capital of the Bluegrass. The Bluegrass lies at Kentucky's geographic and legendary heart. Its 8,000 square miles are encircled by the Knobs and the Ohio River. The region was named for the long-stemmed grass that flourishes there; it shows a bluish tint when rippled by the wind that blows across the uplands. The underlying limestones are rich in phosphates and have created pasturage for some of the world's most famous horse farms.

The Pennyrile, or Pennyroyal. The 7,800 square miles of the Pennyrile touch every other region except the Bluegrass. On the east it joins the Mountains; to the north its irregular boundaries are the Knobs, the Ohio River, and the Western Coal Field; in the west it joins the Purchase; and on the south it is bounded by Tennessee. Its unusual name derives from the local pronunciation of pennyroyal, a plant of the mint family that is abundant in the area.

The Pennyrile encompasses wooded rocky hillsides, small stock farms, cliffs, and an area once known as the Barrens—a condition caused by the Indians' continuous burning off of forest cover to make grasslands for the buffalo. Above all, it is a region of caves. Abundant waters, both surface and underground, and the limestones deposited during the Mississippian Period of geologic history (more than 300,000,000 years ago), have combined to create the area known as the Land of Ten Thousand Sinks and such famous subterranean passages as Mammoth Cave. The vast underworld cavern includes three rivers, two lakes, and a sea, and it covers more than 150 miles on five distinct levels. Its temperature remains constant at 54° F (12° C) throughout the year. Many other caves underlie the Pennyrile.

Cave region

Western Coal Field. Surrounded by the Pennyrile and the Ohio River and crossed by the Green River, the Western Coal Field's 4,680 square miles comprise less than half the area of the eastern coal beds and only a little more than half of that of the Bluegrass. The region has coal deposits throughout its extent, however, and it is fertile on some of its rolling uplands and river bottom lands. Hence it is both a mining and farming area.

The Purchase. Bounded by the Tennessee, Mississippi, and Ohio rivers, the Purchase, also called Jackson Purchase, encompasses only 2,569 square miles in the southwestern corner of the state. It was purchased in 1818 from the Chickasaw Indians, and Andrew Jackson, the seventh president of the United States, was one of the signers of the treaty. The Purchase is the lowest topographic area of Kentucky, but it is not uniformly flat. Wide floodplains are broken by low hills that may have been sandbars in ancient oceans. Bluffs, swamps, and lagoons form part of the terrain, and soft rocks of the region erode rapidly, altering the landscape. The area is one of the most fertile sections of Kentucky and is widely known both for crops and for its fine stands of poplar, hickory, and oak.

THE PEOPLE OF KENTUCKY

Patterns of settlement. Early settlers of Kentucky, who were predominantly English and Scots-Irish, came from North Carolina, Virginia, Maryland, and Pennsylvania. The migrations of Daniel Boone reflected those of many of his fellow countrymen. He moved from Pennsylvania, where he was born, down the Great Valley of Virginia into North Carolina, where he lived until he led new settlers through Cumberland Gap to Kentucky. Despite the horrors of backwoods warfare during and following the Revolution, the migration into the Bluegrass country continued. In addition to the Cumberland Gap route, the Mississippi brought early French emigres from New Orleans, particularly to the Louisville area, while dunng the mid-19th century the Ohio River carried many German settlers and other migrants, via Pittsburgh, from New England and the Middle Atlantic states. There was also a large black population in Kentucky, though the proportion decreased after 1833. Just prior to the Civil War, the Underground Railroad flourished in Kentucky to help transport escaped slaves to free soil, and there was considerable black emigration during and after the war. By 1980 only 7.1 percent of Kentucky's population was black.

Demography. From the beginning, Kentucky has been a strongly rural state of small towns and crossroads. Only Louisville and Lexington had a population greater than 100,000 by the early 1980s. Part of the urban population lives in small cities such as Lexington, Covington, Paducah, and Frankfort. Since World War II many of the younger people have left rural counties for cities both within and outside the state, creating severe economic, educational, and cultural problems.

Traditions. Distinctive traditions of life have attracted Kentucky historians and sociologists. There is no such thing, it is claimed, as a halfway Kentuckian. Politics occupies a central position in the lives of most residents, with political speeches and barbecues having a special Kentucky flavour. The pace of living for most Kentuckians remains more leisurely than in many other areas of the country, even in the urban areas. Strong family ties and personal relations dominate major aspects of individual and community life. The popular image of the Bluegrass Kentuckian includes leisurely colonels and fast horses, cold mint juleps and a hot game-meat stew called burgoo, while the mountaineer Kentuckian might be idealized as a dulcimer-strumming weaver of old English rhymes and homespun wisdom. Each is rooted, to some degree, in fact and, though Kentuckians are losing many of their distinctive traits, part of the colour and flavour of individualism remains.

THE STATE'S ECONOMY

Principal crops

Agriculture. In the first half of the 19th century, Kentucky was a leading producer of the nation's hemp, corn (maize), hogs, oats, rye, tobacco, wheat, and beef. Demands of Kentucky hemp growers brought adoption of the protective-tariff system in the United States. When sailing vessels were replaced by steamships, however, the need for hemp rope declined. Farm products have remained basic to the state's economy; after years of declining, the amount of farm acreage increased slightly during the late 1970s. Tobacco, corn, hay, and soybeans are major crops, Lexington is the heart of the light burley tobacco country, while the extreme western and southwestern parts of the state, known as the Black Patch, produce dark tobacco. Louisville, once the nation's largest tobacco market, has extensive cigarette factories. Corn and hay are essential to livestock production. Though cattle, sheep, and hogs are important, the state ranks first nationally in the breeding of Thoroughbred horses, both as saddle horses and as racehorses. About one-half of the thoroughbred saddle horses in the nation are bred in Kentucky. Production of bluegrass and orchard-grass seed is also important.

Mining, forestry, and manufacturing. Kentucky ranks second only to West Virginia in coal production. The high death toll in underground mines and the ecological disasters of strip-mining have brought national attention to the state's coal industry and its problems. Petroleum, natural gas, asphalt, and iron ores are also among Kentucky's resources. Despite rapid depletion of the forests, lumbering and allied furniture and woodworking industries remain important. Manufactured products include whiskey, which has been produced in Louisville since 1783, textiles, foods, paints and varnishes, machinery, and iron and steel products.

Mine labour and strife

Labour. A strong labour union tradition exists in the Ohio River towns, and the United Mine Workers of America is influential in the coal region. Early struggles between the UMWA and coal operators in eastern Kentucky gave rise to tragic violence. The name of Bloody Harlan commemorates that county's labour wars during the 1920s and 1930s, highlighting working and living conditions that became popularly identified with those of the state as a whole. Numerous ballads recount the history of conflict and death surrounding work in the coal mines.

Transportation. The first road engineers in Kentucky were the buffalo, whose traces to and from salt licks provided roadbeds for the earliest settlers. But, when Daniel Boone and his party hacked a passage through Cumberland Gap in 1775, Kentucky's transportation system was born. The Ohio and Mississippi rivers, bordering Kentucky for more than 700 miles, and the various tributaries of the Ohio provided Kentucky with one of the most extensive systems of navigable waterways in the nation. The waterways supported a colourful era of flatboats, rafts, and steamboats in the 19th century. Along the trails and toll roads, livestock drovers, stagecoach drivers, and hardy horsemen found travel difficult, and rough terrain hampered road building in mountainous areas. Railroads were slow to develop in Kentucky, but in 1859 the first train began to run between Louisville and Nashville, Tennessee.

Highway construction has diminished the isolation of the eastern area. The Mountain Parkway in the east, the Kentucky Turnpike from Louisville south, and the Western Kentucky Parkway are among the scenic highways making each area of the state accessible. Several cities had airline service by the early 1980s.

ADMINISTRATION AND SOCIAL CONDITIONS

Governmental branches

State and local government. The state government is comprised of the executive, legislative, and judicial branches. The governor is elected for four years and may not seek a second term. The General Assembly includes a Senate of 38 members, half to be chosen every two years for four-year terms, and a House of Representatives of 100 members, who are elected for two-year terms. All tax bills must originate in the House, which has sole power of impeachment. There are several levels in the state judiciary, from the Court of Appeals to local police courts. All judges are elected by popular vote. In 1955 the state constitution was amended to lower the voting age to 18 years. The major units of local government are the county and the municipality. All county officers are elected to four-year terms. Municipalities are divided into six classes according to population. In 1970 Area Development Districts were created by the state to provide more local participation in community-planning decisions. Each of the 15 multicounty districts is governed by a board of directors, which consists of county judges, a mayor from each county, and citizens.

Federal installations. During World War I four military training camps were established in Kentucky. The most famous, Fort Knox, home of the gold bullion depository of the United States since 1936, contains billions of dollars in gold bricks. During World War II Bowman Field and Camps Breckinridge and Campbell were national training sites for the air force.

Education. A long struggle was required to establish and support a public school system in Kentucky. In 1838 the legislature passed a law establishing the first public school system, but efforts at implementation found only scattered support, and during the Civil War the movement received a setback. A state school fund, distributed on a per-pupil basis, provided the sole support of public schools until 1908, when the legislature established each county as a unit of taxation to provide educational funds. Gradually, improvements were made: school laws were codified in 1934, the Council of Higher Education was formed, and in 1954 a base salary for teachers was set in public schools. Increases in appropriations to the program were made possible by a state sales tax. The one-room, one-teacher schools that long characterized many of Kentucky's educational districts have disappeared, and the consolidation of schools has changed the way of life in many remote areas. School attendance of children aged seven to 16 is required. Kentucky adjusted to school-desegregation requirements with varying degrees of speed and success, and Louisville was the first major Southern city to integrate its schools.

Transylvania University (chartered in 1780 as a college), the oldest institution of higher learning west of the Allegheny Mountains, was moved to Lexington in 1788. It had the first library, football team, medical school, and law school in the West. Its enrollment of less than 1,000 students during the early 1980s contrasted sharply with that of the University of Kentucky (1865), also in Lexington, with almost 40,000 students. The University of Louisville, founded by the city council in 1798, is the oldest public college in the state. Other state and community colleges are located in various sections of Kentucky, as are a number of vocational schools. There are numerous private colleges at both the two- and four-year level. Among the most distinctive is Berea College, founded in 1853 to serve youths with small financial resources. Many leaders from the mountain communities were educated there.

Health and welfare. A state board of health was first established in 1878, and the state Department for Human Resources now includes numerous divisions and bureaus that handle such areas as preventive medicine, medical inspection and licensing, medical care for the needy, chronic disease control, sanitation in water supply and sewage, public assistance, and child welfare programs. Despite federal aid granted for the construction of hospitals and other health centres, much improvement in the state's public medical facilities is needed.

CULTURAL LIFE

The horses. One of the more colourful aspects of the social and cultural life of Kentucky is that of the Kentucky Thoroughbred. Lexington is the centre of the world of horse breeding, and its boosters point out that, though they may race the horses at New York or California tracks, they breed, rear, and train them in the rolling Kentucky countryside. There are more than 300 horse-breeding farms around the city; many are open to visitors. At Louisville's Churchill Downs, one of the most colourful traditions of the sports world takes place, as it has since 1875, on the first Saturday of each May, when the Kentucky Derby is run.

The arts. Louisville is the centre of much of the cultural life of the state. It has an outstanding symphony orchestra, known for its performances and recordings of contemporary music; a ballet group; the J.B. Speed Art Museum; an active little theatre; and the Kentucky Opera Association. Kentucky continues to make a special contribution to the national culture with its folk arts, especially in the rural areas. Haunting ballads from Elizabethan days and mournful songs relating recent tragedies or desertions combine to create a distinctive musical life among mountain people. Crafts handed down through generations still produce handsome homespun cloth, hand-carved furniture, patchwork quilts, and sturdy pottery. Surrounded by a mechanized, standardized world, Kentucky folksongs and handicrafts preserve a link with earlier days. Among the nationally recognized writers identified with Kentucky are Robert Penn Warren and Irvin S. Cobb.

Folk arts

Park system. An extensive system of state parks, from Kenlake in the west to Jenny Wiley in the east, provides ample recreation for Kentuckians and tourists. In addition to providing attractive lodge and camping accommodations and a variety of sports facilities, the parks capitalize on local features of landscape or history. Restoration projects, exhibits, and outdoor pageants recreate specific episodes of pioneer days, the Civil War, and the life of Lincoln and other famous residents. Unique natural wonders, such as Mammoth Cave, are preserved and often made a focal point for the parks.

Communications. ***The Courier Journal*** of Louisville achieved distinction under one of the South's most noted editors, Henry Watterson. It is nationally recognized as an outstanding newspaper, and its voice and influence reach throughout most of Kentucky. WAVE-TV, Kentucky's first television station, and WAVE radio are located in Louisville; WHAS-TV and radio are operated by ***The Courier Journal*** and ***The Louisville Times.*** Other television and radio stations are located throughout the state.

Prospects. In Kentucky the past is reflected in many areas of life. The result is a people, a society, an economy, and a culture that are at once fascinating and baffling. The potential mineral wealth remaining in the state is enormous, yet the problems of severe poverty persist and multiply. The needs of all public agencies continue to grow and providing adequate health services, education, law enforcement, and welfare for the whole population is a continuing challenge.

The scenic wonders of the state are varied and spectacular, and the abundant natural resources can support a wide range of agricultural and industrial pursuits. Such natural assets, however, will require strong conservation efforts and careful planning in the future if they are to be preserved.

BIBLIOGRAPHY. THOMAS D. CLARK, ***Kentucky: Land of Contrast*** (1968), a lively and readable history and evocation of the state, and ***The Kentucky,*** rev. ed. (1969), an interesting discussion of Kentucky's chief river and its history; LEWIS COLLINS, ***History of Kentucky,*** rev. and enlarged by RICHARD H. COLLINS (1874, reprinted 1979), an early, valuable history of the state; GEORGE MORGAN CHINN, ***Kentucky Settlement and Statehood, 1750–1800*** (1975); and HAMBLETON TAPP and JAMES C. KLOTTER, ***Kentucky: Decades of Discord, 1865–1900*** (1977), two volumes of a projected four-volume study published by the Kentucky Historical Society; JOSEPH O. VAN HOOK, ***The Kentucky Story,*** 4th ed. (1974), a general text; MOSES E. LIGON, ***A History of Public Education in Kentucky*** (1942); WILMA DYKEMAN and JAMES STOKELY, ***The Border States*** (1968), an account of Kentucky, past and present, and its four neighbouring states; HARRY M. CAUDILL, ***Night Comes to the Cumberlands*** (1963), the authoritative work on strip-mining in Kentucky's mountains; FEDERAL WRITERS' PROJECT, ***Kentucky: A Guide to the Bluegrass State*** (1939, reissued 1973), an excellent book on all aspects of Kentucky life; HARRIETTE SIMPSON ARNOW, ***Flowering of the Cumberland*** (1963), and ***Seedtime on the Cumberland*** (1960), thorough, detailed studies of the development of the Cumberland River country; JOHN FETTERMAN, ***Stinking Creek*** (1967), an excellent, factual account of life in one mountain community of Kentucky; ALLAN W. ECKERT, ***The Frontiersmen*** (1967), recreation of life on America's frontiers, covering much of the history of early Kentucky; P.P. KARAN and COTTON MATHER (eds.), ***Atlas of Kentucky*** (1977), a thematic atlas.

(W.D.)

Kenya

An independent republic within the Commonwealth of Nations, Kenya is situated in East Africa astride the Equator. It is bordered on the north by Ethiopia and The Sudan, on the west by Uganda, on the south by Tanzania, and on the east by the Indian Ocean and Somalia. It has a total area of 224,960 square miles (582,646 square kilometres) and, in 1980, had a population of almost 16,400,000. The pronunciation of the national name as Kēn-ya is associated with British colonialism and Kĕn-ya is preferred by African nationals. Nairobi is the national capital and the country's largest city.

Since attaining independence in 1963, Kenya has been marked by its national slogan of "*Harambee,*" or

"Pulling together." It has tried to strengthen its traditional agricultural base and to augment it with other economic activities, particularly industry and tourism. Renowned for its abundant wildlife, Kenya has set aside vast preserves as protected domains for endangered species. Kenya's post-colonial development has centred around communal efforts and self-help projects that include large farming units, schools, and even independent churches. The nation's peoples are independent and proud of their cultural heritages, yet at the same time they are aware of the need to forge a strong national identity of which cooperation is a basic ingredient.

The beauty and variety of the landscape, which includes snowcapped mountains, lakes, grassy plains, and gently rolling hills, as well as the pleasant and sunny climate, exert a compelling attraction that has often found a reflection on the cinema screen or in the pages of literature. (*Green Hills of Africa,* a tale by the American writer Ernest Hemingway; *Out of Africa,* an autobiographical memoir by the Danish writer Isak Dinesen; and *Born Free,* a filmed story about raising a lion cub by Joy Adamson, an Englishwoman, are examples.) In addition, Kenya's diverse tribal life styles--exemplified by the colourful cattle-raising Masai people who, although small in numbers, are internationally famous for their Spartan ways—have also contributed to an only partly mythical image of life in Kenya. As a result, the country attracts more tourists than any other African country (see the city article NAIROBI; for associated physical features, see EAST AFRICAN LAKES; EAST AFRICAN MOUNTAINS; for history, see EAST AFRICA, HISTORY OF).

THE LANDSCAPE

Five geographic regions

Relief. The country is divided into five contrasting geographic regions. They are the Lake Victoria Basin, the central Rift Valley and its associated highlands, the eastern plateau, the semiarid and arid areas of the north and south, and the coastal region.

The Lake Victoria Basin. The southwestern corner of Kenya borders on Lake Victoria, which is shared by Kenya, Uganda, and Tanzania. The basin region extends eastward from the lake to the Mau Escarpment and is composed of a level plateau between 3,000 and 4,000 feet above sea level. The rolling grassland is cut almost in half by an eastward extension of the lake known as the Kavirondo (Winam) Gulf that penetrates inland for 50 miles. The region is very fertile and is intensively cultivated with cotton, sugarcane, and subsistence crops.

The central Rift and the highlands. The part of the Rift Valley that runs through Kenya from north to south splits the highland region into two sections—the Mau Escarpment to the west and the Aberdare Range to the east. The valley is from 30 to 80 miles wide, and its floor rises from 1,500 feet in the north around Lake Rudolph to over 7,000 feet at Lake Naivasha but then drops to 2,000 feet at the Tanzanian border in the south. It is occupied by a chain of shallow lakes that are separated by inactive or extinct volcanoes. Lake Naivasha is the largest; the others include Lakes Magadi, Nakuru, Hanrington, and Baringo. Between Lake Baringo and the southern tip of Lake Rudolph the Suguta Swamp lies in an area where the valley flattens and loses its distinctive formation.

The Mau Escarpment rises over 9,000 feet and stretches for more than 200 miles northward from the Tanzanian border to the Cherangany Hills. Important plateau areas of the western highlands include Trans-Nzoia, Uasin Gishu, Nandi, and Kericho, all between 6,000 and 7,500 feet above sea level. Mt. Elgon rises to 14,178 feet (4,321 metres) on the Uganda border. The Aberdare Range, which forms the eastern border of the Rift Valley, rises to nearly 10,000 feet. The eastern highlands extend from the Ngong Hills near Nairobi northward to the Laikipia Plateau. Snow-capped Mt. Kenya, the nation's highest mountain, rises to 17,058 feet (5,199 metres) and is linked to the highlands by the Nyeri saddle. The relief of both highlands is complicated and includes plains, deep valleys, and mountains. The soils are fertile; the area formerly attracted European settlement.

The eastern plateau. Sandwiched between the eastern highlands and the coastal strip, the vast Nyika plain slopes gently toward the sea. Its surface is covered with thorn scrub and is dotted with striking elevated formations, including the Taita, Kasigau, Mach, and Kitui Hills. The entire region suffers from periodic drought and is considered the country's traditional famine region. The human population is concentrated in the isolated hilly areas, where the climate is somewhat less severe.

The semiarid and arid regions. To the north and northeast, the Nyika plain merges with an area of increasing aridity that covers most of northern Kenya. A similar area is located in the southern part of the Rift Valley around Lake Magadi. There are scattered trees and grasses, but the meagre and unreliable rainfall usually supports only dwarf shrubs and bush. Areas of real desert are limited to the region east of Lake Rudolf.

The coast. The coastal belt runs for about 250 miles along the Indian Ocean. Here the shore is indented by bays, inlets, and river mouths and is dotted with islands, including Mombasa and Lamu islands. The coastal plain is limited on the west by a series of low hills. In the south the plain consists of a narrow strip two to ten miles wide; to the north the plain broadens, forming the 100-mile-wide Tana Lowlands that extend along the Somali border.

The river system

Drainage. Kenya is drained by a network of small rivers and streams that are generally shallow and exhibit seasonal changes in flow. Its two most important rivers, the Tana and the Galana, rise in the eastern highlands and flow roughly southeast into the Indian Ocean. Neither is navigable above its lower reaches, but the Tana River is being developed as a source of irrigation and hydroelectric power. The western highlands generally drain into Lake Victoria through the Nazoia, Yala, Mara, and Nyando rivers. The headwaters of the Ewaso Ng'iro are divided from those of the Tana by the Nyeri saddle; the river drains the southern portion of the eastern highlands into Lake Natron. Smaller rivers in the north and east are no more than intermittent streams or marshes.

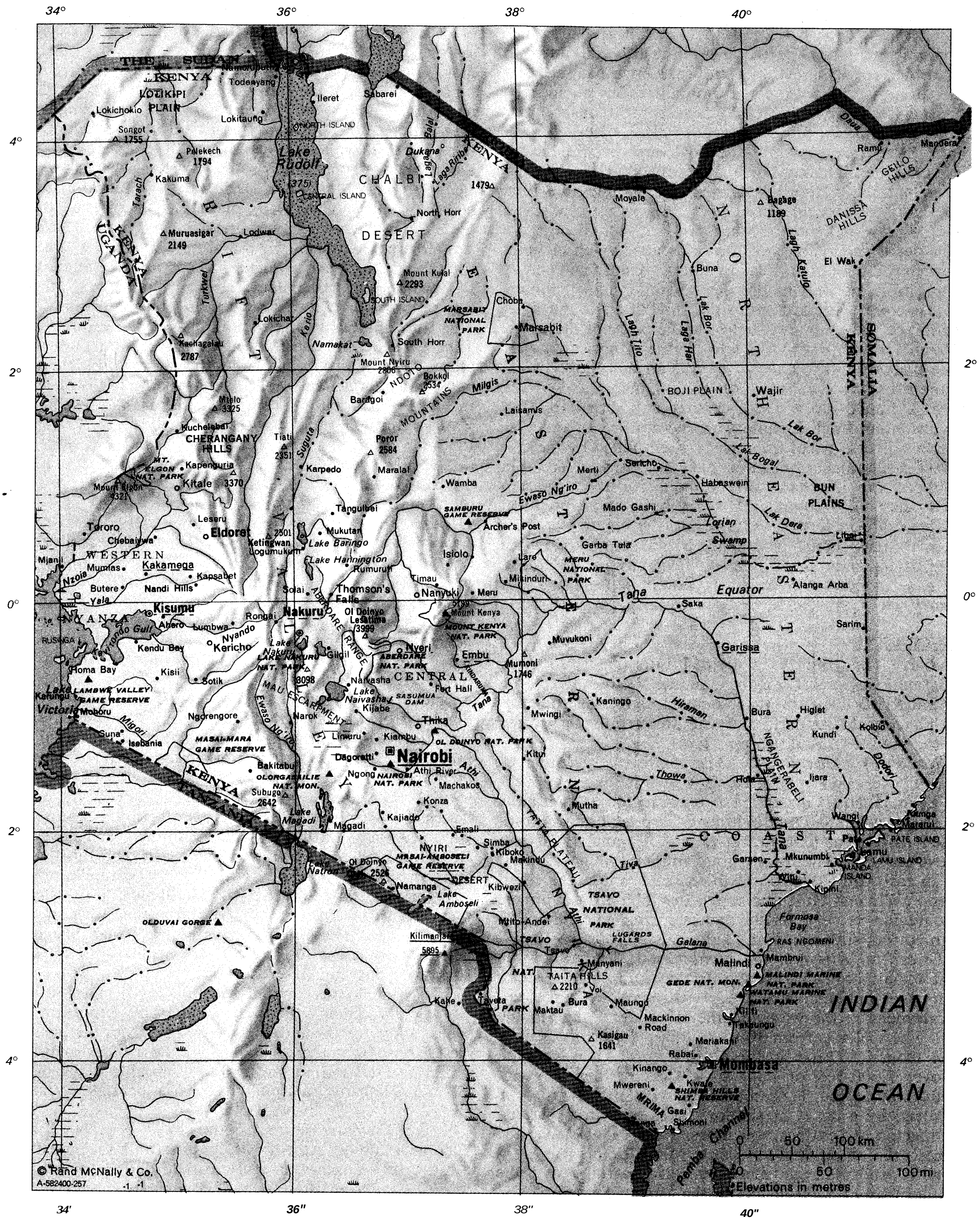
Soils. The most fertile soils, consisting of dark brown loams, occur in the highlands. Loams, sandy loams, and clays are found in the plateau lands of west Kenya, the Rift Highlands, and parts of the coast. The most widespread soils are the sandy loams of the semiarid regions between the coast and the Rift Highlands. To the north are vast areas covered by the red desert soils, mainly sandy loams. The arid north and northwest are characterized by areas of shallow stony soils with frequent rock outcrops and desert soils.

The tendency to impeded drainage in parts of the country has given rise to a number of important soil types. Clay loams and brown loams occur in parts of west Kenya and the areas east of the rift. Also in this group are the red to strong brown friable clays of the Eldoret Plateau, Laikipia, and parts of Lake Victoria Basin.

In the lowlying parts of the Lake Victoria Basin, Eldoret Plateau, Athi-Kapiti plains, and the coast are a range of black cotton soils. In addition, alluvium and recent lacustrine depressions cover parts of the Rift Valley and the lower parts of some large rivers.

Climate. Seasonal climate rhythms are controlled by the large-scale pressure systems of the western Indian Ocean and adjacent landmasses. From December to March, northeast winds predominate north of the Equator, and south to east winds south of it. During the same months the dominant air mass is comparatively dry, although rain may occur locally. The rainy season extends from late March to May, with air flow from the east in both hemispheres. From June to August the southeast winds gradually prevail in the south; north of the Equator, however, they blow from the southwest. During this season of little rainfall, weather is stable, especially in the highlands, where it tends to be dull and cold. October marks the beginning of the transition period, which ends with the return of northeast winds in December, north of the Equator.

Local variations in temperature and rainfall occur ac-



KENYA

MAP INDEX

Political subdivisions

Central	0·45s	37·00e
Coast	3·00s	39·30e
Eastern	0·05n	38·00e
Northeastern	1·00n	40·15e
Nyanza	0·30s	34·30e
Rift Valley	0·30n	36·00e
Western	0·30n	34·35e

Cities and towns

Ahero	0·11s	34·55e
Alanga Arba	0·07n	40·25e
Archers Post	0·39n	37·41e
Athi River	1·27s	36·59e
Bakitabu	1·29s	35·34e
Baltia	2·14s	40·50e
Baragoi	1·47n	36·47e
Buna	2·47n	39·31e
Bura	1·06s	39·57e
Bura	3·30s	38.18e
Butere	0·13n	34·30e
Chebaiywa	0·33n	34·48e
Choba	2·26n	38·03e
Dagoretti	1·18s	36·46e
Eldoret	0.31 n	35·17e
El Roba	3·57n	40·01e
El Wak	2·49n	40·56e
Emali	2·05s	37·38e
Embu	0·32s	37·27e
Fort Hall	0·43s	37·09e
Garba Tula	0·32n	38·31e
Garissa	0·28s	39·38e
Garsen	2·16s	40·07e
Gasi	4·25s	39·30e
Gilgil	0·30s	36·19e
Gimali	3·08n	40·32e
Habaswein	1·01n	39·29e
Hara	1·55s	40·05e
Higlet	1·04s	40·19e
Hola	1·29s	40·02e
Homa Bay	0·31s	34·27e
Ijara	1·36s	40.31e
Ileret	4·19n	36·13e
Isiolo	0.21 n	37·35e
Kajiado	1·51s	36·47e
Kakarnega	0·17n	34·45e
Kakuma	3·43n	34·52e
Kaningo	0·49s	38·32e
Kapenguria	1·04n	35·07e
Kapsabet	0·12n	35·06e
Karpedo	1·10n	36·06e
Karungu	0·51s	34·09e
Kendu Bay	0·22s	34·39e
Kericho	0·22s	35·17e
Kiambu	1·10s	36·50e
Kiboko	2·15s	37·42e
Kibwezi	2·25s	37·58e
Kijabe	0·56s	36·34e
Kilifi	3·38s	39.51e
Kinango	4·08s	39·19e
Kipini	2·32s	40.31e
Kisii	0.41s	34·46e
Kisumu	0·06s	34·45e
Kitale	1.01n	35·00e
Kitui	1·22s	38.01e
Kiunga	1·45s	41·29e
Kolbio	1·10s	41·15e
Konza	1·45s	37·07e
Kuchelebai	1·29n	35·01e
Kundi	1·08s	40·41e
Kwale	4·11s	39·27e
Laisamis	1·36n	37·48e
Lamu	2·16s	40·54e
Lare	0·20n	37·56e
Leseru	0·35n	35·10e
Limuru	1·06s	36·39e
Lodwar	3·07n	35·36e
Logumukum	0·27n	36·05e
Lokichar	2·23n	35·39e
Lokichokio	4·12n	34·21e
Lokitaung	4·16n	35·45e
Lumbwa	0·12s	35·28e
Machakos	1·31s	37·16e
Mackinnon Road	3·44s	39·03e
Mado Gashi	0·44n	39·10e
Magadi	1·54s	36·17e
Makindu	2·17s	37·49e
Maktau	3·24s	38·08e
Marnbrui	3·07s	40·09e
Mandera	3·56n	41·52e

MAP **INDEX** (continued)

Manyani	3·05s	38·30e
Maralal	1·06n	36·42e
Mararui	1·56s	41·18e
Mariakani	3·52s	39·28e
Marsabit	2·20n	37·59e
Maungu	3·33s	38·45e
Merti	1·04n	38·40e
Meru	0·03n	37·39e
Mikinduri	0·07n	37·50e
Mkunumbi	2·18s	40·42e
Mohoru	1·01s	34·07e
Mombasa	4·03s	39·40e
Moyaie	3·32n	39·03e
Mtito Andei	2·41s	38·10e
Mukutan	0·38n	36·16e
Mutha	1·48s	38·26e
Muvukoni	0·30s	38·15e
Mwereni	4·20s	39·08e
Mwingi	0·56s	38·04e
Nairobi	1·17s	36·49e
Naivasha	0·43s	36·26e
Nakuru	0·17s	36·04e
Nandi Hills	0·07n	35·11e
Nanyuki	0·01n	37·04e
Narok	1·05s	35·52e
Ngong	1·22s	36·39e
Ngorengore	1·02s	35·30e
North Horr	3·19n	37·04e
Nyeri	0·25s	36·57e
Omara	2·44s	39·00e
Pate	2·08s	41·00e
Rabai	3·58s	39·37e
Ramu	3·56n	41·13e
Rongai	0·10s	35.51e
Rumuruti	0·16n	36·32e
Sabarei	4·20n	36·55e
Saka	0·09s	39·20e
Sarim	0·23s	40·58e
Sericho	1·05n	39·05e
Shimoni	4·39s	39·23e
Simba	2·10s	37·36e
Solai	0·02n	36·09e
Sotik	0·41s	35.21e
South Horr	2·06n	36·55e
Suna	1·05s	34·26e
Takaungu	3·41s	39·51e
Tangulbei	0·48n	36·17e
Taveta	3·24s	37.41e
Thomsons Falls	0·02n	36·22e
Timau	0·05n	37·14e
Todenyang	4·32n	35·56e
Tsavo	2·59s	38·28e
Vanga	4·39s	39.13e
Voi	3·23s	38·34e
Wajir	1·45n	40·04e
Wamba	0·59n	37·19e
Wangi	2·00s	40·55e
Witu	2·23s	40·26e

Phvsical features
and points of interest

Aberdare National Park	0·30s	36·45e
Aberdare Range, mountain range	0·25s	36·38e
Amboseli, Lake	2·37s	37·08e
Athi, river	2·59s	38·31e
Bagage, mountain	3·23n	40·12e
Balal, Laaa. river	3·25n	37·15e
Baringo, Lake	0·38n	36·05e
Bogal, Lak, *river*	0·45n	40·50e
Boji Plain	1·30n	39·45e
Bokkol, *mountain*	1·50n	37·02e
Bor, Lak, *river*	1·18n	40·40e
Bun Plains	0·44n	40·42e
Central Island	3·30n	36·03e
Chalbi Desert	3·00n	37·20e
Cherangany Hills	1·15n	35·27e
Chiamboni, Ras, *headland*	1·36s	41·36e
Danissa Hills	3·17n	40·59e
Daua, *river*	4·00n	41·59e
Dera, Lak, *river*	0·30n	41·00e
Dodori, *river*	1·52s	41·02e
Dukana, *well*	3·59n	37·16e
Elgon, Mount. mountain	1·08n	34·33e
Formosa Bay	2·45s	40·20e
Galana, river	3·09s	40·08e
Gede National Monument	3·19s	40·03e
Geilo Hills	3·36n	41·17e
Har, Laga, river	1·40n	39·36e
Hannington, Lake	0·15n	36·06e
Hiraman, river	1·07s	39·55e
Indian Ocean	4·00s	41·00e
Kachagalau, mountain	2·19n	35·03e
Kasigau, mountain	3·50s	38·40e
Katulo, Lagh, river	2·08n	40·56e
Kavirondo Gulf, *bay*	0·15s	34·35e
Kenya, Mount, mountain	0·10s	37·20e
Kerio, river	2·59n	36·07e
Ketinawan, mountain	0·40n	35·50e
Kindaruma, hill	0·48s	37·48e
Kulal, Mount, mountain	2·43n	36·56e
Lake Nakuru National Park	0·20s	36·05e
Lambwe Vailey Game Reserve	0·37s	34·15e
Larnu Island	2·17s	40·52e
Lesatima, Ol Doinyo, mountain	0·19s	36·37e
Lorian Swamp	0·40n	39·35e
Lotikipi Piain	4·20n	34·45e
Lugards Fails, waterfall	3·03s	38·42e
Magadi, Lake	1·52s	36·17e
Malindi Marine National Park	3·15s	40·10e
Marsabit National Park	2·15n	38·00e
Masai-Amboseli Game Reserve	2·30s	37·00e
Masai-Mara Game Reserve	1·15s	35·15e
Mau Escaroment	0·40s	36·02e
Meru National Park	0·00	38·15e
Migori, river	0·59s	34·15e
Milgis, river	1·48n	38·06e
Mount Elgon National Park	1·05n	34·45e
Mount Kenya National Park	0·09s	37·19e
Mrima, hill	4·29s	39·16e
Mtelo, mountain	1·39n	35·23e
Mumoni, mountain	0·31s	38·01e
Munda Island	2·15s	40·57e
Muruasigar, mountain	3·08n	35·02e
Nairobi National Park	1·24s	36·50e
Naivasha, Lake	0·46s	36·21e
Nakuru, Lake	0·22s	36·05e
Namakat, salt *lake*	2·14n	36·34e
Ndoto Mountains	1·45n	37·07e
Ngangerabeli Plain	1·30s	40·15e
Ng'iro, Ewaso, river	0·28n	39·55e
Ngomeni, Ras, headland	2·59s	40·14e
North Island	4·04n	36·03e
Nyando, river	0·15s	35·32e
Nyika, physical region	3·37s	38·44e
Nyiri Desert	2·25s	37·20e
Nyiru, Mount, mountain	2·08n	36·51e
Nzoia, river	0·03n	33·57e
Ol Doinyo National Park	1·09s	37·12e
Olorgasailie National Monument	1·40s	36·22e
Orok, Ol Doinyo, mountain	2·29s	36·46e
Pate Island	2·07s	41·03e
Pelekech, mountain	3·48n	35·04e
Poror, mountain	1·14n	36·37e
Ririba, Laga, river	3·34n	37·15e
Rudolf, Lake	3·30n	36·05e
Rusinga Island	0·24s	34·10e
Sarnburu Game Reserve, wildlife refuge	0·45n	37·30e
Sasurnua Dam	0·45s	36·40e
Shimba Hills National Reserve	4·15s	39·25e
Songot, mountain	3·59n	34·28e
South Island	2·38n	36·36e
Subugo, mountain	1·40n	35·49e
Suguta, river	2·03n	36·33e
Taita Hills	3·25s	38·20e
Tana, river	2·32s	40.31e
Tarach, river	4·09n	34·56e
Thowa, river	1·33s	40·03e
Tiati, mountain	1·19n	35·56e
Tito, Lagh, river	1·34n	39·24e
Tsavo National Park	3·00s	38·40e
Turkwel, river	3·06n	36·06e
Victoria, Lake	1·00s	34·00e
Watamu Marine National Park	3·23s	40·00e
Yala, river	0·04n	34·09e
Yatta Plateau	2·00s	38·00e

cording to changes in altitude and physiography. In the Lake Victoria Basin, annual maximum mean temperatures range from 72° to 93" F (22" to 34° C), and the mean minimum from 50° to 64° F (10° to 18° C); precipitation varies between 40 and 70 inches annually. The higher areas of the Rift Valley experience mean temperatures of 60° to 80° F (16" to 27" C) and rainfall of from 20 to 60 inches. The highlands often experience frost and mist; mean temperatures vary from 55° to 60° F (13" to 16° C); and about 45 inches of rainfalls yearly from March to September in the west, and from 35 to 60 inches in the east, occurring from March to May and from October to December. Annual maximum mean temperatures over the eastern plateau range from 72" F (22" C) to less than 64° F (18" C) and the minimum mean temperatures from 50" F (10" C) to less than 43° F (6° C). Over the arid regions maximum mean temperatures are well above 79" F (26" C) and the minimum mean well over 57° F (14" C). The eastern plateau receives only 20 to 30 inches (500 to 750 millimetres) of precipitation annually, while the arid regions receive less than 20 inches. The coast has a mean annual temperature of more than 80" F (27" C) and rainfall of between 40 and 60 inches (1,000–1,500 millimetres).

Vegetation and animal life. In the highlands the characteristic landscape consists of patches of evergreen forest separated by wide expanses of short grass. Above the forest, a zone of bamboo extends to about 10,000 feet, beyond which the mountain moorland includes tree heaths, tree groundsel (a foundation timber of the *Senecio* genus) and giant lobelia (a widely distributed herbaceous plant). East and west of the highlands, the vegetation gives way to low trees that are scattered through grass five to eight feet high. On the northern and southern margins of the highlands, flat-topped trees are scattered through an even cover of shorter grass. Semidesert conditions below 3,000 feet give rise to thornbush, interspersed with acacias and thick-trunked baobabs. In the still drier areas of the north, desert scrub occurs, exposing the bare ground. In the coastal belt, dense high bush alternates with limited areas of forest and extensive open glades. Coconut palms are characteristic at the high-tide line, and stands of mangrove occur in the tidal estuaries and lagoons.

Large mammals

Kenya is noted for its big game, which ranges especially over the dry thornbush country in numbers that vary greatly in accordance with migrational movements. Large carnivores include the lion, leopard, cheetah, and wild dog. The commonest animals are herbivorous, however. They include the elephant, buffalo, and rhinoceros, which are found in open country as well as in the mountain forests, and huge herds of zebras, antelope, and gazelles, which range over the open plains. The wildebeest, hartebeest, impala, Thomson's gazelle, Grant's gazelle, oryx, eland, and giraffe are some of the other numerous herbivores. The hippopotamus and crocodile are common in most of the large rivers, swamps, and lakes. Hyenas and jackals play a useful role as scavengers. Baboons frequent the open country, while Sykes' and colobus monkeys live in the forests.

The many species of snakes include the mamba, cobra, puff adder, and python. Among Kenya's birds are ostriches, storks, several kinds of eagles and vultures, guinea fowl, weavers, and hornbills. By the lake shores flamingos, pelicans, herons, ibis, and ducks abound. There are also many species of migrant birds from Europe. The main mountain streams are well stocked with trout. The ngege (a nest-building freshwater fish of the genus *Tilapia*) is the principal fish in Lake Victoria, and Kenya's sea fisheries include big game fishing.

The landscape under human settlement. ***Rural*** *settlement.* More than 90 percent of the population lives in dispersed rural settlements. In the greater part of the country, land is held by traditional communal ownership. Virtually no community settlements, such as villages or towns, have developed except along the coast. Those that do exist in the interior have resulted primarily from the introduction during the colonial period of mission stations, administrative centres, and markets. Even where land reforms have resulted in the enclosure of individual land, new settlements have not been formed. Village life, where it exists, tends to centre around the activities of foreign missions.

Urban settlement. Pre-European urbanization was confined to Arab trading ports and towns established in the narrow coastal strip; a number of these, which often have a long history, still exist. More recent urban centres developed during the period of European colonization. As the railway penetrated into the interior, towns grew up along the route, as well as along the feeder roads that linked the rail line to the rest of the country.

The small town with a population of 3,000 to 10,000 is the dominant form of urban settlement. Such towns are, however, overshadowed by the capital city of Nairobi (*q.v.*), with its population of about 500,000, and by the Indian Ocean port of Mombasa; together the two cities account for approximately 69 percent of the total urban population. Apart from these, the only cities with populations of between 30,000 and 50,000 are Nakuru and Kisumu. Although most of the towns were founded for administrative reasons, their roles as centres of economic growth have taken on importance, and as more people move to the cities they play a dynamic role in the changing balance between rural and urban populations. Most of these towns are located in the highland region where European settlement was mostly concentrated. Since independence, however, the tempo of urbanization has been increasing on the coast and in the Lake Victoria Basin.

THE PEOPLE

Ethnic composition. Almost the entire population is composed of Africans. Other groups—including Asians, Europeans, and Arabs—amount to only 2 percent of the country's total inhabitants. The 10,700,000 Africans are divided into 70 different tribal groups that are distinguished from each other by language and culture. The overwhelming proportion of these peoples are agriculturalists; except for city dwellers, the rest are either herdsmen, hunters, or fishermen. Although each group tends to populate its traditional territory, different groups often coexist in a complex ecological relationship that is most developed in the southwest.

Language and tribal groups

The African peoples can best be distinguished according to the three major Kenyan language groups: Bantu, Nilotic, and Cushitic. The Bantu generally inhabit central and western Kenya; the Nilotic speakers are found in the west and northwest; and the Cushitic peoples occupy the northeast. The 11 largest tribal groups compose 92 percent of the African population; these are the Bantu-speaking Kikuyu, Kamba, Meru, Gusii, Luhya, and Mijikenda (Nyika); the Nilotic-speaking Luo, Masai, Kalenjin, and Turkana; and the Cushitic-speaking Somali. Of these 11 groups, the Kikuyu and Luo are the most numerous, and both tend to hold a great deal of political power.

Most of the Kenya tribes are well represented in the rapidly growing urban centres. The accelerating rate of urbanization and the tendency of the various ethnic groups to retain their identity against this process are well-known problems. Although urbanization tends to minimize tribal differences, the majority of the ethnic groups still retain strong links with rural origins. The dangers of tribal rivalries and consequent conflicts are problems to which political leaders have constantly drawn attention.

The Indian population of about 140,000 is found in the urban centres throughout the country. The Indians are traders; in the cities they are also engaged in large businesses. The Indian communities are cohesive, with their social activities usually restricted to their own groups. They retain their own languages—Hindustani, Punjabi, Gujarati, and Konkani—and adhere to the Islāmic or Hindu religions.

The white, or European, community numbers about 41,000. Most Europeans are from the United Kingdom, but there are also small numbers of continental Europeans as well as some Americans. The white population is variously employed, mostly as farmers, businessmen, civil servants, and missionaries. Because of the colonial legacy, their education, and their comfortable economic status, the white community enjoys an elite status. The Europeans tend to preserve their Western culture, and almost all speak English.

Arabs have continued to occupy the coastal strip since medieval times; they were—and often still are—engaged in trade between Africa and Asia. They number about 28,000, still speak Arabic, and adhere to Islām. The community is divided between "Old" or "True" Arabs—respectively, those who are descended from the early traders and those whose ancestors migrated to Kenya after the Portuguese arrived in the 16th century.

Prevalence of Swahili

About 75 different languages are spoken in Kenya. The lingua franca is Swahili, a multipurpose language that evolved on the coast and is spoken in several dialects throughout the country. Essentially a Bantu language, it contains an enormous Arabic vocabulary, as well as some Portuguese, Hindi, and English loanwords; it has evolved a standard written form. It is the language of trade, newspapers, and courts and is taught as a second language in the school system. In December 1971 Swahili became Kenya's only official language; previously English also had that status.

There is no state religion. Koman Catholicism, Anglicanism, and their reformist offshoots represent the main Christian religions. **Islām** is well established along the coast and in urban areas.

Kenya, Area and Population

	area*		population	
	sq mi	sq km	*1962* census	*1969* census
Area				
Nairobi	264 264	684 684	344,000	509,000
Provinces				
Central	*5,092* *5,093*	*13,188* *13,191*	*1,335,000*	*1,676,000*
Coast	*32,062* *32.279*	*83,041* *83,603*	*741.000*	*944.000*
Eastern	*59,668* *61,734*	*154,540* *159,891*	*1,557,000*	*1,907,000*
North-Eastern	*48,997* *48.997*	*126,902* *126,902*	*269,000*	*246,000*
Nyanza	*4,836* *6,240*	12,525 *16,162*	*1,634,000*	*2,122,000*
Rift Valley	*65.694* *67,125*	*170,147* 173,854	*1.740.000*	*2,210,000*
Western	*3,175* *3.228*	*8,223* *8,361*	*1,014,000*	*1,328,000*
Total Kenya	*219,788* *224,960*	569,251† 582,646†	8,636,000‡§	10,943,000§

*Where two figures are given. the first is the land area, the second the total area. †Converted area figures do not add to total given because of rounding, ‡Includes *3,000* people in transit at the time the census was taken. §Figures do not add to total given because of rounding.
Source: Officialgovernment figures.

Demography. The regional distribution of population exhibits marked inequalities. The northern half of the country is only sparsely inhabited; most of the population is concentrated in the Lake Victoria Basin, the Rift Highlands, and the coastal belt. On the eastern plateau isolated settlements occupy the more elevated areas, leaving extensive stretches of relatively empty country between them. The average population density was about 47 persons per square mile at the 1969 census, while a maximum of 1,400 persons per square mile is reached in Western Province and a minimum of less than 10 per square mile occurs in the arid areas of the north.

The population is young; almost 74 percent are under 30 years of age. The total population of almost 12,000,000 is rapidly expanding, generated by a birth rate of 50 per 1,000 and a death rate of 17 per 1,000. There are more males than females.

THE ECONOMY

Resources **and** resource exploitation. ***Agriculture.*** Agriculture is the mainstay of Kenya's economy, contributing over one-third of the gross domestic product, about 60 percent of the nation's exports, and one-third of all wage employment. About four-fifths of the country is unsuitable for cultivation, however, and farming is restricted to the highlands, the western plateaus, and parts of the coast. Irrigation projects have opened some additional lands to cultivation along the Tana River and at Ahero in western Kenya. The main cash crops are coffee, tea, sisal, wattle (bark of the wattle tree, used for manufacturing tannin), and sugar, which are grown on both plantations and small farms. Grains, pyrethrum (chrysanthemums whose leaves are used in the manufacture of insecticides), cotton, tropical fruit, coconuts, and cashew nuts are also grown for market. Staple crops include maize (corn), pulses (leguminous plants, such as peas, beans, and lentils), millet, bananas, sweet potatoes, and potatoes. Stock raising is also important, including cattle for dairying and for meat. Most of the country's livestock is kept in the dry regions by subsistence pastoralists, or herdsmen, who seldom market their animals. Commercial stock raising is practiced on small farms as well as on large ranches. Hides, skins, and wool are produced both for domestic consumption and for export.

The farm-cooperative program

Agricultural development since independence has been marked by two basic programs: the Africanization of land ownership, and the encouragement of farm cooperatives. About 1,000,000 acres previously owned by Europeans had been transferred to about 30,000 African farmers by the early 1970s, and further transfers of ownership from a willing European seller to a willing African buyer are encouraged. The approximately 900 agricultural cooperatives, whose members are mainly small African farmers, are responsible for marketing, production, and maintaining quality standards. These cooperatives are most popular among the growers of pyrethrum, coffee, and sugar, as well as among dairy farmers.

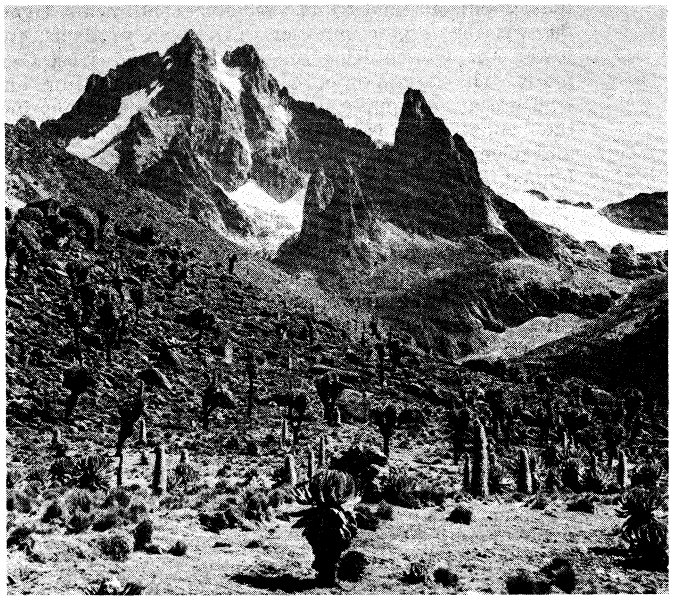
Mining and quarrying. The mining industry, though relatively small, has a potential for expansion. Geological surveys have revealed small deposits of minerals including copper, gold, silver, iron ore, lead, zinc, nickel, barite (barium sulfate), cyanite (an aluminum silicate), and wollastonite (calcium silicate). Development is restricted by the necessity of large capital outlays, the remoteness of some of the deposits, and the limited internal market. There are promising possibilities for the recovery of niobium (a metallic element used in alloy steels) and europium (a rare-earth metallic element with light-pink salts used in the production of colour television) at Mrima Hill in the coastal region. Similar possibilities exist for lead and zinc near Mombasa, as well as the limestone deposits on the coast, near Nairobi, and at Homa Bay. Sodium carbonate (soda ash) is now being worked.

Forestry. The total area of forest reserve of about 4,330,000 acres represents approximately 3 percent of the total area of the country. Most of this reserve is natural forest bush bamboo and grass, and the remainder consists of planted softwoods. The forests help to conserve soil and water and are also used used as sources of fuel and timber. Resources are estimated to amount to 17,600,000 cubic feet of timber.

Fisheries. Fish form an important supplement to the diet of subsistence farmers, but fishing contributes less than 1 percent to the nation's total income. The considerable fishing potential of Lake Rudolf, regarded as one of the largest—and virtually untapped—freshwater fisheries in Africa, has a potential annual production estimated at from 20,000 to 100,000 tons. So far, however, three-quarters of the freshwater fish landed in Kenya come from Lake Victoria. There are also plans to expand the inshore exploitation of crustaceans and to develop deep-sea fishing. The development of Lake Rudolf, located in far northern Kenya, will further necessitate the provision of adequate transportation and processing facilities.

Tourism. The nation's wildlife and natural beauty form the basis of an expanding tourist industry. Because much of the country is unsuited to agriculture, vast areas have been developed as game reserves. In the early 1970s tourism accounted for more than 2 percent of the national income, and there were well over 400,000 annual visitors to the national parks. The most popular tourist centres are Nairobi National Park, which is conveniently located close to the capital city, and Tsavo National Park, the largest game reserve. Other important attractions include Mt. Kenya, Aberdare, Lake Nakuru, and Meru National parks. There are also rapidly growing sporting and holiday facilities along the coast, both north and south of Mombasa. Malindi is an important coastal resort that attracts a growing number of overseas as well as local visitors.

Authenticated News



Peaks of Mt. Kenya, the highest mountain in Kenya. Although near the Equator, it has a permanent covering of ice and snow.

Fuel and power. Natural-fuel resources are limited to timber, from which charcoal is also made. Prospecting for natural gas and petroleum had not met with any success by the early 1970s. Crude oil imported from the Persian Gulf is refined at Mombasa.

In the early 1970s, Kenya produced about 513,000,000 kilowatt-hours of electricity yearly by thermal (oil-fuelled) or hydroelectric generation. As this did not meet domestic demand, 218,000,000 kilowatt-hours had to be imported from Uganda and Tanzania. The East African Power and Lighting Company distributed electricity to consumers, and the Kenya Power Company purchases as well as generates electric power. The Seven Forks hydroelectric scheme, administered by the Tana River Development Company, was established to increase domestic output. The first stage of the Tana River development—which began to operate in 1968 at Kindaruma—will have an initial capacity of 40,000 kilowatts. It is planned that the completed scheme will produce 300,000 kilowatts.

Industrial development

Manufacturing. Manufacturing industries were introduced into Kenya after World War II, and by the early 1970s about 1,000 such enterprises had been established, mainly in Nairobi and Mombasa. In the late 1960s the construction and food-processing industries grew the fastest. The production of textiles declined, but the manufacture of footwear rose. Products include meat preparations and frozen fish, chemicals and petroleum, pyrethrum, beer, and tobacco.

Financial services and foreign trade. *Financial services.* The Central Bank of Kenya, established in 1967, regulates the issue of notes and coins; assists in the development of a sound monetary, credit, and banking system; and serves as banker and financial adviser to the government. In addition the Commercial Bank of Kenya grants short-term or seasonal loans. The nation is also served by six international banks. The raising of funds and investment is promoted by the Nairobi Stock Exchange, which also floats public loans. An exchange control regulates the flow of private capital out of the country.

The unit of currency is the Kenyan shilling (KShs.), which is divided into 100 cents. Twenty Kenyan shillings equal one Kenyan pound (K£); (KShs. 7.14 equal $1 U.S.; KShs. 17.14 equal £1 sterling on April 1, 1971).

Foreign trade. In 1971, exports were valued at about K£111,000,000, and imports at about K£198,000,000. Coffee, tea, and sisal constituted about half of the exports; all of them were subject to fluctuating agricultural prices. Other important exports included petroleum products, chemicals, and fruits, vegetables, and meats. Uganda was the largest importer of Kenyan products, followed the United Kingdom, Tanzania and West Germany. The largest imports were crude petroleum and transportation equipment. Other products brought into the country included paper, synthetic fibres, medicines, and scientific instruments. Major import sources were the United Kingdom, West Germany, Uganda, Japan, and the United States.

A unified customs and excise service—a department of the East African Community—is in operation in Kenya, Uganda, and Tanzania. Nondiscriminatory customs dues are fixed at levels appropriate to the commodity; capital equipment is usually admitted free. Uganda and Tanzania are Kenya's major trading partners in Africa.

Attempts to secure foreign investment

Management of the economy. Economic policy recognizes that domestic resources available for development are limited and therefore encourages the investment of foreign capital. The investment of private capital in projects that will increase both production and employment is welcomed. The government also seeks to save foreign exchange and to ensure that development projects are given an equitable geographic distribution.

The central government, local authorities, and the East African Community all participate in various economic activities. Between 1960 and 1970, central government expenditure more than doubled, and in the late 1960s, government investment contributed about 27 percent annually to the gross domestic product.

Taxation. Income tax is levied upon incomes that are earned in or are derived from East Africa, providing the government with one-quarter of its revenue. Subject to certain conditions, personal allowances may be deducted according to marital status and age. Deductions are allowed for capital expenditure. A graduated personal tax is levied upon every adult living in Kenya in order to help pay for local public services. Companies are subject to a corporation tax or to an undistributed income tax.

Employment and trade unions. In the 1960s there were 625,000 employees and self-employed persons out of a total labour force of 3,200,000. The normal working week consists of 45–48 hours, depending upon the trade or profession. The Kenyanization Bureau is responsible for increasing the employment of Kenyans in all sectors of the economy. In 1965 the government enacted the National Social Security Fund and National Social Security Fund Act, which provide insurance coverage for employees in return for wage deductions.

Workers' and employers' unions

The national constitution guarantees the right of workers and employers to form trade unions for purposes of collective bargaining on wages and terms of employment. The Central Organisation of Trade Unions (COTU) was established in 1965, amalgamating the smaller unions then in existence. In the early 1970s there were 47 registered trade unions affiliated to the Central Organisation of Trade Unions. The Kenya Federation of Employers is composed of 16 employers' unions. Provision is made for settlement of a trade dispute through voluntary bargaining by the parties concerned, or by referring the dispute to arbitration.

Transportation. *The railways.* Almost 1,300 miles of railways are operated in Kenya by the East African Railways Administration. The main line runs northwestward from Mombasa through Nairobi, Nakuru, and Eldoret to the Uganda border at Tororo. In addition, there are branch lines from Nakuru to the Lake Victoria port of Kisumu; from Voi to Kahe in Tanzania; and from Konza to the soda deposits at Lake Magadi. Other lines running to agricultural regions include those from Gilgil to Thomsons Falls, from Rongai to Solai, from Leseru to Kitale, and from Kisumu to Butere.

Harbours. Mombasa, the principal port of Kenya, handles all the import and export traffic of Kenya, Uganda, and the northern Tanzanian mainland. At Mombasa, which is situated on an island, harbour activities are carried on at Kilindini—which extends to Kipevu—as well as at the old Mombasa Harbour on the eastern side of the island. The Mombasa-Kilindini Harbour has 13 deepwater berths, together with a jetty for the oil tankers that supply the refinery. In 1971 about 3,900,000 tons of cargo, including oil, were landed annually at Mombasa, and about 1,700,000 tons loaded. The ports of Lamu and Malindi provide additional but limited facilities; they mainly serve the coastal trade and fisheries.

Inland waterways. The services operated on Lake Victoria provide important links between Kenya and Uganda and Tanzania. From Kisumu, lake steamers serve the main ports, including Mwanza and Port Bell. In addition, lighter services serve the lesser ports. There is a marine base with work shops and a dry dock at Kisumu.

Road building

Roads. Kenya has an extensive road network covering the more populous parts of the country. There are plans to extend the system to the more arid regions. A highway from Nairobi to Addis Ababa in Ethiopia is under construction. A large proportion of development funds have been devoted to improvements in road communications. Between 1960 and 1971 paved road mileage almost trebled, increasing from 700 to 1,900 miles. Since the mid-1960s total receipts from the road transport industry consequently increased by 61 percent, exceeding the receipts from railway transport. Registration of new vehicles has increased by about 10 percent a year; with a sharp rise in the registration of trucks and 'motorcycles. Priority is placed by the government on the development of feeder roads as part of the rural-development program.

Aviation. Nairobi Airport, one of the main international airports of East Africa, serves 20 international airlines. Its runways are being extended to accommodate the largest jets. The airport is situated more than eight miles from the centre of Nairobi in the Athi-Kapiti plains.

Domestic routes link Nairobi, Mombasa, Malindi, Kisumu, and Kitale; there are connections to international air routes through Nairobi, as well as through Dar es Salaam in Tanzania and Entebbe in Uganda. In addition, several private airlines provide charter service to smaller centres in East Africa.

ADMINISTRATION AND SOCIAL CONDITIONS

Government structure. *The constitutional framework.* Kenya is a republic and a member of the Commonwealth of Nations. The constitution consists of a British Order in Council, made at the time of independence in 1963, which has since been radically amended by the Kenya parliament. It provides for a parliament consisting of the president and the National Assembly. The president is the head of state and commander in chief of the armed forces. The unicameral assembly consists of 158 elected members and 12 members who are elected by the National Assembly sitting as an electoral college. Elections are by universal adult suffrage, and are held at least every five years. The speaker and the attorney general are ex officio members of the National Assembly. The president appoints the vice president from among the ministers, who are all elected members of the assembly. The Cabinet consists of 23 ministers, including the president, vice president, and attorney general. There are 29 assistant ministers appointed by the president from among the members of the National Assembly.

Local government. Kenya is divided into seven provinces: Central, Rift Valley, Nyanza, Western, North-Eastern, Eastern, and Coast. The administrative district of Nairobi forms a separate unit. Local government consists of appointed provincial commissioners and elected county councils and township or municipality authorities. The commissioners are responsible for education, transport, and health in their provinces, while the councils are concerned with services. An appointed district commissioner serves as an ex officio member of the county council. Local advisory committees composed of the councillors of a specific area advise the councils on local matters. Mayors and county chairmen are appointed from names submitted by the elected councillors.

The Kenya African National Union

The political process. Although the constitution provides for a multiparty system, Kenya is at the moment a one-party state. The ruling part is the Kenya African National Union (KANU), which is headed by the president. Party officials include the president, provincial vice presidents, and district chairmen. Participation in the party's activities is through membership. The election of party candidates for the National Assembly is by secret ballot.

De facto political developments. The former opposition party, the Kenya People's Union, is banned, and no other party was registered by the early 1970s. Political representation tends to reflect the geographical distribution of ethnic groups. There is, however, a strong national identity among members of KANU.

Justice. The judiciary is independent and is headed by the chief justice and 11 puisne (associate) judges of the High Court, which has full civil and criminal jurisdiction. There are also resident magistrates' courts and district magistrate courts. Judges are appointed by the president. All legal practitioners are advocates of the High Court and may appear in all courts. The final court of appeal is the Court of Appeal for East Africa in Nairobi. Although there is only one judiciary, Islāmic customary law is accepted in the courts.

The armed forces. The army totals about 6,500 men and is headed by a major general; it is divided into four battalions and a paratroop company. In addition, there is a 2,000-man paramilitary organization, the General Service Unit, which is organized on a territorial basis and which assists in the maintenance of law and order. The Kenya Air Force and the Kenya Navy each consist of a small nucleus capable of further expansion. The air force possesses a number of training and transport planes. The small navy operates patrol boats.

The police. Kenya has the largest and best equipped police force in East Africa. Its 12,000 members are headed by a commissioner of police who is appointed by the president; he is assisted by a number of senior assistant and provincial commissioners. Responsibility for the police, immigration, and prison administration lies with the vice president, in his, capacity as minister of home affairs.

Social conditions. ***Education.*** A single national educational system consists of three educational levels. At the first level, there are seven years of primary education; followed by a further six years of secondary education; the third level is represented by university education, which requires from three to five years of study in order to sit for degree examinations. When primary schooling is completed, entrance to the secondary level is contingent upon passing an examination.

Secondary education has expanded markedly since independence, and greater emphasis has been placed upon technical education. As accommodations available in government-supported secondary schools are limited, self-help secondary schools, which now outnumber government secondary schools, have accounted for much of the expansion. The self-help schools receive liberal government assistance, both in the provision of teachers and learning materials.

The University of Nairobi, which, until 1970, was part of the University of East Africa, is now Kenya's independent national university. University education is oriented toward the promotion of Kenya's social and economic development.

Housing. Because of a rapidly expanding population, the provision of housing is a critical national problem. In the rural areas it is a question of improving the quality of the existing housing, but in the cities and towns the influx of population has created a demand for more housing units. Investment in housing, both public and private, is increasing. Government housing programs for the construction of rental flats and private units in the cities are the responsibility of the National Housing Corporation, an agency of the Ministry of Housing. Inproved rural housing is promoted by government housing loans. To safeguard the public against excessive rental charges, the government has established a Rent Restriction Department; rent tribunals meet periodically in Nairobi, Kisumu, and other towns.

Rural housing

Rural housing consists of circular or rectangular buildings of mud and thatch. Tin roofing, cement, bricks, and tiles are commonly used as building materials. Groups of buildings, or compounds (enclosures with dwellings), are separated from the surrounding fields by fences, hedges, stone walls, or courtyards. Housing in the towns is of various types, and sanitation facilities are provided in varying degrees. The town dwellings occupied by recent African migrants are similar to those in the rural areas, and are not well suited to crowded urban conditions. Other buildings include shop-houses, apartment buildings, and bungalows. Most homes are rented, but since the mid-1960s the government has promoted private ownership.

Health and welfare services. Health services are provided by the central government, local authorities, voluntary agencies, and private practitioners. There are almost 1,800 medical practitioners, as well as dentists, pharmacists, and a large number of registered and enrolled nurses and midwives. Government hospital facilities slightly exceed those available at private or mission hospitals.

Emphasis is changing from curative to preventive medicine, and the government plans a rapid expansion in paramedical personnel. Maternal and child health care and family-planning services are provided. The government participates in a wide range of community and social welfare services to promote the organization of self-help activities, such as the development of family- and child-welfare services, the care and rehabilitation of the disabled, providing assistance to the handicapped, and the promotion of adult education.

Cultural life. Kenya, a country of diverse and rich cultural traditions, seeks to cultivate and develop those traditions to ensure that its valuable cultural assets are not irretrievably lost and that social cohesion is not undermined in the process of change to newer ways.

The national museum

A National Archive Service has been established, and it is saving an increasing number of documents. A national library service board has also been established to equip, maintain, and develop libraries in Kenya, including a branch library service. Kenya has a national museum of world repute; it contains collections of wildlife, archaeological remains, and objects of Kenya's material culture. The museum cooperates closely with the University Cultural Division of the Institute for Development Studies. The division promotes and conducts research in history, archaeology, social anthropology, musicology, literature, traditional arts and crafts, and systems of belief.

Performing arts and literature. The Kenya National Theatre is incorporated in the Kenya Cultural Centre. A theatre school was founded in 1968 to provide professional training in theatrical techniques, which include the writing of plays by Kenyan authors and the performance of traditional music and dance. Music and dance play an integral role in social and religious life. Rhythm, all-important, is largely provided by the drum, supplemented by wind and stringed instruments. Swahili literature, both oral and written, is traditional in form and content. Contemporary novelists, including James Ngugi and Mugo Gatheru, deal with the social frictions between traditional and modern society. Visual arts are largely confined to the mass production of wood sculpture for the tourist trade. Elimo Njau and Ronal Rankin are popular Kenyan painters.

The media. The national press consists of daily and weekly newspapers published either in Swahili or English. It is independent and uncensored, but it is expected to support government developmental policies. Radio services are provided by the government-owned Voice of Kenya, which broadcasts from Nairobi in Swahili, English, and several African vernaculars, as well as in Hindustani and Gujarati. There are transmitters at Mombasa and Kisumu. News broadcasts and public affairs themes are popular; other programs cover religion, sports, music, and drama. Television, also government-sponsored, is available in Nairobi and in the Rift Valley, West, and Coast provinces. Programs are broadcast simultaneously in Swahili and English; many are imported from the United States and the United Kingdom. In the towns, movie theatres show full-length feature films that largely come from the United States, the United Kingdom, and India. Educational and documentary films are popular and are shown in rural areas also by mobile units.

Prospects. The continuation and extension of development programs aiming at Kenya's social and economic advancement may be expected to provide the focus of national effort in the future. Economic progress, which in Kenya is dependent upon the participation of the individual through community activities, is also dependent to some extent upon foreign investment. National income is also, to some extent, subject to fluctuations in commodity prices and to the threat of drought. Health, sanitation, and housing improvements are being made, but gains in these domains are challenged by the increasing birthrate, although family-planning programs are also being introduced. Progress is dependent upon the preservation of an atmosphere of sustained national unity.

BIBLIOGRAPHY. F.F. OJANY, *The Physique of Kenya* (1966), is a short but comprehensive paper on the Kenya landscape; S.H. OMINDE, *Land and Population Movements in Kenya* (1968), a comprehensive geographical account of the country and its population in the context of planning; KENYA COLONY AND PROTECTORATE, *Survey of Kenya: Kenya Atlas*, 3rd ed. (1971), contains texts and photographs. Development aspects of Kenya are discussed in S.H. OMINDE, "Kenya," in *Developing Countries of the World* (1968); and R.B. OGENDO, *Some Aspects of the Geography of Electricity in Kenya* (1965). Useful information on population trends and problems is contained in *United Nations Mission to Kenya on Housing* (1965); *Family Planning in Kenya: Report of the Mission of Popzrlation Council of the United States of America* (1967); and in *East Africun Royal Commission Report 1953–1955*, Cmd 9475. Important sources of population data include the following volumes from the Kenya Population Census 1962: *Population Census Advanced Report:* vol. 1 and 2 (1964); vol. 3: *African Population* (1966); and vol. 4, *Non-African Population* (1966). The most recent population data is published in the *Kenya Population Census* (1969, 1970). Current industrial policy and conditions are treated in a summary from *A Guide to Industrial Investment* (1968), published by the MINISTRY OF COMMERCE AND INDUSTRY. W.T.W. MORGAN (ed.), *East Africa: Its Peoples and Resources* (1969), is a useful collection of 20 essays on East Africa including Kenya.

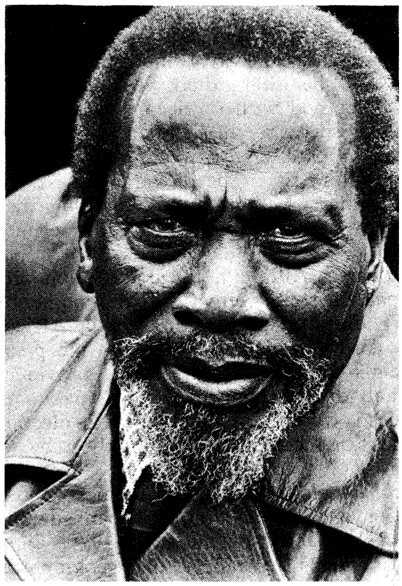
(S.H.O.)

Kenyatta, Jomo

The career of Jomo Kenyatta, an African statesman and nationalist leader, spans the entire period of colonial rule in British East Africa. A heavy-set, powerful man with considerable personal magnetism and a gift for oratory, he became, in 1963, the first prime minister of independent Kenya.

Early life

He was born as Kamau, son of Ngengi (c. 1894), at Ichaweri, southwest of Mount Kenya in the East African highlands. His father was a leader of a small agricultural settlement; his grandfather was a *murogi*, or diviner, whose knowledge of medicine and magic set him apart from other people. Like all Kikuyu boys, Kamau learned hunting skills, close observation, memory discipline, social obligations and responsibilities, family and clan history. From his grandfather he learned herbal remedies and gained a respect for the spiritual knowledge and powers of the diviner.



Kenyatta.
John Moss—Black Star

At about the age of ten Kamau became seriously ill with jigger infections in his feet and one leg, and he underwent successful surgery at a newly established Church of Scotland mission. This was his initial contact with Europeans. Fascinated with what he had seen during his recuperation, Kamau ran away from home to become a resident pupil at the mission. He studied the Bible, English, mathematics, and carpentry and paid his fees by working as a houseboy and cook for a European settler. In August 1914 he was baptized with the name Johnstone Kamau. He was one of the earliest individualists and outcasts among the Kikuyu to run away from the confines of his own culture. And, like many others, Kamau soon left the mission life for the bright lights of Nairobi.

There, he secured a job as a clerk in the Public Works Department. During World War I Kamau enjoyed his relative affluence, impressing people with flashy clothes and a general savoir-faire. He added the name Kenyatta, the Kikuyu term for a fancy belt, which he affected. After serving briefly as an interpreter in the High Court, Kenyatta transferred to a post with the Nairobi Town Council. About this time he married and began to raise a family. According to his younger brother, he was "not interested in politics."

The first African political protest movement in Kenya against a white-settler-dominated government began in 1921—the East Africa Association (EAA), led by an educated young Kikuyu named Harry Thuku. Kenyatta joined the following year. In March 1922 Thuku was arrested; he was later deported and overt protest was silenced. Kenyatta, however, continued to work privately for the EAA as propaganda secretary. As a government employee, he was supposed to avoid politics, but he managed to remain inconspicuous. In 1925 the EAA disbanded as a result of government pressures, and its members re-formed as the Kikuyu Central Association (KCA). Three years later Kenyatta became its general secretary, though he had to give up his municipal job as a consequence.

Entrance into full-time politics

In May 1928 Kenyatta launched a monthly Kikuyu-language newspaper *Mwigithania* ("He Who Brings Together"), aimed at gaining support from all sections of the Kikuyu. The paper was mild in tone, preaching self-improvement, and was tolerated by the government. But soon a new challenge appeared. A British commission recommended a closer union of the three East African territories (Kenya, Uganda, Tanganyika). Settler leaders supported the proposal, expecting that internal self-government might follow. To the KCA such a prospect looked disastrous; in February 1929 Kenyatta went to London to testify against the scheme.

In London the secretary of state for colonies refused to meet Kenyatta, but several groups critical of British colonialism aided him; the League Against Imperialism arranged a brief trip to Moscow for him, from August to October 1929. The following July Kenyatta attended the International Negro Workers' Conference at Hamburg. His hosts urged the unity of the black proletariat as a worldwide exploited class, but Kenyatta's interests remained rivetted on the sufferings of his own people. On March 26, 1930, he wrote an eloquent letter in ***The Times*** of London setting out five issues championed by the KCA: (1) security of land tenure and the return of lands alienated to Europeans, (2) increased educational facilities, (3) repeal of hut taxes on women, which forced some to earn money by prostitution, (4) African representation in the Legislative Council, (5) noninterference with traditional customs.

He concluded by calling the lack of these measures "a short-sighted tightening up of the safety valve of free speech, which must inevitably result in a dangerous explosion—the one thing all sane men wish to avoid."

Again in 1931 Kenyatta's testimony on the issue of closer union was refused despite the help of liberals in the House of Commons. In the end, however, the government abandoned its plan for union for the time being. Kenyatta did manage to testify on behalf of Kikuyu land claims in 1932 at hearings of the Carter Land Commission. The commission decided to offer compensation for some alienated territories but maintained the "white-highlands" policy, which restricted the Kikuyu to overcrowded reserves. Kenyatta again visited the Soviet Union and eastern Europe, and on his return he supported himself from 1933 to 1936 as a phonetics informant at University College, London. He studied anthropology in 1936 at the London School of Economics. His thesis was revised and published in 1938 as ***Facing Mount Kenya,*** a study of the traditional life of the Kikuyu characterized by both insight and a tinge of romanticism. It defended a way of life that was already deeply eroded, and it ignored the extensive adaptations Kikuyu had made to European culture. The book signalled another name change, to Jomo (Burning Spear) Kenyatta.

During the 1930s, Kenyatta briefly joined the Communist Party, met other black nationalists and writers, and actively organized protests against the Italian invasion of Ethiopia. The onset of World War II temporarily cut him off from the KCA, which was banned by the Kenya authorities as potentially subversive. Kenyatta maintained himself in England by lecturing for the Workers Educational Association and working as a farm labourer. But he continued to produce political pamphlets publicizing the Kikuyu cause.

Return to Kenya

Kenyatta helped organize the fifth Pan-African Congress, which met in Manchester on October 15–18, 1945, with W.E.B. Du Bois of the United States in the chair; Kwame Nkrumah, the future leader of Ghana, was also present. Resolutions were passed and plans discussed for mass nationalist movements to demand independence from colonial rule. Kenyatta returned to Kenya in September 1946 to take up leadership of the newly formed Kenya African Union, of which he was elected president in June 1947. From the Kenya African Teachers College, which he directed as an alternative to government educational institutions, Kenyatta organized a mass nationalist party. But he had to produce tangible results in return for the allegiance of his followers, and the colonial government in Kenya was still dominated by unyielding settler interests. The "dangerous explosion" he had predicted in 1930 erupted in Kenya as the Mau Mau rebellion of 1952.

On October 21, 1952, Kenyatta was arrested at his home at Gatundu. Police seized documents and arrested 98 other African leaders. Despite government efforts to portray Kenyatta's trial as a criminal case, it received worldwide publicity as a political proceeding. In April 1953 Kenyatta was sentenced to a seven-year imprisonment for "managing the Mau Mau terrorist organization." He denied the charge then and ever afterward, maintaining that the Kenya African Union's political activities were not directly associated with Mau Mau violence.

The British government responded to African demands by gradually steering the country toward African majority rule. In 1960 the principle of one man, one vote was conceded. Kenya nationalist leaders such as Tom Mboya and Oginga Odinga organized the Kenya African National Union (KANU) and elected Kenyatta (still in detention despite having completed his sentence) president in absentia; they refused to cooperate with the British while Kenyatta was detained. In a press conference Kenyatta promised that "Europeans would find a place in the future Kenya provided they took their place as ordinary citizens."

Kenyatta was released in August 1961, and, at the London Conference early in 1962, he negotiated the constitutional terms leading to Kenya's independence. KANU won the pre-independence election in May 1963, forming a provisional government; Kenya celebrated its independence on December 12, 1963, with Kenyatta as prime minister.

A year later Kenya became a one-party republic with a strong central government under Kenyatta as president. Kenyatta, who had spent his youthful years herding his father's goats on the ridges of Kikuyuland, quite unaware of the greater world outside, presided over the complex problems of postindependence development and instructed his United Nations delegates on the subtleties of a "nonaligned" foreign policy. Always—in spite of his imprisonment by the British authorities—the most pro-British of African leaders, Kenyatta made Kenya the stablest black African country and one that attracted foreign investment on a broad scale. Under his leadership the economy prospered: agriculture, industry, and tourism all expanded. He died at Mombasa on August 22, 1978.

BIBLIOGRAPHY. GEORGE DELF, ***Jomo Kenyatta: Towards Truth About "The Light of Kenya"*** (1961), written while Kenyatta was still detained, defends Kenyatta from charges of leading Mau Mau. A later biography, by JEREMY MURRAY-BROWN, is ***Kenyatta*** (1972). ANTHONY HOWARTH, ***Kenyatta: A Photographic Biography*** (1967), assembles information on Kenyatta's early life from a variety of published sources and includes excellent pictures. Two works by KENYATTA are ***Facing Mount Kenya*** (1938), a study of traditional Kikuyu life from a sympathetic and even romantic viewpoint but containing much insight; and ***Kenya: Land of Conflict,*** with an introduction by GEORGE PADMORE (1944), a pamphlet that signals Kenyatta's shift toward more aggressive demands on behalf of Kenya Africans. CARL ROSBERG and JOHN NOTTINGHAM, ***The Myth of "Mau Mau"*** (1966), is a survey of the rise of nationalist politics in Kenya.

(Jo.R.)

Kepler, Johannes

The Renaissance astronomer and astrologer Johannes Kepler is best known for his discovery of the three principles of planetary motion, by which he clarified the spatial organization of the solar system. Moreover, he founded modern optics by presenting the earliest correct explanation of how human beings see. He was the first to set forth accurately what happens to light after it enters a telescope, and he designed a particular form of that instrument. His ideas provided a transition from the ancient geometrical description of the heavens to modern dynamical astronomy, into which he introduced the concept of physical force.

Early life. On December 27, 1571, in the German town of Weil der Stadt, then a "free city" within the Holy Roman Empire, Johannes was born prematurely, the offspring of an unhappy marriage. His father was a ne'er-do-well mercenary soldier, his mother the quarrelsome daughter of an innkeeper. Small in stature, Johannes never enjoyed robust health, but his superior intelligence was recognized even when he was a young child. Coming from a poor family, he would have received no education had not the dukes of Wiirttemberg adopted the enlightened policy of providing generous scholarships for the bright sons of their impoverished subjects.

With such help Kepler in 1587 was able to attend the University of Tiibingen, where he had the good fortune to study astronomy under Michael Mästlin, a professor who may have been unique in his day, for he was convinced that the astronomical system propounded by Nicolaus Copernicus was basically true: the Earth is a planet that rotates daily around its own axis and revolves annually around the Sun. Kepler's youthful acceptance of Copernican astronomy profoundly affected the subsequent course of his life.

Kepler, engraving by an unknown artist, c. 1730, after a contemporary painting.
Archiv fur Kunst und Geschichte

Major achievements. After obtaining the B.A. in 1588 and the M.A. in 1591, Kepler planned to become a Lutheran minister. But in 1594, during his last year of training in theology at Tubingen, the teacher of mathematics in the Lutheran high school of Graz, in Austria, having died, Kepler was strongly recommended by the Tiibingen faculty to fill the vacancy. Kepler did not finish the theology course at Tiibingen but went to Graz the same year. On a summer day in 1595, while he was teaching a class, a spectacular idea flashed through his mind. Ancient Greek geometry had proved that there were five regular solids, or "Platonic bodies": tetrahedron (pyramid), cube, octahedron (formed by eight equilateral triangles), dodecahedron (12 pentagons), and icosahedron (20 equilateral triangles). The ancients knew that these five solids could be enclosed in a sphere, and that there can be no additional regular solids. Sustained by a vision of mathematical harmonies in the skies, a vision he derived from the philosophy of Plato and the mathematics of the Pythagoreans, Kepler tried to relate planetary orbits with geometrical figures.

The notion of cosmic harmony

According to Copernican astronomy there were six planets, whose orbits were regulated by the turning of invisible spheres. But why were there only six planets and not nine or 100? Was the cosmos so constructed that one of the five regular solids intervened between each pair of the unseen spheres that carried the six Copernican planets? This nest of alternating planets and regular solids constituted the main theme in Kepler's *Prodromus Dissertationurn Mathematicarum Continens Mysterium Cosmographicum* ("Cosmographic Mystery"), which he published in 1596 under the auspices of the Tübingen faculty. The Platonic and Pythagorean components in Kepler's conception of celestial harmony, however mystical in origin, helped to lead him to the three principles of planetary motion now known by his name.

Kepler sent copies of his first major work to a number of scientists, including Tycho Brahe, who was soon to become the imperial mathematician of the Holy Roman Empire. Although Brahe did not agree with the underlying Copernican foundation of Kepler's *Mysterium Cosmographicum,* he was so impressed by the author's knowledge of astronomy and skill in mathematics that in 1600 he invited him to join his research staff in the observatory at Benatek (now Benátky nad Jazerou), outside Prague. When Brahe died the next year, Kepler was promptly appointed his successor as imperial mathematician. His first publication at Prague, *De Fundamentis Astrologiae Certioribus* (1601; "The More Reliable Bases of Astrology"), rejected the superstitious view that the stars guide the lives of human beings. Nonetheless, his deep feeling for the harmony of the universe included a belief in the harmony between the universe and the individual, and his skill in astrological prediction was much in demand.

While Kepler was watching a rare conjunction of Mars, Jupiter, and Saturn in October 1604, a supernova appeared that remained visible for 17 months. This event was evidence that the realm of the fixed stars, considered since ancient times as pure and changeless, could indeed experience change. He published the results of his observations in 1606 as *De Stella Nova in Pede Serpentarii* ("The New Star in the Foot of the Serpent Bearer").

Contributions to optics

Kepler now had access to Brahe's incomparable collection of astronomical observations, the results of decades of unremitting and painstaking toil by the greatest naked-eye observer of the heavens and the leader of a highly qualified team of astronomers. As a member of the team, Kepler had been assigned to investigate the planet Mars. But before he could use the raw observations, Kepler felt that he had to solve the problem of atmospheric refraction: how is a ray of light, coming from a distant heavenly body located in the less dense regions of outer space, deflected when it enters the denser atmosphere surrounding the Earth?

Kepler incorporated his results in a book that he modestly entitled *Ad Vitellionem Paralipomena, Quibus Astronomiae Pars Optica Tradirur* (1604; "Supplement to Witelo, Expounding the Optical Part of Astronomy"); Witelo (Latin Vitellio) had written the most important medieval treatise on optics. But Kepler did much more than add to Witelo's work. He made an analysis of the process of vision that provided the foundation for all of the advances in the understanding of the structure and function of the human eye. Kepler wrote that every point on a luminous body in the field of vision emits rays of light in all directions, but that only those rays can enter the eye that impinge on the pupil, which functions as a diaphragm. He stated that the rays emanating from a single luminous point form a cone, the circular base of which is in the pupil. All of the rays are then refracted within the normal eye to meet again at a single point on the retina, identified by Kepler as the sensitive receptor of the eye. If the eye is not normal, the second short interior cone comes to a point not on the retina but in front of it or behind it, causing blurred vision. For over three centuries eyeglasses had helped older persons to see better. But nobody before Kepler was able to explain how these little pieces of curved glass had worked.

After the invention of the telescope had been reported to Galileo, who promptly proceeded to make his astounding discoveries, Kepler applied the same ideas concerning optics to the explanation of how the telescope works. Although Galileo's findings were received in general with skepticism and ridicule, Kepler acknowledged the Italian's accomplishments in his *Dissertatio cum Nuncio Sidereo Nuper ad Mortales Misso a Galilaeo Galilaeo* in 1610.

Kepler's three principles of planetary motion

Galileo did not return the compliment. He chose to ignore the epoch-making results Kepler had published the preceding year. In his *Astronomia Nova* ("New Astronomy") of 1609, Kepler had demonstrated that the orbit of the planet Mars is an ellipse. Although it had been believed since antiquity that the planets, being heavenly bodies, were perfect and therefore could move only in perfect circles or combinations of circles, Copernicus had correctly classified the Earth as one of the planets; and it was fully accepted that the Earth was far removed from perfection. Kepler extended Copernicus' reasoning to the other planets and was the first to declare that the other planets resemble the Earth in being material bodies. That a material body, being imperfect, need not travel in a perfectly circular orbit was a conclusion made by Kepler after he tried unsuccessfully to fit the orbit of Mars to Brahe's observations in every possible combination of circles his ingenuity could devise. Because none of them worked, he tried noncircular paths until he found the true solution: Mars revolves in an elliptical orbit with the Sun occupying one of its two focuses.

The pre-Keplerian dogma that permitted only circular paths entailed the concept of uniform motion—*i.e.*, the moving body or point must traverse equal arcs in equal intervals of time. Such a conception of uniform motion as measured along an arc was, of course, incompatible with an elliptical orbit. But Kepler found an alternative form of uniformity. This new uniformity equated equal areas with equal times. With the Sun remaining stationary in one focus of the ellipse, the planet, while revolving along the periphery of its elliptical orbit, would sweep out, in equal intervals of time, equal areas of the ellipse, not equal arcs along the periphery of the ellipse.

In 1619, ten years after Kepler published these first two principles of planetary motion (the elliptical orbit and equality of areas), he published the *Harmonices Mundi* (*Harmonies of the World),* in which he expounded his third principle, which related a planet's mean distance from the Sun to the time it takes to complete its elliptical orbit around the Sun. The cube of the distance proved to be in a constant ratio to the square of the time required for all the planets to complete such an orbit. The enunciation of this rule (sometimes called the 3/2 ratio) completed Kepler's contribution to the understanding of planetary motion and helped to prepare the way for Sir Isaac Newton's idea of universal gravitation, affecting all of the material bodies in the physical universe.

Later life. Meanwhile, Kepler's patron, the Holy Roman Emperor, had been compelled by his brother to abdicate, and Kepler himself had found it desirable to leave Prague, then the capital of the empire. Although he was reappointed imperial mathematician by the new emperor, Kepler moved to Linz, in Austria. His first wife had died in Prague; Kepler remarried in 1613. Once, when buying supplies for his new home, Kepler became unhappy about the rough-and-ready methods used by the merchants to estimate the liquid contents of a wine barrel. Because the curved containers they used were of various shapes, Kepler sought a mathematical method for determining their volumes. Following the model established by Archimedes, the most talented mathematician of antiquity, Kepler, in his volumetric researches, investigated the properties of nearly 100 solids of revolution—made by rotating a twodimensional surface on one of its axes—that had not been considered by Archimedes. Starting with an ordinary wine barrel, Kepler enormously extended the range of Archimedes' results. He did so by refusing to confine himself, as Archimedes had done, to cases in which a surface is generated by a conic section—a curve formed by the intersection of a plane and a cone—rotating about its principal axis. Kepler's additional solids are generated by rotation about lines in the plane of the conic section other than its principal axis.

While he was in Linz, Kepler published his *Epitome Astronomiae Copernicanae* (1618–21; *Epitome of Copernican Astronomy*). He modelled this title after the highly successful introduction to astronomy published by his former Tiibingen professor in a number of editions. But, whereas Mastlin had deemed it prudent pedagogical practice to keep Copernicanism out of an elementary textbook, which he therefore entitled simply "Epitome of Astronomy," Kepler emphasized his open espousal of the new cosmology by inserting the provocative label "Copernican."

Witchcraft trial of his mother

In Linz in 1620, Kepler heard that his mother had been indicted on the charge of being a witch. Such a defendant was often subjected to torture and, if convicted, was usually burned at the stake. If his mother had suffered this fate, Kepler's own status as imperial mathematician of the Holy Roman Empire and mathematician of Upper Austria might have been irreparably impaired. He rushed to her defense, therefore, not only out of filial devotion but also out of prudent self-interest. Only his skillful intervention saved her from torture and a fiery death.

Kepler had planned to publish his *Tabulae Rudolphinae* (*Rudolphine Tables),* named in honour of his first imperial patron, Rudolph II, in Linz. But this work, the final outcome of long years of unceasing reflection and tireless calculation, could not be printed there because of a rebellion by the peasants, who were infuriated by a combination of being forced to return to Catholicism and to pay heavy taxes. Kepler had to find another home and a new patron.

Albrecht von Wallenstein, duke of Friedland and Żagań, a successful soldier of fortune who had put his private army at the disposal of the empire in the Thirty Years' War, accepted the responsibility of satisfying Kepler's financial needs. The astronomer moved to Żagań in Silesia, where he was able to establish his own printing press. The *Rudolphine Tables* were printed at Ulne, Germany, in 1627, before Kepler went to Żagań in 1628. But Wallenstein turned out to be someone on whom Kepler could not rely.

Leaving his family behind in Żagań, Kepler went west to collect the interest due on two promissory notes he held in exchange for money he had deposited in Austria. On his way he stopped at Regensburg, where the Imperial Diet was in session. He fell acutely ill and died on November 15, 1630. The tremendous upheavals suffered by Germany in the Thirty Years' War later obliterated his grave.

BIBLIOGRAPHY. MAX CASPAR, *Bibliographia Kepleriana,* 2nd ed. by MARTHA LIST (1968), a complete enumeration of the publications by and about Kepler; *Johannes Kepler,* 3rd ed. (1958; Eng. trans., 1959), the best biography; JOHANNES KEPLER, *Gesammelte Werke* (1937–), a magnificent edition of Kepler's works still in progress; *Joannis Kepleri astronomi opera omnia,* ed. by CHRISTIAN FRISCH, 8 vol. (1858–71), still a useful collection of Kepler's works; *Epitome of Copernican Astronomy,* bk. 4–5; and *Harmonies of the World,* bk. 5, trans. by CHARLES GLENN WALLIS in "Great Books of the Western World," vol. 16 (1952); ROBERT SMALL, *An Account of the Astronomical Discoveries of Kepler* (1804, reprinted 1963); EDWARD ROSEN (ed.), *Kepler's Conversation with Galileo's Sidereal Messenger* (1965), on Kepler's enthusiastic reaction to Galileo's spectacular discoveries with the recently invented telescope, and *Kepler's Somnium* (1967), on Kepler's imaginary space flight to the moon. Specialized studies on Kepler by EDWARD ROSEN are: "In Defense of Kepler," in ARCHIBALD R. LEWIS (ed.), *Aspects of the Renaissance: A Symposium,* pp. 141–158 (1967); "Galileo and Kepler: Their First Two Contacts," *Isis,* 57:262–264 (1966); "Kepler's *Harmonics* and His Concept of Inertia," *Am. J. Phys.,* 34:610–613 (1966); and "Kepler and Witchcraft Trials," *Historian,* 28:447–450 (1966).

(Ed.R.)

Kerala

Kerala, one of the smallest states in the Republic of India, was formed in 1956. It has an area of 15,007 square miles (38,869 square kilometres) and a population of more than 21,000,000 (1971). With only about one percent of the area of India, Kerala has to support almost four percent of its population.

The state stretches for about 360 miles along the Malabar Coast on the western side of the Indian peninsula; its width varies from 20 to 75 miles. It is bordered by Mysore state to the north, Tamil Nadu (Madras) state to the east, and the Arabian Sea to the west. The capital is Trivandrum. Isolated from the Deccan Plateau by the mountainous belt of the Western Ghāts, but with a long coastline that opens it to foreign influences, Kerala has evolved a unique culture. It is a highly politicized region, but it has a long tradition of religious amity. It is an educationally advanced state with its own language, Malayalam, and has the highest rate of literacy among India's states. Women in Kerala enjoy a high social status, thanks perhaps to its historic matrilineal system. But the hills of Kerala are inhabited by some of India's least advanced tribes. (For a related physical feature see ARABIAN SEA.)

History. The first known mention of Kerala occurs on one of the rock inscriptions left by Aśoka, emperor of India during the third century BC. It was then an independent kingdom known as Keralaputra, which was ruled by the powerful Cēra dynasty until the fifth century AD. In ancient times it traded with Egypt, Babylonia, Greece, Rome, and China. This was the period when Aryans migrated into Kerala and introduced Hinduism, Buddhism, and Jainism into this Dravidian kingdom. It is believed that St. Thomas the Apostle founded the Syrian Church—and Jewish immigrants established a colony—at Muziris (Cranganore) in the first century AD.

The Malabar era

The history of Kerala from the sixth to eighth centuries AD is obscure. It was probably during this period that Arab traders introduced Islām. From AD 825—the year in which the Kollam era, regarded as still in progress, began—Kerala rose to prominence under the Kulasekhara dynasty. The ninth and tenth centuries formed a golden age marked by a remarkable development in philosophy, literature, education, and the fine arts. This period marks the origin of Malayalam as a separate language.

During the 11th century a hundred years' war with the Cōḻas led to the disintegration of the Cēra kingdom into innumerable petty principalities. By the beginning of the 14th century, however, the ruler Ravi Varma Kulasekhara established a short-lived supremacy over the whole of southern India. His empire fell to pieces, however, soon after his untimely death in 1314, after which Kerala became a conglomeration of warring chieftaincies, among which the most important were Calicut in the north and Travancore in the south.

The year 1498—when Vasco da Gama landed at Calicut—marked the beginning of the era of foreign intervention. In the 16th century the Portuguese superseded the Arab traders and dominated the commerce of Malabar. Their attempt to establish sovereignty, however, was thwarted by the *zamorin* of Calicut. The Dutch ousted the Portuguese in the 17th century but met with a crushing defeat at the hands of King Martanda Varma of Travancore in 1741. The destructive invasions of Kerala between 1766 and 1790 by Hyder Ali and Tippu Sultan, both of Mysore, paved the way for the establishment of British power. Both Cochin and Travancore were obliged to accept British suzerainty in 1791 and 1795, respectively; the *zamorin*'s territories were annexed by the British and constituted into the Malabar district of the Madras presidency under their direct rule. Unsuccessful revolts against British rule were led by Pazhassi Raja of Malabar from 1800 to 1805 and by Velu Thampi of Travancore in 1809, both of whom were martyred as a result. Two years after India's independence was achieved in 1947, the two states were united as the Travancore-Cochin state. In 1956 the present Kerala was constituted on a linguistic basis, thus also including other Malayalam-speaking areas in the Malabar region.

The landscape. Kerala is a region of great natural beauty. From the majestic heights of the Western Ghāts the country undulates westward, presenting a vista of silent valleys clothed in the richest green. Among the many rivers that glide across the plains to merge their waters with the Arabian Sea, the more important are the Periyār, the Pamba, and the Bharatapuḻha. Along the coast, sand dunes shelter a linked chain of lagoons and backwaters the still waters of which are studded with sea gulls and country canoes plying at a snail's pace. The silence of their clear skies is broken only by the coos of koels, a type of cuckoo, and the frequent flutter of cranes perched on the embankments.

Climate. The climate is equable and varies little from season to season. The temperature normally ranges from 80° to 90" F (27" to 32° C) in the plains but drops to about 70° F (21" C) in the highlands. The state gets its due share of both the southwest and northeast monsoons, and the rainfall is heavy, averaging about 118 inches annually.

Land reclamation

Vegetation and animal life. The alluvial lowlands are covered with coconut groves and rice fields. A unique feature of agriculture in the Kuttanad area of Travancore is the annual reclamation of land from deep lakes and backwaters for rice cultivation. The midland region between the coast and the plateau is rich in a variety of crops such as rice, tapioca, and millet, and in palm, mango, jackfruit, and cashew trees. The highlands are rocky in the higher reaches, but at lower levels there are extensive plantations of pepper, rubber, coffee, tea, and cardamom. Primeval forests cover the mountainous belt and are inhabited by a variety of wildlife including elephants, tigers, panthers, bison, and cobras. The game sanctuary at Thekkady is a centre of tourist attraction.

Population. *Ethnic composition.* The racial distribution of the state's population is representative of the subcontinent's history. Several of the hill tribes exhibit Negroid features, while others appear to be related to the Veddas, the forest nomads of Sri Lanka (Ceylon). These racial types occur only in southern India and are probably remnants of the region's earliest peoples. The bulk of Kerala's inhabitants, however, are of the Dravidian stock. The Nairs and Ezhavas, who form the leading Hindu communities, belong to the Dravidian group of the early Cēra kingdom, while the Nambudiri Brahmins are descendants of the later Aryan migrants from the north. The Malayālis—Malayalam-speaking people—are a product of race admixture between the predominant Dravidian and the smaller Aryan groups.

Demography. Kerala is the most densely populated state in India. The average density of population in 1961 was 1,126 persons per square mile---over three times the Indian average. About 50 percent of the talukas (subdistricts), however, had a density of 1,000 to 3,000, and 18 percent were above 3,000. According to the 1971 census, the average density increased by 26 percent during the decade. The proportion of urban population in 1971 was 16 percent, but this low figure is deceptive as the villages in Kerala are semi-urban and form a continuous chain. The major industrial complexes are at Alwaye, Quilon, Alleppey, Calicut, and Cannanore. Other demographic features reflect the status of a burgeoning population. A high birthrate—which, thanks to the family-planning campaign, shows signs of declining—a high female sex ratio of 1,019 females to 1,000 males (as against the national ratio of 932), and a declining death rate have contributed to a decennial growth rate of more than 25 percent between 1961 and 1971. Despite the highest average marriage age of females in India, Kerala has a high fertility rate, which is concentrated in the age group of 20 to 39.

Religious harmony

Religions. Kerala has a unique record in India of harmonious coexistence of diverse religions. The majority of the Malayālis belong to the Hindu community, which has absorbed Buddhists and Jains into its fold. There is no conflict between the Dravidian *naga* (serpent-god) worship and that of Kālī (the mother goddess) on the one hand and the Hindu pantheon on the other, nor between the rival sects of Saivism (worship of the god Śiva) and Vaiṣṇavism (worship of the god Viṣṇu). The concept of unity in diversity is exemplified in the various Kerala temples by giving all these deities their allotted places. The Jewish community remains a small, exclusive sect. The Christians, forming about a fourth of the population, belong broadly to the Orthodox Syrian, Roman Catholic, and Protestant churches, but the denominations among them are no less numerous than the subcastes among Hindus. The Muslim minority have lived in habitual amity with the Hindus.

Administration and political life. The administration conforms to the uniform pattern adopted by states in the Indian federation, with a governor at the head, a single-chamber legislature, a cabinet executive presided over by a chief minister, and an independent judiciary. The legislative assembly has a membership of 134. The structure of the judiciary is hierarchical, with a high court at the apex from which appeals to the Supreme Court of India may be made. For provincial administration the state is divided into 10 districts subdivided for revenue purposes into 56 talukas and about 1,650 villages. Local self-governing institutions consist of three city corporations, 28 municipalities, and 952 *pañcāyat*s (ancient institutions revived as the lowest administrative units).

Kerala's modern political experience has been one of unstable government, with popular ministries rising and falling in quick succession. Between 1948 and 1970 as many as 12 ministries held office at Trivandrum and four midterm elections were conducted, one of which proved abortive. There were also three interludes of presidential rule, extending in all to about four years. In 1957 Kerala returned the local Communist Party to power. Though this government was overthrown two years later, the Communists again rose to power after the 1967 elections. The subsequent proliferation of political parties, however, added a new dimension to Kerala politics and made coalition governments inevitable.

Social conditions. *Customs.* The Aryan-Dravidian ethnic synthesis created a caste-ridden society characterized by peculiar social customs, manners, and institutions. The matrilineal joint family has only recently given way to innumerable nuclear families (each composed of father, mother, and children) with paternal inheritance. Women in Kerala enjoy an honoured place in society and are to be found in all walks of life. Both polygamy and polyandry were widely prevalent but today are illegal except for Muslim polygamy. The marriage ceremony of the Nairs, consisting merely of the presentation of cloth, is simple, but it is an accepted and respected bond of union.

A high level of education

Education and welfare. In education, Kerala is the most advanced state in India. The level of literacy, about 66 percent among males and about 54 percent among females, is substantially higher than the national average. Elementary education is free and compulsory between the ages of 6 and 11. There are over 11,000 primary, basic, and secondary schools, about 50 polytechnical and industrial training institutes, more than 100 arts and science colleges, about 30 professional colleges, and 3 universities. The state maintains a relatively high standard of health service, with one hospital bed for every 1,000 persons. There is a comprehensive health insurance plan for factory workers and free medical treatment in many hospitals and dispensaries.

Economic conditions. The economy is largely circumscribed by geographic and geologic constraints. The amount of arable land is far below the needs of the crowded population. The state is devoid of major deposits of minerals and fossil fuels, except for such minerals as ilmenite (the principal ore of titanium), rutile (titanium dioxide, used for coating welding rods), and monazite (a mineral composed of cerium and thorium phosphates), which are found in the beach sands. It has, however, enormous hydroelectric potential in its rapid rivers. Total installed hydroelectric capacity, which in the early 1950s was about 30 megawatts, had risen to almost 550 megawatts by the early 1970s. The educational system, a developed banking system, and excellent transport facilities provide conditions favouring further economic development.

Agriculture is the state's main economic activity. Cash crops, which account for 40 percent of the total land under cultivation, earn a sizable amount of foreign exchange but have also resulted in a shortage of food for local consumption. Though livestock is of poor quality, commercial poultry farming is the best developed in India, and eggs are exported. The forests yield industrial raw materials such as bamboo (used in the rayon industry), wood pulp, charcoal, gums, and resins, as well as ivory and rare timbers such as teak and rosewood. Kerala also accounts for 20 percent of the fish production in India; fishing trawlers have been bought, and the industry has been mechanized with Norwegian assistance.

The majority of the population is unaffected by industrialization. The traditional low-wage cottage industries, such as the processing of coconut fibre and cashews or weaving, employ about 60 percent of industrial workers. Most of those employed by larger industrial enterprises are engaged in food processing. The state participates in most major industrial concerns; products include fertilizers, chemicals, electrical equipment, titanium, aluminum cables, plywood, ceramics, and synthetic fabrics. More than one-fourth of the working force is employed in providing services. Unemployment, however, is acute, and the problem is accentuated by a high level of education among the jobless. The per capita income of Kerala in 1967 was only about 300 rupees a year, compared with the national average of about 320 rupees.

Transport and communications. In proportion to its area the state has the most developed road system and traffic flow in India; it is connected with the states of Tamil Nadu, Mysore, and Mahiiriishtra by the national highway and by a west coast road. The railway stretches for 551 miles from the northern border near Mangalore down to Trivandrum, connecting the principal towns. There are three major ports—Cochin, Alleppey, and Calicut—which handle both coastal and foreign traffic, and nine minor ports. More than 1,100 miles of inland waterways constitute the main arteries of commerce and carry the bulk of freight from and to the ports. Trivandrum has an international airport and Cochin a domestic one.

Cultural life. Kerala's cultural heritage contains elements of ancient Hindu culture that have been enriched by centuries of contact with both East and West. Literature and learning, in both Tamil and Sanskrit, flourished from the 2nd century AD onward. The Malayalam language, though an offshoot of Tamil, has absorbed the best in Sanskrit and has a prolific literature. Notable names in Malayalam poetry are Thunchath Ezhuthachan and Kunchan Nambiar, among classical poets, and Vallathol and Kumaran Asan in the present century.

Dance and drama

In recent times dramatic productions in the Kerala style that combines poetry, music, and dance have entered folk tradition. The themes are usually stories from the great Indian epics. The Kalamandalam, a recently established cultural institution, seeks to revive and promote traditional arts. There are a number of folk dances. A display of various visual arts takes place during Ōṇam, the national festival of Kerala. Ōṇam is also the occasion for an exciting snake-boat race at Aranmulai, in which hundreds of participants in each boat row rhythmically to the accompaniment of music and dancing.

In the fine arts, Swati Thirunal, the 19th-century royal composer, made signal contributions to classical Karnatak music of South India, while the artist Ravi Varma was renowned for his oil paintings. The splendid murals that adorn the walls of palaces, temples, and churches reveal a tradition and style that is distinct from classical Indian painting. Wood carving has attained a high level of excellence. The "Malabar gable"—a triangular projection at the top extremities of tiled or thatched roofs—is the characteristic feature of the Kerala style of architecture, which is noted for its beauty and utilitarian simplicity.

BIBLIOGRAPHY. A. SREEDHARA MENON, *A Survey of Kerala History* (1967), a succinct history based on recent research; L.A. KRISHNA IYER, *Social History of Kerala,* vol. 1, *Pre-Dravidinns* (1968), vol. 2, *Dravidians* (1970), an anthropological study; W. LOGAN, *Malabar,* 3 vol. (1887–91, reprinted 1951), a classic; GEORGE WOODCOCK, *Kerala: A Portrait of the Malabar Coast* (1967), an admirable monograph with a contemporary focus; NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH, *Techno-Economic Survey of Kerala* (1962); BUREAU OF ECONOMIC STUDIES, KERALA, *Kerala: An Economic Review* (annual); STELLA KRAMRISCH, *The Arts and Crafts of Kerala* (1970), fully illustrated.

(V.R.P.)

Kertanagara

Kertanagara, the last king of Tumapel (or Singhasiiri) in Java, is still venerated by Indonesians as one of their greatest rulers. The information on this 13th-century historical figure is sketchy and is primarily based on the two Javanese chronicles—the *Pararaton* ("Book of Kings") and *Nāgarakertāgama* (the epic of Majapahit), which give contradictory pictures of the King.

Pararaton says only that the King was a drunkard and fond of good food. He dismissed his able chief minister Raganatha (Kebo Arema) and appointed Aragani, who could serve him delicious food every day. Aragani is also known as Kebo Tengali, though some scholars say these were two separate men. He drank palm wine and held orgies, which eventually led to his death—he was killed by his enemies during one of his wild parties.

Nāgarakertāgama describes Kertanagara as a great and wise king. He was the son of respectable families—King Vishnuvardhana of Janggala and a princess of Kaḍiri—so that by birth he was a reuniter of the two moieties of the realm. Even his name Kertanagara, meaning Order

Stone sculpture thought to be Kertanagara in the form of *hari-hara ardhanari,* c. early 14th century. Acquired by the Museum fur Volker Kunde, Berlin, in 1865.

Consecration as king

in the Realm, might refer to reunion, which was the achievement of a great king according to the Javanese dualistic cosmology. Probably for this reason, he was consecrated a king in 1254 before his father's death. For the ceremony people from Janggala and Kadiri went to Kutaraja, the capital city of Tumapel, which was then renamed Singhasāri. From that time Tumapel was also called Singhasāri. Kertanagara was too young to rule the country; therefore, his father was still the de facto ruler. Kertanagara came to power only in 1268.

Nāgarakertāgama further describes Kertanagara as a zealous follower of Tantric Buddhism, which dealt with magic and demons and encouraged drinking and orgies of a ritual sort having nothing to do with pleasure-seeking. He believed that he was living in the Kali *yuga,* in Hindu cosmology the last of the four periods of world history, which was full of confusion, fear, and disaster, and that he, as a ruler, should save the world. According to at least one 20th-century scholar, however, "he knew the practical difficulties of protecting the world in Kali *yuga,* therefore he zealously upheld his religious duties and vows, in order to strengthen the Buddhist faith and to follow the example of former rulers with a view to securing the welfare of the world."

Many scholars tend to accept the *Nāgarakertāgama* because it fits in the historical context and is in accordance with the historical evidence.

Kertanagara's reign (1268–92) coincided with the expansion of Kublai Khan's realm in Southeast Asia. Kertanagara thought that he could stop an invasion only by aligning himself with or conquering neighbouring states and by strengthening himself with religious rites. He believed that Kublai Khan had already acquired a type of secret knowledge, which he himself could also gain. His chief minister Raganatha disagreed with Kertanagara's policy and, as a result, was removed from the post; Aragani, who was an ardent supporter of the policy, therefore came to power.

Kertanagara married a princess of Champa (southern Vietnam), a country that was dangerously exposed to Kublai Khan. Kertanagara sent envoys to Malayu (Sumatra) in 1275, and his statue as Amoghapāśa (Dhyāni Buddha Avalokiteśvara) was erected in the area (Jambi) in 1286, an indication that his influence had reached Sumatra. He also conquered Java and took Bali in 1284. A statue of a meditative Buddha, known as Jaka Dolog, discovered in Surabaja (eastern Java), was also erected by Kertanagara. According to its inscription, the statue was erected for the benefit of the King, the royal family, and the unity of the kingdom. Because the statue was erected on the hermitage of Bharada, the great wizard who had been asked to divide east Java into two kingdoms during the reign of Airlangga (1019–49), it must also have been aimed at neutralizing the evil effects of Bharada's division of his land.

Both the Jaka Dolog and the Amoghapāśa statues indicate that Kertanagara was very religious, as suggested by the chronicle *Nāgarakertāgama.* This does not mean, however, that he was unskilled in politics and government; in fact, he often used religion as an aid in gaining political ends.

Insult to Kublai Khan

When an envoy of Kublai Khan, Meng Ch'i (or Meng K'i), arrived in Java in 1289 to ask for tribute, Kertanagara refused to pay. This infuriated the Khan, who sent armed forces to punish Kertanagara. Before the troops arrived, however, Kertanagara was killed by the ruler of his vassal state, Jayakatwang of Kadiri, in the course of a Tantric ritual drinking bout.

Nāgarakertāgama indicates that toward the end of his life, Kertanagara became more involved in religious practices. It can be argued that he was too involved in his religion to provide for security of his throne and his country. Thus, although he thought of himself as being strong enough to resist Kublai Khan and hence refused to pay him any homage, he was not able to defend himself against an internal rival. Nonetheless, he was worshipped as a great king by many later Javanese leaders.

BIBLIOGRAPHY. D.G.E. HALL, A *History of Southeast Asia,* 3rd ed., pp. 70–79 (1968), a detailed and comprehensive history of Java; THEODORE G.T. PIGEAUD (ed.), *Java in the Fourteenth Century: A Study in Cultural History,* 5 vol. (1960–63), includes the complete translation of the Javanese chronicle *Nāgarakertāgama,* together with its original text, commentaries, and recapitulation, by a leading Dutch scholar in the field—a major source for the history of Singhasāri and Majapahit; BERTRAM SCHRIEKE, *Indonesian Sociological Studies,* pt. *2* (1957), a detailed and stimulating study of premodern Indonesian government and politics by a leading Dutch sociologist; B.H.M. VLEKKE, *Nusantara:* A *History of Indonesia,* rev. ed., pp. 61–65 (1965), a well-researched section on Kertanagara in a standard textbook on the history of Indonesia.

(L.Sur.)

Keyboard Instruments

The large and important group of keyboard instruments includes all musical instruments on which different notes may be sounded by pressing a series of push buttons or parallel levers. This group has assumed its great importance because the keyboard enables a performer to sound as many notes as he has fingers, either at once or in close succession. This unique capability enables him to play a reasonable approximation of any work in the entire literature of Western music, whether it involves complex harmonies or the interplay of independent contrapuntal parts. Moreover, the capabilities of keyboard instruments have influenced the composition of music for other media, because virtually every major composer from William Byrd (c. 1543–1623) to Béla Bartók (1881–1945) and beyond has been at least an accomplished keyboard performer, if not a renowned virtuoso.

In its broadest sense, the term keyboard instruments may be applied to any instrument with a keyboard and thus may refer to organs, accordions, percussion instruments such as the celesta and the carillon, and many electronic instruments—for example, the Moog synthesizer and the Ondes Martenot. In this article the term is applied in a narrower sense and is restricted to instruments in which sound is produced from strings, whether by plucking, striking, or rubbing. For other instruments operated by a keyboard, see ELECTRONIC MUSIC AND INSTRUMENTS; ORGAN; and articles on instrument families, such as PERCUSSION INSTRUMENTS.

DEVELOPMENT OF THE KEYBOARD

Evolution from early forms. Long before the appearance of the first stringed keyboard instruments in the 14th century, the keyboard was developed and applied to the organ. A keyboard of the kind familiar today—a series of parallel levers balanced or pivoted so that they can be pushed down by the fingers—first appeared on the hydraulis, an organ used by the Romans at least as early as the 1st century AD. This type of keyboard seems to have disappeared after the fall of Rome, and the organs of the early Middle Ages generally had sliders that were pulled out to sound different notes; some may have had keys that turned like the key for a lock. Keys of the last type were certainly used on the organistrum, a large medieval hurdy-gurdy operated by two players: one turned a crank rotating a wheel that rubbed against one or more strings to make them sound, while the other produced different notes by turning the key-shaped levers that stopped the strings at various points (much as guitar strings are stopped against the fingerboard).

Key arrangements and colours

Some small portable organs had push buttons instead of keys as late as the 1440s, but a keyboard very similar to the modern type existed in the 14th century, although the arrangement of naturals and sharps (corresponding to the white and black keys on the modern piano) was only gradually standardized. The arrangement of the keys depended in part on the music played and partly on the current state of musical theory. Thus, early keyboards are reported with only a single raised key in each octave (B♭), and at least one organ exists that has both B and B♭ as "natural" keys, with C♯, D♯, F♯, and G♯ as raised keys. The colours of the keys—white for naturals and black for sharps—became standardized much later, c. 1800, depending on fashion or on the relative cost of ivory and ebony. Flemish instruments had bone naturals and black sharps by 1580; English instruments generally had either brown boxwood naturals and black sharps or ebony naturals and ivory sharps until c. 1720; French and German instruments had ebony naturals and ivory sharps until the 1790s.

Special key arrangements. *The* short octave. Even after the present arrangement of five raised keys and seven natural keys per octave had become standard in the 15th century, two special exceptions existed. The first of these—the "short octave"--concerned only the lowest octave at the bass end of the keyboard. In the short octave, not all keys actually sounded notes of the expected pitch; their respective strings were tuned to lower notes. In the earlier form (Figure 1A), the keyboard apparently started on E, but the string for this key note was tuned to the C below. The apparent F♯ was tuned to D and the G♯ to E, so that the notes of the entire octave from C to c were encompassed within an apparent key span of only E to c. With this arrangement C♯, E♭, F♯, and G♯ were not available in the bass octave, but these notes were rarely required in the bass when this form of short octave first appeared (c. 1540). When the missing F♯ and G♯ became necessary in later music, short-octave keyboards were made with these keys divided into two parts, the fronts sounding D and E and the back parts sounding F♯ and G♯. Later still, a second short octave was developed (Figure 1B) in which the keyboard apparently began on low B. This key actually sounded the G below, and the apparent C♯ and E♭ were tuned to A and B (or B♭). Eventually, as musical styles changed, the two retuned sharps were divided in this arrangement as well, providing C♯ (or sometimes B♭) and E♭ at the back of these keys while retaining A and B at the front.

Divided sharps. The second type of exceptional keyboard arrangement was originally required by the so-called meantone tuning system generally used in the 16th–18th centuries. Meantone tuning provided significantly purer tuning for a relatively small number of tonalities than does the system of equal temperament now in use (in which all tonalities are somewhat out of tune; see TUNING AND TEMPERAMENT, MUSICAL), but only at the expense of making it virtually impossible to use any of the remaining tonalities. This characteristic arose because in meantone tuning each of the raised keys could be used in only one way: for example, if the key between D and E was tuned to E♭, it could not be used as D♯ without retuning. One solution was to build keyboards with the raised keys divided, the front half of the key sounding the appropriate sharp while the back half sounded the equivalent flat. The keys most commonly divided were A♭/G♯, E♭/D♯, and B♭/A♯. Instruments with up to three divided keys in each octave were commonly made in the 16th and 17th centuries, especially in Italy. Since the 1590s still more complicated keyboards have been built, permitting even more refined tunings; some in the 19th century had more than 50 keys per octave. Instruments have also been made with the octave divided into 24 rather than 12 equal parts to permit playing music utilizing quarter-tone intervals.

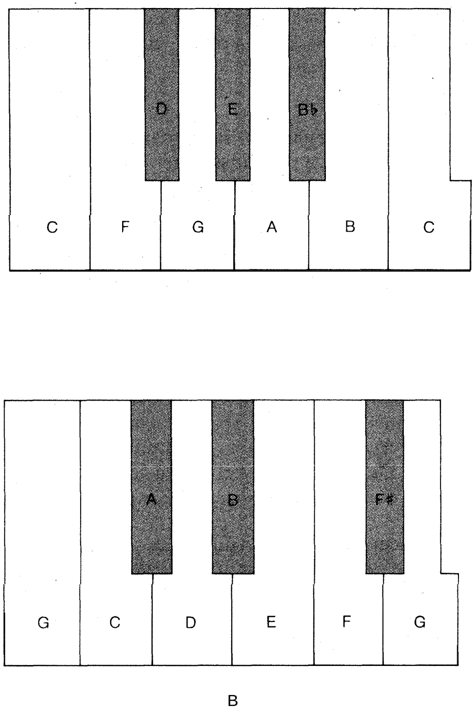


Figure 1: Special key arrangements.
Short octave (A) beginning on C and (B) beginning on G (see text).

Impact of tuning systems

Keyboard size and range. Although some early organs had very wide keys that could be played only with the fists, stringed keyboard instruments seem always to have had natural keys no more than an inch wide, yielding an octave span of seven inches. The octave span on the modern piano is about 6½ inches, much the same as on Flemish and Italian harpsichords of the 16th-18th centuries, whereas that of English keyboards was generally 6⅜ inches. On most French and German instruments of the 18th century, the octave span was even narrower (6¼ inches), permitting the playing of tenths—such as C to the second E above—by a hand of average size.

The range of the keyboard gradually expanded from a single octave for some early organs to 2% or 3 octaves in the 15th century and 4 or 4% octaves in the 16th century. By the early 18th century, except in Italy and Spain, a range of five octaves was standard: from the F below low C to the F above high C (F_1–f‴). This range began to be expanded only at the very end of the century, usually upward toward c'''' (C above high C) but occasionally downward to C_1 (C below low C). A few pianos with a range of six octaves (from C, to c'''') were built before 1800, and Beethoven's Hammerclavier Sonata, Opus 106 (completed, 1818), requires 6% octaves from C_1 to f''''. A seven-octave range was reached before

1830, and the modern piano keyboard consisting of 88 keys provides the only slightly greater range seven octaves and a third, from A, to c′′′′′.

THE CLAVICHORD

Some kind of stringed keyboard instrument is known to have existed by 1385, when an instrument called the *eschiquier* was described as "resembling the organ but sounding by means of strings." There exists no description of the eschiquier, however, and it is not known whether it was a variety of clavichord, in which the strings are struck by blades of metal that must remain in contact with them as long as they are to sound; a harpsichord, in which the strings are plucked; or a type of keyboard-equipped dulcimer, in which—as in the piano—the strings are struck by small hammers that immediately rebound from them. All three types of instruments were described and illustrated in the first half of the 15th century by Henri Arnaut of Zwolle, personal physician of Philip the Good, Duke of Burgundy, and later physician and astrologer to King Charles VII and King Louis XI of France.

Relation to the monochord

Despite the uncertainty regarding the eschiquier, it seems probable that the clavichord was the earliest stringed instrument having keys that could be pushed down by the fingers. Its principle of operation resembles that of the medieval organistrum, and it is apparently closely related to the monochord, an instrument consisting of a shallow closed box over which one or two strings were stretched. The monochord was in continuous use by theorists from ancient Greece onward as a tool for explaining and measuring musical intervals. Monochords were also used as performing instruments at least as early as the 13th century, and the kinship of the clavichord to the monochord was so close that, as late as the 16th century, clavichords were often called monocordia.

Shape and layout

Principle of operation. The clavichord is rectangular in shape, and its strings run from left to right across the keys, which are placed along one of the longer sides of the rectangle. The soundboard of the instrument is at the right-hand end of the case, and the vibrations of the strings are communicated to it by a bridge on which the strings rest. The soundboard amplifies the sound of the strings by permitting them to set a large mass of air into vibration rather than the very small mass of air that can actually be in contact with the string itself. (This is the same principle that makes a tuning fork sound louder when its stem is held in contact with a tabletop.)

The clavichord's principle of operation is extremely simple. A brass blade rather like the end of a screwdriver is driven into the top surface of each key near the back of the key (Figure 2); a smaller piece of wood, whalebone, or horn is driven into the back end of the key. (This piece fits into a fixed slot behind the key and prevents the key from moving from side to side as it moves up and down.) When the front end of the key is pushed down by the finger, the back end rises, and the brass blade, called a tangent, strikes the strings (which in most clavichords are arranged in pairs), causing them to vibrate. To the left of the tangent a strip of cloth is woven between the strings. When the key is struck, only the portion of the strings to the right of the tangent—*i.e.*, between the tangent and the bridge—sounds; the cloth prevents the string section to the left of the tangent from sounding. As soon as the key is released, the tangent falls away from the strings, which are then entirely silenced by the cloth. Because the sounding portion of each string is the segment between tangent and bridge, the tangent serves not only to set the strings in vibration but also to determine their sounding length. Thus, a series of tangents striking a given pair of strings at different points will produce a series of different notes, and all the earliest clavichords were designed to take advantage of this fact. Arnaut of Zwolle's clavichord used only nine or ten pairs of strings to produce all the 37 notes of its three-octave keyboard, and the clavichord represented in an Italian intarsia (picture in wood inlay) of about 1480 (Palazzo Ducale, Urbino) used only 17 pairs of strings to produce 47 notes in a four-octave range.

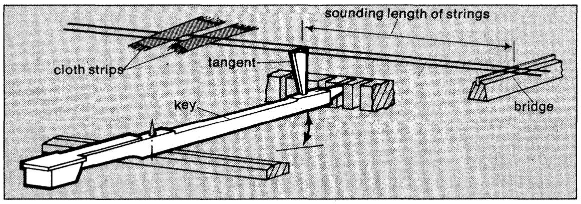

From N. Bessaraboff, Ancient European *Musical* Instruments (1941); Harvard University Press

Figure 2: Clavichord action (see text).

Making a single pair of strings serve several keys had two important disadvantages. Because each pair of strings can sound only one note at a time, it is impossible to play any two notes sounded from the same strings simultaneously, making it impossible to play certain chords. Furthermore, an unpleasant clanking sound is likely to result if the performer attempts legato playing of successive notes sounded from the same strings, making it necessary to play in a semi-detached fashion.

As early as the time of Arnaut of Zwelle, the first of these disadvantages was minimized by allowing no more than four keys to sound from the same pair of strings and by carefully choosing the points at which such groups of four keys were placed, so that only dissonant chords would be unplayable. The second problem could be solved only when a maximum of two keys were served by the same strings, so that each natural key shared its strings only with the sharp or flat next to it. G, for example, was paired with G♯, and in the normal music of the period the two notes were never needed at the same time or in immediate succession. Of course, if one wanted to use the G♯ key as an A♭, the problem would reassert itself; but, as long as meantone tuning was in use, the G♯ could not serve as A♭ in any case.

Fretted and unfretted clavichords

Eventually, however, it was felt necessary to be able to play in all tonalities without restrictions either of style of playing or in the use of dissonant chords, and clavichords began to be built with one pair of strings for each key. Such clavichords are called "unfretted," in contrast to those having several keys for each pair of strings, which are called "fretted." Although the unfretted clavichord was apparently known before 1700, fretted clavichords were being made well into the 1780s; they had fewer strings to go out of tune, and the smaller number of strings permitted all the keys to be shorter and more equal in length, giving the instrument a superior touch. In addition, the smaller number of strings imposed a smaller downward force on the soundboard, resulting in a brighter, clearer tone.

Tone quality. The greatest disadvantage of the clavichord is its extremely soft tone. Because it arises directly from the way in which the sound of the instrument is produced, this disadvantage cannot readily be overcome. It is impossible to impart very much energy to a string by striking it at one end (it is for this reason that a guitarist makes less sound when he strikes the strings against the fingerboard with his left hand than when he plucks them with his right, even though the pitches produced are the same). In compensation, the clavichordist alone of all keyboard-instrument players has control over a note once it has been struck. As long as a note is sounding, he has contact with the string through the tangent and key, and by changing his pressure on the key he can vary the pitch of the note, produce a controlled vibrato, or even create the illusion of prolonging or swelling the tone. Although the maximum loudness of which a clavichord is capable is not great, its softest pianissimo is very soft indeed, and the clavichordist has complete control over an enormous number of gradations in loudness between these two extremes. As a result of this touch sensitivity, the clavichord was highly valued as a teaching and practice instrument. In addition, its relative cheapness made it the normal domestic keyboard instrument in Germany and Scandinavia.

The soft tone of the clavichord made it impossible to use the instrument for any kind of ensemble music, except for providing a very discreet accompaniment for a single singer. Although much of the solo keyboard music of the 16th–18th centuries can be played on the clavichord, it cannot be stated that any of it before the latter part of the 18th century was especially composed with the clavichord in mind. At that time, however, the clavichord experienced a great revival in Germany, and music composed with its singing tone and unique capabilities of dynamic shading and vibrato was written for it by such masters as Carl Philip Emanuel Bach (1714–1788).

Clavichords in the 19th century

Clavichords continued to be made in Germany and Scandinavia well into the 19th century, long after the piano was both well known and popular. Indeed, many instrument makers built both clavichords and pianos (and harpsichords as well). The continued demand for the older instruments may have been a consequence—among other things—of musicians' recognition of the three instruments' differing capabilities.

The clavichord owes its modest modern revival largely to the efforts of Arnold Dolmetsch in England, who began building clavichords and performing on them in public in the 1890s. Both his style of playing the clavichord and the design of his instruments have remained a strong influence today. Only German makers, however, are presently producing clavichords in any great numbers.

THE HARPSICHORD

Principle of operation. ***Plucking mechanism.*** The sound of the wing-shaped harpsichord and its smaller rectangular, triangular, or polygonal relatives, the spinet and virginal, is produced by plucking their strings. The mechanism that accomplishes this is called a jack (Figure 3). It rests on the key and consists of a narrow slip of wood with two slots cut into it. The larger slot holds a pivoted tongue from which protrudes the quill, plastic, or leather plectrum that does the actual plucking; the smaller slot holds a piece of cloth that rests on the string and silences it when the key is not depressed. When the harpsichordist pushes down on a key, the back end rises, lifting the jack and forcing the plectrum past the string, plucking it. When he releases the key, the jack falls, and when the plectrum reaches the string on the way down, it forces the pivoted tongue backward so that the plectrum can get past the string again without plucking it. Once the plectrum has passed the string, a light spring made of bristle, plastic, or metal pushes the tongue forward again. Finally, when the key is completely at rest, the cloth damper touches the string, silencing it. A wooden bar covered with felt on its underside is placed over the jacks. The purpose of the wooden bar is to prevent the jacks from flying out of the instrument and to limit the depth to which the keys can be depressed.

From N. Bessaraboff, *Ancient European Musical Instruments* (1941); Harvard University Press

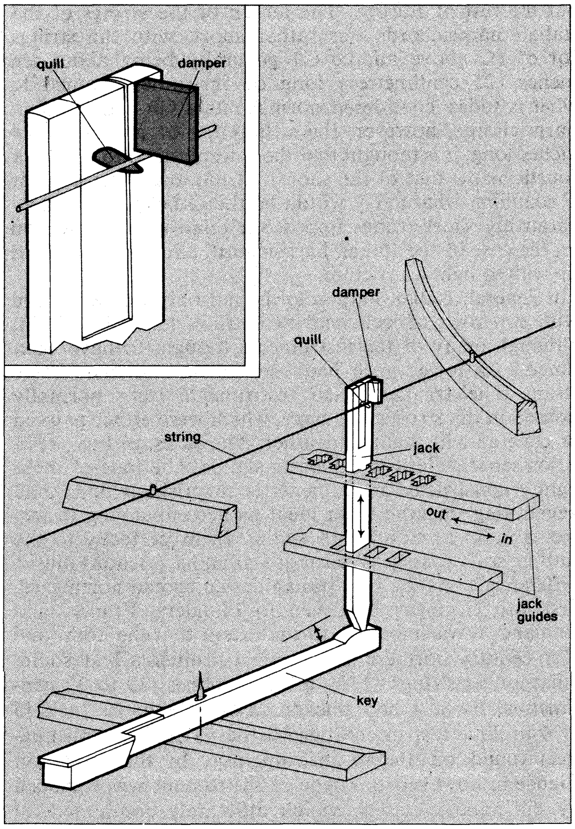


Figure 3: Harpsichord action (see text).

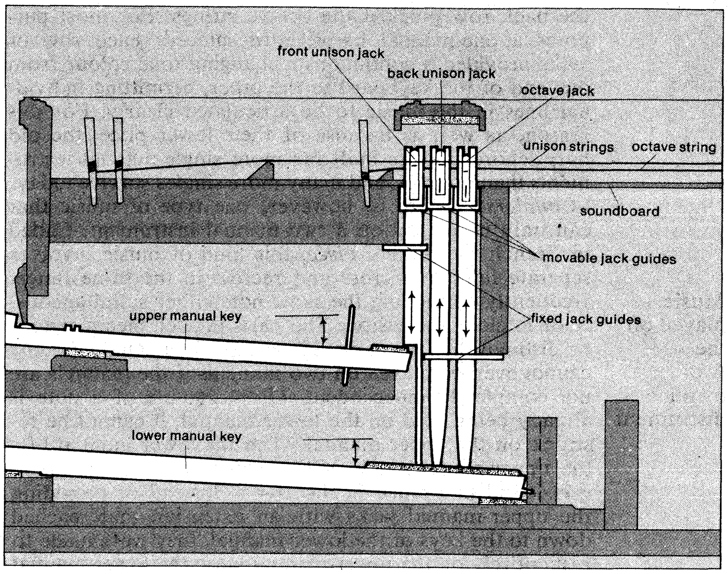


Figure 4: Flemish two-manual harpsichord action (see text).

Variations in loudness and timbre

Although minor variations in loudness and timbre, or tone colour, can be obtained by differences in the firmness with which the harpsichordist depresses the kevs. no great differences in loudness and no sustained crescendos are obtainable by the action of the fingers alone. For this reason, most harpsichords made since about 1550 have had at least two strings and two jacks for each key. Each can be engaged or disengaged at will by a slight shift of the uppermost of two slotted guides through which the jacks pass. Moving the guide in one direction brings the entire row of jacks close enough to the strings for the quills to pluck them; moving the guide in the opposite direction takes the jacks far enough from the strings so that the quills cannot reach them. Two rows of jacks can provide three different levels of loudness or three differing tone colours, depending on whether the performer uses each row separately or both together.

Two-manual instruments. Even given two rows of jacks, it would not ordinarily be possible to produce the rapid changes in loudness required for pieces in echo style, for example, or to play loudly with one hand while providing a soft accompaniment with the other. To accomplish this, it is necesssary to have two keyboards or "manuals," one of which operates a single row of jacks while the other operates two or more. It then becomes possible to play loudly on one keyboard and softly on the other, [illegible]r simultaneously or in rapid alternation. Two [illegible]l [illegible]si[illegible]d of th[illegible] kind were [illegible]ted at some point before 1620 in Flanders and gradually became known throughout the rest of Europe during the 17th century. These instruments had three sets of strings, two unison strings at normal pitch (called 8' pitch because the low C at this pitch is produced by an organ pipe eight feet long) and a third set of shorter strings tuned an octave higher, or at 4' pitch; this shorter set passed over its own bridge and was fastened to pins driven through the soundboard into a rail fixed to its underside. There were three rows of jacks (Figure 4). The front row plucked one set of unison strings and was made in such a way that it would be moved by the keys of both the upper and the lower keyboards. Both the middle and back rows operated from the lower manual only; the second row plucked the second set of unison strings, and the back row plucked the octave strings. For most purposes a one-manual harpsichord sufficed: each row of jacks provided a continuously changing tone colour from one end of the keyboard to the other, permitting individual lines in the music to be articulated clearly. For this reason, as well as because of their lower price, the old harpsichord makers built far more single-manual instruments than doubles, and many more singles survive today.

Music played on the two-manual instrument

Couplers. There is, however, one type of music that can only be played on a two-manual instrument. Called in French the ***pilce croisée***, this kind of music involves separate lines that cross and recross in the same range, frequently employing the same note either simultaneously or in close succession. The parts in such pieces cannot be distinguished when played on a single manual and they cannot even be played on two manuals if the manuals are not completely independent. (For example, if a note is already being held on the lower manual, it cannot be restruck on the upper manual when the lower manual lifts the upper-manual jacks.) The solution to this problem was found in France in the 1640s. Instead of providing the upper-manual jacks with an extension that reached down to the keys of the lower manual, they were made to rest entirely on the upper-manual keys; the lower-manual keys were then fitted with small upright pieces-of wood called coupler dogs, which reached upward toward the underside of the upper-manual keys. The upper manual was constructed to slide forward and back by about ¼ inch. When it was pushed into the instrument, the coupler dogs were positioned below the back ends of upper-manual keys. As a result, when any lower-manual key was pushed down and its back end rose, the coupler dog would push up on the underside of the corresponding upper-manual key, lifting its jack as well. When one wished to uncouple the two keyboards in order to play ***pilces croisées***, one could do so by pulling the upper manual outward. The coupler dogs then passed slightly beyond the ends of the upper-manual keys, so that they were not lifted when the lower-manual keys were depressed.

Two-manual harpsichords of this kind permit one to exploit the difference in the tone colours produced by the two rows, or "registers," of unison jacks. This difference depends on the distance along the string at which it is plucked. The closer the plucking point is to the end of a string, the brighter and more nasal the sound; the farther away from the end that a string is plucked, the fuller and rounder the tone becomes, until one approaches the centre; plucking near the centre of a string produces a sweet, flutey, but somewhat hollow sound. In order to emphasize the difference in tone colours produced by the two rows of unison jacks French harpsichord builders put the row of octave jacks between them, thereby increasing the difference between the two unison plucking points and the difference in tone of the two unison registers.

Special effects. A set of jacks plucking very close to the end of the string yields a very brassy, nasal sound. This type of register, called a lute stop, was first used in Germany in the 16th century and later spread to Flanders and to England, where it was added to the normal three registers on two-manual instruments. It did not have its own set of strings but, rather, plucked those of one of the existing unison registers. In England the lute stop plucked the same set of strings as the set of jacks operated by both keyboards; but, because the lute-stop jacks rested only on the upper-manual keys, they could also be used to provide a completely independent register on the upper manual. It was thus possible to play *pièces croisées* by taking off the unison register controlled by both manuals, using the lute stop for the upper manual and leaving the lower manual with its own unison register. Many harpsichords of all countries were also equipped with a buff stop (sometimes also called a lute stop), a device that presses pieces of soft leather against one of the sets of unison strings, producing a muted, pizzicato tone.

In Germany in the 18th century, harpsichords were made with still more strings and jacks for each key. Some had three unison strings in addition to an octave string; some had two unisons, an octave, and a suboctave (or 16') register; and some even had a 2' register, with very short strings tuned two octaves above the unisons. Harpsichords with three keyboards were apparently built throughout the 17th and 18th centuries, although only one authentic three-manual harpsichord is known today.

Resources of the 18th-century harpsichord

It should be emphasized, however, that the harpsichord of the 16th–18th centuries normally had only one or two keyboards and only two or three sets of strings and jacks per note. In the 16th and early 17th centuries, one-manual instruments usually had only two registers (either two unisons or a unison and an octave) with or without a buff stop; in the second half of the 17th century a second unison register became common, increasing the number of jacks and strings to three per note. Two-manual instruments, likewise, had no more than three sets of strings (two unisons and an octave) and three sets of jacks throughout the 17th century. In the 18th century, a fourth row of jacks was sometimes added. In England and Flanders, this row was the close-plucking lute stop; in France, if a fourth row was added, it was placed behind the other three and equipped with plectra of soft buff leather that provided a gentle, flutey tone, which was highly prized in the rather decadent period of the harpsichord's decline. Until the last half of the 18th century, it was possible to change registers only by moving knobs at the side of the instrument or above the keyboards, which could be done only when one hand or the other was not playing. This fact and the surviving written evidence suggests that the harpsichordists of earlier times changed registers relatively infrequently, preventing monotony of sound by relying on variations of touch and the changes of texture and pitch level written into the music.

History. The harpsichord was described by Arnaut of Zwolle in the mid-15th century and was apparently known throughout Europe by the end of the century, although no 15th-century examples have survived. The harpsichords depicted in sculptures, paintings, and miniatures of the period all appear to be shorter and to have thicker cases than the earliest surviving 16th-century examples, all of which are Italian and are constructed of very thin cypress.

Italy. The thin-cased style of harpsichord construction appears to have been developed in Italy about 1500, and it rapidly influenced the design of harpsichords throughout the rest of Europe. The length of the strings of the Italian harpsichords was rather short, with the strings for c'' (C above middle C) generally being about ten inches (25 centimetres) long on instruments tuned to what is today considered normal pitch. On some Italian harpsichords, however, the strings for c'' are about 14 inches long; it is thought that these were tuned to a pitch a fourth below that of the shorter-strung ones, the key for C sounding what today would be the G below. The comparatively short strings imposed a relatively low tension on the case of the Italian harpsichord, allowing it to hold up with so light a structure.

In general, Italian harpischords had only one keyboard with two rows of jacks and two strings tuned in unison, although many of the instruments thought to have been tuned to a lower pitch had one unison and one octave string. The fragile Italian instruments were normally housed in thick softwood cases, which were either painted or covered with stamped leather. The cases, in turn, rested on separate legs or elaborate stands. The tone of these lightly constructed instruments is surprisingly loud and penetrating, making them ideal as accompanying instruments in an orchestra and suiting them perfectly to the rattling scale passages typical of Italian harpsichord music.

Flanders. As the new Italian design spread northward, first into Germany and then to Flanders, France, and England, it was modified to the extent that the 16th- and 17th-century northern European instruments had somewhat longer strings (11% to 12% inches [29 to 32 centimetres] for c'') and thicker cases (3⁄16 to ¼ inch [5 to 6 millimetres] in contrast to the ⅛ inch [3 millimetres] found on Italian instruments). In the 1560s in Flanders, however, this type of instrument was replaced by still another design, which ultimately dominated all northern European harpsichord making. These instruments had long strings (about 14 inches for c'' at normal pitch) and thick cases with substantial internal bracing to withstand the greater tension imposed by the greater string length. Because the longer strings made it unfeasible to double the string length for each octave below middle C, harpsichords of the newer Flemish design have less gracefully curved bentsides and wider tails than either Italian harpsichords or the intermediate instruments built elsewhere north of the Alps.

Harpsichord building of the Ruckers family

The name most often associated with Flemish harpsichord building is that of the Ruckers family, which for four generations (a period spanning the century from about 1580 to 1680) dominated Flemish harpsichord making and whose instruments were exported to all parts of Europe—one was even shipped as far as Peru. At first sight, Ruckers harpsichords appear to be rather crude compared to their Italian counterparts, and their thick softwood cases give the impression of being clumsily cobbled together on the inside. Nonetheless, the tone of unaltered or properly restored examples is extraordinarily good, and it is easy to see why Ruckers instruments were so highly prized that a lively business in making forgeries of rhem flourished in the 18th century.

In addition to a wide variety of virginals (discussed below), the Ruckers family made several different harpsichord models. The most popular was apparently six feet long, having a four-octave keyboard from *C* to c''', with a short octave in the bass; one unison and one octave register; and, occasionally, a buff stop on the unison. They were typically painted in imitation of marble on the outside and decorated on the inside with block-printed paper on which a Latin motto was painted. The soundboard was usually decorated with paintings of flowers, leaves, and birds. (This decoration should be contrasted with that of the Italian harpsichords, which was generally lavished on the outer case, leaving the instrument unadorned except for finely profiled top and bottom moldings.) Flemish harpsichords were set directly on any of various types of fairly massive stands, examples of which may be seen in the numerous Dutch paintings of musical groups of the period. Similar harpsichords were made in smaller sizes tuned a fifth or an octave above normal pitch (the key c' sounding either g' or c''). By the mid-17th century, some single-manual instruments had a range of 4½ octaves from F_1 or G_1 in the bass to c'''.

The Ruckers family appear to have been the first to make two-manual harpsichords. These were of two types: in one (which may have been the earlier type and was not built after *c.* 1650), both keyboards were served by a single set of unison and octave strings and were not meant to be played at the same time. Instead, the keyboards were so arranged that c''' on the upper keyboard was placed over f''' on the lower keyboard, which meant that playing a piece on the lower keyboard automatically transposed it to a pitch a fourth below that of the upper keyboard. Whether this arrangement was used to facilitate routine transpositions or whether it was intended to provide in a single instrument the same resources as those available from both an Italian instrument with a 10-inch c'' and one with a 14-inch c'' is still a subject for controversy. The second type of two-manual harpsichord built by the Ruckers family was basically the type one finds today, with keyboards aligned over one another and intended to provide contrasts in loudness. Because the only set of upper-manual jacks was also played directly from the lower manual, it was not possible to play *pièces croisées.*

France. During the 17th century, instruments of the Ruckers type gradually influenced those being built throughout northern Europe; and by the early 18th century France, England, and Germany all had developed their own national variations on the thick-cased Ruckers design, replacing the thinner cased and shorter strung instruments of their earlier native schools. The sound of a typical 18th-century French harpsichord is delicate and sweet compared to the rather more astringent sound of a Ruckers. Those examples by the Blanchet family and their heir Pascal Taskin (1723–1793) are noted for their extraordinarily high level of craftsmanship and the lightness and evenness of their touch. Eighteenth-century French harpsichords were almost always painted and rest on elaborate carved and gilded cabriole (curved-leg) stands. As with Flemish harpsichords, the French soundboards are decorated with painted flowers and birds, and the maker's mark appears in the form of a cast ornament in the sound hole. In the 1760s, Taskin added a fourth row of jacks with soft plectra of buff leather as a special solo stop and also devised a highly ingenious system of knee levers that permitted the harpsichordist to play crescendos and decrescendos and to change registers without taking his hands from the keyboard. By the time of these inventions, however, the great Baroque composers of harpsichold music, such as François Couperin, J.S. Bach, Jean-Philippe Rameau, and Domenico Scarlatti, were dead, and these devices have no relevance to the historically accurate performance of virtually any of the harpsichord music one is likely to hear today.

Blanchet and Taskin harpsichords

England. In England the making of harpsichords in the 18th century was dominated by two London families, the Kirkmans and the Shudis. Both families made instruments for several generations and eventually moved on from harpsichord building to piano building. Their harpsichords are very similar, and the two-manual instruments all have a close-plucking lute stop in addition to the usual two unisons and octave. They are invariably veneered in walnut or mahogany and rest on simple stands, usually with straight or tapered legs. The tone of a Kirkman or Shudi harpsichord is both more robust and more brilliant than that of a French or Flemish instrument, making it a superb instrument for filling in the harmonies in the orchestral music as well as for the performance of the solo harpsichord literature.

Germany. Two German schools appear to have existed in the 18th century. One in the southern part of the country has left very few surviving instruments, which is unfortunate because these are the kind probably played by J.S. Bach. As far as is known, the southern German instruments were fairly plain, veneered ones, having only three registers and a rather darker tone than either French or Flemish instruments. The second German school was centred in the city of Hamburg and is best represented by the work of the Hass family. The Hass instruments are among the most elaborate ever made, both in decoration and complexity. They are the only 18th-century harpsichords with 16' and 2' registers, and some have lute stops as well. Their tone does not, unfortunately, live up to the quality of their craftsmanship or the ingenuity of their design, seeming overly brilliant and too thick in all of the surviving examples that have been restored to playing condition.

Hass family instruments

Decline of the harpsichord. Although many of the finest surviving harpsichords date from after 1750, few composers of the first rank were writing for the instrument by that time. Furthermore, the emergence of a newer, lighter style of music and an increased interest in crescendo and decrescendo effects led to the addition of various new devices foreign to the essential nature of the instrument. These include the knee- and foot-operated contrivances for the rapid changing of registers or for producing crescendos and decrescendos. Such devices represent the harpsichord builders' response to the same musical needs that eventually caused the harpsichord's replacement by the piano; but they were created before the real rise in the piano's popularity and must not be thought of as attempts to stave off the competition of the newer instrument.

As with the clavichord, builders continued to make harpsichords side by side with pianos. In England, Shudi's son-in-law, John Broadwood (see below), continued to make harpsichords until after 1800 (although in decreasing quantity), producing at the same time an ever-increasing number of pianos. There is even a small but interesting group of compositions by English, German, and French composers calling for both instruments.

Modern revival. The harpsichord had all but vanished except as a curiosity or in very rare historical concerts when the modern revival began in the 1890s with the building of new harpsichords by the piano firms of Erard and Pleyel in Paris. Almost immediately, the full brunt of 19th-century piano technology was applied to the manufacture of the revived instruments, and they became increasingly massively strung and framed as time passed. Pedals for changing registers were included from the beginning, and Pleyel first added the 16' stop in 1911. The Pleyel's sound, as preserved in the recordings of the Polish virtuoso Wanda Landowska and her numerous pupils, typified the harpsichord for most music-lovers until the 1950s, and it is for a heavy, metal-framed instrument of this type, with pedals for changing registers and a 16' stop, that most of the 20th-century harpsichord music has been composed.

Pleyel instruments

In 1905 modern harpsichord building was begun in Germany, initially taking the new Pleyel and Érard instruments as inspiration. Subsequent German building has produced a highly characteristic instrument somewhat reminiscent of the harpsichords of the 18th-century Hamburg school in sound. Taking as their model an improperly restored instrument falsely said to have belonged to Bach, these instruments generally have the unhistorical stop arrangement of one 8' and the 4' on the upper manual, with the second 8' and a 16' on the lower manual.

Arnold Dolmetsch, who began the modern revival of the clavichord, also built harpsichords, working in Paris and Boston as well as in England. He deserves to be considered the "godfather" of not only the present British school of harpsichord making but also of the flourishing U.S. school, most of whose members are, however, building a very different and far more historically based instrument than any that Dolmetsch made after about 1910.

Harpsichords specifically intended for electronic amplification have been built since the 1960s. Although some were meant for the performance of Baroque concertos with modern orchestras and in large halls, such instruments have had their greatest success among the other amplified instruments used by rock groups.

The virginal, spinet, and clavicytherium. The virginal, spinet, and clavicytherium are all varieties of harpsichord that differ from it primarily in size, shape, and musical resources. Virginals and spinets usually have only a single set of strings and a single row of jacks. The clavicytherium is basically a harpsichord set upright so that its soundboard is vertical. Instruments of this form were made from the 15th through the 18th centuries. Generally, their mechanism must be fairly complicated because the jacks must move horizontally rather than vertically and cannot therefore return to their rest position solely by the action of gravity. As a consequence of this complexity of its mechanism, the touch of the clavicytherium tends to be heavy, although the instrument takes up less space, and it is claimed that the vertical soundboard projects the sound outward far more effectively than the horizontal soundboard of the conventional harpsichord.

The virginal and spinet are small varieties of harpsichord, but the precise usage of the terms differs. Some writers reserve the term virginal for rectangular instruments and call all small triangular or polygonal instruments spinets. Others apply the term virginal to all plucked-stringed keyboard instruments whose strings run more or less from left to right across the keys (a usage followed in this article), reserving the term spinet for instruments in which the strings run obliquely away from the player. The terminological question is complicated by the fact that the word virginal in 16th- and 17th-century England referred to all plucked stringed keyboard instruments, including harpsichords and spinets as well as those today termed virginals. The term *épinette* ("spinet") had a similarly broad usage in France.

Italian builders of the 16th and 17th centuries made virginals and spinets employing a thin-cased construction similar to that of their harpsichords, and, like Italian harpsichords, these smaller instruments were kept in stout outer cases. The typical Italian virginal was either rectangular or polygonal in shape, with its keyboard projecting from the front of the case, and many of the surviving examples are sumptuously decorated with inlay or intarsia. Most Italian spinets are constructed as an irregular quadrilateral, but in the 17th century a new form was developed that more closely resembles a small harpsichord in having a bentside at the player's right and a long straight back slanting away from him. The new form was copied throughout Europe and became the standard domestic keyboard instrument in England in the late 17th century.

In Flanders, early virginals were polygonal and resembled Italian ones except that their keyboards were inset rather than projecting. In the 1560s, at the same time as the thick-cased harpsichord is believed to have emerged, thick-cased rectangular virginals made their appearance. By the end of the 16th century, two distinct types existed. They can readily be distinguished by the position of their keyboards: off-centre either to the left or to the right in one of the long sides of the rectangular case. Virginals with the keyboard at the right were far more common. They produce a characteristic flutey tone because the placement of the keyboard causes the strings to be plucked near their centre for most of the instrument's range. In virginals with the keyboard at the left, the strings are plucked off-centre except in the extreme treble, and the tone changes gradually from reedy in the bass, through full in the middle register, to flutey in the treble, much as on a harpsichord. Flemish builders produced virginals of both types in several sizes, the smaller ones being tuned to higher than normal pitches. They also made "double" virginals, consisting of a large virginal at normal pitch and a smaller one tuned an octave higher, which could be stored in a recess next to the keyboard of the larger instrument. The two virginals could be coupled together by placing the smaller instrument on top of the larger one.

Types of Flemish virginals

The virginals made in England were of the left-keyboard type. Those made elsewhere in Europe (some having the keyboard centred) were also built with strings plucking off-centre.

THE PIANO

Principle of operation. Although the basic principles of the piano's mechanism are simple, the refinements required in developing the powerful yet sensitive modern piano make it also the most complex of all instruments except the organ. The strings of the piano are struck by a felt-covered hammer that must rebound from the strings instantaneously or it will dampen their vibrations in the very act of initiating them. The hammer must thus be thrown toward the strings and allowed to fly freely. For the pianist to retain maximum control of loudness and softness, the distance of the hammer's free flight must be as small as possible; but, if the distance is too small, the hammer will bounce back and forth between the strings and the part of the mechanism that pushed it, producing a stuttering sound whenever the keys are struck firmly. As a consequence, all truly simple piano mechanisms — those in which, say, a rigid rod at the back of the key simply pushes the hammer upward until the key is stopped by a rail and the hammer flies free — must be adjusted to provide a large distance for free flight and can therefore give the pianist only limited dynamic range and control.

History. *Invention.* Piano mechanisms as unsophisticated as that described above continued to be devised and built throughout the 18th century. Nevertheless, the first successful piano — made in Italy by Bartolomeo Cristofori — solved the problems inherent in such simple mechanisms, as well as nearly every other problem facing piano builders until well into the 19th century. Cristofori reportedly started experiments on a "harpsichord with hammers" in 1698. By 1700 one of these instruments, together with six of his harpsichords and spinets, was included in an inventory of instruments belonging to the Medici family in Florence. In 1711 the instrument was described in detail in the Venetian *Giornale* de' *letterati d'Italia* by Scipione Maffei, who called Cristofori's invention *gravicembalo col piano e forte* ("harpsichord with soft and loud") — whence the present names pianoforte and piano.

Cristofori's pianos

In the three surviving examples of Cristofori's pianos, which date from the 1720s, the mechanism, or "action," differs somewhat from that described and pictured by Maffei; however, rather than merely representing an earlier phase of Cristofori's work, Maffei's diagram may be in error. In the surviving instruments a pivoted piece of wood is set into the key (Figure 5). The pivoted piece (which in a modern piano would be called a jack and should not be confused with the jack in a harpsichord) lifts an intermediate lever when the key is depressed. The lever, in turn, pushes upward on the hammer shaft near the hammer's point of attachment to a rail fixed above the keys. When the key is pressed completely down, the jack tilts and disengages itself from the intermediate lever, which then falls back, permitting the hammer to fall most of the way back to its rest position, even while the key is still depressed. This feature, called an escapement, is the heart of Cristofori's invention; it makes possible a short free flight for the hammer, after which the hammer falls so far away from the string that it cannot rebound against it, even when the keys are struck firmly. Cristofori provided a check (a pad rising from the back of the key) to catch and hold the falling hammer. At the end of the key he included a separate slip of wood, resembling a harpsichord jack, to carry the dampers that silence the string when the key is at rest.

From R.E.M. Harding, *The* Piano-Forte (1933); published by Cambridge University Press

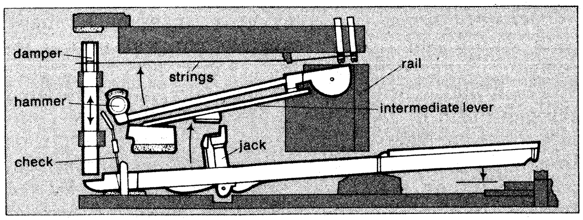


Figure 5: Cristofori's piano action (see text).

Utilizing an intermediate lever to act on the hammer near one end of its shaft provides an enormous velocity advantage, and the hammer flies upward toward the string much faster than the front end of the key descends under the pianist's fingers, adding to the crispness and sensitivity of Cristofori's action. The sound of his instruments is strongly reminiscent of the harpsichord, owing to the thinness of the strings and to the hardness of the hammers. The dynamic range is surprisingly wide, but it should be emphasized that the instrument's loudest sounds are softer than those of a firmly quilled Italian harpsichord and do not begin to approach the loudness of a modern piano.

Gottfried Silbermann's instruments

German and Austrian pianos. After Cristofori's death, piano making in Italy appears to have languished, but word of his invention became known in Germany through a translation of Maffei's account published in 1725. Before 1720 there had been independent attempts in France as well as in Germany to devise hammer mechanisms, although none was comparable to Cristofori's in sophistication or practicality. In the 1730s Gottfried Silbermann, of Freiberg in eastern Germany, a builder of organs, harpsichords, and clavichords, began constructing pianos patterned on Cristofori's. The surviving ones, probably from the 1740s, appear to have been directly copied from an instrument imported into Germany rather than derived from Maffei's description, but the ones he made earlier (and of which Bach is said to have disapproved in 1736) may have owed their failure to an attempt to follow Maffei's diagram exactly. By 1747 Silbermann had sold several of his pianos to King Frederick the Great of Prussia, and one of these is reported to have met with Bach's approval in 1747.

Subsequent German piano building did not follow the path charted by Silbermann. Instead, various German builders attempted to devise actions that were simpler than Cristofori's, generally adapting them to the clavichord-shaped instruments now called "square" pianos. In the most characteristic German actions, the hammers point toward, rather than away from, the player, and, instead of being hinged to a rail passing over all the keys, they are attached individually to their respective keys. As the front of the key is depressed, the back rises, carrying the hammer with it. A projecting point at the rear of the hammer shank catches on a fixed rail above the back of the keys, so that the hammers are flipped upward as the keys are stopped by a second rail set just above them. This action had no escapement, and (on the evidence of a letter of 1777 from Mozart to his father) many German instruments of the 1770s still lacked this highly important feature.

A pupil of Silbermann, Johann Andreas Stein of Augsburg in southern Germany, is generally credited with devising the first German action to include an escapement. As a replacement for the fixed rail that caught the projecting points at the rear of the hammer shanks, Stein provided an individually hinged and sprung piece for each key (Figure 6). As the back of the key reaches its highest point, this piece (the escapement) tilts backward on its hinge and releases the point at the back of the hammer shank. The hammer is then free to fall back to rest position even when the key is still depressed. This is often called the "Viennese" action because it was used by all the important 18th- and early-19th-century piano makers in Vienna, including Stein's daughter and son-in-law, Nannette and Johann Andreas Streicher; Anton Walter, Mozart's favourite piano builder; and Conrad Graf, maker of Beethoven's last piano. It was used in German-speaking countries until the mid-19th century, when it was replaced by mechanisms derived from a Cristofori-based action developed in England.

From R.E.M. Harding, The Piano-Forte (1933); published by Cambridge University Press

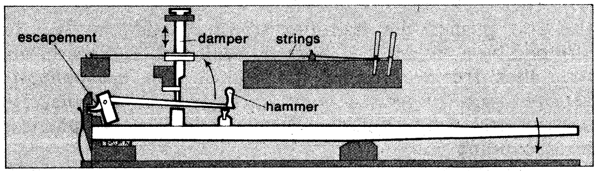


Figure 6: The German action (see text).

Although the tone of a piano by Stein or Walter is not loud, it is very sweet, with a singing treble and a clear tenor and bass that blend superbly with the sound of stringed instruments. The touch is extremely light and shallow: the force required to depress a key is only one-fourth that required on a modern piano, and the key need only be depressed half as far. In their sensitivity to the finest differences in touch and their singing tone, the Viennese pianos suggest the tone of a clavichord, although producing a louder sound.

Special effects

Austrian and German pianos of the late 18th and early 19th centuries often feature a great array of pedals. Only one of Cristofori's surviving pianos has any special effects: a lever on the underside of the instrument permits the player to shift the action sideways so that the hammers strike only one of the two strings provided for each note. By the time Silbermann built his pianos for Frederick the Great, a second special effect had been introduced —a mechanism to lift the dampers from the strings so that they could vibrate freely whether or not the keys were depressed. (These two effects, the sideways sliding of the action—to produce a softer sound and different tone colour—and the lifting of the dampers—to produce a louder, more sustained sound and another variation in tone colour—are the only ones found on all modern grand pianos.) Silbermann's pianos had hand levers for raising the treble and bass dampers separately and an additional hand lever for muting the strings. Stein's pianos normally had two knee levers for raising the treble and bass dampers and a third knee lever that interposed a strip of cloth between the hammers and the strings to produce a velvety pianissimo. Later instruments might have five or more pedals that, for example, pressed a roll of parchment against the bass strings to produce a buzzing sound or rang small bells and banged on the underside of the soundboard in imitation of the cymbals and drums of the then-fashionable "Turkish" music.

The English action. In the late 1750s a number of German piano builders emigrated to England, and one, Johann Christoph Zumpe, invented an extremely simple action for the square pianos he began building in the mid-1760s. Zumpe's action goes back to the Cristofori–Silbermann system in which the hammers point away from the player and are hinged to a rail over the keys. A metal rod tipped with a padded button is driven into the back of the key (Figure 7). When the key is depressed, the rod pushes the hammer upward; and the key is stopped by a padded rail over its back end, and the hammer flies freely. Despite the lack of an escapement, Zumpe's square pianos were an enormous commercial success and were copied in France, the Low Countries, and Scandinavia.

From R.E.M. Harding, *The Piano-Forte* (1933); published by Cambridge University Press

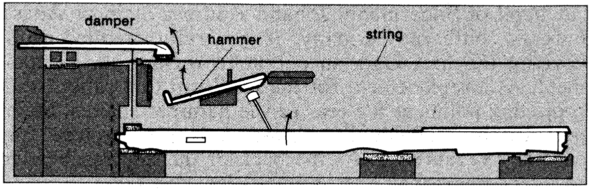


Figure 7: Zumpe's square-piano action (see text).

Zumpe had worked for the harpsichord builder Burkat Shudi when he first came to England, and around 1770 three other workmen in Shudi's shop, John Broadwood, Robert Stodart, and Americus Backers, devised for grand pianos an adaptation of Zumpe's action that included an escapement. This important development made London a major centre of piano building and created a characteristic English piano of fuller and louder sound than the Viennese piano but with a heavier, deeper touch and a consequent inability to play repeated notes as rapidly. In the English grand-piano action (Figure 8), the fixed rod of Zumpe's square-piano action was replaced by a pivoted jack, similar to that in Cristofori's action. The upper end of the jack fits into a notch at the base of the hammer shank, slipping out of the notch as the back of the key reaches its highest point; the hammer then flies free, strikes the string, and falls back to be caught by a hammer check even when the front of the key is still held down by the pianist. The tone of a typical 18th-century English grand piano is surprisingly reminiscent of the tone of an English harpsichord, suggesting that the English piano makers were, like Cristofori, seeking to make an expressive harpsichord, unlike the German builders who, in effect, appear to have been trying to build a louder clavichord.

From R.E.M. Harding. The *Piano-Forte* (1933) published by Cambridge University Press

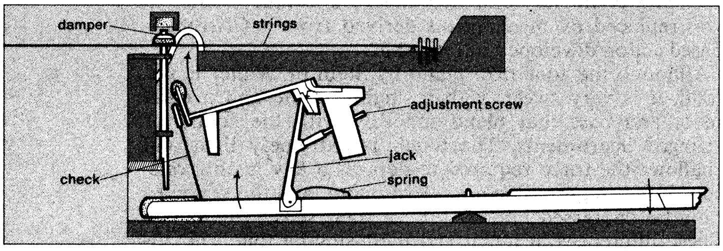


Figure 8: The English action (see text).

Unlike their Austrian and German counterparts, English pianos had two or, at most, three pedals. One of the two ordinary pedals shifted the keyboard sideways so that the hammers struck two or only one of the three strings provided for each note. The second pedal raised all the dampers. It was sometimes replaced by two pedals--one for the treble dampers, the other for the bass dampers—or, occasionally, by a single damper pedal divided into two parts that could be depressed separately or together with one foot, as on the piano presented by Broadwood to Beethoven in 1817.

Although the pianos of the late 18th and early 19th centuries were perfected instruments ideally suited to the music of their period, the increasing popularity of public concerts in large halls and concertos with large orchestras stimulated attempts by piano builders to produce an instrument of greater brilliance and loudness. Their efforts gradually created today's vastly different piano. In recent years, the special merits of the earlier instruments (sometimes called "fortepianos" to distinguish them from modern pianos) have come to be appreciated, and several builders have begun to make replicas of them.

Other early forms. As previously mentioned, many 18th-century pianos were "squares," built in a form resembling the clavichord. More compact and less expensive than wing-shaped grands, the square piano continued through much of the 19th century to be the most common form of piano in the home. But as square pianos became larger and larger, these advantages diminished, and the square piano was eventually replaced by the upright. In the 18th and early 19th centuries, upright pianos (*i.e.*, pianos with vertical strings and soundboard) took three different forms. In the "pyramid piano" the strings slanted upward from left to right, and the case above the keyboard took the form of a tall isosceles triangle. Or a grand piano was essentially set on end with its pointed tail in the air, producing the asymmetrical "giraffe piano." Placing shelves in the upper part of the case to the right of the strings yielded the tall rectangular "cabinet piano." Because the lower end of the strings, which ran nearly vertically, was about at the level of the keyboard, all such instruments were very tall. Although there were attempts to construct lower instruments by, in effect, positioning a square piano on its side, the American builder John Isaac Hawkins made the first truly successful low uprights in 1800 by placing the lower end of the strings near floor level. Robert Wornum in England built similar small uprights in 1811, and in 1842 he devised for them his "tape check" action, the direct forerunner of the modern upright action.

Pyramid, giraffe, and cabinet pianos

Development of the modern piano. In the early 19th century, piano makers were principally concerned with two problems whose solutions led to the modem piano. These were the relatively small volume of sound that could be produced from the thin strings then in use and the difficulty of producing a structure that could withstand the tension even of such light strings once the range of the instrument exceeded 5½ octaves.

Bracing and frame. Like 18th-century harpsichords, the pianos of the 18th and early 19th centuries were constructed entirely of wood, with the case (supported by a structure of internal wooden braces) sustaining the entire stress exerted by the strings. The only metal bracing in such instruments appears in the form of flat pieces bridging the gap through which the hammers rise to strike the strings. These braces eventually proved insufficient when the walls of the case itself and the pinblock (the long piece of wood into which all the tuning pins were driven) were incapable of withstanding the increasing tensions placed upon them. For this reason, ever-increasing quantities of metal bracing came into use, first in the form of individual bars running parallel to the strings from the side of the case to the pinblock but finally in the form of a single massive casting that took the entire tension of the strings upon itself. The one-piece cast-iron frame was first applied to square pianos by Alpheus Babcock of Boston in 1825, and in 1843 another Bostonian, Jonas Chickering, patented a one-piece frame for grands. With the adoption of such frames, the tension exerted by each string (about 24 pounds for an English piano of 1800) rose to an average of approximately 170 pounds in modern instruments, the frame bearing a total tension of 18 tons.

Babcock and Chickering pianos

Overstringing. The strings in early pianos, like those in harpsichords or clavichords, ran parallel to one another, causing the grand pianos of the 18th and early 19th centuries to retain much of the graceful shape of the harpsichord. In the 1830s it was realized that the bass strings could be made longer and their layout improved if they were made to fan out over the treble strings. This idea was first applied to square pianos, but in 1855 Steinway & Sons built a grand piano with a complete cast-iron frame embodying this "overstrung" plan, in which the strings of the treble and the middle registers fan out over most of the soundboard and the bass strings cross over them, forming a separate fan at a higher level. Because the bass strings fan out, the tail of the modern grand piano is far wider than that of the earlier "straight-strung" instruments.

Modifications in the action. The gradual strengthening of the piano's structure to permit the use of heavier strings eventually gave rise to hitherto unforeseen problems. The thicker strings could yield the louder sound of which they were capable only if they were struck by heavier hammers; because of the great velocity advantage provided by the early piano actions, any increase in the weight of the hammer required a manyfold increase in the force required to depress the keys. This difficulty was present to a minor extent even in the 18th-century English grand-piano action, and the touch on these instruments was both deeper and heavier than on Viennese pianos. Moreover, the deeper touch meant that it took longer for a key to return to rest position so that a note could be restruck. Consequently, English pianos were not capable of the rapid repetition of Viennese instruments. This problem became quite severe as the hammers grew heavier and as musicians wished increasingly to use tremolo effects in imitation of orchestral music.

The double-escapement action of Érard

What was necessary was an action that would permit a note to be restruck before the key returned to rest position. The first successful action of this type was devised by the Frenchman Sébastien Érard, who as a young man had built a harpsichord with a particularly elaborate system of pedals and knee levers and in 1810 devised the system of pedals still in use on the harp. Brard's first "repetition" or "double-escapement" action was patented in 1808, and an improved version that is the basis of the modern action was patented in 1821.

A further consequence of the use of thicker strings was that, if the sound of the instrument were not to become unduly harsh, the hammers had to be softer than those used on 18th-century instruments—light slips of wood covered with a few layers of thin leather. Felt-covered hammers were patented by the Parisian builder Jean-Henri Pape, who also contributed a number of other ingenious and important improvements, but the use of felt instead of leather did not become universal until after 1855.

With the adoption of the one-piece cast-iron frame, overstringing, and felt hammers, the piano achieved its modern form in all but a few details. One was the invention in 1862 by Claude Montal of Paris of a pedal that kept the dampers off the strings only for notes already held down. Individual notes could thus be sustained without the overall blurring caused by raising all the dampers by the ordinary damper pedal. This device is not generally found on European pianos, but it is included in American instruments as the middle pedal with the damper ("loud") pedal at the right and the action-shifting (una corda, or "soft") pedal at the left.

Types of modern piano. Since the abandonment of the square piano, only upright and grand pianos are manufactured. The grands range in length from a minimum of about 5 feet (150 centimetres) for a "baby" grand to a maximum of about 9 feet (270 centimetres) for a "concert" grand, although both shorter and longer instruments have been constructed. Among upright pianos, the models over 4 feet (120 centimetres) tall—which frequently had an excellent tone because of their relatively long bass strings—have largely been superseded by the lower models, the "console" (about 40 inches [100 centimetres] high) and the "spinet" (about 36 inches [90 centimetres high]), which look less bulky in the smaller rooms of present-day homes. Because the spinet's case rises such a small distance above the keyboard, it usually has "drop" action, most of which lies below the level of the keys.

Console and spinet models

Modern piano actions. In 1636 Marin Mersenne, the author of the treatise ***Harmonie universelle,*** quoted a remark that the harpsichord of his time contained 1,500 different parts. The modern piano contains 12,000, most of which are found in the action. The modern grand-piano action (Figure 9) is a simplified version of Brard's double-escapement action of 1821, and, although different manufacturers' actions differ in detail, they all work in much the same way. When the key is depressed, its back end rises, lifting the wippen. The wippen raises a pivoted L-shaped jack that pushes the hammer upward by means of a small roller attached to the underside of the hammer shank. The hammer flies free when the back of the L-shaped jack touches the adjustable regulating button. At the same time, the upper end of the repetition lever—through which the upright arm of the jack passes—rises until it is stopped by the drop screw. When the hammer rebounds from the string, the roller falls back until it is stopped by the intermediate lever, enabling the tip of the jack to return to position beneath the roller, even if the key is still partially depressed. The jack is then ready to raise the hammer again should the player restrike the key before it returns to rest position. In the meantime, the hammer is prevented from bouncing back up toward the strings by the padded hammer check, and the damper is raised above the strings by a separate lever lifted by the extreme end of the key.

From E.Q. Norton. Tuning and Care of the Pianoforte

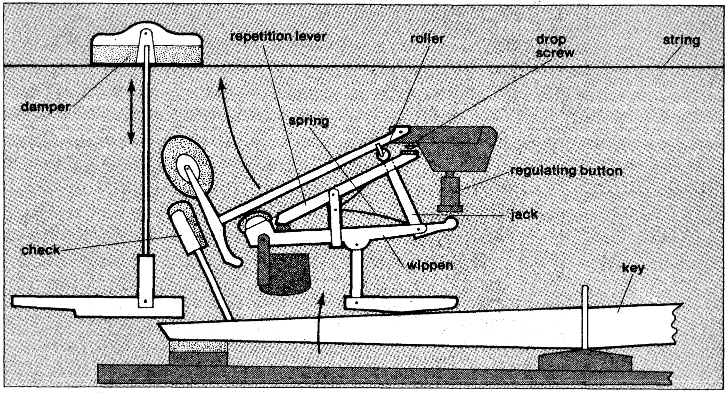


Figure 9: Modern grand-piano action (see text).

Action of the modern upright piano

In the modern upright-piano action (Figure 10), depressing the key also lifts a wippen, which in turn carries an L-shaped jack like that in the grand action. The tip of the jack engages a notch in the hammer butt, thrusting it forward until the regulating button causes the jack to pivot and escape from the notch. Rapid repetition is ensured by the bridle, a strip of cloth that tightens as the hammer moves toward the strings and helps to bring the hammer back to rest position. The dampers are moved away from the strings by a small metal pin, or "spoon," at the back of the wippen.

Player pianos. The history of automatically playing keyboard instruments dates at least to the 16th century. The inventory of musical instruments owned by King Henry VIII at his death in 1547 included "an instrument that goethe with a whele without playing upon," and three spinets equipped with a pinned barrel like that of a music box or barrel organ survive from the workshop of the Augsburg builder Samuel Bidermann (1540–1622). The most common type of player piano operates by means of a roll of punched paper that controls a pneumatic system for depressing the keys. Its heyday was the 1930s, and it was largely rendered obsolete by the increasing popularity of the phonograph and the radio.

Electronic pianos. Since the 1920s there have been experiments with instruments using electric or electromagnetic pickups, amplifiers, and loudspeakers instead of the conventional piano soundboard. Many such pianos do not use strings at all for the generation of sound, employing instead either electronic circuits or tuned pieces of metal. Small, inexpensive instruments of this type are sometimes used for practice, the student listening to himself through earphones. Although some electronic pianos approach the sound of a conventional instrument, their principal advantage seems to lie in producing effects that the conventional piano cannot, by means of electronic prolongation and alteration of their tone.

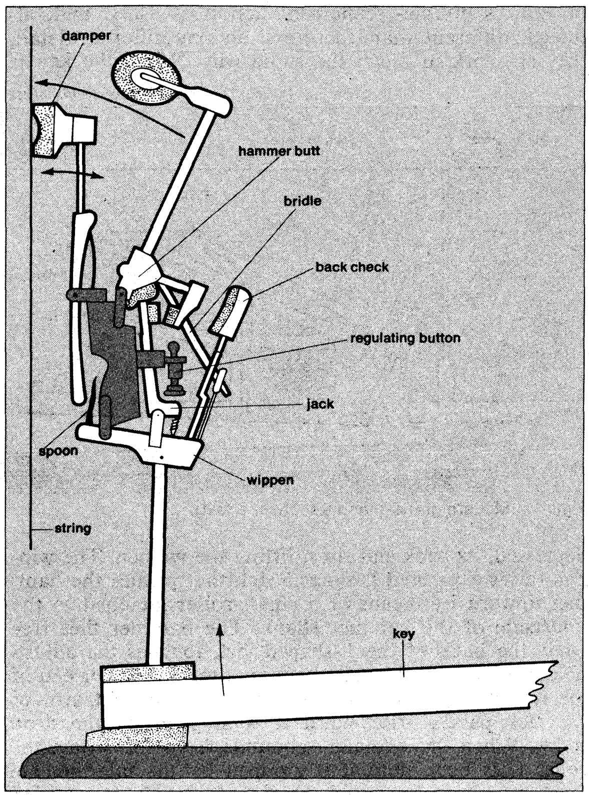


Figure 10: Modern upright-piano action (see text).
From E.Q. *Norton, Tuning and Care of the Pianoforte*

OTHER STRINGED KEYBOARD INSTRUMENTS

Stringed keyboard instruments have as their principal defects an inability, first, to sustain a tone indefinitely and, secondly, to alter the tone's loudness once a key has been depressed. There have been various attempts to build stringed instruments sounded by other means than plucking or striking—including vibrating the strings by blowing a current of air past them, as in the ***piano éolien*** of 1837. The most successful of these other instruments in one way or another adopted the principle of the hurdy-gurdy: ***i.e.***, vibrating the strings by friction.

Heyden's *Geigenwerck*

An instrument of this kind appears in several diagrams in the notebooks of Leonardo da Vinci (1452–1519). Some apparently highly successful ones (none of which, unfortunately, have survived) were made by the Nürnberg builder Hans Heyden, who described them at length in pamphlets published in 1605 and 1610. These instruments had a series of rosined wheels that rubbed the strings when they were drawn against them by the action of the keys. According to Heyden, the instrument, which he called a ***Geigenwerck***, was capable of recreating the sound of an ensemble of viols and produced sounds of different loudness depending on the force with which the keys were depressed. The sole surviving instrument of this type, made in Spain in 1625, is far less sophisticated than Heyden's and gives only a faint inkling of what his must have sounded like.

In 1772 a device called a celestina was patented by Adam Walker of London; it employed a continuous horsehair ribbon (kept in motion by a treadle) to rub the strings of a harpsichord. Thomas Jefferson, who ordered a harpsichord equipped with a celestina in 1786, commented that it was suitable for use in slow movements and as an accompaniment to the voice. Similar devices, some using rosined rollers, were applied to pianos by various ingenious inventors throughout the 19th century.

BIBLIOGRAPHY

Development of the keyboard, the clavichord, and the harpsichord: FRANK HUBBARD, ***Three Centuries of Harpsichord Making*** (1965), the only book on the subject to clarify the history of the instrument and the techniques of earlier harpsichord makers; RAYMOND RUSSELL, ***The Harpsichord and Clavichord*** (1959), a general history, less technical than Hubbard's, illustrated with excellent photographs; DONALD H. BOALCH, ***Makers of the Harpsichord and Clavichord, 1440–1840*** (1956), known data on all makers whose names have survived as well as the locations of their extant instruments; EDWIN M. RIPIN (ed.), ***Keyboard Instruments*** (1971), a collection of 10 articles on specialized topics, including Italian, Flemish, and English harpsichords, 15th-century instruments, and the ***Geigenwerck;*** HANNS NEUPERT, ***Das Klavichord*** (1956; Eng. trans., 1965) and ***Das Cembalo,*** 3rd ed. (1956; Eng. trans., ***Harpsichord Manual,*** 1960), elementary treatises with emphasis on modern instruments—much of the historical information has now been superseded; FRANZ J. HIRT, ***Masterwerk des Klavierbaus*** (1955; Eng. trans., ***Stringed Keyboard Instruments,*** 1968), a splendid picture book with interesting but somewhat unreliable supporting text; EMANUEL WINTERNITZ, ***Keyboard Instruments in the Metropolitan Museum of Art*** (1961), a useful, small, and inexpensive picture book with a fair text; WOLFGANG J. ZUCKERMANN, ***The Modern Harpsichord*** (1969), the standard work on modern harpsichords and their makers, although unreliable for historical information and technical aspects of instrument design and building; WILLI APEL, "Early History of the Organ," ***Speculum,*** 23:191–216 (1948), an excellent account, especially useful for the details it gives on the evolution of the keyboard; EDWIN M. RIPIN, "The Early Clavichord," ***Musical Quarterly,*** 53:518–538 (1967); "A Reassessment of the Fretted Clavichord," ***Galpin Society Journal,*** 23:40–48 (1970); "The Two-Manual Harpsichord in Flanders Before 1650," ***ibid.,*** 21:33–39 (1968); and "Expressive Devices Applied to the Eighteenth-Century Harpsichord," ***The Organ Yearbook,*** 1:65–80 (1970), a series of articles presenting new information or interpretations not available in book form; RAYMOND RUSSELL, "The Harpsichord Since 1800," ***Proceedings of the Royal Musical Association,*** 82:61–74 (1955–56), an account of the early period in the modern revival of the harpsichord that fills some of the gaps in Zuckermann's work.

The Piano: R.E.M. HARDING, ***The Piano-Forte: Its History Traced to the Great Exhibition of 1851*** (1933), the standard work in the field—the diagrams are excellent, but the text yields its information only reluctantly; ALFRED J. HIPKINS, ***A Description and History of the Pianoforte and of the Older Keyboard Stringed Instruments*** (1896), a classic that has never been superseded as a brief account of the subject, includes discussion of all stringed keyboard instruments; ARTHUR LOESSER, ***Men, Women and Pianos*** (1954), a lively account of all keyboard instruments, their music, and their place in society—the author's bias in favour of the modern piano makes the book unreliable for details on instruments before 1840; WILLIAM L. SUMNER, ***The Pianoforte*** (1966), an acceptable elementary history; DANIEL SPILLANE, ***History of the American Pianoforte*** (1890, reprinted 1969), still the standard work on American piano building; WILLIAM B. WHITE, ***Piano Tuning and Allied Arts,*** 5th ed. rev. (1946), the standard technician's manual in America; A.W.J.G. ORD-HUME, ***Player Piano*** (1970), an excellent history as well as a technical manual; THEODORE E. STEINWAY, ***People and Pianos*** (1953), contains interesting details and human-interest material; HUGH GOUGH, "The Classical Grand Pianoforte, 1770–1830," ***Proceedings of the Royal Musical Association,*** 77:41–50 (1950–51), a good discussion of the musical and technical aspects of the instruments, especially valuable for its comparison of English and German pianos.

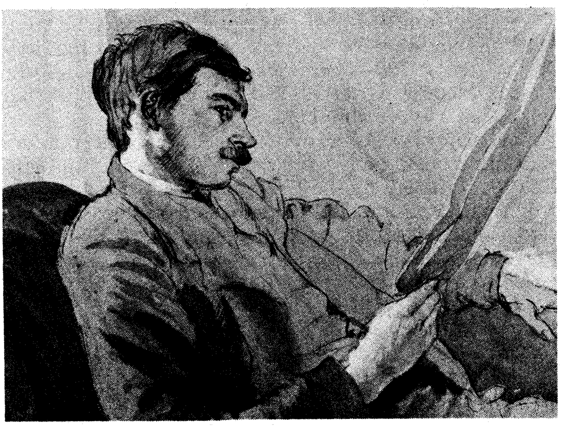
(E.M.R.)

Keynes, J.M.

John Maynard Keynes, critic and architect of economic policies, political journalist, and successful financier, first made his mark as an iconoclastic economist with ***The Economic Consequences of the Peace,*** a compelling indictment of the Versailles Peace Treaty. His enduring fame, however, rests on his ***General Theory of Employment, Interest and Money,*** which showed that recovery from a recession can best be achieved by a government-sponsored policy of full employment.

Early life and education. Keynes was born in Cambridge, England, on June 5, 1883, into a moderately prosperous family of the professional elite. His father was an esteemed economic scholar and later academic administrator in Cambridge University. His mother, the daughter of a minister endowed with advanced opinions, was one of the first female graduates of the same university.

An obviously gifted youngster, Keynes travelled the customary educational route provided for the upper and upper middle class child. From his preparatory school, he



Keynes, watercolour by Gwen Raverat (born 1885). In the National Portrait Gallery, London.
BY courtesy of the National Portrait Gallery, London

Keynes at Eton

went to Eton, the most prestigious public (*i.e.*, private) boy's school in the land. In sharp contrast to the bitter school memories of such gifted Englishmen as the writers George Orwell and Cyril Connolly, Keynes appears to have enjoyed every moment of his Eton experience. He was somewhat delicate of constitution and no athlete, but he excelled both as a mathematician and as a scholar of the classics. His essays on such varied themes as Bernard of Cluny, the character of the Stuart monarchs, and the differences between East and West won numerous prizes. As his greatest feat, he secured the respect and admiration of his irreverent contemporaries by a general affability and a pronounced talent for the management of student societies. His classmates honoured him by election to Pop, the student society that traditionally participates in the government of Eton affairs.

At Cambridge, which Keynes entered in 1902, he fared even better. At King's College, one of the oldest and richest of Cambridge endowments, he "quickly established," as one biographer put it, "the ascendancy that he had achieved over his contemporaries at Eton." He blossomed as a polemical speaker, becoming president of the University Union Society, the undergraduate counterpart of the House of Commons. Gradually, as he came under the influence of Alfred Marshall, the leading Cambridge economist, his mathematical interests diminished, and his attachment to politics and economics increased.

The Bloomsbury group

Cambridge introduced Keynes to an important circle of writers and artists. The early history of Bloomsbury, that exclusive circle of the cultural elect, which counted among its members E.M. Forster, the philosopher G.E. Moore, Virginia Woolf, the painter Duncan Grant, and the art critic Clive Bell, centres upon Cambridge and the remarkable figure of Lytton Strachey. Strachey, who had entered Cambridge two years before Keynes, inducted the younger man into the exclusive private club known simply as "the Society." Its members and associates were the leading spirits of Bloomsbury. Throughout his life, Keynes was to cherish the affection and respond to the influence of this tribe of free souls. The group was more than extraordinary in its achievements: it was also utterly fearless in its opinions and courageous in its conduct. No practice, from homosexuality (Keynes was later reported to have vied at one time with Lytton Strachey for the affections of Duncan Grant) to conscientious objection, was the occasion for reprobation in this circle, provided only that it could logically be justified as promoting one or both of the major group values—friendship and beauty.

Early career. After graduating from Cambridge, Keynes became a civil servant and moved to the India Office in Whitehall. His experience there formed the basis of his first major work, which is still the definitive examination of pre-World War I Indian finance and currency. He then returned to Cambridge, where he taught economics with great success until 1915. The onset of World War I brought him back to government employment, this time in the Treasury, an agency even more powerful than its American counterpart. Here he was daily concerned with the economic management of the war. His special responsibility covered relations with allies and the conservation of England's scant supply of foreign currencies.

His scintillating performance marked Keynes for a great public career. But the Versailles Peace Conference was to transform his prospects. Accompanying his prime minister, Lloyd George, to France as an economic adviser, Keynes soon became so distressed at the political chicaneries of the proceedings and the unrealistic character of the reparations policies to be imposed upon a defeated Germany that his health deteriorated, and he resigned his post, oppressed, to quote from a letter to his father, at the impending "devastation of Europe."

Always an activist, Keynes translated personal distress into public protest. In two summer months he composed the indictment of the Versailles settlement that by Christmas 1919 reached the bookstores as *The Economic Consequences of the Peace*. The permanent importance of this great polemical essay lies in its cogent economic analysis of the excessive weight of reparations upon the German economy and the corresponding lack of probability that they ever would be paid. What guaranteed the book general success were the blistering sketches of Woodrow Wilson, Clemenceau, and Keynes's old chief Lloyd George. Wisely, Keynes began the book with these portraits in acid.

***The** Economic Consequences of the Peace*

To Keynes's former superiors and colleagues in the Treasury, the performance was even worse than a scandalous betrayal of a privileged position: it was a raving success into the bargain. As a consequence, in some Whitehall circles Keynes remained to the end of his life a man not quite to be trusted, an iconoclast perfectly willing to rock any boat into which he had imprudently been invited.

In Cambridge, to which Keynes now returned, his reputation was rather different. He was quite simply esteemed as the most brilliant student of Alfred Marshall and A.C. Pigou, the two Cambridge economists who between them had produced the authoritative explanation of how competitive markets functioned, business firms operated, and consumers spent their incomes. Although the tone of Keynes's major writings in the 1920s was occasionally skeptical, he did not directly challenge that conventional wisdom of the period that held laissez-faire, only slightly tempered by public policy, the best of all possible social arrangements.

Two of Keynes's opinions did foreshadow the theoretical revolution over which he presided in the 1930s. In 1925 he opposed England's return to the gold standard at the prewar dollar–pound ratio of $4.86; and long before the Great Depression, Keynes was worried about the persistent unemployment of British coal miners, shipyard workers, and textile operatives. Reconciled by this time to Lloyd George (who was never to return to office), he supported the Liberal Party's program of public works to take the unemployed off welfare and place them in useful jobs. But respectable economists still relied upon the expected automatic adjustments of the free market, and the Treasury was attached to the still more doctrinaire view that public works were necessarily useless because any increase in the government deficit would surely reflect itself as an equal decline in private investment. Since Keynes could not yet offer a theoretical refutation of his colleagues' opinions, his agitation for public works had little political effect.

Marriage and prosperity

His private life was more satisfactory. In 1925 he married the talented and charming ballerina Lydia Lopokova. Despite the doubts expressed by Cambridge conservatives of his wisdom in courting a "chorus girl," the marriage was happy. Indeed, in the judgment of Mrs. Alfred Marshall, the economist's widow, it was "the best thing Maynard ever did." His friendships with the Bloomsbury circle prospered, and he easily financed a comfortable London apartment from the proceeds of highly successful stock speculations. Always a late riser, Keynes conducted the management of his investments in civilized fashion while he breakfasted in bed. As a journalist he extended his reputation, writing in the leftist New *Statesman* (whose Board of Directors he later joined) as well as the major newspapers, *The Times* (of London) and the *Manchester Guardian* (now the *Guardian).*

Keynes and the Depression. What altered Keynes's public reputation a second time and swept him to permanent fame was his response to the economic tragedies of the 1930s. With a few noteworthy exceptions politicians and economists were universally bewildered by events and shorn of confidence in traditional policies. Conventional economic wisdom still held, against mounting evidence to the contrary, that time and nature would restore prosperity if government did not try to manipulate the economy. After all, the unemployed could always get jobs if they only were willing to work for lower wages. Businessmen could always restore their sales by the parallel strategy of slashing their prices. In the process a few of the weaker brethren would indubitably be wiped out. Nevertheless, if all groups accepted the discipline of competitive adjustment, soon recovery, prosperity, and higher wages would return.

The General Theory of Employment, Interest and Money

Unhappily, the remedy worked nowhere. In the United States Franklin Roosevelt's 1932 landslide presidential victory over Herbert Hoover attested to the political bankruptcy of laissez-faire. New explanations and new policies were urgently needed. These Keynes supplied. His enduring triumph was his composition in the early 1930s of the most influential treatise composed by an economist thus far in this century, comparable to the *Wealth of Nations* as an intellectual event and to Malthus' *Essay on Population* as a guide to public policy. *The General Theory of Employment, Interest and Money,* which appeared in England at the very end of 1935, is a highly technical, even abstruse exposition of new ideas that had been partly foreshadowed in his *Treatise on Money* (1930). The central message is readily translated into two powerful propositions. The first declared the existing theory of unemployment nonsense. In a depression, according to Keynes, there was no wage so low that it could eliminate unemployment. Accordingly, it was wicked to blame the unemployed for their plight. The second proposition proposed an alternative explanation about the origins of unemployment and depression. This centred upon aggregate demand—*i.e.*, the total spending of consumers, business investors, and public agencies. When aggregate demand was low, sales and jobs suffered. When it was high, all was well.

From these generalities there flowed a powerful and comprehensive view of economic behaviour. Because consumers were limited in their spending by the size of their incomes, they were not the source of business cycle fluctuations. The dynamic actors were business investors and governments. In depressions the thing to do was either to enlarge private investment or to create public substitutes for private investment deficiencies. In mild economic contractions, monetary policy in the shape of easier credit and lower interest rates just might stimulate business investment and restore the aggregate demand caused by full employment. Severer contractions required as therapy the sterner remedy of deliberate public deficits either in the shape of public works or subsidies to afflicted groups.

Impact of Keynes

In an extremely short space of time, Keynes converted most of his professional colleagues. During World War II and after, one Western democracy after another affirmed its new commitment to maintain high employment. In the United States the Employment Act of 1946 formally imposed upon presidents and congresses the duty of maintaining prosperity. In the United States the 1964 tax cut, initiated by President John F. Kennedy and enacted in his successor's administration, quite openly enlarged an existing national deficit so as to promote faster expansion and high employment. This was truly the triumph of the Keynesian idea and vision.

Keynes's late career. The *General Theory* was Keynes's last major written work. In 1937 he suffered a severe heart attack. Two years later, though not completely recovered, he returned to teaching at Cambridge, wrote three influential articles on war finance entitled *How to Pay for the War,* and served once more in the Treasury as an all-purpose adviser. As the war drew to a victorious conclusion, Keynes turned his thoughts to the design of international financial institutions calculated to limit the spread of depression. At the Bretton Woods Conference in 1944 he played a prominent part. But the institutions that resulted from the conference, the International Monetary Fund and the World Bank—two agencies that survive into the 1970s—bear much stronger marks of the orthodox theories of the United States Treasury of that time than of Keynes's thinking.

His last major public service was his brilliant negotiation in the autumn and early winter of 1945 of a multibillion-dollar loan granted by the United States to England. By this time the British establishment was more than ready to welcome home its errant son. He was now both a member of the Court of the Bank of England (as its governing board is termed) and, as Lord Keynes of Tilton, a peer of the realm. In the old Bloomsbury tradition of disrespect for the conventional, he wryly commented, "I am not sure which of us is being made an honest woman—the Old Lady [as the Bank of England is popularly known] or me." The worry was groundless. Keynes's old enemies had by now capitulated to one of their country's supreme intellects.

On April 21, 1946, overstrained by wartime exertion, Keynes died in Firle, Sussex. Intimate alike of artist, scholar, and politician, cultural entrepreneur, gifted teacher, loyal friend, and original thinker, this child of the philosophical tradition of Locke, Hume, and Mill late in his life expressed as a single regret that he had not drunk more champagne. Bloomsbury has always preferred effervescence to pomposity.

BIBLIOGRAPHY. R.F. HARROD, *The Life of John Maynard Keynes* (1951), is an authoritative biography written by one of Keynes's students and colleagues. R. LEKACHMAN, *The Age of Keynes* (1965), contains a short biography, a summary of Keynesian economics, and an account of its use in the U.S. An exceptionally clear exposition of Keynes's major work may be found in D.D. DILLARD, *The Economics of John Maynard Keynes* (1948). Keynes's opinions of public issues and of other economists are best examined in his graceful essays contained in the following collections: *Essays in Biography* (1933); and *Essays in Persuasion* (1931).

(R.L.)

Khoisan Cultures of Africa

The term Khoisan is derived from the Hottentot roots *khoi* ("man," in the generic sense) and *sa* ("Bushman"). It is therefore an apt portmanteau word to designate the indigenous southern African peoples known as Hottentots and Bushmen. The nature of the relationship between the two Khoisan groups has given rise to much discussion. They closely resemble each other in physical appearance and are markedly different from any of the other southern African populations. The prevailing view is that they are of the same genetic stock and differ only in their cultures, the Hottentots being pastoral while the Bushmen were formerly all hunter-gatherers. The Central Bushman languages, Naro (Nharo), G/wi, and G//ana (/ and // represent click sounds in the Hottentot language), have such close lexical and structural resemblance to Hottentot languages that all have been classified as belonging to the same language family.

Physical appearance of the Khoisan peoples

Bushmen and Hottentots are lightly built with thin limbs and wiry muscles. They are shorter in stature than most southern African peoples; Bushman males average about five feet two inches (1.57 metres), women being an inch or so shorter, and Hottentots are perhaps two inches taller than Bushmen. Their skin colour at birth is coppery yellow, darkening to yellowish brown with exposure to the sun. They have a rather flat, protuberant forehead (a feature more marked in Bushmen), narrow, deep-set eyes, and prominent cheekbones. The nose is rather flat, with arched nostrils, and the lips are turned slightly outward. The chin is small and slightly pointed. Their small, often lobeless ears sit close to the head beneath black, tightly kinked "peppercorn" hair. Their dark-brown eyes are narrowed by the fullness of the upper lids and the pronounced epicanthic folds, a feature that has led some erroneously to link them with Mongoloid peoples.

Facial hair is sparse, and few men develop beards before middle age. There is little body hair, even in the armpits and genital region. The skin loses its elasticity early, supposedly because of the thinness of the layer of subcutaneous fat, and heavy wrinkles often form on the face, neck, and abdomen, giving the individual a misleadingly aged appearance.

Steatopygia (protuberant buttocks) and "Hottentot apron" (elongated labia minora) are common among Khoisan women. Steatopygia is caused partly by a characteristic outward curvature of the spine about the lumbar–sacral region, which pushes the buttocks outward; the prominence of the buttocks is increased by the racial tendency to accumulate fat on the hips and buttocks rather than about the abdomen. The penis of Bushman males and of many Hottentot males extends outward almost horizontally in a semi-erect position even when not tumescent.

The now-extinct Bushmen south of the Molopo River were smaller and lighter coloured than are the present-day Bushmen of Botswana, South West Africa, and Angola. The latter are thought to have acquired Negroid characteristics from interbreeding with Bantu-speaking peoples.

THE KHOIKHOIN, OR HOTTENTOTS

Origin of the name Hottentot

The origin of the name Hottentot is uncertain. The most plausible theory is that it was invented in 1647 by the survivors of a Dutch shipwreck who camped in the vicinity of present-day Cape Town for a year before being rescued. They established friendly relations with the local inhabitants but were quite unable to learn their language. The supposition is that the Dutchmen took the refrain autentou from a ¾-time dance of the people and used it in friendly, joking reference to them. In the form hottentot it was later adopted by the Europeans and subsequently by the Hottentots themselves. The original Hottentot self-designation had been Khoikhoin ("people people," or "proper people").

Knowledge of traditional Hottentot culture is fragmentary and imprecise. The only records come from travellers and missionaries who were untrained observers and therefore strongly influenced by their own values and preconceptions. By the time capable and interested observers came to the southern African scene, war and disease had taken their toll, and only vestiges of the traditional cultures remained.

Subsistence and economics. The Khoikhoin were pastoral, grazing their fat-tailed sheep and long-horned cattle over much of the drier, western half of the subcontinent. The native sward and the meagre supplies of water were unsuited to sustained stock raising, and the Khoikhoin had periodically to trek their flocks and herds in search of new pasture and water. No crops were cultivated. The diet of milk and meat was supplemented by game and wild plants.

Their settlement pattern appears to have varied with the conditions of the locale, but a typical community may have consisted of 200 or 300 people, together owning upward of 1,000 head of stock. The temporary settlement was enclosed by a circular fence of thornbush, within which were the kraals (corrals) for lambs and calves. No special enclosures were made for adult beasts, which were driven into the main enclosure at night among the huts of their owners. These huts were hemispherical structures made by planting a ring of wands in the ground and bending the tops over, tying them together to form a dome that was covered with woven grass matting. Weatherproof and cool, the huts could be assembled or taken down in a few hours, loaded onto oxen, and carried to the next settlement site. The oxen were used as mounts as well as pack animals.

The hut was the property and domain of the wife; she had complete authority over it, to the extent of being able to exclude her husband if she wished. It appears that the Khoikhoiti woman's status was equal or nearly so to that of the man; women could own stock and were economically important in their role as gatherers of plant food. The men were herders and hunters and, among some tribes, metalworkers. The principal weapons were arrows and spears tipped with bone or iron. An extensive trading network seems to have existed through which scattered settlements exchanged aromatic, medicinal, and narcotic herbs; pottery; and metalware such as iron and copper beads, arrowheads, and knife and spear blades. Some Khoikhoin, particularly in northeastern Botswana, had developed rather advanced mining techniques; the old shafts indicate a high order of geological insight.

Religion of the Hottentots. The Khoikhoi deity Tsuni-g // oab is celebrated in myths as the first of that people from which all others take their origin. A great chief and a powerful magician, he worked many wonders, including the creation of men and women. He died several times and was resurrected, eventually becoming a purely spiritual being. As such he was omniscient and omnipresent and the giver of life and of rain. Supplication was made to him in an elaborate ritual to persuade him to send the rain that was essential for survival of man and stock.

Death and rebirth

Another deity, Heitsi-eibib, was the hero of a cycle of myth. He could change his form and had died and been reborn many times. He was worshipped at his many graves scattered about Hottentot country; the passerby would add a stone to the heap marking the grave and pray to Heitsi-eibib for prosperity and success.

G // aunab was a god with varying attributes who, with appropriate phonemic changes of his name to fit a particular language, appeared in the theologies of a number of Khoisan peoples. In the myths of the Hottentots he is the adversary of Tsuni-g // oab and the bringer of death, disease, and other misfortunes. In some accounts he appears as the personification of the malevolent spirits of the dead. Among those peoples exposed to Christianity, he has been equated with the devil.

The Khoikhoin are also said to have worshipped the moon, although not much is known about this. Certainly the moon figures in their very rich mythology.

The practice of magic does not seem to have been very extensive. Curing and divination were the main activities of magicians, although they occasionally misused their powers for evil purposes. When they did, their influence could be negated by the application of cold water, either to the sorcerer or to his victim—an illustration of the high value placed on water by the Hottentots.

Political organization. Hottentot communities were linked by language and kinship into a loose confederation of clans that formed a nation. Several nations existed in the 17th century. Each had its own territory, its own name, and apparently its own dialect. (The dialectal differences must have been minor, as there was great similarity in the Hottentot languages across their total range of nearly 2,000 miles [3,000 kilometres] from the eastern Cape Province to the highlands of South West Africa.) The clans comprising a nation were ranked in order of seniority, the chief of the nation being drawn from the senior clan. Marriage within the clan was strictly forbidden; the need to find spouses in other clans helped to keep the nation together. The nation as a whole had exclusive rights to grazing and hunting land, water holes, and mineral rights. In the larger nations there seems to have been some apportionment of land among the clans.

Hottentot confederation of clans

Leadership, despite its clan basis, was not strong; it was persuasive rather than autocratic. Many instances are recorded of individual clans hiving off from the main body, attracting a following from other clans, and setting up as an independent nation with a new rank order of constituent clans. It has been suggested that this anarchistic tendency was reinforced by social and environmental pressures. Among a people lacking a strong and coercive political system and totally dependent on a scarce resource such as water, the use of that resource can be effectively controlled only if harmony prevails in the community. Harmony is threatened if the community reaches an unwieldy size and imposes too heavy a demand on the limited supply of water. The only solution then is for the community to divide. The "daughter" component must put sufficient distance between itself and the "parent" community to avoid predation by the latter. The effect is to limit the size of communities and to encourage their wide dispersal over habitable country.

Hottentot nations did not develop great strength or solidarity. As nomadic pastoralists they were forced to give ground before the incursions of more cohesively organized peoples: the Bantu who moved down the southeastern coastal areas (probably about the 10th century AD) and European settlers who began moving inland from Table Bay in the 17th century. By the 19th century some of the Hottentot nations felt themselves desperately pressed. Strong leaders were able to unite fragments of clans and nations under them, and large groups moved across the Orange River out of Cape Colony into South West Africa, thus retracing the route that may have been taken by forebears, who are thought to have migrated into southern Africa from the central African lakes. In South West Africa some of the Hottentots had troubles with the Herero group of the Bantu, who were roving farther and farther south. Caught between the Europeans and the Bantu, the Hottentots fought for nearly 100 years, giving up the struggle only after the suppression of an uprising at Bondelswart in 1922.

Little now remains of a distinctively Khoikhoi culture, despite the efforts of missionaries and governments to soften the impact of alien cultures. The Nama and Garikwe communities retain their identities as Hottentots, and the former have kept their language and some of their customs. Many other communities live on reserves, where they make a living from their livestock, supplemented by jobs as unskilled or semiskilled labourers, but theirs is more the culture of poverty than that of the traditional Khoikhoin people.

THE BUSHMEN

The term Bushman is an anglicization of the 17th-century Dutch *bosjesmans* ("bush dwellers"; *i.e.*, people without fixed abode). The historical evidence concerning Bushmen is incomplete and often unreliable. As hunters and gatherers they were unobtrusive in their occupation of the land. Few early explorers took much interest in them, and fewer had the time and means to study the small, widely dispersed communities that shifted their encampments every few weeks. The Bushmen occupied most of the country south of the Zambezi River that was not in the hands of the Hottentots or Bantu. Skeletal and cultural remains, particularly rock paintings and engravings, indicate that the yellow hunters once ranged over the whole subcontinent.

Distribution of modern Bushmen

Present-day Bushmen. The Bushman population in the 1970s numbered between 45,000 and 50,000. Some 25,000 or 30,000 lived in the Republic of Botswana, 17,000 of them in the western Kalahari districts. South West Africa had about 15,000, and it is estimated that another 4,000 lived in southeastern Angola. A small number of them moved among Zambia, Rhodesia, and Botswana. Two groups lived in the Republic of South Africa, one in the Kalahari Gemsbok National Park in Cape Province and the other, a very small community, near Lake Chrissie in the eastern Transvaal.

It is often stated that the surviving Bushmen migrated to their present inhospitable desert habitats in recent times after being driven out of other territories. But there is ample evidence that they occupied their present locations in the mid-19th century, when other Bushmen were still living in parts of southern Africa where they are now extinct. Secondly, there is no evidence that the latter groups migrated into the desert areas; rather, it is clear that they remained where they were until they were overwhelmed by war and disease or absorbed by interbreeding with Europeans and Bantu. Thirdly, the distribution of dialects among the present-day population is consistent with their having been in their present locations over a very long period.

The majority of today's Bushmen have been dispossessed of their territories by European and Bantu cattle raisers. The intrusion of cattlemen has reduced the supply of game animals and esculent plants that constituted the greater part of their diet. Most Bushmen have therefore turned to livestock raising for their living, either by hiring themselves as labourers to ranchers or by entering into master–client relationships with Bantu owners of cattle posts. A few of them still follow the traditional hunter-gatherer style of life.

Ranch labourers. On the marginal cattle ranches in the arid regions of southern Africa, Bushmen compete with the Bantu for available jobs. Because of their relative lack of strength, sophistication, and marketable skills and their common ignorance of languages other than their own, they can get only the lowest paid jobs. Any man who finds employment soon attracts a number of dependent kin and friends who come and live with him on the ranch. Most ranchers recognize the plight of the Bushmen and increase the labourer's issue of rations (the usual supplement to wages) to accommodate the train of dependents. Some ranchers give an annual bonus of small livestock, intended to be kept and allowed to multiply on the ranch; more often than not, however, the Bushmen slaughter the bonus for Christmas or some other special occasion.

Ranch Bushmen, although retaining their languages and some aspects of social organization such as the kinship structure, have lost much of the rationale of their old culture. In the traditional life-style, women had great economic importance as the gatherers of food plants, which constituted the main part of the diet. On the ranches, however, a labourer's wife may be something of a liability when he is competing with other men for a job. Deserted wives and children are numerous; they turn to already overburdened kinsmen for help or to prostitution and petty crime.

Hunting-gathering Bushmen. Independent Bushmen are confined mainly to the western part of Botswana. In the northwest are the K!ung (! represents a click sound), who inhabit an area with some permanent water holes and numerous esculent plants, including extensive groves of mugongo trees (*Ricinodendron* rautanenii), which bear pleasant-tasting and nutritious nuts. The !xõ live in the west central part of Botswana, a region without natural supplies of permanent water but rich in game and esculent plants. In the central Kalahari region of Botswana are the G/wi, /aba, and G//ana groups. This region does not have the seasonal abundance of game that is seen farther west and is also waterless except during the wet season. Other hunter-gatherer Bushmen are found elsewhere in Botswana, South West Africa, and Angola, but they are not independent and autonomous, and their way of life is therefore less representative of the traditional Bushman cultures.

Cultural variety of Bushmen

Bushmen are not culturally homogeneous, although the cultures of some of the groups resemble one another. The K!ung, !xõ, and G/wi, though all hunter-gatherers, live in rather different habitats containing different resources, and the three groups show marked differences in their technologies and their patterns of exploiting resources. Organizationally, these groups also exhibit marked contrasts. The lexical and structural differences among their languages are so great that they have been classed in three different language families. Generalizations about the Bushmen can therefore be made only at the level of a "southern African hunter-gatherer culture" and limited to such superficial common features as the use of bows and poisoned arrows in hunting; a low population density and small communities; generally diffuse and ephemeral leadership; egalitarian communities, the government of which is by consensus; an absence of the lineage principle in most if not all, kinship systems; and a lack of formal legal and judicial systems. Both the breaking up and coalescence of communities are common. Violence and warfare, although frequently reported in the literature of now-extinct Bushmen, are not characteristic of the present-day hunting-gathering Bushman peoples; on the contrary, their ethics stress cooperation.

A detailed description of one of the Bushman peoples follows.

The G/wi people of the Kalahari. The G/wi inhabit the western half of the Central Kalahari Game Reserve and the eastern fringe of the Ghanzi ranches. About 800 G/wi live on the ranches as labourers and their dependents. Another 500 or so move into the ranching area in time of drought. Sporadic trading visits are made at other times to exchange the desert produce of game-hide leather for tobacco, iron rods, and fencing wire—used in the manufacture of tools and weapons. The remaining 2,000–2,500 G/wi are permanent desert dwellers who have only sporadic contact with Bantu and Europeans and have retained their traditional hunter-gatherer culture.

Environment of the G/wi

The habitat. The G/wi country lies in the centre of the Kalahari Basin, a depression in the vast inland plateau of southern Africa. The basin is covered by a mantle of fine-grained sand, up to **400** feet (120 metres) deep, poor in plant nutrients and of a texture that renders it susceptible to wind erosion, leaching, and scorching. Rainfall is confined to ten or 12 weeks after midsummer. The country responds with startling rapidity to the first good rains, and within four days the landscape is transformed by a luxuriant growth of grasses and herbs and by the flush of foliage on shrubs and trees. Rain is the key to all life in the Kalahari. The extreme seasonal variations in rainfall, humidity, and temperature lead to corresponding seasonal contrasts in the amounts of plant and animal food available to hunter-gatherers. There are a few localities in which resources are sufficient to meet the year-round needs of the hunters and gatherers, and these are the areas that the G/wi communities occupy as their territories.

The G/wi world view. In G/wi theology the universe was created by N!adima. He is the owner of all that is and may dispose of it as he wishes. His actions are, however, bound by the natural systems he is believed to have ordained; he cannot suspend or reverse their operation. N!adima is remote from his creatures. They cannot communicate with him or influence his will to their own advantage or favour. G/wi therefore have no religious rites such as prayer, worship, or sacrifice and no priests. They stoically accept N!adima's occasional caprice; *e.g.*, when "he grows tired of a man's face" and sends a marauding lion or some other misfortune to kill him.

G // amama is a less powerful being than N!adima. He sporadically attempts to do harm to man and is often successful, but he can be frustrated by a number of means that N!adima created and man has been able to discover. Several herbal medicines have the power to counter G // amama's influence; certain communal dances also serve to exorcize the evil he is believed to send in the form of invisible, magical slivers of wood.

The logic of the G/wi world view is that N!adima, having created man, intends him to survive and thus permits him to make use of what is available in the environment to that end, subject to the restriction that N!adima will be angered by wastefulness or greed on the part of those using his property. Man, at least G/wi man, must devise for himself the best means of survival that he can, including the regulation of interpersonal behaviour. Social usages and customs are therefore seen as man-made and not sacrosanct. The stability of behaviour and the predictability of others' actions require agreement by all concerned. Such agreement can be won only by persuasion and ensured by the fulfillment of obligation. It is therefore imperative that individuals and the community remain on good terms with everybody in the circle of social contact.

Social and political organization. The communities, or bands, usually have between **40** and 60 members, the extremes being 25 and 85. They are open communities, in the sense that there are no particular qualifications for membership, but members tend to stay with a group over long periods. Eventually, some catastrophe such as severe drought or epidemic disease (*e.g.,* the smallpox outbreak of 1950) may lead some bands to merge, in order to ensure their members' survival.

G/wi kinship and social structure

The G/wi kinship system is partially classificatory (*i.e.,* it subsumes several different biological relationships under a smaller number of categories). There is also a high degree of equivalence among those classified by the same kinship term. These features allow the kinship system to be applied to persons other than actual kin, including the whole membership of the band. The kinship system can thus be used as the organizing principle of society and, by further extension, to govern relationships between people in different bands.

The kinship system permits not only extension but also truncation: groups can be enlarged by inclusion of new members and also easily split when necessary by substituting other members for those who have left. This is of great advantage to the G/wi. Not only are their bands subject to periodic disasters, but they also grow; in some cases they outgrow their territorial resources and need to be able to split up. The bands normally respond to the decreased density of food resources in winter and summer by reducing their population density. This is achieved by fragmenting into constituent households, each of which goes off into a separate part of the territory and remains there until the approach of the wet season, when food supplies increase again. The organization of the band must therefore compromise between making the community cohesive enough to hold the members together and keeping it loose enough to allow them to separate when necessity dictates.

The kinship system also provides sets of rules for appropriate behaviour. All kin of any individual are either his avoidance or his joking relatives. Behaviour between avoidance relatives is reserved and polite. The avoidance category includes those who are in a relationship involving submission to or exercise of authority—firstly, the parent–child relationship and, by extension, relationships between all those who are classificatorily equated with parents and children. Siblings of opposite sex are in the avoidance category and, by extension, all those whose classificatory relationship prohibits intermarriage. All other kin are joking relatives, among whom behaviour is much less restrained: possessions are freely shared; ribald, bawdy joking is in order; and pointed but friendly criticism of behaviour is permitted in public without offense.

The regular retreat into isolation requires that each household be capable of independent existence for three or four months of the year. This precludes the development of any centralized authoritarian political system either within the band or among bands. Leadership is diffuse and ephemeral, arising in response to specific situations. The function of a leader is to make proposals for action that are acceptable to the members, and he has little power to force agreement. Because of its lack of firm leadership and its equalitarianism, the band would have only tenuous solidarity were it not for the emphasis placed upon harmonious relationships and the fact that social relationships are structured so as virtually to exclude competition.

Means of subsistence

The G/wi economy. About **30** plant species provide the basis of G/wi subsistence. None is available the year round, the number available at one time varying from 29 in autumn to only four in early summer—of which only two are sufficiently plentiful to make a significant contribution to the diet. The work required in gathering food for a household therefore varies greatly with the season; the more species there are to be taken, the easier and shorter is the search and the greater the number of people able to subsist on the resources of a given area. It is for this reason that the band separates in winter, when the frosts blight the plants and sharply reduce the amount of plant food available. Dispersed in individual households, each exploiting a particular area, the relative load on the food resources is reduced, as is also the amount of work required to gather enough to eat. Just before the onset of the wet season the variety and amount of food plants increase; the band can again reform and resume its normal life as one community, shifting its encampment every five or six weeks to a new supply of plant food.

The work of gathering is done mainly by the women and girls. They range over a radius of up to five miles, picking fruits and berries and digging the roots and tubers that comprise about **80** percent of the G/wi diet. A three-foot stick sharpened at one end serves as digging tool and carver. An antelope-hide cloak, worn over the shoulders and fastened around the neck, does duty as a blanket at night. The lower points of the cloak are tied about the waist, forming a bag.

Hunting and trapping is the work of men. They use a light bow and flimsy, unfeathered arrows tipped with poison prepared from the pupae of the beetle *Diamphidia simplex.* Arrowheads were formerly of bone, but most are now made of fence wire hammered by stone into precise shape; the poison is applied just behind the cusps of the arrowhead. Careful stalking is necessary if the hunter is to get close enough to the antelope and small desert mammals to place his arrow effectively and inject a dose of poison. The poison requires between six and 96 hours to take effect, depending on the animal, during which time the animal must be tracked and followed; when killed, its meat has to be carried many miles back to the encampment.

The beetle grubs for making poison can be gathered only in summer, and, since by the middle of winter supplies are usually exhausted, recourse must be had to traps. These are snares, loaded by attaching the free end of the line to a bent-over sapling. The noose is held in place by a trigger, which, when stepped upon, releases the sapling and causes the noose to tighten around the leg of the antelope and hold it until the trapper returns, which he does about four times a day. The monthly intake of meat per person varies from about four pounds (two kilograms) in October, the end of winter, to 15 pounds (seven kilograms) in January, when fresh arrow poison is available and the antelope population is increased by the arrival of migrating herds.

The G/wi have drinking water only for about eight weeks of the year. They meet their fluid requirements at other times from the juices of plants and by drinking the blood and rumen liquor (squeezed-out cud) of antelope; they also drink the amniotic fluid of pregnant does.

Technology

Tools are light and multipurpose. The desert does not provide many raw materials, and migrations between campsites impose a need for easy portability. Iron rods obtained from the ranches are beaten out on stone anvils to make the heads and blades of spears, knives, adzes, and axes. Pestles and mortars are adzed out of wood. Bowls and other receptacles are also carved from wood. Tin cans and enamelware are imported and used as cooking pots, replacing the fragile pottery that was formerly obtained from the Boteti (Botletle) River area and has now become rare. Antelope hides, pegged out to dry in the sun, are scraped clean of fat and other waste and then tanned, using a mixture of plant juices, bone marrow, urine, and rotted brains. The leather of the steenbok and duiker is worked into a fine kid having the softness of cloth and is one of the main commodities of exchange with the bands nearer civilization. Other antelope hides are made into cloaks and skirts or sewn to make bags or sandals.

Shelters vary with the season. The dry-season structure is a roofless windbreak of shrubs, branches, and grass. The wet-season hut is of a beehive shape, made from a frame of branches thatched with clumps of grass. When a band migrates, the huts or shelters are abandoned and new ones set up within several hours after arrival at a new campsite.

BIBLIOGRAPHY. The classic reference work is I. SCHAPERA, *The Khoisan Peoples of South Africa* (1951), a review and synthesis of the literature up to 1928 with a full bibliography. J.A. ENGELBRECHT, *The Korana* (1936), is an account of the history and customs of a now virtually extinct Hottentot people. ELIZABETH MARSHALL THOMAS, *The Harmless People* (1959), is a popular account that conveys better than any other book in English the quality of K!ung Bushman life. A series of papers based upon sustained anthropological fieldwork among Bushmen by LORNA MARSHALL were published in various numbers of *Africa:* vol. 29, no. 4 (1959); vol. 30, no. 4 (1960); vol. 31, no. 3 (1961); and vol. 32, no. 3 (1962). GEORGE B. SILBERBAUER, *Report to the Government of Bechuanaland on the Bushman Survey* (1965), describes a six-year survey of the Bushmen of Botswana. RICHARD B. LEE, "What Hunters Do for a Living, or How to Make Out on Scarce Resources," in R.B. LEE and I. DEVORE (eds.), *Man the Hunter* (1968); and "!Kung Bushman Subsistence: An Input-Output Analysis," in D. DAMAS (ed.), *Contributions to Anthropology: Ecological Essays* (1969), analyze the ways in which one group of Bushmen adapt to their environment.

(G.Si.)

Khomyakov, Aleksey Stepanovich

Khomyakov was the founder of the 19th-century Slavophile movement, which extolled the superiority of the Slav or Russian way of life, as well as-a poet and influential lay theologian of the Russian Orthodox Church. His importance lies in that—at the time of the triumphant march of Western civilization—he recognized the West's serious defects and suggested radical changes.

Novosti—Sovfoto

Khomyakov, engraving by Ivan Pozhalostin, 1879.

Early life and literary career

Born on May 13 (May 1, old style), 1804, in Moscow, Khomyakov came from a family that had for many generations served the Russian tsars. His father, Stepan, lacking practical ability, brought the family's financial'affairs to the brink of ruin; his mother, Maria Kireyevskaya, however, saved the situation because of her combined business sense and intelligence. Deeply devoted to the Orthodox Church, she brought up her two sons and her daughters in the spirit of traditional Russian piety and enabled them to receive an excellent education. Aleksey had among his private tutors a French refugee priest, a Greek scholar, and some well-known Russian professors. He acquired a mastery of French, German, English, Greek, and Latin; he also learned Sanskrit and compiled a Russian–Sanskrit dictionary. Though he did not enroll as a student, Aleksey passed his final examinations in mathematics at Moscow University. A visit to France for 18 months completed his education.

During the Russo-Turkish War (1828–29) he served with distinction and spent the remainder of his life in Moscow involved to a great extent in intellectual pursuits, though frequently visiting his family estates of Bugucharovo and Lipitsy. He was reputed to be happily married and had several children.

Among his many accomplishments, Khomyakov was known as a gifted writer and a brilliant controversialist; he dealt with a wide range of subjects, composed poetry, wrote philosophical and political essays and treatises on economics, sociology, and theology. A successful landlord, he won prizes in England for the construction of agricultural machinery. Being a remarkable marksman he made important improvements in the design of guns. He was also a self-taught doctor and treated many of the peasants on his estates.

A rigid censorship in tsarist Russia permitted only some of his articles to appear in print during his lifetime. His major theological and historical works were thus published posthumously by his friends and associates. Even today his writings are proscribed in the U.S.S.R.

Khomyakov belonged to what has been called the Golden Age of Russian literature. At that time (first half of the 19th century) an intellectual elite of exceptional ability created a striking contrast to the proponents of the reactionary rule of Nicholas I (1825–55). The best Russian thinkers were preoccupied with the problem of the political and social orientation of their people. Russia had been brought out of its previous isolation by Peter the Great's reforms (1682–1725), and contacts with the West had stimulated the upper classes, but under Nicholas I they felt frustrated and dissatisfied. Two main groups appeared: the Westernizers and the Slavophiles. The Westernizers saw in the political institutions of the West and in liberal and socialist ideas a pattern for imitation. The Slavophiles, led by Khomyakov, insisted that Russia should follow its own path of development based on the pre-Petrine (prior to Peter the Great) culture inspired by the Eastern Orthodox Church.

Slavophile views

Though Khomyakov felt at home in the Western world, he also knew and loved Russia's past, a sentiment rare among his upper class contemporaries. His aim was to combine the best elements of both traditions, but the social system he advocated was opposed to the individualism of the West. He did not believe that secularized and selfish man, denying the existence of a divine creator, could establish a satisfactory political and social order. Criticizing both capitalism and socialism, he regarded them as coming from the same Western outlook. He believed that the root of man's problems lay in what he viewed as the defects of Western interpretations of Christianity, and he blamed Roman Catholicism and Protestantism equally for their failure to solve the problem of the relationship between authority and freedom. "Rome kept unity at the expense of freedom. Protestants had freedom but lost unity." He was convinced that the Orthodox Church possessed a more balanced presentation of Christian teaching than any of the Western churches.

In Khomyakov's system the key term was ***sobornost*** (togetherness-symphony). This word, in the Slavonic version of the Nicene Creed, corresponds to "catholic." It does not, however, mean "universal" but denotes a perfect organic fellowship of redeemed people united by faith and love. Khomyakov believed that a person could best achieve spiritual and intellectual maturity in an organic community that respected the liberty of its members and that true progress would depend not on competition (as in the West) but on cooperation. He thus claimed that the task of the church was to teach mankind how to live in unity and freedom. In Khomyakov's estimation, the Christian West, after its separation from the East, had not been able to fulfill this role.

In another aspect of his Slavophile thought, Khomyakov idealized the Russian peasants, exalted their humility and sense of brotherhood, and regarded them as better fitted to realize a Christian social order than the more aggressive Western nations.

Khomyakov's vast erudition, his literary gifts, his integrity, and strength of convictions should have procured for him a distinguished political and academic career. But he lived under the oppressive rule of Nicholas I and had no opportunity of using his talents for the public good. He remained until his death a mere retired cavalry captain. He died on October 5 (September 23, O.S.), 1860, at the age of 56, from cholera, which he caught from the peasants whom he treated.

BIBLIOGRAPHY. Khomyakov's works were published in Moscow in 1862. Enlarged editions in 8 volumes appeared in 1882 and 1900. Only a few of his writings, such as ***The Orthodox Doctrine of the Church*** (1864), ***L'Église Latine et le Protestantisme . . .*** (1872), and ***The Church Is One*** (1948, 1968), have been translated.

Secondary Sources: W.J. BIRKBECK (ed.), ***Russia and the English Church During the Last Fifty Years*** (1895), includes Khomyakov's correspondence with William Palmer dealing with the reunion between the orthodox Anglican churches; P. CHRISTOFF, ***A.S. Xomjakov*** (1961), contains full bibliography; A. GRATIEUX, ***A.S. Khomiakov et le mouvement slavophile,*** 2 vol. (1939), the best study of Khomyakov in the west; N. RIAZANOVSKY, ***Russia and the West in the Teaching of the Slavophiles*** (1952); N.M. ZERNOV, ***Three Russian Prophets: Khomiakov, Dostoevsky, Soloviev*** (1944), traces the common background of three of the greatest Russian Christian thinkers of the 19th century.

(N.Z.)

Khosrow I of Persia

Khosrow I of Persia, also called Chosroes in classical sources, ruled the **Sāsānian** Empire from 531 to 579. He is known in Persian sources as **Anūshīrvan**, or Of the Immortal Soul, since he is remembered as a great reformer and possessed of a noble character. He was also given the surname the Just. Because of his fame the Arabs later called all rulers of the **Sāsānian** Empire by his name, in the Arabic form **Kisrā.**

Khosrow I, crystal medallion, 6th century. In the Bibliothèque Nationale. Paris.
J.E. Bulloz

Little is known of the early life of Khosrow beyond legends. One story says that when Khosrow's father, King Kavadh, took refuge with the Hephthalites, eastern neighbours of Iran, on the way (near the town of Nishapur) he married a peasant's daughter, who gave birth to Khosrow. At his father's death Khosrow did not at first succeed him, but in a struggle for the throne he was successful and put to death his brothers. At the end of his father's reign, great social disorders had occurred because of a religious revolution of a sect called the Mazdakites. Khosrow first restored order and then launched reforms to transform the declining **Sāsānian** Empire.

Reforms

The reform of taxation was the most important of his actions, and it was probably copied from the Roman system inaugurated by Diocletian. Previously in the **Sāsānian** Empire taxes had been levied on the yield of land; Khosrow established a fixed sum rather than a yearly variation. Other taxes were introduced that brought stability to the income of the state and were also fairer to those who paid the taxes. Khosrow's program of taxes lasted into Isltimic times.

Khosrow also reorganized the **Sāsānian** bureaucracy, and the system of ministries, or divans, under a prime minister is said to have been initiated by him. He was fortunate during most of his reign in having a capable prime minister called Bozorgmehr, who became famous in story and legend for his wisdom and abilities. Khosrow also reorganized the army and appointed four chief commanders to guard the four frontiers of Iran. On the frontier against the Byzantines and their Arab allies in the Syrian Desert, against the peoples of the steppes of south Russia at the town of Derbent between the Caucasus Mountains and the Caspian Sea, and to the east of the sea in the present Turkmen steppes, Khosrow built defensive walls. The army, however, did not remain on the defensive, for Khosrow's reign is noted for his wars against the Byzantines. In 540 the great city of Antioch was conquered and held for a short time by Khosrow. He brought great numbers of prisoners from Antioch and settled them in a new town near his capital of Ctesiphon, built after the model of old Antioch. In the east, Khosrow, in alliance with the Turks, a new power in Central Asia, crushed the Hephthalites, and he established a hegemony over many of their principalities. The **Sāsānian** frontier in the east reached the Oxus River during his reign.

He also fought extensively in Armenia and Lazica in the Caucasus. Furthermore, under his reign a **Sāsānian** army conquered Yemen. Khosrow relied on a professional army more than his predecessors, who had summoned feudal levies when they set out to war.

Under Khosrow the process of decentralization of the power of the monarch was reversed, and the lower aristocracy, or knights, called dihqdns, grew in importance at the expense of the great feudal lords, who had been more powerful under Khosrow's predecessors. It is difficult to know how many changes really can be attributed to Khosrow's reign and how many are arbitrarily assigned to him because of his place in history. Whether the religiously sanctioned division of society into priests, warriors, bureaucracy, and common folk was codified under his reign, as claimed by some sources, is difficult to determine.

Patron of culture

Khosrow was also a great patron of culture, and in 529, when the ancient academy of Athens was closed, a number of Greek philosophers migrated to the **Sāsānian** Empire, where they were well received by the **ruler.** The later famous medical school of **Gondēshāpūr** was probably started in Khosrow's reign, and the famous physician Burzoe is supposed to have been sent to India by Khosrow to gather Sanskrit books of learning to be translated into the Middle Persian language. The game of chess reportedly was also brought by him from India. Astronomy and astrology flourished at the court of Khosrow, and one star table (called the *zīj-i* Shahriydr), which was the basis of many later **Islāmic** tables, is said to have originated during the reign of Khosrow. Several works of Middle Persian, such as the Book of Deeds of Ardashir (*Kārnā*mak), are attributed to this period. Likewise, some scholars claim that the codification of the Avesta, the sacred book of the Zoroastrian religion, as well as the creation of a special Avestan alphabet to record the text, took place at the order of Khosrow. Further, it is supposed that the stories and legends of ancient Iran were gathered into a *Khwatāy-nāmak* ("Book of Kings") in the time of Khosrow and thus provided the source for Ferdowsi's immortal epic much later. Some of the names found in Ferdowsi's *Shāh-nāmeh* appear among the royal family of Khosrow, which indicates at the least an interest on the part of the monarch in ancient legends.

Perhaps more than fact, the stories told about Khosrow have made his name famous in history. Almost any **pre-Islāmic** structure in Iran the origin of which is unknown will be attributed to Khosrow by the simple folk. Undoubtedly he built many bridges, roads, and palaces, but much more is assigned to him in legend. The famous palace with the huge arch, called **Ṭāq** Kisrā, in Ctesiphon, near modern Baghdad, is said to date from Khosrow I, but this is uncertain. Several collections of wise sayings of this monarch, as well as stories about him, have been preserved in Arabic or New Persian versions. In them his reputation for justice as well as wisdom is constantly cited. The splendour of the court and the glory of his reign provided models for the later **ʿAbbāsid** court in Baghdad, and many of the institutions established by Khosrow were maintained in **Islāmic** times, when Khosrow was hailed as the model **pre-Islāmic** ruler to be emulated by Muslim princes.

BIBLIOGRAPHY. The principal study on Khosrow is in ARTHUR E. **CHRISTENSEN,** *L'Iran sous les* Sassanides, 2nd **ed.** rev. (1944).

(R.N.F.)

Khosrow II of Persia

Khosrow **II Parvīz** (the Victorious) was a late **Sāsānian** king of Persia (Iran). His glorious court and the bitter contrast of his failure and death so impressed themselves on the poetic tradition of that country that it is impossible to be certain about his personal history or the social conditions during his reign. Under him the **Sāsānian** Empire achieved its greatest expansion, while its art and culture flourished as never before.

Expansion of the empire. Son of Hormizd IV, Khosrow was proclaimed king in AD 590 in turbulent times. Hormizd's general, **Bahrām** Chiibin, after defeat by the Byzantine army at Lazica, had been openly insulted by the King. During a subsequent palace revolt led by **Bostām** and **Bindōē** (brothers-in-law of **Hormizd**), which culminated in the King's assassination, **Bahrām Chūbīn** renounced the allegiance of his army to the monarchy and forced the new **king** Khosrow to flee to **Mesopotamia.** Khosrow's pursuers were held off by the military tactics of his uncle **Bindōē,** until eventually the Byzantine emperor Maurice provided Khosrow with forces to defeat his adversary. **Bahrām** Chiibin was subsequently assassinated.

Khosrow II, coin, 590–628 AD. In the collection of the American Numismatic Society.
By courtesy of the American Numismatic Society

Insecure and unpopular, Khosrow now eliminated those connected with his father's murder, including **Bindōē,** on whose support he had relied. Although he retained a bodyguard of Byzantine legionaries, he resented the Byzantine presence in Armenia, which he had been forced to cede.

Invasion of the Byzantine Empire

Using the murder of Maurice (602) and his replacement as emperor by Phocas as a pretext and encouraged by the fact that Narces, who had commanded the Byzantine force that established Khosrow on the throne, refused to recognize Phocas, Khosrow's armies invaded Armenia and Mesopotamia. The Byzantine forces in Mesopotamia were weak, and the towns of Dara, **Amida,** and **Edessa** soon fell (604). Crossing the Euphrates, Khosrow took Hierapolis and Beroea (Aleppo). Internal dissensions made the eastern Byzantine provinces an easy prey, and Armenia and central Asia Minor were overrun by the Persians—though apparently not permanently occupied or administered. Nor was the Persian advance checked when Heraclius became emperor in 610 and sued for peace.

A second invasion of Mesopotamia, by Khosrow's ablest general, **Shahrbarāz,** took place in 613. Damascus was taken in that year, and in 614 Jerusalem fell. The Holy Sepulchre was destroyed and the true cross carried to Ctesivhon. Although Khosrow himself was generally tolerant of **Christianity, Shahrbarāz** permitted **thousands** of Christian prisoners to be tortured by his Jewish aides. In 616 **Alexandria** was captured, and **in** 617 Chalcedon (opposite Byzantium), which had long been under siege by another of Khosrow's generals, **Shāhīn,** finally fell to the Persians.

This tide of conquest was turned by Heraclius in a series of brilliant campaigns between 622 and 627. Since he retained command of the sea, Heraclius was able to sail to Issus and rout the Persian army near the Armenian border. In alliance with the **Khazar kingdom** north of the Caucasus, he invaded Armenia again in 623, gaining victory over the King's army near Canzaca (Takht-e **Sulaimān,** Shiz). The town and fire temple were destroyed, together with the temple at Lake **Urmia,** traditionally associated with Zoroaster. The campaigns of 624 and 625 **ranged** across northern Syria and Mesopotamia and culminated in a reversal for **Shahrbarāz'** forces on the river Saras.

Khosrow rallied his forces in 626 and, in alliance with the Avars, a people who were also in conflict with **Byzantium** at this time, sent one army to besiege **Constantinople** and another to oppose Heraclius. Constantinople held, and **Shāhīn** was defeated; the Persian second force was outmanoeuvred in 628 by Heraclius' brave dash to Dastagird, the royal residence 70 miles north of **Ctesiphon.** An important but indecisive battle was fought near Nineveh, but, as the Byzantine army reapproached **Dastagird,** Khosrow fled. His letters calling **Shahrbarāz** to his aid had been intercepted, and, although his resources were by now drastically reduced, he refused peace terms.

Khosrow's execution

Khosrow's prestige was shattered, and he was now sick. The execution of Shahrbariiz and the desecration of Shāhīn's corpse were followed by revolution in the royal household. Khosrow was condemned to death and executed (628), and his youngest son and heir, Mardānshāh, was murdered before his eyes. His eldest son, Qobād, II Shērōē, signed the peace.

Cultural and economic influence. Khosrow was a serious patron of the arts; silver working and carpet weaving reached their peak during his reign. Sources tell of the enormous "Spring of Khosrow," a carpet designed like a garden. A splendid silver dish in the Bibliothbque Nationale is thought to depict him in the traditional Sāsānian royal hunt. Most authorities attribute to Khosrow II the grottos at Taq-e Bostan (Kermanshah), taking them as evidence of a renaissance of rock sculpture in his reign. The reliefs depict the king in hunting scenes and standing motionless listening to a group of harpists—a reminder of the famous musicians Biirbad and Sarkash, who were kept at Khosrow's court. His architectural work is chiefly known from the ruins of the enormous palace Imirat-e Khosrow near Qasr-e Shirin (near Khānaqīn) and at nearby Hawsk-Kuri. A provincial palace exists at Qaṣr al-Mushattii in Transjordan.

Booty and taxes brought Khosrow enormous wealth, including thousands of elephants, camels, horses, and women. The 9th-century Arab historian aṭ-Ṭabarī describes his golden throne supported by legs of rubies, as well as such curios as a piece of malleable gold and an asbestos napkin. But despite widespread trade connections and the amassing of individual fortunes, there is no evidence that the economy flourished. High taxation and the uncertainties of war did nothing for the merchant class. By creating a military aristocracy, Khosrow II had weakened the authority of the king, while his administrative reforms and bureaucratic centralization removed the power of regional dynasties and their feudal armies, which might have resisted the invasion of the Arabs 12 years after Khosrow's death. Already in 611 the Arabs had inflicted a defeat on the Sāsānian army at Dhu-Qar. The destruction by Khosrow II of the Christian Arab states of the Lakhmids and Ghassānids in Syria and western Iraq was a further factor exposing Iran to Arab attack.

The love of Khosrow for his Armenian Christian wife Shirin was celebrated by the poets, especially by the 12th-century poet Neẓāmī in *Khosrow-va-Shirin.*

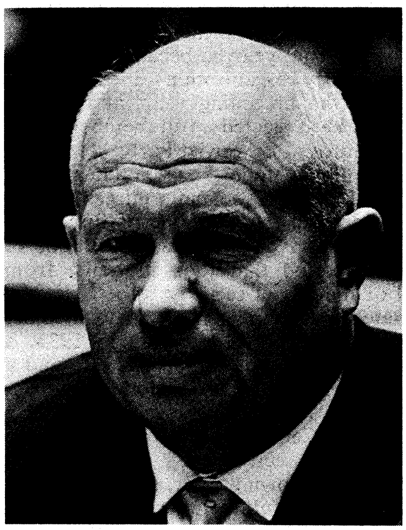
BIBLIOGRAPHY. The fullest account in English is in SIR PERCY SYKES, *A History of Persia,* 3rd ed., vol. 1 (1969). The two basic works dealing with the Arabic literary sources are T. NOELDEKE, *Geschichte der Perser und Araber zur Zeit der Sasaniden: Aus der Chronik des Tabari . . .* (1879); and A.E. CHRISTENSEN, *L'Iran sous les Sassanides* (1936), particularly good on the use of Byzantine sources. The national poet FERDOWSI in the *Shah-Nameh,* vol. 8-9 (Eng. trans. 1908), deals in detail with the war with Bahrām Chūbīn, treasures, and buildings. For literary references to Khosrow Parvīz, see E.G. BROWNE, *A Literary History of Persia,* vol. 1 (1928–29). An authoritative recent discussion of social conditions will be found in R.N. FRYE, *The Heritage of Persia* (1963).

(W.Cu)

Khrushchev, Nikita Sergeyevich

For the Soviet Union and indked for the entire world Communist movement, Nikita Khrushchev was the great catalyzer of political and social change. In his seven years of power as first secretary and premier, he broke both the fact and the tradition of the Stalin dictatorship and established a basis for liberalizing tendencies within Soviet Communism. Khrushchev was a thoroughgoing political pragmatist who had learned his Marxism by rote, but he never hesitated to adapt his beliefs to the political urgencies of the moment. His experience with international realities confirmed him in his doctrine of peaceful coexistence with the non-Communist world—in itself a drastic break with established Soviet Communist teaching. He publicly recognized the limitations as well as the power of nuclear weapons, and his decision to negotiate with the United States for some form of nuclear-testing control was of vast importance. At the same time, Khrushchev's rough empathy with the Soviet people resulted in concessions to a consumer economy and in a general relaxation of security controls, which had equally far-reaching effects. Despite his repression of the Hungarian uprising in 1956, his acceptance of "different roads to Socialism" led to growing independence among European Communist parties, but his Russian nationalism and his suspicion of Mao Tse-tung's Communism helped create an unexpectedly deep fissure between China and the Soviet Union that changed the balance of world politics. By the time he was removed from office in 1964, he had set up certain guidelines for and limitations to Soviet policy making that his successors were hard put to alter.

Werner Wolf—Black Star



Khrushchev, 1960.

Early life. Khrushchev was born on April 17, 1894, at Kalinovka in the Russian province of Kursk, close to the Ukrainian border. Unlike Stalin, Lenin, and most other Soviet leaders, who had generally middle class backgrounds, Khrushchev's father was a miner; his grandfather had been a serf who served in the tsarist army. After a village education, Khrushchev went with his family to Yuzovka (later named Stalino, now Donetsk), a mining and industrial centre in the Donets Basin, where he began work as a pipe fitter at the age of 15. Because of his factory employment, he was not conscripted in the tsarist army during World War I. Even before the Russian Revolution of 1917, he had become active in workers' organizations, and in 1918, during the struggle between Reds, Whites, and Ukrainian nationalists for possession of the Ukraine, he became a member of the Russian Communist Party (Bolsheviks).

Membership in the Communist Party

In January 1919 he joined the Red Army and served as a junior. political commissar, ultimately in the campaigns against the Whites and invading Polish armies in 1920. After he was demobilized in 1921, he returned to work in Yuzovka. His first wife, Galina, died during a famine that year, leaving him with two children.

At Yuzovka in 1922, Khrushchev secured admission to one of the new Soviet worker's schools, where he received a secondary education along with additional party instruction. He quickly became a student political leader and was appointed secretary of the Communist Party Committee at the school. At Yuzovka, also, he married his second wife, Nina Petrovna, a schoolteacher, in 1924. She bore him three children.

Political career under Stalin. In 1925 Khrushchev went into full-time party work, as party secretary of the Petrovsko-Mariinsk district of Yuzovka. He distinguished himself by his hard work as well as his down-to-earth knowledge of mine and factory conditions. His work soon came to the notice of Stalin's close associate, Lazar M. Kaganovich, secretary general of the Ukrainian Party

Central Committee. He accompanied Kaganovich as a nonvoting delegate to the 14th Party Congress in Moscow in December 1925, where Stalin celebrated his final victory over Trotsky.

For the next four years—in Yuzovka, then in Kharkov and Kiev—Khrushchev was active as a party organizer. In 1929 he received permission to go to Moscow to study metallurgy at the Stalin Industrial Academy. There, too, he made a name for himself through intense political activity and was appointed secretary of the academy's Party Committee. In 1931 he went back to full-time party work in Moscow. By 1933 he had become second secretary of the Moscow Regional Committee, again directly under Kaganovich.

During the early 1930s Khrushchev consolidated his hold on the Moscow party cadres that would ultimately secure his ascendancy. He personally supervised the completion of the Moscow subway, which became a showplace of the Soviet Union and for which he received the Order of Lenin in 1935. That year he became first secretary of the Moscow party organization—in effect, the mayor of Moscow. In the preceding year, at the 17th Party Congress, he had been elected a full member of the 70-man Central Committee of the Soviet Party.

Membership in the Central Committee

Khrushchev was a zealous supporter of Stalin in those years and participated in the sweeping purges of the Communist Party leadership. He was one of three provincial secretaries who survived the mass executions of the *Yezhovshchina*, a period that took its name from Yeshov, then head of the Soviet security forces. He was promoted appropriately, becoming an alternate member of the federal legislature's ruling Presidium in 1935, a member of the Constitutional Committee in 1936, and a member of the Foreign Affairs Commission of the Supreme Soviet in 1937. In 1938 Khrushchev was made a candidate member of the Central Committee's Politburo and sent to Kiev as first secretary of the Ukrainian party organization. The following year, on the eve of World War II, he was made a full member of the Politburo.

In 1940, after Soviet forces had occupied eastern Poland, Khrushchev presided over the "integration" of the former Polish areas into the Soviet Union. Throughout the Ukraine his principal objective was to liquidate both Polish and Ukrainian separatist movements, as well as to restore the Communist Party organization, which had been shattered in the purge. This work was disrupted by the German invasion of the Soviet Union in June 1941.

Khrushchev's first wartime task was to evacuate as much of the Ukraine's industry as possible to the east. Thereafter, he was attached to the Soviet Army and received the rank of lieutenant general; his principal task was to stimulate the resistance of the civilian population and maintain liaison with Stalin and other members of the Politburo in Moscow. He was political adviser to Marshal Andrey I. Yeryomenko during the defense of Stalingrad (now Volgograd), and he later participated in the huge tank battles at Kursk in 1943 as political adviser to Lieut. Gen. Nikolay F. Vatutin.

After the liberation of the Ukraine in 1944, Khrushchev worked to restore the civil administration and to bring that devastated country back to a subsistence level. A famine in 1946 was probably the worst in the Ukraine's [illegible] Khrushchev fought to restore g[illegible] p[illegible] and to distribute f[illegible] supplies, [illegible] St[illegible] on greater production from the Ukraine for use in other areas. During this period Khrushchev made firsthand acquaintance with the problems of Soviet agricultural scarcity and planning, with which he was to grapple throughout his political life.

Temporary demotion

Largely because of his failure to achieve prompt collectivization of Ukrainian agriculture, Khrushchev was demoted in 1947 and replaced by his former boss Kaganovich; but he was later restored to favour. In 1949 Stalin called him back to Moscow, where he took over his old job as head of the Moscow City Party and concurrently was appointed secretary of the All-Union Central Committee. In Moscow, over the next four years, Khrushchev continued to consolidate his power.

The period 1949–53 was far from pleasant for Khrushchev and other members of the Soviet leadership, who found themselves pawns in Stalin's paranoiac palace politics. "All of us around Stalin," Khrushchev said later, "were temporary people." Khrushchev moved more and more into agriculture, where he began his schemes for the *agrogorod* ("farming town") and larger state farms at the expense of the conventional collectives. On this and other issues he clashed with Georgy Malenkov, Stalin's heir apparent. After Stalin died, in March 1953, Khrushchev cooperated with Malenkov and other members of the Politburo in the arrest and execution of the deputy prime minister and state security chief, Lavrenty Beria.

Leadership of the Soviet Union. Later Khrushchev engaged in a power struggle with Malenkov and gained the decisive margin because of his control of the party machinery. In September 1953 he replaced Malenkov as first secretary and in 1955 was able to remove Malenkov from the premiership and replace him with his handpicked nominee, Marshal Nikolay A. Bulganin. Through this period Khrushchev adopted a conservative party position, attacking Malenkov for not developing heavy industry and for neglecting party organization.

Personal diplomacy

In May 1955, however, when Khrushchev made his first trip outside the Soviet Union—to Yugoslavia with Bulganin—he began to show greater flexibility; he apologized to Tito for Stalin's denunciation of Yugoslav Communism in 1948. Later, in trips with Bulganin to Geneva, Afghanistan, and India, he began to exhibit a brash, extroverted personal diplomacy that was to become his trademark. Although his attacks on world capitalism were virulent and primitive, his outgoing personality and peasant humour were in sharp contrast to the picture of reserved hostility toward non-Socialist countries that earlier Soviet public figures had cultivated.

Destalinization. On February 24–25, 1956, during the 20th Party Congress in Moscow, Khrushchev delivered his memorable secret speech about the excesses of Stalin's one-man rule, attacking the late Soviet ruler's "intolerance, his brutality, his abuse of power." It shook the Communist world to its foundations. The spectacle of the First Secretary of the Communist Party exposing the wrongful executions of the Great Purge of the 1930s and the excesses of Soviet police repression, after years of fearful silence, had far-reaching effects that Khrushchev himself could barely have foreseen. The resulting "thaw" in the U.S.S.R. saw the release of thousands of political prisoners and the "rehabilitation" of many thousands more who had perished. Although some argue that Khrushchev had been forced into his denunciations by party pressures, he will go down in history as the man who ultimately destroyed not only the "cult of personality" but also the idea, fostered since the Revolution, of Soviet Communism as an infallible monolithic authority.

Rebellions in eastern Europe. The destalinization movement had repercussions in the Communist countries of eastern Europe. Poland revolted against its Stalinist government in October 1956. Hungary followed shortly afterward. Faced with open revolution, Khrushchev flew to Warsaw on October 19 with other Soviet leaders and first tried to face down, then temporized with the new Polish national Communist leadership. He ultimately acquiesced in the Polish leader Władysław Gomułka's national Communist solution, which allowed the Poles a great deal of freedom. Khrushchev's behaviour in dealing with the Hungarian Revolution was quite different; his shared decision to crush the Hungarian Revolution by force, after much hesitation, came largely because of the Hungarian premier Imre Nagy's decision to withdraw from the Warsaw Pact. With this sanguinary exception, however, Khrushchev allowed a considerable amount of freedom to the European Communist parties. Indeed, his slogan of "many roads to Socialism" produced a strong revival in various Communist parties, whose leaders found themselves permitted by Moscow to adopt more nationalist positions.

Opposition at home. The stresses in eastern Europe helped crystallize opposition to Khrushchev within the Soviet Party. In June 1957, he was almost overthrown from his position of primacy by a coalition led by Malenkov, representing the managerial class, and by Vyacheslav M. Molotov and Kaganovich, leading the conservative pro-Stalinist bloc. Although a vote in the Presidium actually went against him, Khrushchev managed to reverse this by appealing to the full membership of the party Central Committee. In the end he secured the permanent disgrace of Malenkov, Molotov, and others, who were labelled members of the antiparty group. In October he dismissed Marshal Georgy Zhukov from his post as minister of defense, and in March 1958 he assumed the premiership of the Soviet Union himself, as sign and seal of his victory.

Assumption of the premiership

Relations with the United States and China. Confirmed in power, Khrushchev widely asserted his doctrine of peaceful coexistence, which he had first enunciated in a public speech at the 20th Party Congress. In opposition to old Communist writ, he stated that "war is not fatalistically inevitable." At the 21st Party Congress in 1959 he said: "We offer the capitalist countries peaceful competition." His visit to the United States in 1959, where he toured cities and farms with the ebullience of a politician running for office, was a decided success, and the "spirit of Camp David," in Maryland, where he conferred with Pres. Dwight D. Eisenhower, brought Soviet–American relations to a new high. A long-planned summit conference with Eisenhower in Paris in 1960 broke up, however, with Khrushchev's announcement that a U.S. plane (a U-2 reconnaissance aircraft) had been shot down over the Soviet Union and its pilot captured. In 1961 his Vienna conference with the new U.S. president, John F. Kennedy, led to no agreement on the pressing German question; the Berlin Wall was built shortly thereafter.

Soviet success in lofting the world's first space satellite in 1957 had been followed by increased missile buildups. In 1962 Khrushchev attempted to emplace Soviet medium-range missiles in Cuba. During a famous confrontation in October 1962, when the U.S. and the U.S.S.R. apparently stood on the brink of war, Khrushchev agreed to remove the missiles on the promise that the United States would make no further attempts at overthrowing Cuba's Communist government. The U.S.S.R. and Khrushchev in particular were bitterly attacked by the Chinese Communists for this settlement, and relationships between the two great Communist countries, already strained, grew worse. The Sino-Soviet split, which began in 1959 with Khrushchev's second trip to Peking, reached the stage of public denunciations at international party meetings in 1960. China's ideological insistence on all-out "war against the imperalists" and Mao Tse-tung's annoyance with Khrushchev's coexistence policies were exacerbated by Soviet refusal to assist the Chinese nuclear weapon buildup and to rectify the Russo-Chinese border. The Soviet Nuclear Test-Ban Treaty with the United States in 1963, although generally welcomed throughout the world, served only to intensify Chinese denunciations of Soviet "revisionism."

Internal pressures. During Khrushchev's entire time in office, he had to rudder constantly between opposed tendencies. There were continuing popular pressures toward a consumer-oriented society and agitation by intellectuals for greater freedom of expression; these were offset by the growing fear of the Soviet bureaucracy that destalinization reforms would get out of hand. Khrushchev himself was uneasy with intellectuals, and he sanctioned the repression of Boris Pasternak's ***Doctor Zhivago*** (1957) within the Soviet Union, culminating in the Soviet government's refusal to allow Pasternak to accept the Nobel Prize in 1958. On the other hand, Khrushchev permitted the 1962 publication of Aleksandr Solzhenitsyn's ***One Day in the Life of Ivan Denisovich,*** with its sweeping denunciation of Stalinist repression. Meanwhile, for the first time, Soviet tourists were permitted to go overseas, and Khrushchev often seemed amenable to widening exchanges with both Socialist and capitalist countries.

Khrushchev's desire to reduce conventional armaments in favour of nuclear missiles—partly a result of budgetary problems—was bitterly resisted by the Soviet military. In addition, his freewheeling, often high-handed methods of leadership and his attempted decentralization of the party structure antagonized many in the very cadres that had supported his rise to power. But the central crisis of Khrushchev's administration was agriculture. Perennially an optimist, he based many plans on the bumper crops in 1956 and 1958, which fuelled his repeated promises to overtake the United States in agricultural as well as in industrial production. ("If we catch up with the United States in per capita production of meat, butter and milk, we will have hit the pillar of capitalism with the most powerful torpedo yet seen.") He opened up more than 70,000,000 acres of virgin land in Siberia and sent thousands of young labourers from Russian cities to till them; but his plan was unsuccessful, and the Soviet Union soon again had to import wheat from Canada and the United States to keep up its own precarious balance of grain imports and exports.

Khrushchev's central crisis

Fall from power. The failures in agriculture and the China quarrel, added to his arbitrary administrative methods, were the major factors in Khrushchev's downfall. A Soviet critic wrote, "Khrushchev suffered from an incurable organizing itch, an inability to leave well enough alone." By October 1964 opponents were strong enough to bring him down. On October 14 the Central Committee accepted Khrushchev's request to retire from his position as the party's first secretary and chairman of the Council of Ministers of the U.S.S.R. because of "advanced age and poor health."

Last years. The succession passed without any violence. For almost seven years thereafter, Khrushchev lived quietly in Moscow and at his country dacha outside the city, as a special pensioner of the Soviet government. He was mentioned in the Soviet press only occasionally and in a derogatory manner; he appeared in public only to vote in Soviet elections. The one break in this ordered obscurity came in 1970 with the unprecedented publication of his memoirs in the United States and Europe, although not in the Soviet Union. He denied granting permission to print them, but most foreign observers felt them to be basically genuine in both style and content.

Publication of memoirs

On September 11, 1971, Khrushchev died of a heart attack. His wife, one son, and three daughters survived him. Almost 48 hours elapsed before his death was announced to the Soviet public. He was denied a state funeral and the concomitant interment in the Kremlin wall, although he was allowed a quiet burial at Novodevichye Monastery Cemetery in Moscow.

Appraisal. The cautious handling of his death announcement reflected his increasing popularity in his last years, both in the U.S.S.R. and the outside world, as many contrasted his consistent, if occasionally stormy, peaceful-coexistence diplomacy with the more restricted and conservative policy of his successors. At the time of his death it was widely felt that the basic changes in Soviet life made under his regime would be hard to uproot and might indeed result in ultimate changes in the pattern of Soviet society and world power relationships. Whatever the view of his personal eccentricities, his boisterousness, his vulgarity, and his bewildering shifts of position, he was accounted a man of stature. As his son Sergey pronounced in a short eulogy at the cemetery, "There were those who loved him, there were those who hated him, but there were few who would pass him by without looking in his direction."

BIBLIOGRAPHY. EDWARD CRANKSHAW, ***Khrushchev: A Career*** (1966), is the best all-around biography; while MARK FRANKLAND, ***Khrushchev*** (1966); and LAZAR PISTRAK, ***The Grand Tactician*** (1961), are also useful. EDWARD CRANKSHAW, ***Khrushchev Remembers*** (1970), is the controversial memoir, generally accepted as genuine, with invaluable notes by Crankshaw. MYRON RUSH, ***The Rise of Khrushchev*** (1958), concentrates on the political phenomena of Khrushchev's assumption of power; as do WOLFGANG LEONHARD, ***Nikita Sergueievitch Khrushchev: Ascencion et Chute d'un homme d'État sovietique*** (1965); HOWARD R. SWEARER, ***The Politics of Succession in the U.S.S.R.*** (1964); BORIS NICOLAEVSKY, ***Power and the Soviet Elite*** (1965); and BERTRAM WOLFE, ***Khrushchev and Stalin's Ghost*** (1957), an excellent treatment of the secret speech and the destalinization movement. ZBIGNIEW BRZEZINSKI, ***The Soviet Bloc: Unity and Conflict*** (1960); and

FRANK GIBNEY, *The Khrushchev Pattern* (1960), are concerned with the impact of Khrushchev's policy outside the Soviet Union and the Moscow–Peking Axis. PRISCILLA JOHNSON and LEOPOLD LABEDZ (eds.), *Khrushchev and the Arts* (1965), is especially interesting for its insight into Khrushchev's relations with the Soviet intelligentsia. For a more general background on Khrushchev and his milieu, see LEONARD B. SHAPIRO, *The Communist Party of the Soviet Union* (1963); and MILOVAN DJILAS, *Conversations with Stalin* (Eng. trans. 1962).

(F.B.G.)

Kiangsi

Kiangsi (Jiang-xi in Pin-yin romanization) is one of the 21 provinces of the People's Republic of China. Occupying a south central location, it is bounded by the provinces of Hupeh and Anhwei on the north, Chekiang and Fukien on the east, Kwangtung on the south, and Hunan on the west. On the map its shape resembles an inverted pear. Situated on the south bank of the Yangtze River (Ch'ang Chiang), the port of Chiu-chiang, 430 miles upstream from Shanghai and 135 miles downstream from Hankow (Han-k'ow), is the province's principal outlet on that great river.

The area of the province is 63,600 square miles (164,700 square kilometres), and its population in 1953 was almost 17,000,000; an estimate from the early 1970s was 25,000,000. The provincial capital is Nan-ch'ang.

The name Kiangsi means West of the (Yangtze) River, although the entire province lies south of it; this seeming paradox is caused by changes made in administrative divisions throughout China's history.

Kiangsi's strategic location

Lying in the midst of a longitudinal depression between China's western highlands and the coastal ranges of Fukien Province, Kiangsi constitutes a corridor linking the province of Kwangtung, in the south, with the province of Anhwei and the Grand Canal, in the north. Throughout China's long history, Kiangsi has played a pivotal role in national affairs because of its position astride the main route of armies, commerce, and population migrations. The beautiful basin of the Kan Chiang (Kan River), together with the valleys of its many tributaries, was one of the nation's most affluent regions before trade patterns were changed by the opening of treaty ports to the Western powers in the mid-19th century.

In the first half of the 20th century, Kiangsi became a focal point for revolution and war; for the entire period from 1911 to 1949, it experienced more than its share of deprivation and suffering. With the establishment of the people's republic in 1949, however, the province entered an era of stability and progress, and many new economic and social developments were pioneered there. These include the establishment of new models for land utilization, for a labour university, and for local initiative in developing and financing rural industries.

History. From 770 to 453 BC, during the Chou dynasty, Kiangsi was a part of the Kingdom of Ch'u. During the period of the Warring States (481–221 BC), the territory east of the P'o-yang Hu (Poyang Lake) was annexed by the Kingdom of Wu. When a unified empire was established under the Han dynasty (206 BC–AD 220), Kiangsi became the western portion of the large province of Yang-chou and grew rapidly in population and culture.

From 221 to 589, large numbers of families from North China, fleeing the Tatar invaders, settled in Kiangsi. Initially, there were clashes between the northern newcomers and the original inhabitants. In time mutual accommodation prevailed, and the province benefitted immensely from the introduction of northern arts, culture, and administrative skills. It was during this period that the Kan Chiang Valley became the main highway of the empire.

Under the T'ang dynasty (618–907) the growth of commerce and population in Kiangsi was even greater than in earlier times. This was caused firstly by the opening of the Grand Canal, linking Lo-yang with the Lower Yangtze, and secondly by a new influx of people from North China. Equally noteworthy was the spread of Buddhism in this period.

Period of Confucianist domination

In the Sung dynasty (960–1279) Kiangsi became a microcosm of the Confucianist state, governed by scholar-officials. The Pai Lu Tung (White Deer Grotto) Academy, near Ku-ling, where Ch'u Hsi, an influential thinker of the 12th century, taught, became a renowned centre of Confucian learning. From 1069 to 1076, Wang An-shih, a native of Lin-ch'uan, southeast of Nan-ch'ang, was prime minister; Wang introduced reforms to curb the rich and help the poor, only to be overthrown by the conservative champions of the traditional order. In the late Sung period, and throughout the era of the Mongol conquest, Kiangsi's cultural and political vigour declined. Such was the obscurantism of the government that it sanctioned a Taoist "papacy" at Lung-hu Shan, near Kuei-ch'i, which lasted into the mid-20th century.

In the early years of the Ming dynasty (1368–1644) Kiangsi produced a number of great statesmen, but after a time the government's despotic tax program alienated the people. From the early 16th century onward, peasant brigands living in the hills fought the government. The widespread unrest was ended after the Ch'ing dynasty (1644–1911) reunified the country. During this period of prolonged peace Kiangsi again became one of the richest regions of China, but its days of prosperity ended in the mid-19th century, when the Yangtze Valley was devastated by the great Taiping Rebellion against the ruling Ch'ing dynasty and when treaties with the Western powers diverted trade to coastal regions.

After the 1911 revolution the province fell victim to warlord rule, until Chiang Kai-shek brought it under Nationalist control in 1926. Chiang's break with the Communists, however, made Kiangsi a bone of contention between the two sides. An uprising was staged in Nan-ch'ang by the Communists in 1927, followed by the establishment of peasant bases in the southern counties (from the Chiang-kang Shan area eastward) under the Communist leaders Mao Tse-tung and Chu Teh. Such was the growth of their strength that, in 1931, Jui-chin, in southeastern Kiangsi, was declared the capital of the Chinese Soviet Republic. In the continuing struggle the Communist guerrillas withstood Chiang's "annihilation campaigns," but his use of an economic blockade forced the Communists to flee Kiangsi on their Long March (1934–35) to northwestern China. Chiang then briefly regained control of southern Kiangsi, but from 1938 to 1945 Kiangsi was under Japanese occupation. The Communists carried on guerrilla activities inside Kiangsi throughout the period.

After the Japanese withdrawal Communist guerrillas dominated the countryside, while the Nationalist government took precarious control of the cities. In 1949 Communist forces crossed the Yangtze from the north and took possession of the province.

Physical geography. Topographically, Kiangsi corresponds to the drainage basin of the Kan Chiang, which runs northeastward in descending elevation from the southern tip of the province to P'o-yang Hu (Poyang Lake) and the Yangtze in the north. This basin is surrounded by hills and mountains that rim the province from all sides. Among the more important ranges are: the Hwai-yü Shan (Chinese *shan*, "mountain range"), to the northeast; the Wu-i Shan, to the east; the Chiu-lien Shan and Ta-yu Ling, to the south; the Chu-kuang Shan, Wan-yang Shan (including Ching-kang Shan), Yun-hsiao Shan, Wu-kung Shan, and Chiu-ling Shan, to the west; and the Mu-fou Shan and Lu Shan, to the northwest. A remarkable feature of these mountains is that they rise in disconnected masses and thus contain corridors for interprovincial communication, especially along the Hunan border. The mountains to the south, too, present no formidable barrier. The Mei-ling Kuan (Plum Range Pass) is a broad and well-paved gap, at an elevation of 1,000 feet, leading to Kwangtung Province.

Other mountains are found in the centre and north of the province. East of the Middle Kan Valley is the Yu Shan. Made up of short and moderate hills separated by a complicated network of streams, the country traversed by this range consists of a succession of small valleys with bottomlands from five to 12 miles wide. The Lu Shan, in the north, rises sharply to 5,000 feet from the alluvial lowlands west of the P'o-yang Hu.

The rivers

The principal river of Kiangsi is the Kan, which traverses the entire province from south to north. Its headwaters are two streams (the Kung Shui and the Chang Shui) that converge to form one river at Kan-chou. Along its course this great river receives such tributaries as Sui Chiang, Ho Shui, Yüan Shui, and Chin Chiang from the west and such smaller tributaries as the Fu-chuan Shui and Wu Chiang from the east.

Besides the major system of the Kan Chiang, other rivers of Kiangsi form distinct basins of their own in the northeastern and northwestern parts of the province. These include the Hsin Chiang, which rises near Yii-shan in the northeast and runs westward through Shang-jao and Yii-kan to P'o-yang Hu; the Ch'ang Chiang and the Lo-an Chiang and its headwater, the Wu Chiang, also in the extreme northeast of the province; and the Hsiu Shui, which, rising in the Mu-fou Shan in the northwest, drains southeastward into P'o-yang Hu via Wu-ning and Yung-hsiu.

Ultimately, all Kiangsi's rivers drain into P'o-yang Hu, which is connected with the Yangtze by a wide neck at Hu-k'ou, a short distance east of the Yangtze port of Chiu-chiang. In summer, when the Yangtze rises, P'o-yang Hu gains in size and depth: its area then covers about 1,800 square miles (90 miles long by 20 miles wide), and its depth averages 65 feet. In winter, when the Yangtze waters recede, it shrinks in size, leaving shallow channels of water in many places. If the high-water stage occurs simultaneously on the Yangtze, the Kan, and other rivers, floods inevitably result. Fortunately, such flooding is an infrequent happening. The lake also serves as a useful reservoir.

Situated in the subtropical belt, Kiangsi has a hot and humid summer lasting more than four months, except in spots of high elevation, such as Lu-shan. Daytime high temperatures in Nan-ch'ang in July and August average 95° F (35° C). In winter the variations in temperature between north and south are greater. January temperatures in the extreme north at times fall to 25° F (—4° C), while those in counties south of Chi-an average 40° F (4° C). Most of the province, nevertheless, has a growing season of ten to 11 months, thus making it possible to raise two crops of rice. Rainfall is plentiful. During May and June, farmers, as a rule, can count on continuous rains to fill the rice fields with needed water. Average annual rainfall is 47 inches in the north and 60 inches in the south. In the Wu-i Shan region, on the Fukien border (where Wu-i, or "Bohea," tea is grown), the annual rainfall is as high as 78 inches.

The soil in the plains of northern Kiangsi is alluvial and permits intensive cultivation. The hilly lands in other parts of the province have red and yellow soils. Here few natural forests have been preserved; commercial trees planted are tea, tung, camphor, bamboo, and pine. On farms with clayey red soils, where rains have washed away the mineral contents, as well as the humus, the soil requires working over and the addition of green manure or chemical fertilizers in order to become productive.

Population. The 1953 census gave the population of Kiangsi as 16,770,000. The average population density at that time was 264 per square mile (102 per square kilometre); there were, however, 530 persons per square mile in the P'o-yang Hu plains and the lowlands of the river valleys but only 120 per square mile in the mountainous regions in the south.

An official estimate placed the population of Kiangsi in the early 1970s at approximately 25,000,000.

Composition of the population

As already noted, Kiangsi received successive waves of migration from North China through the ages. Its population is virtually all Han (or Chinese); the Miao, Yao, and Muslim peoples total only 10,000. The Hakkas, descendants of a unique group of migrants from North China, have maintained their separate identity, with their own dialect and social customs.

The language usually spoken is Mandarin, with a marked Lower Yangtze accent, although it has an admixture of the Fukien dialect in regions south of Kuei-ch'i and is heavily tinged with the Cantonese accent in the Ta-yü region, south of Kan-chou.

Through the centuries all three of China's major religions—Confucianism, Taoism, and Buddhism—had large followings in Kiangsi and left indelible marks on its cultural heritage. Under the people's republic, however, Buddhist monks and Taoist priests have lost their landed property to the state and have abandoned their profession to take up other employment, but, despite the regime's ideological campaigns, the influence of Confucianism still endures.

Approximately 14 percent of Kiangsi's population is urban, while 86 percent of the people live in rural areas. The leading city is Nan-ch'ang, with a population of 600,000. Situated on the right bank of the Kan Chiang, a short distance before it enters P'o-yang Hu, Nan-ch'ang is the focal point for rail and river transport, an industrial centre, and a trading centre for agricultural products. Chiu-chiang, on the south bank of the Yangtze 87 miles north of Nan-ch'ang, is the principal port through which the province's products are exported. Six miles south of Chiu-chiang is the beautiful resort of Ku-ling (Ox Ridge), perched 3,500 feet atop the Lu Shan.

From Nan-ch'ang southward up the Kan, Chi-an and Kan-chou are the leading cities, each with a population of more than 100,000. Chi-an, rich in literary lore and the centre of a tea industry, is the commercial metropolis of the Middle Kan Valley, just as Kan-chou is the centre of culture and trade in the Upper Kan Valley. Smaller cities along the river include Feng ch'eng (an industrial centre), Ch'ing-chiang (a trade centre for herbs), T'ai-ho (famous for its scribing paper), and Ta-yü, on the headwaters of the Kan (a centre for tungsten mining). Cities with populations ranging from 30,000 to 70,000 dot the hinterland on both sides of the river. The leading city in the extreme northeast is Ching-te-chen, the porcelain capital of China, with a population of 250,000.

The vast stretch of country east and southeast of Nan-ch'ang contains many cities of historical and commercial importance. Nan-feng, on the Shu Shui (or Hsii Chiang), is noted for its tangerines, and I-huang is known for its hemp, while Lin-ch'uan (Fu-chou), on the lower course of the Fu Ho, is a distributing centre for watermelons. The major cities on the Hsin Chiang are, from east to west, Yii-shan, Shang-jao, Ho-k'ou, Kuei-ch'i, Ying-t'an, and Yii-kan.

In the west and northwest of the province, Hsiu-shui and Wu-ning, on the Hsiu Shui, are important centres of the tea trade, while Jui-ch'ang is a centre of the iron industry. Wan-tsai, on the upper Chin Chiang, is noted for its ramie cloth. On the Chekiang–Kiangsi railroad from Nan-ch'ang westward, Hsin-yü is a new industrial centre; I-ch'un is noted for its ramie cloth and wrapping paper; while Ping-hsiang, which is situated on the Hunan border, is an important centre for heavy industry because of its coal mines.

Administration. As in other provinces, Kiangsi's administrative divisions are arranged in a hierarchy of levels. Immediately below the province there are two municipalities (shih), Nan-ch'ang and Ching-te-chen; two administrative bureaus, Lu Shan and Ching-kang-shan; and six areas (*t'i-ch'ü*), Ching-kang-shan, Chiu-chiang, Fu-chou, I-ch'un, Kan-chou, and Shang-jao. At the next lower level there are five additional municipalities (also called shih) and 80 counties (hsien). The lowest political units are the communes and towns.

People's congresses and councils

People's congresses at all levels are the organs of governmental authority; the deputies at the lowest level are chosen by direct vote of the people, but those at the other levels are elected by the people's congresses at the next highest level. The executive functions of government are performed by the people's councils–elected at each level by their respective people's congresses. The chairman of the Provincial Revolutionary Committee (called the Provincial People's Council before the Great Proletarian Cultural Revolution of 1966 to 1969) is the governor of the province. The Communist Party, with its own congresses and committees at every level, exercises controlling power over this system of government. Candidates are nominated under the supervision of the party, and no more than one list of candidates is voted upon. Many positions are held by party members; further, party cadres are attached to all levels of government.

Kiangsi was one of the most politically progressive of China's provinces during the Cultural Revolution. The Kiangsi Provincial Revolutionary Committee—a new power structure based on the "triple alliance" of the representatives of the revolutionary cadres, the army, and the revolutionary masses—was set up in Nan-ch'ang in 1968. Revolutionary committees were then established for all special districts, municipalities, and counties throughout the province. In subsequent years the consolidation of this power structure continued. Campaigns were launched to develop mass criticism and to study ideology as propounded by Mao-tse Tung. There is evidence that one of the concrete results of these campaigns was the closer cooperation of the government, the army, and the people for national preparedness.

Social conditions. Before 1949 the greatest scourge was the prevalence of malaria. This debilitating disease, carried by the anopheles mosquitoes, which bred in the marshes along the rivers and lakes, annually took a heavy toll of lives. Since 1949 the draining of swamps and of pools of stagnant water and measures taken for epidemic prevention have reduced this disease to a minimum. Another menace to health peculiar to the P'o-yang Hu region was liver fluke (a kind of flatworm). Many thousands of lives were previously lost every year as such worms, incubated in aquatic snails, entered the human body to develop into flukes sucking on the liver. This disease, too, is rapidly becoming a danger of the past, following mass control of the fluke embryo in the lake and surrounding waters.

Curative medicine

In curative medicine, improvements made by the early 1970s were impressive. Clinics providing free medical care were available in every commune and on every state farm, while modern hospitals were found in all cities and counties. A significant trend since the Cultural Revolution has been the rediscovery and promotion of indigenous treatments and drugs. Literally hundreds of indigenous prescriptions for many types of disease have been reported in curing not only common ailments and chronic diseases but also such diseases as jaundice, hepatitis, and gastritis. A great variety of indigenous herbs have been brought into use for the treatment of fractures and wounds, tumours, and other forms of sickness. The regime makes much of the fact that the popularization of indigenous treatments and drugs makes low-cost medical care generally available.

A minimum of social welfare is available. The welfare fund of the commune guarantees care for the sick, disabled, and aged and is also available when drought or flood occurs. For industrial workers there are measures for accident prevention, as well as insurance programs that provide for hospital treatment, sick leave, disability compensation, maternity leave, and old-age and death benefits. In Nan-ch'ang and other industrial towns the government has been improving the housing situation and expanding recreational facilities.

In education, children from three to seven attend nursery schools and kindergartens. Those from seven to 13 attend primary schools, which are run by production brigades in the rural areas and by street committees in the cities. The curriculum of the middle schools—comprising, as a rule, three years of junior and three years of senior grades—stresses proletarian ideology, physical training, and the sciences rather than social studies or humanities. All students in the middle schools are required to work and study at the same time. In the early 1970s there was a conspicuous trend to shift the operating costs of the middle schools from the state to the communes or the factories, while emphasis on the "proletarian revolution in education" brought about radical changes in educational policy. Often a pacesetter of innovations on the national scene, Kiangsi's educational program was striving to serve proletarian politics and to combine learning with productive labour. One of the models of such proletarian education was the Part-Time Tea-Growing and Part-Time Study Middle School in Wu-yüan County, Shang-jao Special District, where leadership was exercised by tea-growing peasants and instruction given by barefoot teachers who participated in the tea production.

Centres of higher learning include the Botanical Garden and Arboretum, in Lu Shan, and the Kiangsi Branch of the Chinese Academy of Sciences, the Kiangsi Library, the Kiangsi Provincial Museum, the Kiangsi Agricultural Institute, and the Kiangsi Medical College, all located in Nan-ch'ang. Perhaps the most significant development in higher education has been the Kiangsi Labour University, founded in 1958. It has its main campus in Nan-ch'ang but operates a network of more than 50 branch campuses, plus 14 affiliated technical schools, throughout the province. Aiming at the development of productive work through the dissemination of advanced education, the branch campuses are pioneering a multiplicity of development projects, including the building of roads in mountain areas, the founding of new villages, land reclamation, the building of factories, and the promotion of afforestation.

Economy. Kiangsi is one of China's richest agricultural provinces. Measures taken to increase agricultural production since 1949 include reclamation of unused land, treatment of red soil to make it more fertile, construction of large numbers of irrigation projects and hydroelectric power stations, and the increased use of chemical fertilizers and mechanization. As a result, arable land now occupies well over 35 percent of the total area of the province.

Principal crops

Food crops produced in Kiangsi include rice, wheat, and sweet potato; of these, rice is by far the most important. The P'o-yang Hu plain, the Lower Kan Chiang Valley, and the Shu Shui Valley are the principal areas of rice production; two crops a year are raised in all parts of the province. Kiangsi also produces a great variety of commercial crops. Tea is grown on hillsides in many regions; Chi-an, Wu-ning, and Hsiu-shui are the major centres of production. Ramie (China grass), used for making a fine, silky grass linen for summer wear, is produced at Wan-tsai, I-ch'un, Lin-ch'uan, and I-huang. Cotton is grown on the plains northeast of P'o-yang Hu (at P'eng-tse and Hu-You), while the Yü-shan and Kuang-feng regions, on the Chekiang border, produce tobacco. Sugarcane is raised in many parts of the province—at Po-yang, Lo-p'ing, and Tung-hsiang, in the northeast, as well as in Kan-chou, in the south. Other important commercial crops include soybeans and peanuts.

Kiangsi is a great provider of fruit. Especially notable are oranges, tangerines, watermelons, pears, persimmons, pomelos, apples, loquats (the fruit of a small tree of the apple family), and kumquats (a small citrus fruit). The hills of the province also supply the nation's apothecaries with such important herbs as the three-foliaged orange, the plantago major, and the gall nut. The lower valleys of the Wu Chiang and Hsin Chiang, east of P'o-yang Hu, are producers of the indigo plant.

Lush forests occupy 10 percent of the total area of Kiangsi. The region from Chi-an southward contains pine, fir, cedar, oak, and banyan. The timber produced here—used for building material and for furniture—is sent down the Kan and other rivers to Chiu-chiang for export to all parts of China. No less important are the camphor tree and the giant bamboo. The timber industry also yields valuable by-products, especially tung oil, resin, turpentine, lampblack (for making Chinese ink sticks), and tea oil.

Livestock raised in Kiangsi includes water buffalo, pigs, chickens, and ducks. Hog bristles and sausage casings are exported in significant volume. Fishing is a major industry in such cities as Hu-You, Chiu-chiang, Hsing-tzu, and Po-yang on P'o-yang Hu. The catches here consist of carp, eel, sturgeon, shad, loach, bream, stickleback, bleak, beltfish, barbel, perch, pike, bass, trout, and about 20 other kinds of fish. In addition, fisheries are found along the numerous rivers and in the almost countless village ponds. It is estimated that a total of more than 67,000 tons is produced every year.

Industry

Although for centuries Kiangsi was known for handicrafts, modern industry had only a limited base when the present regime took power in 1949. By 1970, however, the province had made immense progress in the establishment of both heavy and light industries. Coal and tungsten are the most important minerals mined. The P'ing-hsiang area in the west, with mines at An-yiian and Kao-k'eng, is the coal capital of south central China. The coke produced there supplies the older blast furnaces in neighbouring Hupeh Province and no doubt will supply the new steelworks being built closer by. Another coal-mining centre of rising importance is Feng-ch'eng, 38 miles south of Nan-ch'ang. The region surrounding Ta-yii, on the Kwangtung border, is the centre of tungsten mining, although extensive new deposits have recently been discovered in the Chien-nan region at the extreme southern tip of the province. It is estimated that Kiangsi contains more than 80 percent of China's total known deposits of tungsten. The ore mined there contains 60 percent tungsten; the remaining 40 percent permits the production of sizable amounts of tin, bismuth, and molybdenum. Iron is mined at An-fu, Ching-kang-shan, Sui-chang, and Chiu-chiang, while manganese is mined at Lo-p'ing. The discovery of a salt field in 1970 ended the long accepted belief that salt did not exist in Kiangsi.

The modern factories constructed since 1949 are designed either to aid agriculture or to manufacture goods using local raw materials. Nan-ch'ang is the largest industrial centre; it has plants for farm machinery, electric motors, chemicals, fertilizers, textiles, paper, dyes, pitch, jute, sanitary supplies, and ramie weaving. Chiu-chiang is principally a centre for textile mills and textile machinery. In Kan-chou are located plants for manufacturing paper, sugar, matches, cigarettes, and cotton cloth, as well as tractors, electric supplies, and chemicals. The majority of Kiangsi's industries operate on power from thermoelectric stations, although hydroelectric power has been developed, such as at the Shang-yu Chiang station, southwest of Kan-chou. Food processing is carried on in many localities throughout the province. Camphor is processed in Chi-an and tea in Chi-an, Shang-jao, and Ching-te-chen.

The development of industry, however, has not affected certain handicrafts for which Kiangsi has been famous throughout history. The ramie cloth produced in Wan-tsai, I-ch'un, I-huang, and Lin-ch'uan continues to be the nation's preferred choice for summer wear. Other important local products are the typical Kiangsi varieties of paper—lien-shih paper for printing (made of bamboo), piao-hsin paper for wrapping (also of bamboo), and mao-pien paper for scribing (made of rice and mulberry straw). Hsii-wan, on the Shu Shui, is a major centre of the engraving and printing industry.

The porcelain industry

No account of Kiangsi's economy is complete without reference to the porcelain industry—the industry par *excellence* of the province. In the Ching-te reign (1004–07) of the Sung dynasty the town of Fou-liang, on the Ch'ang Chiang, was by Imperial decree made a centre for fine porcelain. From that time on, Fou-liang was known as Ching-te-chen, after the Imperial patron; for ten centuries it has supplied the Chinese people with porcelain ware of all descriptions—ranging from items of daily use to artistic works of rare beauty made for the enjoyment of emperors and collectors.

The beautiful translucence and hardness of the porcelains from Ching-te-chen are attributable to two ingredients—kaolin and peituntze. Kaolin (or china clay) is a decomposition of granite, and peituntze (or white briquette) is a hard, feldspathic stone; both are fusible at high temperatures. These materials, found in the Ch'ang Chiang Valley and along the east shore of P'o-yang Hu, are pulverized to a fine powder and mixed to form the clay.

Today Ching-te-chen has a population of about 250,000, most of which is employed in one way or another in the making of porcelain. The bulk of the output is for domestic trade, although selected items are shipped abroad to foreign markets. The government is making an effort to revive and preserve the secret formulas of the Ming and Ch'ing potters. But there, as elsewhere in China, the inevitable tendency seems to be away from handicrafts and toward machinery.

Transportation. Kiangsi has an abundance of waterways. Most of the rivers flow diagonally, from east and west toward the centre, emptying into the Kan Chiang and P'o-yang Hu; many are navigable. The Kan Chiang itself is navigable by steamers from the lake upstream to Chi-an and by smaller steam launches from Chi-an to Kan-chou, although along the latter stretch, navigation is often made difficult by a series of 18 rapids. On many shallow streams, as well as on the headwaters of the Kan Chiang, navigation is by junk. Thus, the products of all the 80 counties of the province enjoy adequate transportation; Nan-ch'ang and Chiu-chiang are the main centres for transshipment and distribution. Goods for export are carried by large steamships on the Yangtze, but the labyrinth of towns on P'o-yang Hu forms a commercial empire in itself, knit together by regional navigation lines.

There are three railroads in Kiangsi. One line, built on the eve of World War I, runs north–south, linking Chiu-chiang with Nan-ch'ang. Another, the Chekiang–Kiangsi railroad, completed in 1937, runs east–west, from Yü-shan, on the Chekiang border, westward to P'ing-hsiang, on the Hunan border. This line forms part of a national trunk line that is being extended westward through Hunan into Kweichow to connect with the rail network of the southwest. The third line, built under the people's republic to provide strategic transport from the Yangtze Valley to the Fukien coast, runs from Ying-t'an southeastward to Amoy, opposite Taiwan.

Besides these existing railroads, new lines are either projected or under construction. These include a north–south line from Ch'i-men, in Anhwei, to Ying-t'an, with a branch line connecting Ching-te-chen with Chiu-chiang; another north–south line from Chin-hsien to Yii-tu, taking in the cities on the Shu Shui and the Mei Chiang and branching out from Yii-tu to Chang-t'ing, in Fukien; and still another north–south line from Hsin-yu—via T'ai-ho, Kan-chou, and Ta-yii—into Nan-hsiung, in Kwangtung, over the Mei-ling Kuan (Plum Range Pass), with a branch line from T'ai-ho westward to Ch'a-ling (east of Hengyang), in Hunan.

The highways

Kiangsi's highways were well developed in the Nationalist period. Many new roads have since been added, thanks to the mobilization of labour in the communes. There are five focal centres for the highway system—Nan-ch'ang, Lin-ch'uan, Shang-jao, Chi-an, and Kan-chou. Serving hinterlands of varying sizes, each of these centres represents at once the hub of regional roads and the terminus of interprovincial highways.

Although Nan-ch'ang is not a major centre of China's air transport, it is linked with Shanghai, Hangchow, Peking, Wu-han, and Canton by regular flights. Until the early 1970s, connections were made at Canton or Peking. Within the province, air travel is restricted mostly to government personnel on important assignments.

Cultural life and traditions. For nearly 2,000 years the people of Kiangsi lived under the pervading influence of Confucian culture. With village life rooted in intensive agriculture and government in the hands of the landlord-scholar-officials, the dynamics of society were regulated by Confucian ethics, which upheld learning and merit as the criteria of leadership and stressed family relationships and harmony with nature as the bases for group endeavour and group behaviour. Such a culture gave the province many famous sons that still fire one's imagination. Besides T'ao Yüan-ming (a great Chin poet of the reclusive life), Chu Hsi (the Sung dynasty Neo-Confucian philosopher), and Wang Yang-ming (the Ming philosopher), all of whom either taught or lived there, Kiangsi produced a full quota of statesmen during both the Sung and the Ming dynasties.

Yet, despite the predominance of Confucian learning and culture, peasant rebellions were a strong tradition in the province. Partly due to the temperament of the people and partly due to geographical environment, the poor and oppressed had a history of revolt against social and political injustice.

The present-day history of Kiangsi is the history of the interaction of these dual traditions. There is no denying the fact that the Confucian tradition has been torn asunder by the triumph of the revolution. With the reorientation of production and distribution since 1949, the people of Kiangsi are undergoing unprecedented changes in their life and cultural values. Proletarian ideology has replaced Confucian ideology; the peasant, the worker, the soldier, and the revolutionary intellectual are the heroes of the new society.

Yet, the peaceable character of the current scene in Kiangsi indicates that the basic elements of Confucian culture have not been totally lost. Many of the functions of the traditional family have been taken over by the state; for example, the government now sanctions marriages through civil registration, and the commune-operated kindergarten cares for the children. The party worker now oversees the moral code and discipline. In place of the old family loyalty, there is now substituted a new loyalty to the state. The government imposes hard work on the people but provides state-supported services to ensure a measure of social well-being. Yet, underlying the new equilibrium in the relationship between the government and people is the same emphasis on the leadership of the educated elite and on group endeavour and group behaviour. Reports on conditions in the communes, factories, and schools in Kiangsi give ample evidence that the dual forces of tradition and revolution are not overwhelming each other but imperceptibly reaching an accommodation.

BIBLIOGRAPHY. WANG CHUN-HENG, A ***Simple Geography of China*** (1958); JEN YU-TI, A ***Concise Geography of China*** (1964); and THEODORE SHABAD, ***China's Changing Map,*** 2nd ed. (1972), contain brief but useful chapters on Kiangsi. ***Nagel's Encyclopedia-Guide to China*** (1968) is current and handy. Maps for general reference may be found in the two CIA publications: ***Communist China Map Folio*** (1967) and ***Communist China: Administrative Atlas*** (1969). For the history of Kiangsi, see SHIH CHANG-JU ***et al., Chung-kuo Li-shih Ti-li,*** 3 vol. (1954). Portions of EDGAR SNOW, ***Red Star Over China*** (1937), deal with the rise of the Communists in Kiangsi. For population, economy, and society, see REWI ALLEY, ***Land and Folk in Kiangsi*** (1962); and YUAN-LI WU, ***The Spatial Economy of Communist China*** (1967). The literature on Kiangsi porcelain is extensive; for a general description, see JOHN KNIGHT SHRYOCK, "Kingtehchen the Porcelain City," ***Asia,*** 20:997–1002 (1920). Current developments are covered in issues in ***China Reconstructs, Peking Review, Union Research Service Reports,*** and ***Far Eastern Economic Review.***

(P.-c.K.)

Kiangsu

A province on the east coast of China, Kiangsu (Jiang-su in Pin-yin romanization) is bounded by the Yellow Sea, the province of Chekiang to the south, Anhwei to the west, and Shantung to the north. It occupies an area of 39,500 square miles (102,300 square kilometres), or 1.1 percent of the total area of China. The population in the early 1970s was 47,000,000, 6.5 percent of the national total. The provincial capital is Nanking, which was the southern capital of China during the Ming dynasty (1368–1644 AD) and the capital under the Nationalist Government (1928–49). Kiangsu became a separate province in 1667 (the sixth year of the reign of Emperor K'ang Hsi). The name is derived from the prefixes of Chiang-ning and Su-chou, the names of the two most important *fu* (prefectures) within the province at that time.

The province consists almost entirely of alluvial plains divided by the estuary of the Yangtze River into two sections, Chiang-nan (south of the river) and Sii-pei (northern Kiangsu). Chiang-nan forms a part of the Yangtze Delta and is fertile and well watered, famed for its silk and handicraft, and very densely populated and industrialized. The famous cities of Soochow, Nanking, Wu-hsi, and Shanghai (the chief metropolis of China and the fifth largest city of the world in the early 1970s) are all located in this region. Shanghai is situated at the mouth of the Yangtze River, on the Huang-pu Shui-tao (Huang-pu Channel). Administratively, the Shanghai Municipality is not a part of Kiangsu Province but is controlled directly by the State Council of the People's Republic of China in Peking. Northern Kiangsu is relatively poor in comparison with Chiang-nan. It is watered by the Hung-tse Hu (the "great marsh lake"), by the new North Kiangsu Irrigation Trunk Canal, which drains the lake, by the Huai Ho, which flows to the sea, and by the Grand Canal, which runs through the province from north to south, connecting it with the Yangtze, the T'ai Hu, and Hangchow Bay. During several periods in Chinese history, northern Kiangsu has also been drained by the Huang Ho (Yellow River), which has now been diverted back to its former course north of the Shantung Peninsula. The northernmost section of Kiangsu, from Siichow to the sea, formed by the alluvium of Huang Ho, it is actually a part of the great North China Plain in its physical geography, as well as in its agriculture and general way of living. The poorest section of Kiangsu, it is densely populated.

Historical background

Before the Christian Era, the province, except for the northernmost part, was a part of the ancient state of Wu. The term *sheng* (province) was adopted for the first time during the Ming dynasty (1368–1644), when the province was first placed under the direct jurisdiction of the Imperial capital of Nanking. When the capital was moved to Peking, it was placed under the jurisdiction of the Southern Capital. During the earlier part of the Ch'ing dynasty (1644–1911), it was part of Chiang-nan Province but was later made a separate province of Kiangsu (1667). It has remained thus ever since. The Yangtze Delta was invaded by British forces in the Opium War (1839–1842). It was occupied by Japan during the Sino-Japanese War (1937–1945) and suffered immense damage, from which it has since recovered.

Physical geography. ***Relief and drainage.*** The most important physical characteristic of the province is its wide alluvial plain, stretching from north to south, at a low elevation above sea level. Most of the soils are thus alluvial, both calcareous and noncalcareous, and including some saline soils. There is an intricate network of rivers and canals, lakes and ponds, all protected from floods by dikes built by the farmers. Most of the province is less than 150 feet above sea level. Hills of moderate elevation are found only in the southwestern corner of the province and in the extreme north along the Shantung border. These uplands constitute no more than 5 percent of the total area. Nearly 10 percent of the total area is occupied by shallow lakes and reedy marshes. The silt of the great rivers encroaches constantly on the sea, leaving seaports of former ages high and dry. In coastal areas below the high-water level, cultivation is carried on in polders (areas protected from the sea, mainly by dikes). Extensive canalization and a vast development of polders have been systematically carried out, first under the Nationalist Government, and, since 1949, by the People's Republic. This section of the surface of the earth has been completely altered by human hands.

The Kiangsu lowlands are floodplains formed by the alluvial deposits of the mighty Yangtze, the Huai Ho, the Huang Ho, and their tributaries. Using the Yangtze and Huai Ho channels as convenient landmarks, the area of these plains may be divided into three sections. The Chiang-nan plain south of the Yangtze forms the principal part of the Yangtze Delta, characterized by flatness and lying only 10 to 16 feet above sea level. It is crisscrossed by streams and canals and dotted with ponds and lakes, forming an elaborate network of flowing water, meticulously maintained by farmers. This area actually possesses the highest stream density in China: within it, no place is more than 100 yards from the T'ai Hu drainage system. The canals were all dug by farmers of the area. Isolated hillocks dot the edge of the T'ai Hu area, which adds to its enchanting beauty. The lakes were parts of former shallow bays and inlets of the sea, obstructed and enclosed by the steady advance of Yangtze Delta. After being cut off from the sea, the water gradually decreased in salinity and formed freshwater lakes. T'ai Hu is connected with the Yangtze and its estuary by many distributaries. The Chiang-nan Canal (the name for the section of the Grand Canal south of the Yangtze), which runs through the full length of the T'ai Hu plain from northwest to southeast, cuts across all the distribu-

The three sections of the plains

taries connecting the T'ai Hu basin with the Yangtze, thus forming a vital link of the T'ai Hu system.

Between the Yangtze and the ancient channel of the Huai Ho is what Chinese geographers call the Yangtze-Huai Ho plain, built by the alluvium of the two rivers. The centre of this plain is only 6.5 to 13 feet above sea level, while its periphery stands at about 17 to 33 feet. It is considered to be a section of the Yangtze Delta, as it has the same topographical elements, including alluvial deposits and drainage. Formerly, the Huai flowed into the sea, but when its channel was gradually usurped by the Huang Ho, beginning over a thousand years ago, it was unable to reach the sea and emptied itself into the Hung-tse Hu, which flows into the Kao-pao Hu, the Grand Canal, and the Yangtze. As a sluggish tributary of the latter, the Huai causes widespread floods during the high-water season.

North of the old channel of Huai Ho is the Süchow-Huai Ho plain, built of the alluvium of the Huai Ho and the Huang Ho and standing about 30 to 150 feet above sea level. In the northern part of the plain are low hills with heights of about 650 feet. Yun-t'ai Shan, near the Yellow Sea, is the highest point in the province, at 2,034 feet (620 metres).

The Yangtze enters the province to the southwest of Nanking on the Kiangsu–Anhwei border, flowing east for a distance of about 300 miles to its junction with the Huang-'pu Chiang (Whangpoo River) before reaching the East China Sea. The waters from upstream meet tide-waters at Nanking, where at some points the river is barely one-half mile wide. The river becomes broader at Chen-chiang, widening to over 11 miles at Nan-t'ung, and more than 56 miles at its mouth. It carries about 10,600,000,000 cubic feet of silt to the sea annually, depositing it to form the delta. Tides and currents carry some of the sediment to form sandbars in the estuary and along the coast. The delta itself grows at an average rate of about 82 feet a year.

Yangtze Valley and North China climates

Climate. Within the province, two subtypes of climate may be distinguished: the Yangtze Valley climate, in central and southern Kiangsu, and the North China climate, to the north of the old Huai Ho. The former is humid subtropical, while the latter is cool, temperate continental, with greater extremes of temperature. Nanking in the south has a mean temperature of 36" F (2.2" C) in January and 82.4" F (28" C) in July. For northern Kiangsu, the mean January temperature is below 32° F (0° C), but summer is as hot as in the south. Annual precipitation generally increases from north to south, ranging from 23.6 to 47.2 inches, that of Nanking being 42 inches. Seasons are distinct in both north and south. Between spring and summer, the south receives prolonged rains of cyclonic origin, typical of the Yangtze Valley and extremely useful for rice growing. The coast is often visited by destructive typhoons between late summer and early autumn.

Vegetation and animal life. Northern Kiangsu is vegetated with temperate broadleaf deciduous trees, typical of the North China Plain, while in southern Kiangsu are found subtropical mixed broadleaf deciduous and broadleaf evergreen trees, typical of the Yangtze Valley. As the whole of the province has been cleared for cultivation since ancient times, no primary forest remains today. In terms of natural flora, it is a markedly depleted territory, because of the dense population and intensive cultivation. There is a warmth-loving and moisture-loving fauna characteristic of the monsoon climate of East Asia. The fauna has considerable economic significance, fish, ducks, crabs, and shrimps being important sources of food. Fish raising is highly developed—the numerous ponds, reservoirs, lakes, canals, and streams are stocked with hundreds of millions of fry that are shipped to other provinces and are also exported to other countries.

There are also numerous agricultural pests, such as rodents and insects, which harm cultivated plants and trees. Great strides have been made in the control of the more common pests, but the insects that damage trees have not yet been adequately studied or brought under complete control.

Population. The registered population of Kiangsu, excluding Shanghai, in 1957 had grown to 45,200,000. The growth rate for the 4.5-year span since the census of 1953 was 3 percent per annum, close to the national average. The population estimate for the early 1970s was 47 million, indicating an average growth rate of less than 0.5 percent per annum, which is less than one fifth that of the preceding period. This may be partially explained by the movement of Kiangsu population into Shanghai, which grew from 6,900,000 in 1957 to 10,000,000 in 1959, adding 3,100,000, or 45 percent, in two years during the Great Leap Forward—the period of intensive pressure toward economic development—in 1958 and 1959. Authorities have made a persistent effort to stop the drifting of population into Shanghai from the surrounding areas and to send migrant peasants back to the provinces.

Population densities

The population distribution patterns of Kiangsu Province and Shanghai Municipality (*q.v.*) should be studied together, as they are inseparable geographically and economically. Their total population in the early 1970s was 57,800,000, giving an average of approximately 1,400 persons per square mile. By comparison, The Netherlands—one of the most densely populated regions of Europe—had a population density of about 900 persons per square mile in the early 1970s. Population density is higher in the north, a fact explained by its earlier development, which dates from ancient times, and its importance as a communication link between North China and the Lower Yangtze Valley. Even the hilly district in southwestern Kiangsu has very high population densities in comparison with Europe and the United States. In the valleys, the figure is above 1,000 per square mile.

These figures indicate the fertility of the region and its capacity to support a large population, even in a state of relative economic underdevelopment.

The population of Kiangsu is entirely Chinese (Han), with the exception of a few Hui (Chinese Muslims). The inhabitants of Chiang-nan speak the Wu (Shanghai) dialect, while those of northern Kiangsu and the Nanking area speak the eastern Mandarin (*hsia-chiang-kuan-*hua).

Rural people make up 71 percent of the population of the Kiangsu-Shanghai area. For Kiangsu alone, the figure was about 85 percent in 1957. Social and economic conditions during the eight-year period from 1950 through 1957 were relatively stable. This was followed by a 12-year period of cataclysmic change and mass movements from 1958 through 1969, including the Great Leap Forward (1958–1959), the reorganization of the Chinese countryside into people's communes (1958), a nationwide agricultural crisis of 1959 to 1961, and, lastly, the Great Proletarian Cultural Revolution, from 1966 to 1969. The effect of these social upheavals upon the distribution of population, birth and death rates, and the urban–rural ratio has never been adequately studied by Chinese or Western scholars. Despite this lack and the statistical gap after 1957, an analysis of the population patterns and trends from 1953 through 1957 should be useful as an indication of what a densely populated province in China is like under normal conditions.

Administration and social conditions. Since 1949, under the People's Republic of China, Kiangsu Province has formed one of the 29 primary administrative divisions of the Republic. The province itself is subdivided into 15 secondary administrative units consisting of municipalities under the direct jurisdiction of the province (shih) and areas (*t'i-ch'ü*).

Organization and functions of communes

These secondary administrative units are subdivided into 68 hsien-level or tertiary units consisting of four municipalities (shih) and 64 counties (hsien). In 1958, the latter subdivisions were replaced by people's communes, administered by elected commune committees, which function as instruments of the Chinese Communist Party (CCP). The commune is divided into production brigades, which own the land and the means of production. The brigade branch committee is the grass-roots control body of the CCP apparatus, providing political and economic guidance to the rural inhabitants. The brigade is again subdivided into productive teams, which form the lowest level administrative unit, averaging 20 to 40 households.

Local government at all levels is theoretically run by elected assemblies, but actually it has usually been run by local CCP cadres. This was changed by the Great Proletarian Cultural Revolution, unleashed in the winter of 1965–66, when many local administrations were disrupted and replaced by revolutionary committees, constituted by the People's Liberation Army, revolutionary mass organizations, and "state organs of the dictatorship of the proletariat."

Numerous communes have successfully established labour-intensive rural industries. The profits from these enterprises have helped to pay for rural primary and secondary education; for social welfare, the care and support of the aged, veterans, orphans, and widows; for rural recreation; and for universal medicare.

The Third Provincial Party Congress met in 1970 and elected a committee of 56 and a standing committee of nine, headed by a first secretary, who is also chairman of the Provincial Revolutionary Committee and commander of the Nanking Military District.

Rural Life and agriculture. The villages are distributed very close to each other on the Yangtze Delta and are generally just 300 to 600 yards apart. They are located mostly on the banks of rivers and canals. Villages with several scores of households are the most common. Communication between villages is usually very easy, thanks to canals and barges, rural roads, and the ubiquitous bicycle, again somewhat like the Low Countries of Europe. The houses are usually well built of brick baked in local kilns. Dwelling conditions are fair to good by Chinese standards.

Dispersed rural settlement is the rule along the coast and the rivers of northern Kiangsu. Dwellings are found singly along the river banks, and in groups of two, three, or four among the rice fields, quite close to each other.

Reorganization of the agricultural force

In the newly reclaimed areas along the coast, the population is sparse, and the peasant dwellings are usually located at the centre of rectangular fields laid out in an orderly fashion. After the Communists came to power, they sought to transform Chinese peasant farming into large-scale, socialized agriculture, first through "mutual aid teams" in the village, then through agricultural producers' co-ops and state farms, and finally by means of communes. Before the Communist period, Kiangsu peasants, like their counterparts in the rest of China, were underemployed for most of the year but extremely busy a small part of the year at planting and harvesting time. A great number of poor peasants who could not make a living in the village drifted into the cities to look for work. They became coolies in the big cities, pulling rickshas, pedalling pedicabs, carrying cargo on their backs, or else joining the army of the unemployed, of which there were more than 670,000 after World War II in Shanghai alone. The new government has sought to eliminate this through organized mass employment in rural reconstruction, such as the building of thousands of small reservoirs and water conservancy works, road construction, and the planting of tens of millions of trees each spring to reforest the country. The construction work is done generally with very little machinery, using human muscle alone.

The whole province is cultivated, the principal food crops being rice and wheat. Secondary food crops include barley, maize, and sweet potatoes. The Yangtze Delta is a major cotton region, and silkworms are raised in the T'ai Hu plain. Tea is grown in the southwestern uplands around I-hsing, which produces the famous I-hsing china tea sets. Cattle, pig, and poultry raising provide an important source of food and income, even under the commune system, in which a small private plot is allowed to each household. Fishing and pisciculture are other sources of food. Fishing is carried out along the coast and in the rivers and lakes.

Urban life and culture. Town and city dwellers constituted about 29 percent of the total population of the Kiangsu–Shanghai area in 1957. Shanghai is the largest municipality in China and one of the ten largest in the world; it is almost a modern conurbation (*i.e.*, an aggregation of urban districts). Besides Shanghai, there is another very large city (demographically) in the region—Nanking, with 1,700,000. This is the largest city of Kiangsu proper and is its administrative and cultural centre. There are three other cities in the region with over 500,000 inhabitants, namely Siichow (710,000 as of 1957), in northern Kiangsu; Soochow (651,000); and Wu-hsi (616,000), in Chiang-nan. There are altogether over 60 cities and towns with a population of over 10,000. Most of them are chief towns of the counties *(hsien)*.

There are two categories of cities from the standpoint of historical development—the ancient cities and the modern cities. The former date from ancient or medieval times and include Nanking, Soochow, Yang-chou, Chen-chiang, and Siichow. Several of them are well-known in East Asian history, are rich in cultural heritage, and have a long tradition that has found artistic expression in Chinese traditional architecture, painting, sculpture, flower gardens, stone bridges, and world-renowned handicraft industries, such as silk embroidery and carving of various materials. These cities often possess historical monuments, famous temples, and local shrines and *p'ai-lou* (arches) honouring their illustrious citizens. Many cities have a rich folklore. Nanking, especially, abounds in national monuments and famous historical relics. The most renowned are the simple Ming tombs (of Ming emperors) and the magnificent Sun Yat-sen Mausoleum, both in the eastern part of the huge city, at the foot of Tzuchin Shan (Purple and Gold Hill). The gastronomic specialty of this ancient capital is the renowned Nanking salted duck, which is sold in gourmet food stores throughout the region. The duck is raised in ponds and lakes nearby. Other renowned products from the Nanking area include the hand-woven silk (*tzu-ching*), particularly the cloud brocades, which use every conceivable shade of colour to portray the clouds of sky at sunset.

The modern cities sprang up in the 19th century after the Opium War and the Treaty of Nanking (1842), which opened China to international trade. Such cities as Shanghai, Wu-hsi, Nan-t'ung and Lien-yiin-chiang-shih fall into this category. Most of them are seaports, river ports, or railway junctions.

Industry. Nanking is the most important modern industrial centre of Kiangsu proper. The Nanking Motor Works was reorganized from a military workshop for repairing motor vehicles under the Nationalists (before 1949) and expanded to produce spare parts. In 1952 it was further enlarged to build 2.5-ton trucks and then four-cylinder 50-horsepower and six-cylinder 70-horsepower engines. In 1964 it employed 3,500 workers. A chemical-fertilizer plant at P'u-k'ou, opposite Nanking on the left bank of the Yangtze, is the biggest of its kind in China. The Nanking industrial area also produces steel, machine tools, electrical supplies, cameras, textiles, cement, and sundry building materials. Other manufacturing industries include machinery, food processing, and oil refining.

Wu-hsi on the Shanghai-Nanking railway is a rising industrial centre with good inland waterway connections to all parts of the province. Modern manufactures include cotton textiles, silk reeling, and food processing. Good deposits of iron and coal are reported to have been found at I-hsing across the T'ai Hu. They are used in a local ironworks and steelworks, which, early in the present decade, employed more than 2,000 workers, making iron, manganese iron, and steel; output was 60,000 tons a year, and an output of 200,000 tons of iron and 200,000 tons of steel a year was projected to meet local needs for such material.

The policy of industrial dispersion

It appears that an upsurge of local industrial development has taken place in Kiangsu since 1966. Eight major branches of small industries have been built or expanded throughout the province, including chemical-fertilizer plants, ironworks and steelworks, agricultural-machinery plants, and coal mines. During the period 1966–1970, 35 new coal mines, 68 small hydroelectric plants, and 250 new factories were built in 13 local administrative units alone. The emphasis is upon industries for supporting agriculture. Each of the province's 64 counties now has agricultural-machinery plants that can effect major repairs. Many of them also make rice transplanters, harvesters, threshers, pumps for irrigation and drainage, and mechanized or semi-mechanized farm machinery. Besides county industrial enterprises, there have developed far more commune-run small plants. Nine-tenths of the counties have a local chemical industry. There are 80 small fertilizer plants in the province and 20 factories manufacturing insecticides. Two-thirds of the counties have nitrate-fertilizer plants; half of them have phosphate plants. The dispersion of industry into the countryside is part of China's preparation against nuclear war, which was given impetus by border clashes between the Soviet Union and China in 1969.

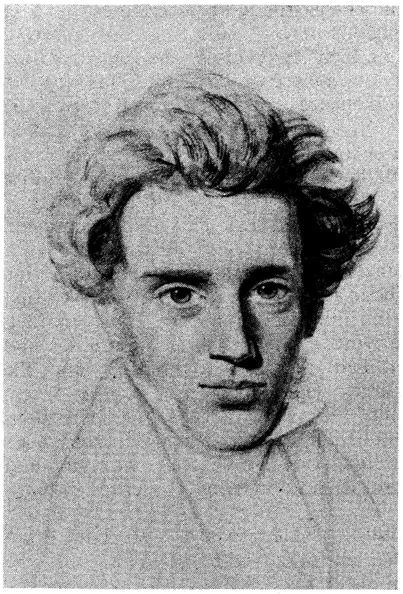
BIBLIOGRAPHY. The most comprehensive contemporary regional economic monograph is SUN CHING-CHIH (ed.), ***Economic Geography of the East China Region (Shanghai, Kiangsu, Anhwei, Chekiang)*** (Eng. trans. 1961). ***Nagel's Encyclopedia-Guide to China*** (1968) includes reports on the observations of young French Sinologists during a visit to China. H.Y. HU, "A Geographical Sketch of Kiangsu Province," ***Geogrl. Rev.***, 37:609–617 (1947), is dated, but shows intimate knowledge of the province. See also V.T. ZAYCHIKOV (ed.), ***The Physical Geography of China***, 2 vol. (1969; orig. pub. in Russian, 1964). A. HERRMANN, ***An Historical Atlas of China***, new ed. (1966), is indispensable, but contains some mistakes.

(F.Hu.)

Kierkegaard, Søren

Søren Aabye Kierkegaard, an eminent 19th-century religious philosopher whose mode of thought inspired many 20th-century thinkers, particularly the Existentialists, and earned him the credit of being the father of Existentialism (*q.v.*), is famous for his critique of systematic rational philosophy, particularly Hegelianism, on the grounds that actual life cannot be contained within an abstract conceptual system. With this stance, he intended to clear the ground for an adequate consideration of faith and, accordingly, of religion—specifically Christianity.

By courtesy of the Royal Danish Ministry of Foreign Affairs



Kierkegaard. drawing by his cousin Christian Kierkegaard, c. 1840. In a private collection.

Early life. Kierkegaard was born in Copenhagen on May 5, 1813. His father, Michael Pedersen Kierkegaard, who had a great influence on his character, had begun his own career as a poor tenant-farmers' helper in the desolate moorlands of western Jutland. One day, desperate with rage at divine indifference to his sufferings and privations, the boy stood on a hill and solemnly cursed God. Soon after, he was sent to Copenhagen, to an uncle who was a dealer in woolen articles, and from that moment he prospered, ending his life as a rich man—the owner of five houses in the capital that all miraculously escaped destruction during the British bombardment of Copenhagen in 1807. Moreover, having placed his entire fortune in gilt-edged securities, he was among the few who escaped ruin in the state bankruptcy of 1813, the year Søren was born. Thus, at his death in 1838, the old man left Kierkegaard and his brother a considerable fortune that enabled Kierkegaard to spend his life writing, unhampered by financial considerations.

Heritage of melancholy

Kierkegaard's psychological heritage was, however, far more important than his financial legacy in its consequences for his development as a man and a writer. His father combined a strict adherence to orthodox Lutheranism with a fondness for the logic of formal argument; and yet the austere religious and intellectual training he devised for the most brilliant of his sons was enlivened by a captivating imagination. Kierkegaard never shook off the influence of his father's overpowering personality nor of the suppressed melancholy that lay so disquietingly below the surface of his father's piety. In his diary he tells the story of a man who one day looked at his son and said, "Poor child, you are living in silent despair." At an early age, Kierkegaard became aware of the heavy burden of guilt that weighed his father down and later learned, in circumstances the traumatic effect of which he designated as "The Great Earthquake," that the reasons for it lay in the boyhood curse his father had hurled at God. Appalled by the knowledge of his father's sin, he threw himself into a life of dissipation yet remained haunted by the elder Kierkegaard's conviction that God's curse lay on the family, a conviction that the death of Kierkegaard's mother and five of his six brothers and sisters seemed to confirm.

He went to the University of Copenhagen to study theology but neglected this in favour of philosophy and fell under the influence of a professor of philosophy, Poul Martin Møller, who was also a writer of some importance. The later characteristics of Kierkegaard as a writer were anticipated in the works of Møller, whose declared aversion to systematic philosophy also made a deep impression on him. Moreover, the other professor of philosophy at the University at that time, Frederik Christian Sibbern (1785–1872), held a viewpoint similar to Møller's and also attempted to explain and illustrate his thoughts by writing fiction.

Relationship with Regine Olsen

The death of his father in 1838 had a sobering effect on Kierkegaard. He resumed his theological studies and two years later took his master's degree. There was, however, another reason for his renewal of purpose; he had fallen in love with a young girl, Regine Olsen, and determined to propose to her. He seems to have exerted a great fascination on everybody, concealing his melancholy under a cover of gaiety and wit, and Regine's acceptance came as no surprise to him. Yet only two days later he began to think he had made a mistake, though he still felt himself deeply in love. It appears that he became increasingly aware of the gulf between the young, innocent, inexperienced girl and himself, weighed down as he was by a feeling of guilt and by his unusual consciousness of the complexities of the human mind, which he would never be able to communicate to Regine. As he wrote in his diary: "I was a thousand years too old for her."

Accordingly, he decided to break the engagement. But Regine was in love with him, and the more he tried to persuade her to let him go, the more she clung to him. In the end he had to break off their relationship himself, but, in order to preserve her reputation, he staged an elaborate show of caddishness so as to make it appear that it was she who had rejected him. This point established, he fled to Berlin, where he lived for half a year. The little romance, novelettish though it may seem in bare outline, had a profound effect on Kierkegaard and furnished him with material for reflection and comment in several of his books.

First philosophical works. He returned from Berlin with an enormous manuscript in his trunk, ***Either/Or: A Fragment of Life***, published in 1843 under a pseudonym. Nearly all Kierkegaard's books were published pseudonymously, with fictitious names suited to the particular work, a peculiarity intended to persuade the reader that the ideas he proposed were not to be taken as the pronouncements of an authority but presented various modes of life for the reader's judgment and, especially, choice. This is, in fact, the meaning of the title *Either/Or*, which offers the alternatives of an aesthetic or an ethical (or ethico-religious) view of life. Kierkegaard's belief in the necessity—for each individual—of making a fully conscious, responsible choice among the alternatives that life offers has become fundamental in all Existential writing and thought.

His unhappy experience with Regine obviously plays a great role in *Either/Or*. Indeed, the final part of the first volume, which bears the provocative title "Diary of the Seducer," recalls his own love story in many details recorded in his diary, and the book can be seen as a secret communication to Regine, intended to explain and justify his attitude to her. Such secret communications run through all his works, and he returns again and again to the question of his responsibility for what he did. It could be argued, in fact, that it was the engagement to Regine and what followed that made Kierkegaard a writer, a writer of the type that his teachers Møller and Sibbern had been before him, a poet–philosopher. *Either/Or* was Kierkegaard's first work in this genre, and it is a work of high artistic value; in addition, it provides an important illustration of the current literary trend when Romanticism was developing some of its later preoccupations—social realism and individual psychology—and was becoming more pessimistic and morbid in its outlook. These elements also occur in the subsequent books, which appeared in rapid succession.

"The teleological suspension of the ethical"

Among them should be mentioned ***Fear and Trembling*** and ***Repetition,*** both of which also were published in **1843.** Both deal with faith and with the idea of sacrifice. The starting point of ***Fear and Trembling*** is the story of Abraham and Isaac. Once more Kierkegaard examines the implications of his break with Regine, a sacrifice, like that of Abraham, performed in obedience to a higher duty, and, like Abraham's readiness to slay his son, an act that contravenes the laws of ethics. The problem is whether situations can be imagined in which ethics can be suspended by a higher authority—*i.e.*, by God, when God himself must be considered the essence of everything ethical. This problem—which Kierkegaard calls "the teleological suspension of the ethical"—led him to the conclusion that faith is essentially paradoxical. ***Repetition*** is associated with ***Fear and Trembling*** since it provides a psychological demonstration of these ideas.

In **1844** ***Philosophical Fragments*** and ***The Concept of Dread*** appeared. They stand in the same relation to each other as do the two previous books, in the sense that the first is a philosophical statement, the second a psychological demonstration. ***Philosophical Fragments*** is an attempt to present Christianity as it should be if it is to have any meaning. It aims particularly at presenting Christianity as a form of existence that presupposes free will, without which everything becomes meaningless. This was an attack on the prevailing Hegelian philosophy, which employed grandiose historical perspectives in which the individual was sucked up as tracelessly as a grain of dust. In fact, by this time Kierkegaard was preparing for a showdown with Hegelian philosophy, but, before he did so, he felt the need to extend his ideas concerning the philosophy of freedom into the sphere of psychology. The result was ***The Concept of Dread.*** Extraordinarily penetrating, it is perhaps the first work of depth psychology in existence.

Here Kierkegaard makes a clear distinction between what he calls ***Angst*** "dread"—a feeling that has no definite object—and the fear and terror that derive from an objective threat (*e.g.*, a wild animal, a gunman). How intimately Kierkegaard's ideas were intertwined with his life can be seen from an extract from his diary:

> But if I had explained things to her [Regine], I would have had to initiate her into terrible things, my relationship with my father, his melancholy, the eternal night that broods over me, my despair, lusts, and excesses, which perhaps in God's eyes were not so heinous; for it was dread which caused me to go astray.

In the last part of the sentence we have the starting point and key to ***The Concept of Dread.*** Kierkegaard perceived that freedom cannot be proved philosophically because any proof would imply a logical necessity, which is the opposite of freedom. The discussion of freedom does not belong to the sphere of logic but to that of psychology, which cannot discuss freedom itself but can describe the state of mind that makes freedom possible. This state of mind is dread. Through experiencing dread, one leaps from innocence to sin, and, if the challenge of Christianity is accepted, from guilt to faith. Dread is thus sin's prelude, not its sequel, as one would think at first.

The three levels of ***Stages on Life's Way***

In **1845** Kierkegaard had a new book ready, ***Stages*** on ***Life's Way,*** a voluminous work and perhaps his most mature artistic achievement. In a way, it reiterates the idea of *Either/Or*, as the two titles indicate, but there is a vital difference—now the religious stage, or sphere, is distinguished not merely from the aesthetic but also from the ethical. This development was, in fact, a logical consequence of the ideas embodied in all the former works, which aimed at exposing the inadequacy of human ethics as a way of life. Accordingly, while in ***Either/Or*** there were only two spheres, the aesthetic and the ethical, in ***Stages on Life's Way*** there are three. In the third and last section of the book, "Guilty?"/"Not Guilty?," Kierkegaard once more dissects the story of his broken engagement but now from a new angle. On the aesthetic plane, a love tragedy signifies that two lovers cannot be united because an extraneous power prevents them; the story of Romeo and Juliet provides a classic example. On the ethical plane, the obstacle consists in their belonging to different spheres of existence, one interpreting love aesthetically, the other ethically. This obstacle can only be overcome by one elevating the other to his own sphere of existence, a thing that very rarely happens. On the religious plane, however, the obstacle lies in the fact that one of the two is constitutionally different, for he conceives his destiny to be one of suffering, and only the acceptance of suffering will enable him to achieve detachment from the here and now and so prepare him for eternity. The aesthetic hero has his opposition outside himself; the religious finds it within. The aesthetic hero becomes great by conquering; the religious hero by suffering. But suffering in the service of "the idea" is precisely the realization of the idea in the religious sphere of existence. This was the argument that Kierkegaard had not himself conceived when he wrote ***Either/Or*** and for whose sake he had to write the book over again.

It is an argument that evinces an increasingly sombre outlook on life and on humanity as a whole. A number of unpleasant experiences had contributed to his changed mood. The girl he loved had married and thus crushed a romantic illusion about their remaining in a sort of divine marriage, raised above the terrestrial level, only waiting for God to make the impossible possible. This, in fact, was the idea underlying both ***Fear and Trembling*** and ***Repetition.*** Now it had all come to nothing, and the disillusionment emerges clearly in the first part of ***Stages on Life's Way,*** called "In Vino Veritas" or "The Banquet," which is modelled on Plato's ***Symposium*** and deals with the same subjects—he, eros, sex, woman—and reflects a biting sarcasm and scathing contempt for women in general.

Attack on Hegelianism. Kierkegaard also had other disappointments. He quarrelled with literary critics who did not see the purport of his writings or, even worse, did see it and still tried to make him a laughing stock. From these skirmishes, he emerged victorious but deeply hurt and filled with an enormous disgust for mankind. This bitterness manifests itself in most of what he wrote afterward. But his next book was an exception. It bears the impressive title "***Concluding Unscientific Postscript to the Philosophical Fragments. A Mimic-Pathetic-Dialectic Composition, an Existential Contribution.*** By Johannes Climacus. Published by S. Kierkegaard." **(1846).**

It is typical of Kierkegaard's irony that his most philosophically important work figures as a postscript to a book only about a fifth its size. And by calling his book "an existential contribution," Kierkegaard gives the reader a strong hint of his own philosophical position; his aim is to settle accounts with the predominant philosophy of his time, the Hegelian philosophy, which had swept Europe. Kierkegaard attacked Hegel's attempt to systematize the whole of existence, declaring that a system of existence cannot be constructed, since existence is incomplete and constantly developing. He further drew attention to the logical error that arose from Hegel's attempt to introduce mobility into logic and so revealed the confusion arising from the mixing of categories. Hegel thought he had created the objective theory of knowledge; Kierkegaard put forward the thesis that subjectivity is truth or, to quote his own definition, "the objective uncertainty maintained in the most passionate spirit of dedication is truth, the highest truth for one existing."

Kierkegaard's view of subjectivity of truth

These tenets, which have become the foundation stones of modern Existentialism, have not only punctured "the system," as Hegel called his own philosophy, but have made all philosophical systems precarious. The system builder will never understand that it is not possible to understand existence intellectually. Hegel equated existence and thought and thus left no room for faith. Accordingly, Christianity appeared as a mere paragraph in the system, an example of the general, and that, according to Kierkegaard, was the scandal. Kierkegaard did not feel himself called upon to persuade people to become Christians, but he certainly did feel an obligation to let his contemporaries understand what Christianity really is. And more than that, he had a feeling that God had designated him for a special task. But what that task was was uncertain. He discussed it with himself in his diaries, particularly after he had finished his last two books. Quite naturally, after five years of enormous concentration and after having completed a production unequalled in the literature of philosophy, he felt the need for a pause for breath, and at times he even thought of giving up writing altogether.

Showdown with the church. But Kierkegaard's retirement was not to be. He could not abstain from writing, and now the "mission" was beginning to crystallize. God had appointed him, he thought, to reveal to his contemporaries the true nature of Christianity and to expose the scandal of the established Church of Denmark, the clergy of which had betrayed their religion by making themselves comfortable in secular society, in short, had become civil servants instead of followers of Christ.

The attack against Christendom

It is clear that Kierkegaard was moving in the direction of even greater austerity in his religious thinking, and in the works that he now produced, particularly *Edifying Discourses in Divers Spirits* (1847), *Works of Love* (1847), *Christian Discourses* (1848), *The Sickness unto Death* (1849), and *Training in Christianity* (1850), he depicted a Christianity sterner and more uncompromising than in any of his other writings. The last book was also a disguised attack on the heads of the Danish Church, particularly Bishop Jacob Pier Mynster, as the highly cultivated, secularized prelate who knew how to combine the hedonism of Goethe with the sufferings of Christ.

Kierkegaard nevertheless retained a great respect for Mynster, who had been his father's spiritual adviser and friend, and therefore hesitated to publish a frontal attack. But when Mynster died in January 1854, and was succeeded by Hans Lassen Martensen (1808–84), a theologian, professor of philosophy, and court preacher whom Kierkegaard did not respect in any way and who was one of the most stubborn Hegelians in Denmark, Kierkegaard did not see any reason why he should not declare war openly on the established church and its clergy. Now he felt convinced that God had authorized him to attack ruthlessly, and he began at once with a great number of small books and pamphlets and even a periodical called *The Moment,* to the ten numbers of which he was the sole contributor.

The strain of this intensely conducted campaign made grave inroads on his health. After nearly two years of it he collapsed and was brought to a hospital, where he died a month later, on November 11, 1855. By that time he had exhausted his fortune. The tenth number of *The Moment,* which he had sent to the printer before he fell ill, had cost the remainder of his inheritance. The few things of value he possessed he left to Regine, the woman he had loved and who by that time lived in the Danish West Indies, married to the governor.

Influence on modern Existentialism. The virulent attack on the state church that was the culmination of Kierkegaard's *oeuvre* was a difficult heritage to administer. It did not change the establishment, but it did cause many individual clergymen to revise their attitude toward it and even to disaffiliate.

It was not until later that the philosophical and artistic value of Kierkegaard's work was fully appreciated. The merit for this belongs to the Danish literary critic Georg Brandes (1842–1927), who in 1877 published the first book ever written about Kierkegaard and gave a brilliant analysis of his thought and life. Brandes was a declared atheist and hater of Christianity, and it was part of his strategy to play Kierkegaard off against the church. It is, in fact, characteristic of Kierkegaard's thought that it appeals as much to people who do not subscribe to its Christian tenets as to persons specifically committed to religion. Thus, of the three editors of the first edition of Kierkegaard's works, one was a convinced Christian, the other two, atheists, and one of these was also a declared enemy of the Christian Church.

Fame in Germany

In Germany interest in Kierkegaard was widespread, and everything of his was translated before World War I. It was not, however, until the years between the two great wars that the knowledge of his work became widespread, partly advanced by Freudian psychoanalysis, which very largely deals with the same phenomena as he does; *e.g.,* "sickness unto death." The theology of the Swiss Protestant theologian Karl Barth also contributed to escalate Existentialist thinking, as did the philosophical thought of Karl Jaspers and Martin Heidegger and the Jewish religious thinker Martin Buber as well.

The crucial understanding of Kierkegaard's writing came in the post-World War II years, which seem to have created a more penetrating realization of such states as *Angst* and suffering. Now the interest in Kierkegaard became universal, reaching both America and Japan, as well as England and France. About a century after his lonely death, Kierkegaard's time had finally come.

MAJOR WORKS

Enten-Eller: Et Livs Fragment (1843; Eng. trans., *Either/Or: A Fragment of Life,* 2 vol., 1944); *Frygt og Baeven. Dialektisk Lyrik af Johannes de silentio* (1843; *Fear and Trembling,* 1939); *Gjentagelsen. Et Forsøg i den experimenterende Psychologi af Constantin Constantius* (1843; *Repetition,* 1942, 1944); *Opbyggelige Taler,* a series of texts published intermittently (1843–44; *Edifying Discourses,* 4 vol., 1943–46); *Philosophiske Smuler, eller en Smule Philosophi. Af Johannes Climacus* (1844; *Philosophical Fragments, or a Fragment of Philosophy,* 1936, rev. 1962); *Begrebet Angest. En simpel psychologisk-paapegende Overveielse i Retning af det dogmatiske Problem om Arvesynden, af Vigilius Hafniensis* (1844; *The Concept of Dread,* 1944); *Stadier paa Livets vei. Studier af forskjellige . . . udgivne af Hilarius Bogbinder* (1845; *Stages on Life's Way,* 1940); *Afsluttende uvidenskabelig Efterskrift til de Philosophiske Smuler: Mimisk-pathetisk-dialektisk Sammenskrift, existentielt Indlæg, af Johannes Climacus* (1846; *Concluding Unscientific Postscript,* 1941, 1968); *Kjerlighedens Gjerninger* (1847; *Works of Love,* 1946 and 1962); *Christelige Taler* (1848); *Sygdommen til Døden. En christelig psychologisk Udvikling til Opbyggelse og Opvækkelse, af Anti-Climacus* (1849; *The Sickness unto Death,* 1941); *Indøvelse i Christendom. Af Anti-Climacus* (1850; *Training in Christianity,* 1941); *Til Selvprovelse* (1851; *For Self-Examination,* 1941); *Dette skal siges: Saa vaere det da sagt* (1854); *Hvad Christus dømmer om officiel Christendom* (1855); *Øieblikket,* 10 pt. (1855; *Attack upon "Christendom,"* 1944 and 1956).

BIBLIOGRAPHY. JENS HIMMELSTRUP (ed.), *Søren Kierkegaard: International Bibliografi* (1962). All Kierkegaard manuscripts are held in the Royal Library in Copenhagen. The Kierkegaard-selskab (society), also in Copenhagen, publishes a yearbook.

Collected works: *Samlede Værker,* ed. by A.B. DRACHMANN, J.L. HEIBERG, and H.O. LANGE, 15 vol. (1901–36); *Samlede*

Værker, rev. and ed. by PETER P. ROHDE, 20 vol. (1962–64). The miscellaneous *Papirer* were published from 1909.

Selections: Selections from the Writings of Kierkegaard, trans. by L.M. HOLLANDER (1923); ROBERT BRETALL (ed.), *A Kierkegaard Anthology* (1946); W.H. AUDEN, *The Living Thoughts of Kierkegaard* (1952); PETER P. ROHDE, *The Diary of Søren Kierkegaard,* trans. *by* GERDA M. ANDERSEN (1953).

Biography and criticism: E.L. ALLEN, *Kierkegaard: His Life and Thoughts* (1935), a general introduction; TORSTEN B. BOHLIN, *Kierkegaard's dogmatische Anschauung in ihrem geschichtlichen Zusammenhange* (1927), a penetrating analysis by a Swedish theologian; T.H. CROXALL, *Kierkegaard Studies, with Special Reference to (a) the Bible, (b) Our Own Age* (1948 and 1956); LEV SHESTOV, *Kierkegaard and the Existential Philosophy* (1970; orig. pub. in Russian, 1939), the first work to underline the Existentialist point of view; HERMANN DIEM, *Die Existenzdialektik* (1950), an important contribution to Existentialist thought; EDUARD GEISMAR, *Lectures on the Religious Thought of Søren Kierkegaard,* 2nd ed. (1938), a clear account of Kierkegaard's thought by one of his leading students; THEODOR HAECKER, *Sören Kierkegaard* (1913; Eng. trans., 1937), a study of a leading German Kierkegaard ideologist; JOHANNES HOHLENBERG, *Søren Kierkegaard* (1949; Eng. trans., 1954), a sober, reliable account; WALTER LOWRIE, *Kierkegaard* (1938), a study by a man who translated most of Kierkegaard's works and whose books bear evidence of an intimate familiarity with the texts; PETER P. ROHDE, *Søren Kierkegaard, et geni i en købstad* (1962; Eng. trans., *Søren Kierkegaard: An Introduction to His Life and Philosophy,* 1963), for the common reader; CHRISTOPH SCHREMPF, *Søren Kierkegaard: Eine Biographie, 2* vol. (1927–28), a fine, scholarly work, now slightly outdated; ADI SHMUELI, *Kierkegaard and Consciousness* (1971), a penetrating account of Kierkegaard's thought from a new angle.

(P.P.R.)

Kiev

The capital of the Ukrainian Soviet Socialist Republic, a port on the Dnepr River, and a large railroad junction, Kiev is one of the largest and most important cities of the Soviet Union. It was the home of some 2,013,000 people in 1976, thus ranking as the third largest Soviet city and first among cities of the Ukraine. Kiev has an ancient and proud history. As the centre of Kievan Rus, the first Russian state, 1,000 years ago, it acquired the title "Mother of Russian Cities." It suffered severe damage during World War II, but by the mid-1950s it was fully restored and by the 1970s had become a thriving, modern capital, with a well-developed economic and cultural life.

The city stands on the Dnepr River just below its confluence with the Desna and 591 miles (952 kilometres) from its mouth in the Black Sea. The original location was on the high and steep right bank, which rises above the river in an imposing line of bluffs, culminating in Batyyeva Hill, 330 feet (100 metres) above mean river level. This precipitous and wooded bank, topped by the golden domes and spires of churches and bell towers and by modern high-rise apartment buildings, makes the city an attractive and impressive sight from across the Dnepr. Since World War II, Kiev has extended on to the wide, low, and flat floodplain on the left bank.

Climate

Kiev has a moderately continental climate. The average January temperature is 21.6" F (—5.8" C), and winter days with temperatures above freezing are not uncommon; in cold spells with a northerly or northeasterly airstream, temperatures may drop sharply, and an absolute minimum of —27.4" F (—33" C) has been recorded. Snow cover lies usually from mid-November to the end of March; on average, the frost-free period lasts 180 days but in some years surpasses 200 days. Summers are warm, with a July average of 67.1" F (19.5" C) and a recorded maximum of 102.2" F (39" C). The mean annual rainfall is 24.5 inches (622 millimetres), with maximum precipitation in June and July. (For information on related subjects, see RUSSIA AND THE SOVIET UNION, HISTORY OF: *Kiev;* and UKRAINIAN SOVIET SOCIALIST REPUBLIC.)

HISTORY

Origins and foundation. Kiev has a long, rich, and often stormy history. Its beginnings are lost in antiquity. Archaeological findings of stone and bone implements, the remains of primitive dwellings built of wood and skins, and large accumulations of mammoths' bones indicate that the first settlements in the vicinity date from the Late Paleolithic Period (some 15,000 to 40,000 years ago). As early as 3000 BC in the Neolithic Period and subsequently at the time of the Cucuteni-Tripolye culture at the end of the Neolithic, tribes engaged in agriculture and animal husbandry lived on the site of modern Kiev. Excavations continue to uncover many artifacts from settlements dating from the Copper, Bronze, and Iron ages. The tribes of the area traded with the nomadic peoples of the steppe to the south, Scythians, Sarmatians, and later Khazars, and also with the ancient Greek colonies that were located on the Black Sea coast.

According to the 12th-century chronicle *Povest vremennykh let* ("Tales of Bygone Years," also known as *The Russian Primary Chronicle),* Kiev was founded by three brothers, Kiy, Shchek, and Khoriv, leaders of the Polyane tribe of the East Slavs. Each established his own settlement on a hill, and these became the town of Kiev, named for the eldest brother, Kiy; a small stream nearby was named for their sister Lybed. Although the chronicle account is legendary, there are contemporary references to Kiev in the writings of Byzantine, German, and Arab historians and geographers. Archaeological evidence suggests that Kiev was founded in the 6th or 7th century AD.

The first Russian capital. Less legendary is the chronicle account of the Varangians, who seized Kiev in the mid-9th century. As in Novgorod to the north, a Russo-Varangian ruling elite developed. Kiev, with its good defensive site on the high river bluff and as the centre of a rich agricultural area and a group of early Russian towns, began to gain importance. In about 882 Oleg (Ukrainian Oleh), the ruler of Novgorod, captured Kiev and made it his capital, centre of the first Russian state, Kievan (or Kiev) Rus. The town flourished, chiefly through trade along the Dnepr, going south to Byzantium and north over portages to the rivers flowing to the Baltic, the so-called "road from the Varangians to the Greeks," or "water road." Trade also went to the Caspian Sea and Central Asia.

Christianization

In 988 the introduction of Orthodox Christianity to Kiev enhanced its significance as the spiritual centre of Rus. By the 12th century, according to the chronicles, the city's wealth and religious importance was attested to by its more than 400 churches. The Cathedral of St. Sophia, parts of the monastery known as the Pecherskaya Lavra, and the ruins of the Golden Gate remain today as witnesses to Kiev at the height of its splendour. The town was famed for its art, the mosaics and frescoes of its churches, its craftsmanship in silver, and the quality of many of its manufactures. One of Europe's major cities, Kiev established diplomatic relations with Byzantium, England, France, Sweden, and other countries. Travellers wrote of its population as numbering tens of thousands.

Throughout the period of Kievan Rus, however, the city was engaged in a succession of wars against the nomadic warrior peoples who inhabited the steppes to the south, in turn the Khazars, Pechenegs, and Polovtsy (Kipchaks). These conflicts weakened the city, but even greater harm was done by the endless, complex internecine struggles of the Russian princedoms into which Rus was divided. In 1169 Prince Andrew Bogolyubsky of Rostov-Suzdal captured and sacked Kiev. Thus by the late 12th century the power of the city had declined, and in the following century it was unable to resist the rising and formidable power of the Mongols. In 1238 a Mongol army under Batu, grandson of Genghis Khan, invaded Rus and, having sacked the towns of central Russia, in 1240 besieged and stormed Kiev. Much of the city was destroyed and most of its population killed. The Italian traveller Giovanni da Pian del Carpini six years later reported only 200 houses surviving in Kiev.

Kiev under Lithuania and Poland. In the 14th century, what was left of Kiev and its surrounding area came under the control of the powerful and expanding Grand Duchy of Lithuania, which captured it in 1362. For a long time thereafter Kiev had little function save as a fortress and minor market on the vaguely defined frontier be-

Kreshchatik, the central thoroughfare [illegible]f Ki[illegible], with apa[illegible] h[illegible] in the [illegible] and the Palace of Culture for O[illegible] ct in [illegible]entre background.

Novosti Press Agency

tween Lithuania and the steppe Tatars, based in the Crimea. It frequently came under attack from the Tatars; in 1482 the Crimean khan, Mengli Giray, took and sacked the town. Almost the only survival of Kiev's former greatness was its role as the seat of an Orthodox metropolitan. A step forward came in 1516, when the grand duke Sigismund I granted Kiev a charter of autonomy, thereby much stimulating trade.

In 1569 the Union of Lublin between Lithuania and Poland gave Kiev and the Ukrainian lands to Poland. Kiev became one of the centres of Orthodox opposition to the expansion of Polish Roman Catholic influence, spearheaded by vigorous proselytization by the Jesuits. In the 17th century a religious Ukrainian brotherhood was established in Kiev, as in other Ukrainian towns, to further this opposition and encourage Ukrainian nationalism. Peter Mogila (Ukrainian Mohyla), major theologian and metropolitan of Kiev from 1633 to 1646, founded in the town the Collegium (later the Academy of Kiev), which became a major focus of the struggle with Roman Catholicism.

The Cossack state

The 17th century also saw increasing unrest among the Zaporozhian Cossacks of the Dnepr downstream of Kiev and an ever-growing struggle between them and the Polish crown. This culminated in the revolt of Bohdan Khmelnytsky, who, assisted by the Crimean Tatars, entered Kiev with his insurgent Cossacks in 1648. Under heavy pressure from the Polish forces, in 1654 Khmelnytsky and the Cossacks offered their allegiance to Moscow (the Pereyaslav Agreement); this was followed by a prolonged and confused period of strife and destruction, leading in 1667 to the Treaty of Andrusovo, by which Kiev and the Dnepr left bank part of the Ukraine became an autonomous Cossack state under the suzerainty and protection of Moscow. Thereafter further struggle ensued against the Turks, with the Cossacks constantly changing sides and engaging in internecine disputes. In 1686 Kiev was finally yielded to Russia by Poland and stood as the sole Russian outpost on the right bank of the Dnepr.

Kiev under the tsars. The Second Partition of Poland in 1793, under Catherine the Great, brought the right bank Ukraine into Russia, and Kiev, assisted by the abolition in 1754 of the tariff barriers between Russia and the Ukrainian lands, began to grow in commercial importance. Catherine's reign saw the abolition of the old administrative system and of the post of Cossack hetman and the division of the Ukraine into new administrative provinces, for one of which Kiev became the centre. Subsequently it became centre of a governor generalship, covering three provinces.

In the first half of the 19th century, Kiev developed as a major focus of Ukrainian nationalism, although severe persecution from the tsarist government forced the movement to shift the brunt of its activities to Lvov in the Austrian Ukraine. In Kiev as in other Russian cities there was clandestine revolutionary activity (beginning with the Dekabrists in the early 19th century) that culminated in a series of strikes and demonstrations leading up to the revolution of 1905. An important role in this revolutionary movement was taken by students of the University of Kiev (now Kiev T.G. Shevchenko State University), which had been established in 1834.

In the 19th century the expanding economic importance of the Ukraine, and especially the growing export of grain, brought further commercial development to Kiev. Modern factory industry appeared; to the Arsenal, set up as early as the 18th century, were added timber working and the building of river craft. The town had significant industries processing agricultural products of the Ukraine —leather, tobacco, distilling, brewing, and textiles. In the late 1860s Kiev was connected by rail to both Moscow and the Black Sea port of Odessa, further enhancing its role as a centre of industry, commerce, and administration. By the outbreak of World War I, the city had a population of some 350,000.

Kiev in the revolutionary period. With the outbreak of the Revolution in 1917, a revolutionary soviet, the Central Rada (rada, "council"), was elected by the city workers, consisting primarily of Menshevik and Social Revolutionary members, with strong support from Ukrainian nationalist groups. In January 1918, the Rada proclaimed an independent Ukrainian state with Kiev as its capital. Minor uprisings by Bolshevik workers, who were mostly concentrated in the Arsenal works, were suppressed, but Red Army troops came to their aid and on February 8, 1918, entered Kiev.

By the Treaty of Brest-Litovsk (March 3) between the Bolshevik government and the Germans, however, the new Soviet government recognized the independence of the Ukraine, which was promptly occupied by German troops. A puppet Ukrainian government was set up in

Kiev by the Germans, but it collapsed with the German surrender to the Allies in November 1918 and the subsequent withdrawal of German troops. Once more an independent Ukraine was declared in Kiev, under the leadership of Semen Petlyura, but its brief and stormy history was a series of struggles between Ukrainian nationalist, White, and Red forces. In November 1919 Kiev was briefly taken by the White armies under Gen. A.I. Denikin before being finally occupied by the Red Army. Peace was still denied the city, with the outbreak of the Russo-Polish War. In May 1920 the Poles captured Kiev but were driven out in a counterattack.

Petlyura government

Kiev in the Soviet period. Kiev's role as the centre for Ukrainian nationalists caused the Soviet government to transfer the capital of the new Ukrainian Soviet Socialist Republic to Kharkov, and it was not until 1934 that Kiev once more resumed its capital status. Meanwhile the restoration of the city's shattered economy was undertaken. During the first Five-Year Plans, between 1928 and 1940, new machine tool, electrical, and chemical industries were established. By 1939 the population of Kiev had reached 846,724. The German invasion in 1941 once again brought severe suffering and destruction to the city. After a fierce 80-day battle for the town, German forces entered it on September 21, 1941. Many of the inhabitants were deported for forced labour and to concentration camps, including almost all the large prewar Jewish segment of the population. In 1943 the advancing Soviet troops forced the Dnepr and, after bitter fighting, liberated Kiev on November 6. The city itself had suffered great destruction, including more than 40 percent of its buildings and some 800 of its industrial enterprises.

For its role in the war Kiev was later honoured by the Soviet government with the Order of Lenin, the title of Hero-City, and the Gold Star medal. In the first postwar Five-Year Plan, rapid reconstruction was undertaken. Since then Kiev has continued to grow and to strengthen its industrial base. Its population passed the million mark in the later 1950s, reaching 1,109,840 in the census of 1959 (within the present city limits) and 1,631,908 in the census of 1970.

THE CONTEMPORARY CITY

The layout of the city. The city limits enclose an area of 300 square miles (707 square kilometres) on both banks of the Dnepr. It is divided into 10 administrative wards: Darnitsky, Dneprovsky, Leninsky, Moskovsky, Pechersky, Podolsky, Radyansky, Shevchenkovsky, Zaliznichny, and Zhovtnevy.

The Upper Town

The focus of Kiev is the area of the ancient Upper Town, crowning the high bluff of the Dnepr. Although very largely of postwar construction, this central area retains its old street pattern, and here are most of the surviving historical and architectural monuments. First among these is the Cathedral of St. Sophia, now a museum. Founded in the 11th century, it remains, despite certain Baroque modifications in the 18th century, one of the finest and most beautiful examples of early Russo-Byzantine ecclesiastical architecture. It has a nave and four aisles, and it is crowned by five domes. The interior is magnificently decorated with frescoes and mosaics; it contains the tomb of Yaroslav, during whose reign the cathedral was built.

Close by is the Baroque Church of St. Andrew (Andreyevskaya), designed by Bartolomeo Rastrelli and built in the mid-18th century; its site on the crest of the steep slope to the river makes it a striking landmark. Other relics of the past in the central area include the ruins of the Golden Gate, also built in the 11th century in the reign of Yaroslav; the Zaborovsky Gate, built in 1746–48; and the five-domed Desyatinnaya church, built in 1828–42 on the site of a church founded by St. Vladimir in the 10th century.

Within and immediately adjacent to the area of the former Old Town are many of the city's museums, theatres, and public buildings as well as the principal shops, including the central department store and the covered market. The axis of the centre is the street known as Kreshchatik, which runs along the bottom of a small valley the sides of which have in part been landscaped with terrace gardens interspersed with tall, modern office and apartment buildings. The greenery of the gardens, the trees lining the street, the squares which it intersects—all combine with the variegated colours of brick, red and gray granites, and decorative ceramic tiles to give Kreshchatik an attractive and colourful aspect, much admired by Kiev's inhabitants. Among important buildings on the street is that of the City Council, where the 800 elected deputies hold their meetings.

Intersecting Kreshchatik at right angles is the wide, poplar-lined Boulevard of Taras Shevchenko, on which stands the university with its eye-catching red-washed walls. Here too is the Cathedral of St. Vladimir (still in use as a church), built in 1850–96 in Byzantine style and containing impressive paintings by Viktor Vasnetsov and other Russian artists. Notable among the many statues in central Kiev are those that commemorate the Cossack leader Bohdan Khmelnytsky and the Ukrainian poet Taras Shevchenko.

North of the old centre is the former trading and Jewish quarter, Podol, with a rectangular pattern of streets and the old merchants' trading exchange, the House of Contracts, built in 1817. Here too is the river port. South of the centre is the Pechersky district, along the top of the river bank. This district contains many of the principal buildings of the Ukrainian republican government, including the glass-domed palace of the Supreme Soviet of the Ukrainian S.S.R., built in 1936–39, and the 10-story block that houses the Council of Ministers. Nearby is the attractive Mariinsky Palace, built in 1752–55 for the tsaritsa Elizabeth and reconstructed in 1870; it is now used for receptions by the Ukrainian government.

Pechersky district

At the southern end of this district is the Pecherskaya Lavra (Monastery of the Caves), one of the most famous and important monasteries in Russian history, founded in the early 11th century. It was at the *lavra* that the monk Nestor wrote the earliest surviving Russian chronicle. Although the Cathedral of the Assumption (inside the walls of the monastery) was blown up by the Germans in 1941, Trinity Church, of the same period, survives. Also within the walls are the 17th-century Church of All Saints and an impressive 18th-century bell tower, rising up 315 feet. A major feature of the monastery is the system of catacombs beneath it in which are the mummified bodies of early monks and saints, including that of Nestor. Although a museum, open to the public, the Pecherskaya Lavra is still in use as a monastery.

South from the *lavra* is yet another monastery, the Vydubetsky, dating from the 11th century; it too suffered severe damage in World War II.

All along the steep river bank, fronting the Upper Town and Pechersky district, an attractively landscaped park has been laid out overlooking the Dnepr. With the views it affords, the park forms one of the most striking features of the city. It contains an open-air theatre, sports stadium, and restaurant, and a funicular railway climbs the 300-foot slope. Within the park are also many memorials; dominating the northern end is the statue of Prince Vladimir, bringer of Christianity to Russia, which marks the place where in 988 the people of Kiev were baptized en masse. The southern end, called the Park of Glory, has an 85-foot granite obelisk rising above the grave of the Unknown Soldier and a memorial garden. Also located in the park are the grave of Gen. Nikolay Vatutin, commander of the Soviet forces that liberated Kiev in 1943, and a rotunda marking the supposed grave of the early Varangian chief Askold.

Around these central districts of Kiev stretch extensive suburbs of factories and residential neighbourhoods. The low priority given to housing during the Stalin period means that the greater part of these suburbs has been built since 1956. The neighbourhood units, known as micro-regions, consist of groupings of apartment buildings housing 2,500 to 5,000 people, together with basic services, local shops, a health centre, cinema, and primary school. Since the late 1960s the apartment buildings have usually been of 12 to 20 stories and of prefabricated construction. Most apartments have only two or

Suburbs

three rooms, and population densities are therefore high, in the new residential developments as much as in the older central areas. The growing ownership of private cars was beginning by the late 1970s to pose problems in the provision of garage space in these new districts. A feature of development since World War II has been the rapid spread of the city on the low left bank of the Dnepr, previously almost devoid of settlement. The left bank is linked to the main part of Kiev by a railway bridge and by the imposing Ye.O. Paton road bridge, which is 4,920 feet (1,500 metres) long and named after its designer.

Between the neighbourhood units are substantial areas of park and green space. These include the very large botanical gardens of the Ukrainian Academy of Sciences, the smaller university botanical gardens (established in the mid-19th century), and in the southwestern suburbs the extensive permanent exhibition of the Ukrainian economy. On the city outskirts are several areas of forest, which are much used for recreation. In the south is the Goloseyevsky Forest Park, dominated by deciduous trees, and to the north are nearly 10,000 acres (4,000 hectares) of the Pushche-Voditsa Forest Park, mainly covered by coniferous species.

Soviet cities on the whole tend to a certain monotony of appearance. A number of factors combine to make Kiev an exception to this rule and one of the most attractive urban places in the Soviet Union—the site, with its sharply contrasted relief and wide views across the Dnepr, the abundance of greenery in and around the city, and the many buildings of historic interest and beauty.

Economic life. Kiev, as capital of the Ukraine, has major administrative functions, with considerable employment in the offices of ministries responsible for the republic's economy. The city is also an important industrial centre, possessing a wide range of manufactures. In the early 1970s some 350,000 persons were engaged in industry, about two-fifths of the total number employed. Factories are found in all quarters of the city, with major concentrations to the west of the city centre and on the Dnepr left bank.

Engineering industries, based on metal from the iron and steel plants of the Dnepr bend region and the Donbass coalfield, take pride of place and include the production of complex machinery and precision tools and instruments. The Bolshevik plant makes equipment for chemical works, such as conveyor lines for vulcanized rubber, linoleum, and fertilizer factories; the Gorky works produce metal-cutting machines. Other engineering products are aircraft, hydraulic elevators, electrical instruments, armatures, river- and seagoing craft, motorcycles, and cinematograph apparatus.

Chemical industry

Another important sector is the chemical industry, making resin products, fertilizers, plastics and chemical fibres, the last at the Darnitsa viscose rayon plant on the left bank. Timber working and the making of bricks and reinforced concrete items are also well developed. Consumer goods manufactured include cameras, thermos flasks, knitwear, footwear, a range of foodstuffs, and, in a plant built in the 1970s, watches. Kiev is a large publishing centre, with 14 printing houses.

Power for the many enterprises is supplied by natural gas, piped from Dashava in the western Ukraine, and by electricity from the Kiev hydroelectric station on the Dnepr. This station, completed in 1968, is at Vyshgorod, just upstream of the city; its 20 turbines have a total capacity of 370,000 kilowatts. The reservoir created by the station's dam is 60 miles (97 kilometres) long. Twenty-five miles southeast of Kiev is the still more powerful Tripolye thermal electric station, with a capacity of 1,800,000 kilowatts.

Transportation for the industries and for the city as a whole is provided by a good communications network. Trunk railways and all-weather roads link Kiev to Moscow, to Kharkov and the Donets Basin (Donbass), to the southern Ukraine and the port of Odessa, and to the western Ukraine and Poland. The navigability of the Dnepr has been improved by a series of barrages and reservoirs. Borispol airport operates direct flights to most major Soviet cities and to many Ukrainian towns, as well as some international connections to Romania and Bulgaria. Within Kiev itself there is efficient bus, streetcar, and trolleybus service, moving more than 2,000,000 persons daily. Construction of an underground railway began in 1960. Transportation accounts for a tenth of the total number of persons employed in Kiev.

Social and cultural life. Kiev's ancient tradition as a cultural centre is still vigorously alive. The Kiev State University, with some 20,000 students, heads an array of 20 institutions of higher education, notable among which are the Polytechnic (founded in 1898), the Agricultural Academy, and the medical, art, and architectural institutes.

There are about 300 general secondary schools, 80 evening schools for adults, and a number of specialist technical schools. A range of research establishments is headed by the Academy of Sciences of the Ukrainian S.S.R., established in 1919, which also maintains the largest of the city's many libraries. Kiev is particularly noted in the Soviet Union for medical and cybernetic research. The emphasis on applied research is illustrated by the academy's Ye.O. Paton Institute of Electrical Welding.

There are a number of theatres, notably the Shevchenko Theatre of Opera and Ballet. Plays are presented at the Lesya Ukrainka and Ivan Franko theatres, which specialize in Russian and Ukrainian drama, respectively; drama is also frequently staged in the 4,000-seat auditorium of the Palace of Culture and in the Palace of Sport, which can seat 12,000 people. In addition there are youth, open-air, and musical comedy theatres. Kiev has a circus and more than 130 cinemas; films are made in a studio in the city. Concerts are regularly given at the Tchaikovsky Conservatory. The most important of the city's many museums are the Kiev State Historical Museum, the Kiev State Museum of Russian Art, and the Kiev State Museum of Ukrainian Art.

Sports

Kiev has good facilities for sports; the largest of its 15 stadiums, the Central Stadium, can accommodate 100,000 people. Aquatic sports take place on the reservoir of the Kiev dam at Vyshgorod and also on Trukhanov Island in the Dnepr opposite the city centre, where there is a fine beach and water sports centre. The city is well provided with health facilities, including general and specialized hospitals and local polyclinics, the latter serving residential neighbourhoods. Since the majority of women are employed, a number of nursery schools and crèches care for children below school age. Around the outskirts of Kiev are a number of health resorts, sanatoriums, and children's holiday camps. (R.A.F.)

BIBLIOGRAPHY. O.E. **Касименко** (ed.), *Історія Киева*, 2 vol. (1959–60), a thorough historical study based on archaeological evidence as well as documentary sources; **І.П. Старовойтенко**, *Соціалістичний Київ* (1958), covers changes in economic, social, and cultural life during the Soviet period; LEONID DAEN, PAVEL POZNYAK, and MARK CHERP, ***Kiev: Travel*** Guide (1970), a well-documented guide in English, and ***Kiev Travel Guide***, published by the Novosti Press Agency Publishing House, is also in English.

King, Martin Luther, Jr.

The Rev. Martin Luther King, Jr., was an eloquent black Baptist minister who, from the middle 1950s until his assassination in April 1968, led the first mass civil rights movement in United States history. He achieved worldwide recognition when he was awarded the Nobel Prize for Peace in 1964 for his application of the principle of nonviolent resistance—patterned after India's Mahatma Gandhi—in the struggle for racial equality in America.

Early years. Born in Atlanta, Georgia, on January 15, 1929, King came from a family steeped in the tradition of the Southern Negro ministry: both his father and maternal grandfather were Baptist preachers. At the age of 15 he entered Morehouse College, Atlanta, under a special program for gifted students, receiving his B.A. in 1948. As an undergraduate his earlier interests in medicine and law were eclipsed by a decision in his senior year to enter the ministry, as his father had urged. Spending the next three years at Crozer Theological Seminary in Chester,

Martin Luther King, Jr.
Julian Wasser—Pix

Pennsylvania (bachelor of divinity, 1951), King first became acquainted with Gandhi's philosophy of nonviolence as well as with the thought of contemporary Protestant theologians. He found himself increasingly dissatisfied with the Social Gospel of Walter Rauschenbusch and yet unable to accept the pessimism of Reinhold Niebuhr's Neo-orthodoxy, which was an attempt to return to the basic theological principles of the 16th century. He was elected president of the student body and graduated with the highest academic average in his class. From Crozer he went to Boston University (Ph.D., 1955), where, in seeking a firm foundation for his own theological and ethical inclinations, he began to focus his attention on conceptions of the relationship of man to God. In his doctoral dissertation, "A Comparison of the Conceptions of God in the Thinking of Paul Tillich and Henry Nelson Wieman," his conclusions were fairly Niebuhrian. King himself conceived of God as an active, personal entity; man's salvation was to be found neither in the quest for social progress (Rauschenbusch) nor in the unaided power of reason (Wieman); faith in God's guidance was the essential thing.

Religious ideas

The Montgomery bus boycott

While in Boston, King met Coretta Scott, a native Alabamian who was studying at the New England Conservatory of Music. They were married in 1953 and had four children. King had been pastor of the Dexter Avenue Baptist Church in Montgomery, Alabama, slightly more than a year when the city's small group of civil rights advocates decided to contest racial segregation on the public buses. On December 1, 1955, Mrs. Rosa Parks had refused to surrender her bus seat to a white passenger and had been arrested for violating the city's segregation law. Black activists formed the Montgomery Improvement Association to boycott the transit system and chose King as their leader. He had the advantage of being a young, well-trained man who was too new in town to have made enemies; he was generally respected, and his family connections and professional standing would enable him to find another pastorate should the boycott fail.

In his first speech to the group as its president, King declared:

> We have no alternative but to protest. For many years we have shown an amazing patience. We have sometimes given our white brothers the feeling that we liked the way we were being treated. But we come here tonight to be saved from that patience that makes us patient with anything less than freedom and justice.

These words introduced to the nation a fresh voice, a skillful rhetoric, an inspiring personality, and in time a dynamic new doctrine of civil struggle. Although King's home was dynamited and his family's safety threatened, he refused to yield his faith in ultimate victory; one year and a few weeks later, the blacks of Montgomery achieved their goal of desegregation of the city's buses.

Rise **to national prominence.** Recognizing the need for a mass movement to capitalize on the successful Montgomery action, King set about organizing the Southern Christian Leadership Conference (SCLC), which gave him a base of operation throughout the South, as well as a national platform from which to speak. King lectured in all parts of the country, discussed problems of blacks with civil rights and religious leaders at home and abroad, visited Ghana and India, and conferred with heads of state. In February 1959 he and his party were warmly received by India's Prime Minister Jawaharlal Nehru; as the result of a brief discussion with followers of Gandhi about the Gandhian concepts of *Satyāgraha* ("soul-force"), King became more convinced than ever that nonviolent resistance was the most potent weapon available to oppressed people in their struggle for freedom.

The Southern Christian Leadership Conference

In 1960 he moved to his native city of Atlanta, where he became copastor with his father of the Ebenezer Baptist Church. At this post he devoted most of his time to the SCLC and the civil rights movement, declaring that the "psychological moment has come when a concentrated drive against injustice can bring great, tangible gains." His thesis was soon tested as he agreed to support the sit-in demonstrations undertaken by local college students. In late October he was arrested with 33 young people protesting segregation at the lunch counter in an Atlanta department store. Charges were dropped, but King was sentenced to Reidsville State Prison Farm on the pretext that he had violated his probation on a minor traffic offense committed several months earlier. The case assumed national proportions, with widespread concern over his safety, outrage at Georgia's flouting of legal forms, and failure of Pres. Dwight Eisenhower to intervene. He was released only upon the intercession of Democratic presidential candidate John F. Kennedy—an action so widely publicized in the black community throughout the nation that it was felt to have contributed substantially to Kennedy's slender election victory eight days later.

In the years from 1960 to 1965 King's influence reached its zenith. The tactics of active nonviolence (sit-ins, protest marches) aroused the devoted allegiance of many blacks and liberal whites in all parts of the country, as well as support from the administrations of Presidents Kennedy and Lyndon B. Johnson. There were also notable failures, as at Albany, Georgia (1961–62), when King and his colleagues failed to achieve their desegregation goals for public parks and other facilities.

In Birmingham, Alabama, in the spring of 1963, King's campaign to end segregation at lunch counters and in hiring practices drew nationwide attention when police turned dogs and firehoses on the demonstrators. King was jailed along with large numbers of his supporters, including hundreds of schoolchildren. His supporters did not, however, include all the black clergy of Birmingham, and he was strongly opposed by some of the white clergy who had issued a statement urging the blacks not to support the demonstrations. From the Birmingham jail King wrote a letter of great eloquence in which he spelled out his philosophy of nonviolence:

The letter from Birmingham jail

> You may well ask: "Why direct action? Why sit-ins, marches and so forth? Isn't negotiation a better path?" You are quite right in calling for negotiation. Indeed, this is the very purpose of direct action. Nonviolent direct action seeks to create such a crisis and foster such a tension that a community which has constantly refused to negotiate is forced to confront the issue. It seeks so to dramatize the issue that it can no longer be ignored. . . . We know through painful experience that freedom is never voluntarily given by the oppressor; it must be demanded by the oppressed.

Near the end of the Birmingham campaign, in an effort to draw together the multiple forces for peaceful change and to dramatize to the nation and to the world the importance of solving the U.S. racial problem, King joined other civil rights leaders in organizing the historic March on Washington. On August 28, 1963, an interracial assembly of more than 200,000 gathered peaceably in the shadow of the Lincoln Memorial to demand equal justice for all citizens under the law. Here the crowds were uplifted by the emotional strength and prophetic quality of King's famous "I have a dream" speech, based on biblical phraseology. In emphasizing his faith that all men, someday, would be brothers, the vision of the philosopher, the faith of the Christian, and the rhetoric of the Southern Baptist preacher all conjoined powerfully to inspire his followers throughout the world.

The rising tide of civil rights agitation produced, as King had hoped, a strong effect on national opinion and resulted in the passage of the Civil Rights Act of 1964, authorizing the federal government to enforce desegregation of public accommodations and outlawing discrimination in publicly owned facilities, as well as in employment. That eventful year was climaxed by the award to King of the Nobel Prize for Peace at Oslo in December. He interpreted his role and his times thus:

> I accept the Nobel Prize for Peace at a moment when twenty-two million Negroes of the United States of America are engaged in a creative battle to end the long night of racial injustice.... Yet when the years have rolled past...men and women will know and children will be taught that we have a finer land, a better people, a more noble civilization—because these humble children of God were willing to suffer for righteousness' sake.

Challenges of the final years. The first signs of opposition to King's tactics from within the civil rights movement surfaced during the March 1965 demonstrations at Selma, Alabama, which were aimed at dramatizing the need for a federal voting rights law. King organized an initial march from Selma to the state capitol building in Montgomery but did not lead it himself; the marchers were turned back by state troopers with nightsticks and tear gas. He determined to lead a second march, despite an injunction by a federal court and efforts from Washington to persuade him to cancel it. Heading a procession of 1,500 marchers, black and white, he set out across Pettus Bridge outside Selma until the group came to a barricade of state troopers. But instead of forcing a confrontation, he led his followers in kneeling in prayer and then unexpectedly turned back. This decision cost King the support of many young radicals who were already faulting him for being too cautious and mocking him with the title "de Lawd." The suspicion of an "arrangement" with federal and local authorities—vigorously but not entirely convincingly denied—clung to the Selma affair. The country was nevertheless aroused, resulting in the passage of the Voting Rights Act of 1965.

King and "black power"

Throughout the nation, impatience with the lack of greater substantive progress encouraged the growth of black militancy. Especially in the slums of the large Northern cities, King's religious philosophy of nonviolence was increasingly questioned. The rioting in the Watts district of Los Angeles (August 1965) demonstrated the depth of the urban race problem. In an effort to meet the challenge of the ghetto, King and his forces initiated a drive against racial discrimination in Chicago at the beginning of 1966. The chief target was segregation in housing. After a spring and summer of rallies, marches, and demonstrations, a Summit Agreement was signed between the city and a coalition of blacks, liberals, and labour organizations, calling for various measures to strengthen the enforcement of existing laws and regulations with respect to housing. But this agreement had little effect; the impression remained that King's Chicago campaign was nullified partly because of the opposition of the powerful mayor, Richard J. Daley, and partly because of the unexpected complexities of Northern racism. His presence in Chicago, however, left a long-term legacy of hope among black people, for he inspired the creation of institutions (such as a West Side tenants' movement and Operation Breadbasket, the economic arm of the SCLC) through which social ills could continue to be attacked even after he was gone.

In Illinois and Mississippi alike, King was being challenged and even publicly derided by young black power enthusiasts. But he found compassion for his dissatisfied brethren when he considered the great disillusionment attendant upon his "radiant promises of progress," as compared with his inability to deliver on those promises. In the face of mounting criticism, King responded by broadening his approach to include concerns other than racism that were equally detrimental to his people's progress. On April 4, 1967, at Riverside Church and again on the 15th at a mammoth peace rally in New York City, he committed himself irrevocably to opposition of the war in Vietnam. Once before, in early January 1966, he had condemned the war, but official outrage from Washington and strenuous opposition within the black community itself had caused him to relent. He then sought to widen his base by forming a coalition of the poor of all races that would address itself to such problems as poverty and unemployment. It was a species of populism, seeking to enroll janitors, hospital workers, seasonal labourers, and the destitute of Appalachia, along with student militants and pacifist intellectuals. "For years I laboured with the idea of reforming the existing institutions of society," he confessed, "a little change here, a little change there. Now I feel quite differently. I think you've got to have a reconstruction of the entire society, a revolution of values." His endeavours along these lines, however, did not engender much support in any segment of the population.

His plans for a Poor People's March to Washington were interrupted in the spring of 1968 by a trip to Memphis, Tennessee, in support of a strike by city sanitation workers. On April 4 he was killed by a sniper's bullet while standing on the balcony of the motel where he and his associates were staying. On March 10, 1969, the white assassin, James Earl Ray, pleaded guilty to the murder and was sentenced to 99 years in prison.

Assessment. The contribution of Martin Luther King to the black freedom movement was that of a leader who was able to turn protests into a crusade, to translate local conflicts into moral issues of nationwide concern. Successful in awakening the black masses and galvanizing them into action, he won his greatest victories by appealing to the consciences of white Americans and thus bringing political leverage to bear on Washington. The strategy that broke the segregation laws of the South, however, was not adequate to solve more complex racial problems elsewhere. King was only 39 when he died—a leader in midpassage who regarded himself as a "drum major" for justice, peace, and righteousness. Although he never wavered in his insistence that nonviolence must remain the essential tactic of the movement or in his faith that all Americans would some day attain racial and economic justice, he did not take for granted the immediate future either of his own leadership or of the causes for which he fought. "Well, I don't know what will happen now," he said characteristically—and prophetically—in his last address in Memphis, "but it really doesn't matter with me now. Because I've been to the mountaintop. ... I may not get to the promised land with you, but I want you to know tonight that we as a people will."

BIBLIOGRAPHY. King wrote a number of books. The most important for an understanding of his career are: ***Strrde Toward Freedom: The Montgomery Story*** (1958); ***Why We Can't Wait*** (1964); and ***Where Do We Go from Here: Chaos or Community?*** (1967). Useful books about King are: LAWRENCE **REDDICK**, ***Crusader Without Violence*** **(1959)**, for his early years; KENNETH L. SMITH and IRA G. ZEPP, JR., ***Search*** for ***the Beloved Community: The Thinking of Martin Luther King, Jr.*** **(1974)**, for his years at Crozer Theological Seminary and at Boston University; DAVID L. LEWIS, ***King: A Biography,*** 2nd ed. **(1978)**, a complete study; C. ERIC LINCOLN (ed.), ***Martin Luther King, Jr.*** **(1970)**, a work that contains varied views of the personality and politics of Dr. King; and WILLIAM ROBERT MILLER, ***Martin Luther King, Jr.: His Life, Martyrdom, and Meaning for the World*** **(1968)**, a sympathetic and competent treatment.

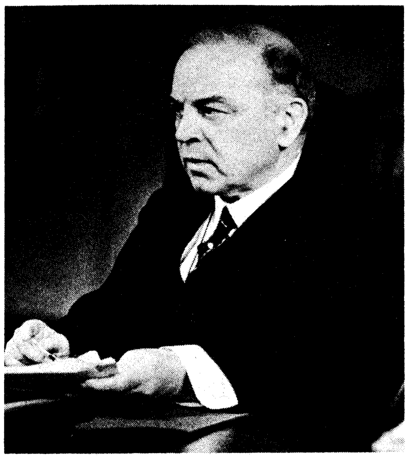
Additional works by and about Dr. King may be found in WILLIAM H. FISHER, ***Free at Last: A Bibliography of Martin Luther King, Jr.*** (1977).

(D.L.Le.)

King, W.L. Mackenzie

One of the leading Canadian statesmen of the first half of the 20th century, William Lyon Mackenzie King held the office of prime minister three times during a record-breaking 21 years. He made the Liberal Party, which he led for 30 years, a truly national party that did much to preserve the unity of the English and the French populations. He championed the independence of Canada in international affairs and led his countrymen through World War II and the difficult postwar adjustment.

Mackenzie King, as he is usually called, was born in Berlin (now Kitchener), Ontario, on Dec. 17, 1874, the son of John King and Isabel Grace Mackenzie, daughter of William Lyon Mackenzie, a leader of the Rebellion of 1837, aimed at establishing self-government in Upper



Mackenzie King.
National Film Board of Canada

Canada. Isabel, born while Mackenzie was in exile after the Rebellion, taught her son from childhood that it was his destiny to vindicate his grandfather. King had an outstanding academic career at Toronto, Chicago, and Harvard universities, broadened by travel in England and Germany. In Chicago (where he stayed at Jane Addams' Hull House) and in London, he engaged in social settlement work that profoundly influenced his later life. He was among the first Canadian politicians to show an active interest in the workers in industry.

In 1900 King declined an academic post at Harvard to take a civil service post as deputy minister of labour in the newly formed government department at Ottawa. In his new position he edited the *Labour Gazette* and showed a remarkable capacity for conciliating industrial disputes. His work brought him favourably to the attention of the Liberal prime minister Sir Wilfrid Laurier. Although King was by nature impetuous, his Presbyterian upbringing and diffident manner gave him an appearance of modesty and a veneer of prudence that became almost second nature. At decisive moments, however, he would overcome his caution and take great risks to further the destiny in which he increasingly believed. Such a risk was his resignation in 1908 from the civil service to stand as the Liberal candidate for Parliament for his native county, North Waterloo, a Conservative stronghold. Elected in 1908, he joined the Laurier government in 1909 as the first full time minister of labour in Canada. King lost his seat when the government was defeated in 1911. For the next three years he occupied himself with party publicity and organization while seeking vainly an opportunity to return to Parliament. In 1914, in need of money, he accepted a post with the Rockefeller Foundation to investigate industrial relations in the United States, resulting in 1918 in the publication of *Industry and Humanity.* When he accepted the Rockefeller post, King had insisted on residing in Canada, and, in the 1917 election, he unsuccessfully contested North York as a Laurier Liberal. After Laurier's death in 1919, King became leader of the Liberal Party. His loyalty to Laurier in 1917 was probably the decisive factor in the leadership contest, though his advocacy of social reform without socialism appealed to many of the younger party members. Leadership of the Liberal Party in 1919 was no assurance of political success. During World War I the party had split over conscription mainly along English–French lines, and several leading Liberals had joined the Conservatives in a Union Government. Moreover, the western base of the party had been sapped by the rise of an agrarian party, the Progressives.

Liberal leader

After the defeat of the Union Government in the election of 1921, King became prime minister on December 29, although his party was just short of a majority in Parliament. The future of King and his party was far from secure. In the election of 1925, he made an appeal for a majority but emerged with fewer seats in Parliament than the Conservatives. Despite this apparent Liberal defeat, the Conservatives also lacked a majority. Instead of resigning, King met with Parliament, where, with the support of Progressive and Independent members, his government won a vote of confidence. The government carried on in 1926 for six months, but, with the emergence of a scandal in the customs department, support in Parliament declined. King decided to end the uncertainty and advised the governor general to dissolve Parliament. When his advice was not taken, he resigned. The Conservative leader, Arthur Meighen, formed a government that was defeated in Parliament two days later. Meighen was given the dissolution that King had been refused. The 1926 election was fought on the constitutional issue. Because of alliances between Liberals and Progressives in many constituencies, King found himself for the first time with a decisive majority in Parliament. He became prime minister again on September 25. Late in 1926, at the Imperial Conference in London, King's was probably the determining voice in securing the declaration of equality of status of the self-governing nations of the empire, thereafter styled the Commonwealth (see **BRITISH EMPIRE** AND COMMONWEALTH).

King's government lost the election of 1930, and he led the opposition through the worst years of the Great Depression and won an overwhelming victory in the election of 1935. From then until his retirement in 1948, King was prime minister and the dominant personality in Canadian public life. It was his leadership of the country through six years of war and three years of postwar reconstruction that gave King a commanding place in Canadian history. During those years, he led a country, long divided over external policy, unitedly into war in 1939; surmounted two political crises over conscription, one nearly fatal to his government; and won the postwar election. The government he led organized a tremendous military, industrial, and financial contribution to the war and at the same time prepared for a smooth and rapid advance in economic development and social welfare afterward. When King retired, his successor, Louis St. Laurent, took over a strong government, a united and effective political party, and a rapidly growing and self-confident country.

Wartime leadership

This remarkable record was achieved by a lonely bachelor, lacking in popular appeal, political eloquence, or the trappings of strong leadership. His success was a compound of acute intuitions of the public mood and a superb capacity for the management of men. He died on July 22,1950, a year and a half after leaving office.

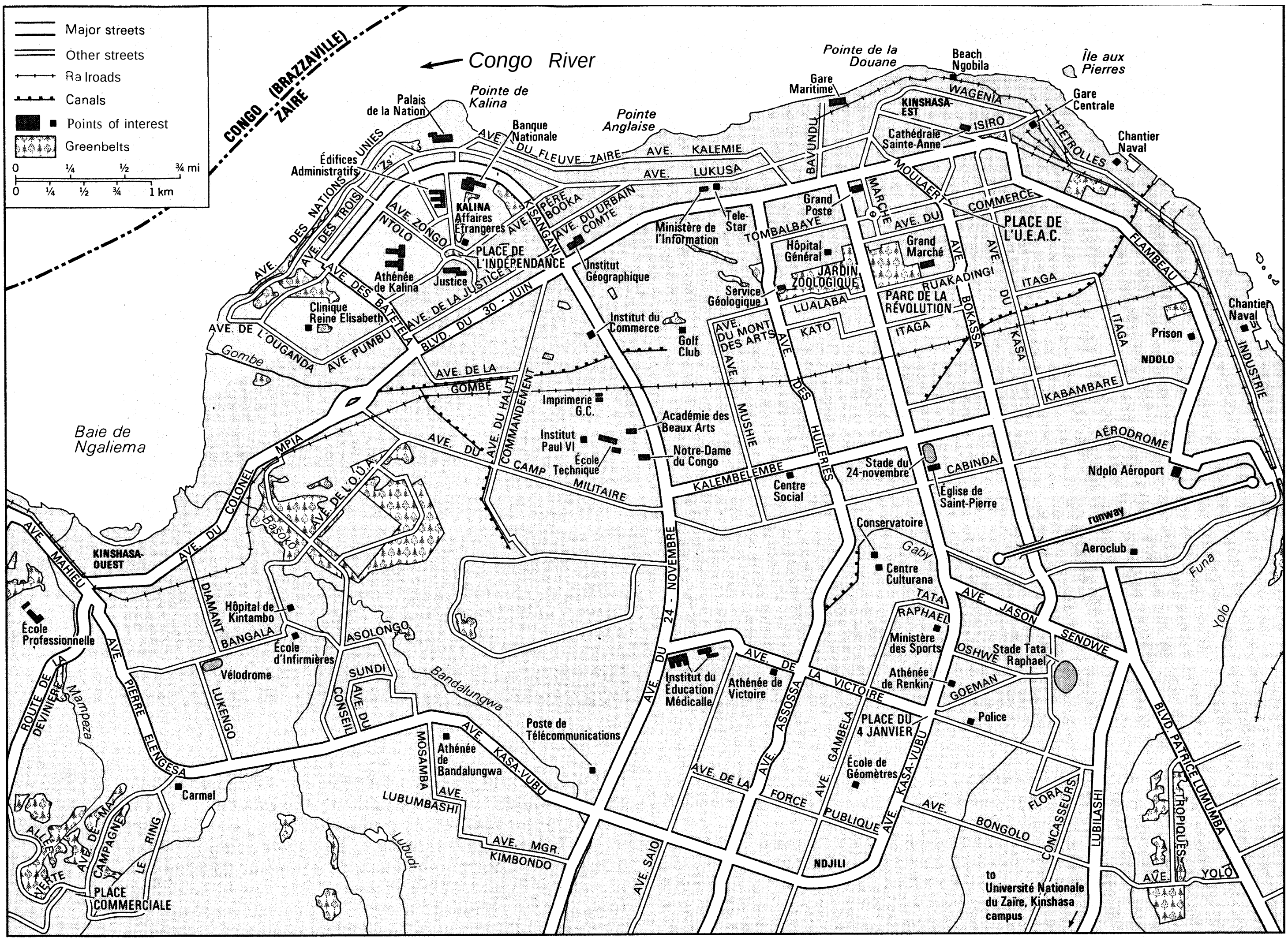
BIBLIOGRAPHY. R.M. DAWSON, *William Lyon Mackenzie King: A Political Biography,* 1874–1923, vol. 1 (1952); H. BLAIR NEATBY, *William Lyon Mackenzie King:* 1924–1932, vol. 2, *The Lonely Heights* (1963); B. HUTCHISON, *The Incredible Canadian* (1953); F.A. MCGREGOR, *The Fall and Rise of Mackenzie King,* 1911–1919 (1963); J.W. PICKERSGILL, *The Mackenzie King Record,* vol. 1 (1960), and with D.F. FORSTER, vol. 2–4 (1968–70).

(Jo.W.P.)

Kinshasa

Kinshasa, the capital of the Republic of Zaire, lies about 320 miles from the Atlantic Ocean on the south bank of the Congo River. It is the *ville neutre* (equivalent to a federal district) of Kinshasa, administered outside the rest of the republic, the outlying regions of which have more in common with their neighbours than with their capital. It is the largest city of sub-Saharan Africa, and it is the administrative, industrial, and commercial hub of the second largest black-African state.

Because of its strategic position and relatively longstanding contacts with Europe, it has far outstripped Zaire's hinterland. As in many other of the republic's main towns, the erstwhile European district is spacious and well planned, with many amenities common to all modem cities, such as electrical and sanitary facilities, a public water system, running water, ice plants, a good food supply, hotels, cafés, cinemas, and social and sports clubs. Well-built houses, cinemas, and restaurants (all the result of sound building projects) also characterize the former African district, which, unlike the satellite towns



The city of Kinshasa,

and some newer parts of Kinshasa, has not been developed by uncontrolled private enterprise.

Few Europeans are to be found in a population that has rapidly expanded from approximately 400,000 at independence in 1960 to almost 1,300,000 in 1970. The city, originally named Léopoldville for the Belgian king Leopold II and, since 1966, Kinshasa, covers an area of approximately 760 square miles (1,968 square kilometres).

History. Like all the shores of Malebo Pool (formerly known as Stanley Pool)—a shallow lake 22 miles long and 14 miles wide formed by a widening of the Congo River (*q.v.*)—the area on which the city stands was inhabited in very ancient times. It was later occupied by the members of the Humbu group who crossed the Congo River to found and colonize the villages of Kinshasa and Kintambo, from which the present city evolved. Sir Henry Morton Stanley, the explorer, who first went to Kintambo in 1877, established such cordial relations with its ruler that on his return in 1881 he was able (despite French efforts to forestall him) to acquire a site there and to found a depot, which he named Léopoldville after his patron. Though Stanley succeeded in opening river traffic as far north as Stanleyville (Kisangani since 1966) by having steamers portaged around the cataracts of the lower river to Stanley Pool, Léopoldville had little or no importance as an international trade centre until the railway line from downstream Matadi was completed in 1898. A pipeline from Matadi to carry crude oil to the upriver steamers at Léopoldville was laid in 1914, and an air service was inaugurated between Léopoldville and Stanleyville in 1920. Both developments so reinforced the city's growing importance that the administrative headquarters of the then Belgian Congo were transferred there from Boma in 1923. Its increasing importance attracted immigrant African labourers, a migration the Belgian administration strove to encourage by providing houses and other social amenities. Léopoldville in 1960 became the capital of the new republic.

The founding of **Léopold-ville**

The modern city. Standing on a plain 1,066 feet above sea level, Kinshasa is partly encircled by higher ground. It stretches about seven miles southward from a 12-mile semicircular stretch of Malebo Pool's southwestern shore. Its climate is that of a typical tropical rain forest. Rain falls throughout the year, with a mean annual total of slightly more than 60 inches. The hottest month is April, with mean daily maximum and minimum temperatures of 89" F (32" C) and 71° F (22° C), respectively. The corresponding figures for July, the coldest month, are 81° F (27" C) and 64° F (18" C). The countryside is heavily farmed rain forest; chief crops are cassava, surgarcane, oil palms, plantain, maize (corn), and beans.

In the extreme west an industrial zone (before 1966 called Old Town or Léo-Ouest) flourishes near the site of Stanley's first depot. To its east lies the riverside residential and administrative suburb of Kalina, which houses most of the European population and the new Congolese elite. The commercial district (known before 1966 as Léo-Est), of which the Boulevard du 30-Juin forms the focal point, occupies the city's eastern sector, which developed from the village of Kinshasa. Its northern waterfront is lined with three miles of quays, behind which large warehouses are stacked like packing cases. Ndolo, on its eastern periphery, comprises a complex of port facilities and industrial plants. The lower working class areas, in what used to be known as the African district, extend southward from the east and west of Kinshasa. Of

The city's **commer-cial** centre

Palais de la Nation in Kinshasa.
Afrique Photo

Kinshasa's two satellite cities, Ndjili contains about 20 percent of the metropolitan population, and upstream Kimpoko is being developed as an outer port.

Transportation. Though Kinshasa is well served by roads, its dense population and concentrated working sector cause much congestion. The railway line from Matadi enters the city in the east and then bisects it north and south; it brings in most of the country's imports, which are mostly thence conveyed by river upstream. At one stage or another almost all inland traffic destined for the Atlantic port of Matadi must be transported by water to Kinshasa, where total river traffic stands at more than 1,500,000 tons annually. An airport at Ndolo provides local and international services.

Demography. The population of Kinshasa, though slow to grow at first (from 5,000 people in 1889 to 23,000 in 1923), increased rapidly after 1940 to almost 1,300,000 in 1970. After the country's political crises of 1960–64, the proportion of European inhabitants drastically declined. Of the black inhabitants, the majority are first-generation immigrants from Zaire at large and from other Central and West African states.

Building types. Notable examples of architecture are found mainly in Kalina and the commercial sector. Chapels, cathedrals, and churches of an obsolete European style stand half hidden by a profusion of ornamental trees, while tall, modern buildings such as banks and public corporation and commercial offices stand out prominently. Multi-story buildings characterize the commercial sector; in the residential areas bungalows and long multi-unit dwellings are common.

Economic life. Its large population has made Kinshasa the most important consumer centre of the republic and has thus ensured its emergence as a major industrial centre. The main industries include food processing (confectionery, beer, and soft drinks), textiles, woodworking, paper, and the manufacture of packing cases, tobacco, and chemicals (perfumes, pharmaceutical products, plastics, and insecticides). Construction and various service industries also contribute to the city's economy.

Kinshasa serves as the headquarters of major public corporations and of industrial and commercial companies. It contains the head offices of the six main banks in Zaire, showing the extent to which Kinshasa dominates the financial and commercial life of the republic.

Government institutions and utilities. The city houses the central government offices. Its urban administration, under a governor nominated by the central government, provides such essential services as water supply (obtained from the river) and sanitation. A special service is the spraying of the town by helicopter to eradicate mosquitoes. Electricity is provided by a company named COLECTRIC. In 1969 there were three major clinics and more than 30 health establishments headed by doctors.

Education. Facilities for primary and secondary education are available. Institutions of higher education include an advanced teacher-training college, a national school of administration and law, a school of telecommunications, and an academy of fine arts, as well as institutes of social training, political studies, medical training, and commerce. Université Nationale du Zaïre, Kinshasa campus (formerly Université Lovanium de Kinshasa), the nation's largest, with a student community of 3,000 and a faculty of about 400 in 1969, lies within the Kinshasa metropolitan district. Apart from its numerous faculties, the university has an archaeological museum and an experimental nuclear reactor. There are also ethnographical and geological museums.

Université
Nationale
du Zaïre

Cultural life. There are four principal daily papers and three weekly newspapers published in French.

The radio station at Kinshasa, one of the most powerful in all Africa, provides a 24-hour service. One state-run station and one commercial station, called Tele-Star, offer television.

Recreational facilities include parks, zoological gardens, a racetrack, two stadiums—one of which can seat 70,000 spectators—and a large number of clubs and hotels.

BIBLIOGRAPHY. CATHERINE HOSKYNS, ***The Congo Since Independence: January 1960–December 1961*** (1965), and H.M. STANLEY, ***The Congo and the Founding of Its Free State: A Story of Work and Exploration, 2*** vol. (1885), two works of historical significance; W.A. HANCE and I.S. DONGEN, "Matadi: Focus of Belgian African Transport," ***A.A.A.G.,*** 48:41–72 (1958), a description of the city as a transport hub, including aspects of urban geography.

(J.O.A.)

Kinship

Kinship—the system of human relationships derived from marriage and descent—plays an important part in all societies. It is a major factor regulating behaviour between individuals and affecting the formation of social, political, and territorial groups. In modern industrial society the domestic family is the most obvious example of an institution based on kinship. In tribal societies kinship is of even more significance, having far-reaching effects on the social and economic life of the community.

CULTURAL AND SCIENTIFIC ASPECTS

Modern research has revealed the nature of the biological continuity between an individual and his genetic parents. But kinship, as a set of social relations, does not depend on knowledge of genetics or physiology. Indeed, some societies with elaborate kinship systems have held beliefs about the development of the human fetus that in no way approximate to the actualities of conception and gestation. Genetic mechanisms are uniform for all mankind, but human groups differ widely in the significance they attach to kinship.

All cultures recognize that the human fetus is born from the womb of its mother, on whom it depends for survival. There is thus a physically based and culturally defined relationship between mother and child. Likewise, all cultures distinguish between male and female individuals and institutionalize, in varying forms, a second relationship: that between a man and a woman who copulate in some acceptable way. This relationship is described in English as "marriage." There are vast differences between cultures in the customary entailments of these two relationships—mother and child, man and woman—and in how they generate other separately identified relationships such as, in English, father, sibling, mother-in-law, or cousin. The study of kinship began with the recognition, from at least classical times, that the names for kin relations in one language cannot always be translated accurately, on a one-to-one basis, into the kin terms of another language. It was stimulated by the discovery by an American ethnologist in 1858 that two Indian languages, Iroquois and Ojibwa, though apparently unrelated, nevertheless possessed common patterns of kin terms so that one-to-one translation was possible. Yet there is more to the study of kinship than the investigation of patterns of names for kin relations. For such study embraces the investigation of (1) the ways in which individuals enter into and leave kin relationships; (2) how they use them in private and public life; (3) how kin relations are made to define social groups and categories; (4) what connection kinship has with other sets of relations between individuals and groups based on political, residential, religious, and other non-kin criteria; (5) how copulation and birth are associated with kin relationship; (6) how ideas about the development of the human embryo, the acquisition of personal characteristics, the fate of the soul after death, and other matters may be linked in any culture with the pattern of kin relations; and (7) what explanations can be given for the genesis, development, maintenance, and decay of these various beliefs and practices.

Kinship and affinity

In some analyses a contrast is made between kinship and affinity—that is, between relations built up from links between parents and children (the so-called consanguineal, or blood, relations) and relations based on one or more marriage links (relations of affinity). In this sense, a man's brother-in-law and his mother-in-law, for instance, are classified as his affines and not as his kin, though not all of the cultures that make the contrast draw the dividing line in the same place. Unfortunately there is no convenient overall term for "kinship and affinity." In this article, unless otherwise indicated, kin and kinship are used in a broad sense to include affines and affinity. A "relative" is a kinsman or affine.

In most industrialized societies the family home, typically inhabited at night and for part of the day by a married couple and their young children, is the site where all members find food and shelter and develop multistranded bonds of affection between and hostility toward one another, where the couple copulate and conceive, and where the children begin to acquire the values, symbols, and skills of their culture. But in industrialized society, education takes place outside the home, and adults work and earn their living elsewhere, bringing back to the home for use and consumption goods produced by strangers. Although links with kinfolk outside the home remain important on festival occasions, at moments of crisis, and particularly in old age, kinship is not the basis of economic, political, or religious activity. At the other extreme of technical development, among peoples who subsist by hunting and by collecting wild plants, kin relationships are dominant within the small local group, but the restricted scale of social relations generally (except in aboriginal Australia) limits the possibility of maintaining an elaborate wide-range kinship system. Most highly developed kinship systems are found in middle-range nonindustrial tribal societies based on agriculture or pastoralism.

Elementary families and genealogies. Relations of motherhood, marriage, and fatherhood are formed in the domestic family and then persist even after the persons linked by them have ceased living under the same roof, and may continue to persist in symbolic form even after some of the persons are dead. Though there are parallels between human and nonhuman primate "domestic" organization, kinship as a symbolic construct seems unique to man. The nuclear family—man, wife, and unmarried children living as a domestic unit—gives rise to one of the basic kinship constructs, the elementary family, consisting of man, wife, and all of their children, wherever they may be living and whether alive or dead. Every person belongs to a so-called family of orientation—*i.e.*, the elementary family containing his mother and father and his full brothers and sisters; he may also belong to one or more "families of procreation" or "reorientation" containing one of his spouses and their children. Elementary families thus overlap and provide a basis for recognizing relations with distant kinsmen traced through several intermediate links. As a man's father has two parents and his mother also has two parents, most cultures recognize a connection between a man and his four grandparents, though not all use four different names for these relations. Similarly, relations with cousins (parents' siblings' children) are recognized, and there is great variation in the ways in which different kinds of cousins are grouped or distinguished.

Families of orientation and procreation

The process of tracing out paths to distant kin via overlapping elementary families can be applied in three contexts. First, the logical product of several culturally distinguished kin relations can be formed to generate still another relation. The product of the relations "father," "brother," and "wife," for example, is "aunt" (that is, any woman who is the wife of one's father's brother is one's aunt). Second, the tracing can be done through individuals, as, for example: "my father Thomas Brown had a brother, now dead, called William, whose widow, now remarried, is known as Mrs. E. Smith, but who is in fact my aunt." Third, tracing may be a way of marking the category of kin from non-kin. In the above example, for instance, one can trace a path to the aunt's second husband James Smith, but one cannot regard him as a kinsman, although he is one's father's brother's widow's second husband.

A written or oral statement of the names of individuals and of their kin relations to one another is a genealogy. In literate societies records of births and marriages can be used to construct large genealogical charts, but these may have little social significance. In nonliterate societies, genealogical knowledge is transmitted orally, and the form and content of a genealogy is more likely to be of social significance and consequence. In some cultures people have little interest in ancestry and may even forget the names of their grandparents. In other cultures, genealogies tracing ascent through one line, though rarely through more than one, may be remembered for 15 or 20 generations. Even if ancestors are forgotten, individuals may be able to trace paths to distant relatives laterally through affinal links. It is often the case that ancestors who do not serve as symbols for defining an individual's place in society are forgotten, and the custom of teknonymy (*i.e.*, the practice in which a parent is identified by reference to the name of his or her child) may have the effect of ensuring that the names of members of higher generations are systematically forgotten. The occasionally encountered prohibition against uttering the name of a dead person clearly has such an effect. In industrial societies, in which extrafamilial kinship is relatively unimportant, women tend to know more than men about kin links, and relatives linked through women tend to be remembered longest.

The process of filiation

Genealogical knowledge provides a basis for identification and connection with relatives. Patrifiliation is the process of identifying with one's father so that one becomes a member of the social groups he belongs to or used to belong to. Matrifiliation is identification with one's mother. The set of all persons connected to an individual by several ascending filial steps (steps between parent and child) constitute his lineal ancestors; those linked to him by several ascending, followed by several descending, cognatic steps (links between blood relatives) form his cognatic or consanguineal kin.

Both patrifiliation and matrifiliation occur in almost all societies, though many emphasize one more than the other; systems in which they are treated equally are called equilateral or bilateral. In some cultures there is no dogma of emphasis or equality, and individuals make their own choice as to which side of their family they will associate with.

Kin terms. All cultures have sets of names for various relations between kin. The number of basic terms in a set varies, averaging about 25, though in some cultures an indefinitely large number of compound terms may be constructed, as in English — "great-great- . . . great-grandfather." Such nomenclatures or sets of relationship terms may be analyzed in several ways. One procedure is to generate a standard genealogical grid consisting of consanguineal and affinal kin types, such as father's brother's wife and mother's sister's son, and to map the terms in a set on to this grid. The various kin types designated by any one term constitute a kin class. Lewis Henry Morgan, the American ethnologist who in the 19th century laid the foundation for the scientific study of kinship terminologies, grouped sets into two categories, descriptive and classificatory. In classificatory sets, lineal and collateral kin and kin belonging to several generations are typically included in a single class and are referred to by the same kin term. All cognatic kin are either lineal or collateral. Two persons are lineal kin to each other if one is an ancestor of the other. All other cognatic kin are collateral. In descriptive sets, close and lineal kin are typically distinguished by separate terms, which are not used for collaterals. In English, for example, "father" designates a single relative, but "uncle" refers to certain others, all collaterals. "Brother" may denote one of several close relatives but designates only a single kin type. Collateral relatives such as parents' siblings' children are all called "cousin." Later writers categorized sets of terms not by characteristics of the sets as a whole but by the way in which certain kin types were sorted into classes. Sets were categorized according to the classes for cousins and according to those for parents' siblings, but commentators have returned to looking at sets as wholes.

Most methods of analysis involve mapping terms on to a standard genealogical grid. Some writers argue that the notion of a grid is not present in all cultures and that these methods do not reveal how kin terms are perceived by the people who utter them. Even if people can trace out kin connections step by step, they may have other ways of deciding what term to apply to a kinsman.

THE DOMESTIC FAMILY

Evidence from other primates. In most societies the majority of people live in domestic family units consisting of one or a few closely related men, wife or wives, and young children. The widespread occurrence of this form of domestic organization suggests that it developed early in human evolution. Despite the fact that man differs in many ways from other primates, some of the mechanisms underlying human domestic arrangements are basically similar to those of apes and monkeys. In all primate species there are bonds between adult females and immature members, and among chimpanzees, recognition and cooperation between a mother and her young sometimes lasts well beyond the period needed for the young to attain biological self-sufficiency. Bonds between adult males and young are also found, but are less prominent. With nonhuman primates as with man, adequate socialization in infancy seems to be a precondition of adequate maternal behaviour in adult life and of the development of paternal bonds. On the other hand, exclusive bonds between a single adult male and a single adult female are found only in gibbons. Some writers on evolution have held that the human hunting band, seen as an aggregation of domestic families, developed from a chimpanzee-like troop in response to increasing separation of function between male and female. Others suggest that isolated hominid (manlike) breeding units, each with one male and several female adults (as found with hamadryas monkeys), learned to associate in order to hunt large animals. Alternatively, it may have been that baboon-like societies with closed membership and a male dominance hierarchy adapted to cooperative hunting by relaxing the hierarchy and forming male–female pair bonds. Whatever evolutionary path was followed, it seems reasonable to assume that early man lived in aggregations of families, with women caring for the young and collecting plant food and with men hunting animals and providing protection against animal predators and human enemies. It is still possible to investigate the material and technical achievements of pre-Neolithic man, but his forms of social organization remain uncertain, and the extent to which he made use of the symbolic notion of kinship quite unknown.

Sexual and filial bonds of primates

Family and household. In anthropological usage the domestic family is any group of kinsmen and spouses that constitutes a household, has a common hearth or kitchen, a common purse or food supply, and, usually, a common roof. There are nonfamily households based on other principles and some which have a family core with the addition of unrelated friends, lodgers, slaves, and servants.

The nuclear family

A very common type of domestic family consists of one adult man, his wife, their unmarried children, and no one else. Some writers have maintained that this nuclear family is universal, but in no society are all households of this type, if only because there are widows, orphans, unmarried mothers, and others who need a home. And on reaching maturity a child may remain at home and be joined by an incoming spouse. When they have children, there will be three generations in the family, and if the process continues there may be four before the original parents die of old age. Family composition is thus determined largely by whether a newly married couple set up a new household or become members of an already existing household occupied by close kin of either the bride or groom.

Extended and expanded families

The terms extended and expanded are used to refer to various types of nonnuclear family. A "stem" family, for example, is an extended family generated by the rule that only one child shall remain at home after marriage. Found in parts of rural Europe and Japan, it is compatible with reliance on a family farm that is adequate to sustain a married couple, plus their aged parents and young children, but not large enough to be subdivided. The child who remains at home is usually the eldest or youngest son. In Japan a young man was sometimes brought into the household if there were no sons; he eventually married the daughter of the couple and took over leadership of family and farm. With a stem-family system, provision has to be made for those forced to leave their natal household. Daughters are provided with dowries at marriage, and sons are either given chattels or money or have to find a place outside the system by pioneering new land or taking nonagricultural employment.

If the economy is less diversified or the agricultural pattern more flexible, other kinds of extended family that allow more or all sons or daughters to remain at home are possible. In studies of industrial societies, the term extended family usually refers to a domestic family plus close relatives living elsewhere, but in nonindustrial societies this grouping is called an expanded family. In nonindustrial societies, an extended family is a single domestic unit. Some extended-family households among the Pandits of Kashmir, for example, consist of an old married couple, all their married sons, their unmarried sons and daughters, their sons' wives, and their sons' children. In this system, women leave their natal household at marriage to join their husbands' households, usually coming under the domestic control of the husband's mother. When the senior couple die, the household may continue intact, forming a fraternal extended family in which the senior men are brothers. A generation later the senior men would be first cousins, but these units often break up before this happens, owing to quarrels over the distribution of resources, the rearing of children, and other domestic issues. Though the domestic unit may split, the group may persist as a land-holding unit or even, as in India in modern times, as a corporation owning industrial property.

Though birth and marriage account for most family recruitment, some societies, particularly in Eurasia, allow for the acquisition of family members by means of the adoption of both children and adults. In ancient Rome a householder without an heir would adopt a boy or unmarried man from another family as his son. The Japanese custom mentioned above was similar, though in this case the "adopted" son and heir, unlike a son by birth, had to marry the man's appointed daughter. In many modern systems of adoption infants are adopted because the mother cannot care for them or because the adopting parents desire children, rather than for reasons of inheritance. The transfer takes place in infancy and the identity of the infant's natural mother may be kept secret. The adopted child thus derives all its social position from the adopting parents. Adoption is distinct from fostering, in which a boy or girl lives with and is cared for by a family but does not abandon its links with its own parents, deriving its social position from them alone.

SEXUAL ACTIVITY AND MARRIAGE

Finding a spouse. Getting married may be a long process. When men and women choose for themselves whom they marry, there is usually some institutionalized way in which those eligible sort themselves into compatible pairs. These customs range from computerized dating to, for example, "carrying leg," as was found among the Kuma of New Guinea, among whom groups of girls summoned young men of their choice to sit in the evening with them in a hut with legs intertwined while they talked, gossipped, and made jokes. Courting activities are often organized by peer groups of youths and girls, though successful courting leads to the formation of individual pairs and the breakup of the group. In some societies a couple may be expected to set up house together; if the trial goes well, the marriage is recognized or ratified. In other systems, as in much of India, bride and groom may not have seen one another before the day they are irrevocably married; their kin have made all arrangements for them. In many societies girls are promised in marriage while still infants, and sometimes a woman may promise a daughter as yet unborn.

Wedding practices

Social recognition of a marriage may depend on the performance of a public wedding ceremony. In peasant societies particularly, weddings are major festivities and may entail greater expense than any contributions that have to be made to endow the couple or to pay for the bride. In contrast, there may be no recognized ceremony or ritual. In some Australian Aboriginal societies, for example, a couple started to live together without marking the event, though their marriage might have been the legitimate consequence of a bestowal made many years previously. In general, ceremonies mark the change of status from unmarried to married, as well as the formation of a particular marriage link; hence first marriages may be celebrated differently from subsequent ones. Living in the same household or copulation may precede or follow the wedding by many years. In the West Indies it is common for couples who have lived together for many years and acquired children to hold an expensive wedding and go through a legal marriage ceremony.

Monogamy and polygamy. In many societies marriage is monogamous: a man may have only one wife at a time and a woman only one husband. This exclusiveness may be permanent, as in those Indian cultures that have prohibited divorce and the remarriage of widows. Polygamous systems, on the other hand, allow either polygyny, the marriage of a man to several women at the same time, or, less commonly, polyandry, the marriage of a woman to several men, often a set of brothers (adelphic polyandry). Both polyandry and polygyny may occur in the same society, and in some Himalayan societies a man who shares one wife with his brothers might also be married exclusively to a second wife of his own.

Some cultures enjoin equality of status for cowives, while others allow for ranking, with one wife perhaps designated as mother of the man's heir. Among the Ngoni of south central Africa, for example, an important man might have several bevies of wives, each containing half a dozen or so women arranged in an order of seniority. Some pairs of cowives in the same bevy, usually sisters, were specially linked, the children of the junior woman being treated as if they had been born to her linked senior.

Concubinage

A distinction more radical than ranking is drawn in some cultures between wives and concubines. Thus, among the Jews in biblical times a man could take a female captive as his concubine or could buy the services of a free girl. Despite her own lower status, the children of such a woman had status equal to the children of a wife. Similarly in China, the sons of concubines enjoyed equal status. In general, the relation of a man to a wife provides the basis for a link between his group and her natal group, whereas with a concubine there may be no significant relation with her natal group, whether or not her children enjoy equal status.

But whatever social customs or practices are used to bring potential spouses together, and however many spouses an individual may take, there are in all societies certain restrictions on the choice of sexual and of marriage partner. These are known as incest and exogamy rules.

Incest and exogamy. In all cultures copulation is the diacritical activity in the marriage relation. In most cultures it is also regarded as an essential or contributory concomitant of the growth of the fetus, and hence underlies the mother–child relation.

Other basic human activities—sleeping, eating, talking, touching, working—may be culturally oriented, but copulation is particularly subject to cultural control. The nearest approach to a universal rule found in all known human cultures is the incest taboo—the prohibition of sexual intercourse between a man and his mother, sister, daughter, or other specified kin. As with the domestic family, this widespread occurrence of incest prohibitions suggests that they developed early in human evolution. And it seems likely that the two developments were interdependent.

Origins of the incest taboo

Evidence for sexual avoidance among wild nonhuman primates is hard to collect, but avoidance between mother and son is reported for Japanese macaque monkeys. Incest prohibitions may have been adopted by early man in response to evolutionary pressures. If, as now, girls reached sexual maturity before they could fend for themselves, avoidance of intercourse between siblings and between father and daughter may have kept births within the family in line with its economic productivity. Inbreeding, some writers argue, is genetically disadvantageous to slow-maturing animals bearing single offspring, so that a human group that, for whatever reason, adopted incest prohibitions would be favoured, even though the genetic selection processes at work remained unnoticed. Unregulated sexual competition within the family would threaten its stability, though some marriage systems allow a woman to share her husband with a daughter by an earlier marriage or a man to share his wife with a son by a different mother. Some writers have argued that socialization cannot be effective unless erotic impulses are frustrated within the family and are directed outside it, while others stress the political and social advantages to the family that follow from the alliances made with other groups as growing children seek sexual satisfaction elsewhere.

The inculcation of incest prohibitions is a part of the socialization process and once established tends to perpetuate itself. Some cultures treat incest as a terrible sin; others regard it lightheartedly. The three basic modes of incest—with mother, sister, or daughter—are not classed together in all cultures, and there may be no indigenous concept corresponding to "incest" as used in English. There is certainly no innate aversion to intercourse with sister or daughter, though some evidence from kibbutzim and elsewhere suggests that common rearing may diminish sexual attraction. Incestuous intercourse occurs in most societies, and very small isolated groups that see their survival threatened may abandon incest prohibitions. Sexual intercourse between siblings or between father and daughter has been permitted or required on ritual occasions or for small elites, as, for example, between the pharaoh and his sister-wife in ancient Egypt.

Exogamic rules

Incest prohibitions are associated with exogamic rules (injunctions against marriage between kin), but they are not necessarily coincident. Copulation may be tolerated where marriage is forbidden. In a few types of relation classified as "marriage," copulation is impossible, as, for example, when the "husband" is dead or is a woman; but in the great majority of marriages copulation is possible and legitimate, and may be a duty as much as a right. Incest prohibitions usually apply to only a portion of those kin whom an individual is prohibited from marrying. Incest rules everywhere extend beyond the basic triad of mother, sister, and daughter, but these extensions vary from one culture to another and cannot be explained by the universalist evolutionary arguments mentioned above. Exogamic rules are often extended further still.

In most cultures marriage is possible only with the cooperation of kinsfolk and may require state approval. Incest rules aim at prevention, whereas exogamic rules direct marriage choice away from kinsfolk; hence the domain and content of the two sets of rules differ. The Tallensi of Ghana, for example, considered copulation with a son's wife as terrible as copulation with a man's own mother. The son's wife was addressed by the same kin term as the man's mother, and if the son died the man could not marry her. A man could also not marry various distant cousins who belonged to his clan and whom he addressed as "sister," yet sexual intercourse with them was not considered incestuous or disgraceful.

Rules of incest and exogamy force men and women to look outside the domestic family for sexual satisfaction and marriage partners. The French cultural anthropologist Claude Lévi-Strauss and others have argued that for this reason the adoption of the concept of incest marks the beginning of human culture. Yet it is exogamy rather than incest that leads men to form enduring links with other groups based on the publicly recognized relation of marriage, rather than on copulation, a private activity.

Marriages between kin. The study of marriages between kin has been a major concern of anthropologists for many decades and is still controversial. In some kinship systems marriage with any relative is prohibited, but in others marriage with a specified relative is required or preferred. Sometimes such marriages are observed to be frequent without being explicitly preferred. The preferred marriage partner is often a "cross-cousin" (mother's brother's daughter or father's sister's daughter) or a parallel cousin on the father's side (father's brother's daughter). Lévi-Strauss and others have explained preferential marriage patterns in terms of the principle of reciprocity, a basic element in many forms of exchange in social life. Marriage is viewed as a form of exchange involving the circulation of women. In a system of restricted exchange, two groups interchange women as wives, while in generalized exchange, group A gives its women as wives to group B, which gives its women as wives to group C. Some societies may thus be broken down into "wife-giving" and "wife-receiving" groups.

Secondary marriages to kin

Secondary kin marriages arise out of interpersonal rights and obligations. In some systems, when a woman's husband dies, she remains married to the dead man but is cared for by his younger brother or other relative; any children she bears by her pro-husband are treated as if they had been fathered by the dead man. In some parts of Central Africa the marriage relation persisted after the husband's death in the custom whereby the "widow" was required to copulate with one of his close kinsmen to cleanse herself of the "pollution" or defilement caused by his death; she was then considered to be divorced and was free to marry again. The term levirate, often used loosely for widow inheritance, is best reserved for the custom, mentioned in the Christian Bible, whereby a younger brother of a married man who died childless is required to copulate with the widow so that the dead man may gain an heir posthumously. This arrangement differs from widow inheritance, whereby a widow becomes the true wife of her dead husband's kinsman, any children she then bears counting as the children of the new husband. Among the Nuer of the southern Sudan a widow who did not wish to live with one of her dead husband's kin usually became a concubine and accepted lovers, any children she bore counting as offspring of the dead husband. The Nuer carried the principle of the levirate a stage further: if a man died unmarried, one of his brothers or brothers' sons was expected to take a woman, not as his own wife but as wife of the ghost. The pro-husband lived with and looked after the woman, but their children were regarded as children of the dead man. These ghost marriages were almost as numerous as ordinary marriages between "live" men and women. Less frequently, a barren Nuer woman owning cattle took a wife or wives as if she had been a man. Someone was called in to beget children and provide masculine assistance around the home, but the children born by the wives called the barren woman "father" and traced connections through her to her kin as if she had been a man.

A man's first marriage may influence his range of choice in subsequent polygynous marriages. Several cultures prohibit a man from marrying two women who are sisters, either simultaneously or serially, whereas in other cultures, this is a preferred arrangement. Sororal polygyny consists in marrying two or more sisters, whereas the term sororate is best reserved for the arrangement whereby the kinsfolk of a married woman who dies or who is barren have to provide another wife to supplement or substitute for the first.

Marriage, alliance, and property. Exogamic and endogamic rules provide a framework within which marriage links are formed. These links may sustain, or may be derived from, distinctive patterns of kin connections between individuals and between groups.

In so-called dynastic marriage, typical of medieval Europe, political advantage was sought through the transfer of dowry and other property and through alliances between rulers created or strengthened by affinity. Similar "political" marriages occur in some nonindustrial societies. In the Trobriand Islands in the South Pacific, a chief married a woman from each of his villages, usually a sister or close relative of the village headman. Among the Lele of Central Africa, about one woman in ten was a "village wife." Rights to receive cooked food from a village wife and to sleep with her in her house were shared by several men in the village; and any villager, but not an outsider, had the right to copulate with her in the forest. A chief who visited a village and collected tribute had to provide a daughter or granddaughter to become a village wife, thereby augmenting the weak political link between himself and the village with a tie of affinity. Elsewhere in Central Africa, relations between villages were expressed in the idiom of kinship. In the Luapula Valley (Zambia) village headmen were said to be brother, sister's son, or son-in-law to one another. Whoever became headman took over unchanged the quasi-kin relations of his predecessor. These relations derived from true kin connections said to have prevailed at the time the various villages were founded and symbolized the contemporary pattern of intervillage alliance.

Marriage is typically accompanied by a prestation (payment) or transfer of property or the performance of services. Transfers may begin with bestowal or betrothal and may continue for years after the wedding. Part of the transfer is horizontal, between the parents or other kinsmen of the groom and the kin of the bride, and although goods may pass in both directions, one side often gains more than the other. Another part of the transfer may be vertical, from one or both groups of kin to the new couple, enabling them to establish a properly endowed household. The payment, prominent in many African systems of marriage, made by the groom or his kin to the kin of the bride is known as "bridewealth," whereas "dowry" usually refers to the contribution made by the bride's kin to the newly married couple or to the bride herself as her personal endowment. When bridewealth is large, a man with many sons is at a disadvantage, whereas if dowries are large, many daughters are a serious burden. Among the Lovedu of the Transvaal, who paid substantial bridewealth, the cattle received at a woman's marriage were used to secure a wife for one of her brothers, establishing an enduring relation between the man and his cattle-linked sister. Some kinds of exchange marriage are favoured as a way of avoiding large transfer payments, as for example among Palestinian Arabs, where a man and his sister might marry a woman and her brother.

Dowry and bride-wealth

Prestations may be fixed by custom, be variable but not a matter for bargaining, or may be subject to bargaining. The various items may relate to specified aspects of the marriage, with specific payments for the virginity of the bride or for the children she has already borne. Prestations may be made to specified relatives, as, for example, to the bride's mother for the pains she felt in childbirth. One set of payments may establish the groom's rights over his bride and his right to claim damages from anyone else who copulates with her, whereas others establish rights attaching to the groom's kin group all the children the bride bears, possibly including those conceived in adultery. Sometimes labour service is performed by the groom or his group for the benefit of the bride's group and may be perceived as compensation for the loss of the bride's labour. In several societies a newly married man has to live with his wife's kin group and work for the group for a few years before he is allowed to take her back to his own group to live.

Adultery and wife lending. Marriage and concubinage, relations that may be combined with mother–child links to generate other kin relations, must be distinguished from other kinds of association for sexual purposes, such as prostitution and various more permanent arrangements whereby a married woman takes a recognized lover. If the marriage bond is perceived as giving exclusive sexual rights to either spouse, then copulation with another person must be viewed as a violation of the marriage (adultery). Many cultures distinguish acts of adultery according to the relation between the adulterer and the woman's husband. Thus, among the Tallensi of Ghana, adultery with the wife of a brother was regarded as incestuous and with the wife of a distant clansfellow as reprehensible but not incestuous, whereas adultery with the wife of a member of an unrelated clan led merely to an attempt by members of the cuckold's clan to seduce a woman married to a member of the adulterer's clan so as to make things even.

Eskimo wife exchange

Sometimes a woman is required, by virtue of her marriage tie to one man, to copulate with another. Among the Eskimos of north Alaska, for example, two men who were trading partners or who belonged to the same whaling crew might exchange wives for a few days. This exchange created a relation of mutual assistance between the children of the two married couples. A more widespread custom is for the younger unmarried brothers of a married man to be allowed to copulate with his wife, any children she may bear being treated as his own.

Divorce. The frequency and ease of divorce varies greatly. It may be nonexistent, as in the Republic of Ireland, or it may be so commonplace that the majority of marriages end in divorce. Among the Bemba of Zambia, for example, it was recognized that in a long succession of divorces and marriages a man might marry the same woman twice; a special verb denoted marrying in this way. In industrial societies the welfare of the children of a divorced couple is often specially provided for, while in nonindustrial societies the important consideration is usually which kin group the children are to join. If children normally join their mother's group, divorce tends to be relatively easy and frequent, whereas if the woman and her children are all fully incorporated into the husband's group, divorce is likely to be difficult. Prestations made at the time of marriage may have to be returned before a divorce can be settled, but large prestations do not necessarily prevent divorce. Divorce is often easier when there are no children, and barrenness may be an adequate ground for divorce. In some societies the children of a divorced woman have at best a weak link to their father, whereas in others this is not so. Among the Ngoni of Zambia and Malawi, for example, a man divorcing his wife could demand back the prestation made at the beginning of the marriage whereby he gained his marriage rights, while at the same time he could offer to the woman's father or brother a larger prestation to secure the affiliation of the children to his kin group.

Legitimacy. Marriage is sometimes defined as a union between a man and a woman such that children borne by the woman are recognized as the legitimate offspring of both parents. This definition fails to discriminate between wives and concubines in those societies in which the children of wives and of concubines enjoy equal status, and it fails to cover those societies in which children regarded as having been conceived in adultery are treated as the offspring of the adulterer rather than the husband. Some cultures, particularly those emphasizing connections with the mother's group rather than the father's, do not draw a clear contrast between the two kinds of status usually described as "legitimate" and "illegitimate," whether this refers to the child's status in society, to the woman's status as mother, or to the kind of membership the child has in a specified kin group. Legitimacy or illegitimacy usually derives from circumstances prevailing at the time of birth, but change later may be possible. In Britain, for example, a child born to an unmarried woman may subsequently be legitimized by her marriage to its physical father.

KINSHIP AND THE COMMUNITY

Kinship and residence. The domestic family is a residential unit, the spatial distribution of which may be determined to a large extent by kin relations, "Virilocal" residence occurs when a newly married couple live near the groom's kin, "uxorilocal" when they live near the bride's kin, and "neolocal" when they choose where to live without reference to where their various kinsfolk live. The older terms patrilocal and matrilocal are still sometimes used for virilocal and uxorilocal respectively but more properly refer to residence near the kin of the father, or of the mother, of either bride or groom; "avunculocal" refers to residence near a mother's brother. The two sets of prefixes may be combined, "matriuxorilocal" referring to residence near the bride's mother, and so on.

The universal prohibition against marriage with a parent or sibling usually entails a move at marriage out of the natal family for at least one of the couple. Regular moves may occur before as well as at marriage. Among the Haida of British Columbia, for example, a young man would leave his parents' house to join the household of one of his mother's brothers. When he married, his wife joined him there.

Rules of residence regulate access to fixed resources such as houses and land. Kinsmen in many societies live together so that they can cooperate readily for hunting and gardening, for defense and attack. Regular patterns of postnuptial residence generate a local community composed of kin and of inmarrying spouses and children. About four out of five married men among the Tiv of Nigeria, for example, lived patrivirilocally, and most of the adult men in any small community were closely related to one another through their father's kin. Among the Yao of Malawi, on the other hand, most married men lived matriuxorilocally, and the core of the local community was made up of married women related to each other as sisters and daughters, plus a resident brother living matrivirilocally with responsibilities toward his sisters. Whatever the form of the larger society within which it is enclosed, the local community is thus typically composed of several domestic families, nuclear or extended, whose members are related to one another by kinship in various ways.

Customary behaviour between kin. Within the family there are tensions of many kinds, and outside the family many peoples have the maxim, "We marry our enemies." Different groups have adopted different ways of attempting to regulate the conflicting interests embedded in the various kinds of kin relations. Particularly widespread are customs formalizing the relation between a man and his wife's mother. In Aboriginal Australia, for example, a man could not touch or talk to his wife's mother. In Polynesia close contact between a man and his sister was usually forbidden. Other traditional specifications of proper behaviour between kin either enjoin respect, as, for example, between a man and his mother's brother, or allow familiarity, as, for example, between a man and his granddaughter. In certain relations, typically between brothers-in-law, joking is encouraged or expected. Similar forms of enjoined respect, avoidance, or joking are customary between groups, particularly between groups based on kinship. The distribution of customary respect and familiarity, avoidance, joking, and the like over the different kinds of kin relations varies from one system to another, but in general each individual finds himself with some relatives he must respect, others who must defer to him, and yet others with whom he can behave with familiarity. Lévi-Strauss argues that the four relations of brother, sister, father, and son form the "true atom" of kinship, constituting a balanced quadruplet, but this analysis has been challenged by others who look for balance in a wider set of relations.

Procreation and paternity. The contrast between the material basis of genetic connection and the cultural or symbolic basis of kinship becomes obvious when attention is turned to kin groups larger than the domestic family. Relations with outsiders are recognized or denied not on grounds of genetic similarity or diversity but because of indigenous principles of social organization. In analysis, three kinds of fatherhood are distinguished. The genetic father is the one who actually supplies the spermatozoon that fertilizes the ovum. The "pater" or social father, usually the mother's husband and also the genetic father, is the person through whom the child acquires, at least in part, its position in society. The "genitor" is a person believed to have contributed essentially to the growth of the fetus in the mother's womb. There may be more than one genitor, particularly among those peoples who believe that continued copulation is necessary to sustain fetal growth. Thus the Iban of Sarawak believed that if a woman copulated with several men during her pregnancy, one would contribute an arm, another a leg, and so on. Similar beliefs have been reported from Melanesia. The genitor of an exceptional individual may not be a man at all, as seen, for example, in the medieval belief that Christ's conception in Mary's womb was achieved by the Holy Ghost through her ear. The alleged ignorance of the fact of physiological paternity among certain groups, particularly the Trobrianders of Papua and the Aboriginal inhabitants of Australia, has created much controversy. Some commentators have argued that the assertion that copulation does not cause conception is an expression of religious dogma rather than of ignorance. Others argue that there was true ignorance even when it was believed that copulation was merely a necessary precondition for the intervention of a spirit genitor or that copulation contributed to fetal growth but did not initiate it. The Murinbata of northern Australia believed that the spirit genitor was directed to the mother by an individual known as the "firestick father"; he was often the mother's husband but was sometimes another man or even a woman. In other cultures, on the other hand, it is asserted that the womb is merely a receptacle for semen and that the mother contributes nothing to the substance of the fetus, a view advanced by Apollo in Aeschylus' play ***Eumenides*** (lines 657–661). In many parts of Asia it was believed that bones were derived from semen whereas the soft parts of the body came from the blood and milk of the mother. The Orokaiva of Papua believe that no genitor contributes to a female fetus but that mother and father both contribute blood to a male fetus.

Concepts of fatherhood

The genitor may have specific rights and duties. Thus, in leviratic marriage among the Nuer, when the pater dies his brother is obligated to become the genitor and foster-father of his children and his widow's pro-husband. Among the Lozi of Zambia, the genitor of a child born in adultery has the right to claim it as his own and to become its pater. In other societies, although the woman's husband is pater of all her children, an adulterer's position as genitor may also be recognized in certain contexts.

Motherhood is simpler. Even if the genetic mother is believed to contribute nothing to the fetus, it is she alone who is recognized as the genetrix (female counterpart of the genitor). Attempts to conceal the true identity of a genetrix do not imply ignorance of physiological maternity in indigenous belief. The genetrix may also be the mater (the social mother), but not necessarily so. In the case of the Ngoni linked cowives mentioned above, for example, the children born by the junior took the senior of the pair as their mater.

Relations with a genitor or genetrix receive some cultural recognition, particularly in the common requirement that genitor and pater must coincide for an individual to obtain full status in society, but the major symbolic constructs of kinship are built up from the links with pater and mater. In discussions of kinship the relations "father" and "mother" refer to pater and mater unless otherwise indicated. An individual may apply a kin term inaccurately translated as "father" to several people, but in most societies he will recognize a stronger connection with one of these; this is his pater.

Quasi-kin relations. Many cultures recognize relations analogous to those of kinship. Quasi-kin relations, also known as fictive or ritual kin, define a connection between individuals or groups that has some of the same content as relations of authentic kinship. The most widespread example of a quasi-kin relationship is usually known as ritual coparenthood ***(compadrazgo*** in Latin America) and is derived from the Christian relation of spiritual kinship formed at baptism with a godparent. The sponsor has responsibilities to his or her godchild analogous to those of a secular pater or mater. The bond of spiritual kinship, combined with authentic kinship, led in medieval Christendom to a prohibition against marriage between the person baptized and a child of his or her godparent. Ritual coparenthood is the basis for enduring alliances between peasant families or between peasants and protectors of higher social status.

Ritual coparenthood

Another form of quasi-kin relation is blood brotherhood, found in Africa and elsewhere. Two men, usually not linked by authentic kinship, establish a formal bond of trust and cooperation. Among the Azande of the Sudan the bond was validated by a ritual act in which each swallowed blood of the other. Among the Chagga of Tanzania, pairs of independent chiefs establish treaties of alliance between their chiefdoms by becoming blood brothers. In Japan the custom known as ***oyako-kankei*** or ***oyabun-kobun*** enabled a man to establish a life-long link to a wealthier patron on whom he could depend for economic assistance. Later, the same relations came to be established between foremen and labourers in building firms and other industrial enterprises, particularly those employing ex-peasants. Not all quasi-kin relations may be combined with those of authentic kinship, but all use the language of kinship to designate a relation involving some of the same moral rights and duties.

Nepotism. In societies in which the boundaries of kinship are clear, one moral code may apply to dealings between kinsfolk and another to dealings with non-kin. Although most cultures recognize that there may be hostility and antagonism between kinsmen, particularly between affines and between cowives, kinsmen are in general expected to stand together against the wider community. Yet in all societies a man may from time to time have to balance his duties to his kin against his obligations to non-kinsmen. This conflict of loyalty is particularly noticeable in industrial societies in which the individual bears many social responsibilities. Nepotism, a term derived from the undue preferment given to the "nephews" (that is, illegitimate sons) of popes, is the practice of favouring one's kin in contexts in which criteria other than kinship are supposed to apply. Some cultures institutionalize procedures for avoiding situations likely to generate a conflict. Thus, many universities do not allow close kin to serve in the same department, and in Norway close kin cannot serve together in the cabinet.

THE WEB OF KINSHIP

Descent systems. By virtue of his acknowledged social parentage, a person may succeed to an office, inherit property, and become a member of a descent group. Often many persons (all sons except the eldest, for example) may be excluded from succession; some persons (daughters, for example) may not inherit; but in general everyone can claim membership in some group or set of kin. The English anthropologist W.H.R. Rivers and later writers have stressed distinctions between succession, inheritance, and descent, but in any society all three are closely related and often run in the same direction.

Everyone has two parents, four grandparents, eight great-grandparents, and so on; the number of a person's ancestors is incalculable, and all the descendants of each of these ancestors are potentially kin to him. If no limitation was placed on the recognition of kinship, everyone would be kin to everyone else. Indeed, in Europe at one time it was fashionable to show how thousands of eminent people could trace descent from Charlemagne. But in most societies some limitation is put on the recognition of kinship, so that a person regards at least some of his associates as not his kin.

In urban Europe of the 20th century, for instance, it is rare to find a man who will associate with his wife's brother's wife's mother, even if they might recognize some link of descent or affinity and even though specific kinship terms may exist for such categories. On the other hand, in Australian aboriginal society, where numbers were small, everyone with whom a person came into contact was regarded as a kinsman; social intercourse was impossible until the kinship connection between them had been ascertained. In China, despite the much larger number of persons who were in social contact with one another, there was a large-range kinship system, and kinship through males was recognized widely.

It is characteristic that kinship may be recognized further in one direction than in another, so that, for example, links through men may be remembered while those through women may be forgotten. Or the reverse may hold. People tend to remember or forget the names of their relatives in conformity with the directions and extent to which they recognize kinship.

Patriliny and matriliny

Unilineal systems. One method of limiting the recognition of kinship is to emphasize the relationships through one parent. Such unilineal kinship systems, as they are called, are of two main types: agnatic or patrilineal systems, in which the relationships through the father are emphasized; and matrilineal systems, in which maternal relationships are stressed. Under patriliny, a man (or woman) is linked to his patrikin—*i.e.*, his sons and daughters, his sons' children, his sons' sons' children, and so on; but he is in certain respects cut off from his daughters' children, his sons' daughters' children, and so on, for these are linked to their own agnatic kin. The people who are linked to him are thereby linked to each other and may form a group persisting through the generations, possessing its own internal organization and specific interests. An enduring group of this kind is called a patrilineal lineage or patrilineage.

Each male member of a patrilineage is the potential founder of a branch or segment of the original lineage. In some societies, each lineage is thus a segmented group, dividing in turn into major segments, each of which is divided into lesser segments, and so on. Groups of this kind may be kept in being by the common interests of their members in rights to land and in the control of marriages of members of the group, by obligations of mutual assistance, or by adherence to an ancestor cult. In some stateless societies, lineages wage war and maintain law and order.

An emphasis on matrilineal descent gives rise to groups of matrikin, consisting of a woman, her sons and daughters, her daughters' children, her daughters' daughters' children, and so on. Matrilineages are usually not so highly segmented as patrilineages. Whereas at least the smaller segments of a patrilineal system are formed around a core of men living in close proximity, two modes of residence are found with matrilineages. The local group may be organized around the women of a matrilineage, consisting of them, their husbands, and their children; or it may be based on the men of the matrilineage, consisting of them, their wives, and their children. In the first instance, the group of local women may require the counsel and leadership of a man of their matrilineage, who must live locally with his family; conversely, some of the women of the matrilineage may be married to men who are acting as local leaders for their own sisters; the kinship basis of the local group is then complicated, and it may split into several units of kin. Similarly, in the second instance, competition among brothers for the allegiance and support of their sisters' sons may split a matrilineal group of organized coresident men. In both instances, the larger matrilineal segments tend to disappear as effective social units. The same arguments do not apply to patrilineages, for a local group of males does not usually need a female leader, while quarrels between brothers merely emphasize the lines of internal segmentation.

In systems of double descent, patrilineal and matrilineal principles are found in the same society, each in its specific context. Thus, among the Yako of Nigeria, a man lives with his male patrikin, after an initial period of living with his wife's relatives, and cultivates land controlled by his patrilineage. He also belongs to a matrilineage, whose members are dispersed but who provide priests for certain cults. He inherits movable property from his matrikin and is most closely connected with their marriages. Unilineal descent, either patrilineal or matrilineal, can generate organized groups persisting through the generations. In double unilineal systems, there are two series of enduring groups, and each person belongs to one or more groups in each series.

Cognatic systems. Unilineal systems differ radically from cognatic systems (those in which everyone has obligations and duties of much the same kind toward both his paternal and maternal kin). In patrilineal descent, an individual is allied to his father's brothers' children but not to his mother's sisters' children; in matrilineal descent, the reverse holds. In a double unilineal system, he is allied to both these kinds of parallel cousins but usually in different contexts. In no kind of unilineal system does he share group membership with his cross-cousins—*i.e.*, his father's sisters' children and his mother's brothers' children. In a cognatic system, however, he is in some sense allied to all his cousins, with a possible division between, on the one hand, paternal or patrilateral cousins (*i.e.*, father's brothers' children and father's sisters' children) and, on the other, the maternal or matrilateral cousins (his mother's siblings' children). The kinship system of most European countries and the United States is cognatic.

Whereas unilineal systems give rise to groups with fixed membership, cognatic systems do not. An individual has his own cousins, but some of their cousins are strangers to him; his set of cognates overlaps but does not coincide with theirs. One of the best known cognatic systems is that of the Germanic peoples, among some of whom compensation paid after homicide was provided by all persons who were fourth cousins or closer cognatic relatives of the slayer and was distributed among similar relatives of the dead person.

Variant systems. A variant of unilineal systems is one in which an individual may choose whether he will belong to his father's group or to his mother's, but he cannot simultaneously belong to both, at least not in the same degree.

Some societies do not apparently display the formal regularities of either unilineal or cognatic kinship. Persons are free to choose whether they ally themselves with their father's relatives or with their mother's or with those of a spouse, and they may shift from one to the other. Omnidirectional kinship systems, as these are sometimes called, generate impermanent groupings of kinfolk, each grouping containing kin of various kinds. The emphasis is on person-to-person links of diverse types through which new members gain a foothold in the unit, rather than on the rights and duties associated with a specified kinship link. Every person has a large number of kinship links that he may or may not utilize; not all people utilize the same kind of link, and every cluster of kinsfolk can recruit its members from a large number of other units.

Control of resources and inheritance. The mother–child relation entails the division of members of the family into different generations; in this sense, families are stratified rather than homogeneous units. Although a son may be older than his father's youngest wife, generation and age usually run parallel. Thus stratified, the family is the centre for the transmission of economic, political, religious, and other attributes from one generation to the next.

Inheritance often occurs only after the death of the person controlling resources, but in some societies, as in rural Ireland and Norway, when the main asset is a peasant farm, it is customary to take over control long before death; the transfer then includes provision for the care and maintenance of the senior person. In literate societies, an individual is usually able to dispose of at least part of his property to nonrelatives by writing a will or testament, but he may be legally prevented from cutting off his wife and children without a penny. If he dies without making a will, his estate is distributed according to formal rules of inheritance. In nonliterate societies, there is little or no provision for testamentation, and all property is subject to rules of inheritance. In many societies, there are constraints on the disposal of property from one living person to another, so as to protect the rights of potential heirs.

Partible and impartible inheritance

Inheritance may be impartible—*i.e.*, in which all a man's or woman's possessions pass to a single heir—or partible, in which possessions can pass to a plurality of heirs, each of whom gets a part. Different kinds of property may be subject to different rules. Among the LoWiili of Ghana, for example, a man's cattle passed at death to certain kinsmen of his mother's brothers, while his house passed to his eldest son. With partible inheritance, heirs may be treated alike or some may be favoured. In some Islamic societies a man's possessions are divided equally among his sons, whereas in others his daughters are entitled to half shares. In Norway, a widow receives her half of the matrimonial estate and inherits a quarter of the other half, the remaining three-quarters of this half being divided equally among the children. All systems have substitution rules so that if, for instance, a man dies childless, some other specified relative becomes his heir. The range of eligible substitutes may be defined, and if there are no available relatives within the range, the property may pass to non-kin—a chief or king, for example.

In patrilineal systems of inheritance a man's property usually passes to his son or sons. In medieval England three versions existed in different parts of the country; gavelkind in Kent, in which all sons inherited equally, and either primogeniture (eldest son as sole heir) or ultimogeniture (youngest son inheriting alone) in central England. Some agnatic systems give priority to a dead man's surviving brother as heir, so that the transmission of property from one generation to the next is delayed until the last member of the senior generation dies. These lateral, rather than vertical, systems are often found together with a rule whereby a man's heirs, sooner or later, are his sister's children rather than his own children. A woman without brothers may, in some agnatic systems, inherit as if she were a man. A married woman is, in many societies, so closely identified with her husband or her husband's group of kin that anything she owns may be treated as if it had been her husband's, but in other societies she may, despite her marriage, remain anchored in her own group and transmit her property to her brother's children. Typical of Africa is a system whereby property owned by men is transmitted to men (either sons or brothers or sisters' sons) while women's property passes to sisters or daughters. More commonly in Europe and Asia, property is inherited across sex lines, by daughters as well as sons.

Consequences of inheritance rules

Various rules of inheritance produce various social consequences. With impartible inheritance, each generation sees a division into those who benefit from inheritance and those who do not, so that economic differentiation is perpetuated, even among close kin. A partible system that rigidly excludes one sex from inheriting may promote the perpetuation of differentiation between large groups of kinsmen but minimizes differences within each group. A diverging system (with no sexual discrimination) gives each individual two sources of inherited wealth, via his two parents, and tends towards equality of distribution. Movable inherited property, whether money, cattle, jewelry, or other chattels, can be distributed spatially, but houses and land can be used only by heirs who live nearby. Hence, if these fixed assets cannot be sold, their inheritance affects family and local organization.

Succession. In kinship studies succession refers to the transmission of office from one incumbent to another, rather than to inheritance, the transmission of control over resources. The same rules of filial or fraternal unigeniture (primogeniture, ultimogeniture, etc.) may apply as with impartible inheritance, but other considerations may also apply, particularly when the office concerns interests wider than those of a domestic unit. Thus, while succession in some monarchies is by filiation, a man being succeeded by his son, in others an office may circulate among several more distant male kin, as in medieval Ireland. Either all certain categories of male kin may be eligible, or succession may be restricted to, say, those whose fathers have held office; choice of successor may be made according to a rule of seniority or by election or armed struggle.

Universal succession, a term taken from jurisprudence, refers to the transfer to someone else of all a man's liabilities, rights, and positions in society. Positional succession refers to the custom whereby every person has a successor. Among the Bemba and other matrilineal peoples of Central Africa, for example, a woman who died was often succeeded by one of her daughters' daughters, who adopted the personal name of the dead woman and referred to her kinsfolk by the terms she used. The guardian spirit of the dead woman was believed to transfer itself to her successor.

Ancestor worship and totemism. In many cultures it is held that after death men and women continue to interact with the living, particularly with members of the kin groups to which they belonged or to which they were linked. If there are unilineal descent groups, shrines to ancestors may serve as markers for the arrangement of lineage segments, as among the Tallensi; in China, ancestral tablets indicated the distinction between one segment and another. Sacrifices and other prestations are made to ancestors. The powers attributed to ancestors vary from one culture to another, but the cult generally provides symbolic legitimacy for the existing social order.

Totemism (*q.v.*), the identification of an individual or human group with a natural species or other feature of nature, is often associated with kinship. In Australia, where totemism was most fully developed, kin groups and categories were associated with totems and with totemic sites at which events of cosmological significance were believed to occur. Some groups believed that the growth of the fetus was caused by the arrival of a "spirit-child" linked to a totemic object which became the "conception totem" of the infant.

Kinship and social evolution. The 19th century saw many attempts to describe the prehistory of social institutions, including kinship. One theorist postulated a progression from promiscuity through gynecocracy (political and social supremacy of women) to father right; another began with promiscuous hordes practicing female infanticide and developing polyandry and matriliny, shifting then to patriliny and finally to monogamy; another emphasized the roots of Indo-European kinship in the patriarchal extended family and claimed that ties of consanguinity gave way to those of territoriality. The American anthropologist Lewis H. Morgan, who used a better analyzed body of ethnographic data than his predecessors, distinguished three main stages of social evolution: savagery, barbarism, and civilization, each with its own distinctive types of social institutions and kin nomenclature.

Although all these schemes came to be discredited, interest in social evolution began to revive in the 1950s. It seems clear that some social forms, such as the nuclear family, are found in societies at all levels of technical development, whereas others, such as highly developed systems of unilineal descent groups, are found neither among hunting and collecting peoples nor in industrialized societies. The types of kin nomenclature associated with unilineal descent groups also are concentrated in middle-range societies. Changes in kinship are not necessarily cumulative, and some writers argue in favour of systematic shifts from one form of filiation or descent or marriage to another, with no irreversible evolutionary changes occurring. Matrilineal institutions, regarded as primitive in most 19th-century schemes, are adaptive to many different social environments, though they are not found with the highest forms of nonindustrial agriculture. They are less frequent among people living by hunting, gathering, and fishing than among agricultural societies and are rare among pastoral peoples. Although industrial development seems to be incompatible with large matrilineal groups, the matrifocal household, with no permanent husband, is common under conditions of industrial poverty. The persistence of nepotism in all areas of modern life suggests that interpersonal kin ties continue to be significant even after kin groups larger than the nuclear family have disappeared. Social mobility does not greatly diminish contacts with extrafamilial kin, and contacts may be maintained by mail or telephone when kin are separated. Differences in religious affiliation are more disruptive of kin ties.

BIBLIOGRAPHY. R. FOX, ***Kinship and Marriage*** (1967), is a good introduction; P. BOHANNAN and J. MIDDLETON (eds.), ***Kinship and Social Organization*** (1968), and ***Marriage, Family, and Residence*** (1968), reprint many relevant articles. W.H. GOODENOUGH (ed.), ***Explorations in Cultural Anthropology*** (1964); and S.A. TYLER (ed.), ***Cognitive Anthropology*** (1969), are also useful. Three collections by single authors bearing on many aspects of kinship are M. FORTES, ***Time and Social Structure and Other Essays*** (1970); J.R. GOODY, ***Comparative Studies in Kinship*** (1969); and E.R. LEACH, ***Rethinking Anthropology*** (1961). The best known collections of descriptions of specified kinship systems include: A.R. RADCLIFFE-BROWN and D. FORDE (eds.), ***African Systems of Kinship and Marriage*** (1950); J.R. GOODY (ed.), ***The Developmental Cycle in Domestic Groups*** (1958); M. FORTES (ed.), ***Marriage in Tribal Societies*** (1962); G.P. MURDOCK (ed.), ***Social Structure in Southeast Asia*** (1960); J.M. MOGEY (ed.), ***Family and Marriage*** (1963); and R.M. GLASSE and M.J. MEGGITT (eds.), ***Pigs, Pearlshells, and Women*** (1969). There are several articles on kinship in J.A. PITT-RIVERS (ed.), ***Mediterranean Countrymen*** (1963); and among innumerable monographic studies L.R. HIATT, ***Kinship and Conflict*** (1965); T.N. MADAN, ***Family and Kinship*** (1965); and D.M. SCHNEIDER, ***American Kinship*** (1968), stand out. The easiest lead into the complex arguments about Trobriand kinship is via H.A. POWELL, "Genealogy, Residence and Kinship in Kiriwina," ***Man***, 4:177–202 (1969). Sophisticated discussion of kinship in general is found in M. FORTES, ***Kinship and the Social Order*** (1970); I.R. BUCHLER and H.A. SELBY, ***Kinship and Social Organization*** (1968); and R.M. MARSH, ***Comparative Sociology***, pt 2, ch. 3 (1967).

Among the special topics in the field of kinship, evidence from primates is assessed in V. REYNOLDS, "Kinship and the Family in Monkeys, Apes and Man," ***Man***, 3:209–223 (1968). Inheritance is dealt with in R.F. GRAY and P.H. GULLIVER (eds.), ***The Family Estate in Africa*** (1964); J.R. GOODY, ***Death, Properly and the Ancestors*** (1962), and "Sideways or Downwards? Lateral and Vertical Succession, Inheritance and Descent in Africa and Eurasia," ***Man***, vol. 5 (1970); and G.C. HOMANS, ***Sentiments and Activities*** (1962). Adoption is treated in J.R. GOODY, "Adoption in Cross-Cultural Perspective," ***Comparative Studies in Society and History***, 11:55–78 (1969). A recent contribution to the debate about incest is A.P. WOLF, "Childhood Association and Sexual Attraction," *Am. Anthrop.*, 72:503–515 (1970); while divorce and genealogies are discussed by J.A. BARNES in A.L. EPSTEIN (ed.), ***The Craft of Social Anthropology*** (1967); a good discussion of pedigrees is in E.L. PETERS, "The Proliferation of Segments in the Lineage of the Bedouin in Cyrenaica," ***Journal of the Royal Anthropological Institute***, 90:29–53 (1960). M.F. ASHLEY-MONTAGU, ***Coming Into Being Among the Australian Aborigines*** (1937), deals with procreative beliefs; as does E.R. LEACH, "Virgin Birth," ***Proceedings of the Royal Anthropological Institute 1966*** (1967). The kindred is examined in N. YALMAN, "The Structure of the Sinhalese Kindred," ***Am. Anthrop.***, 64:548–575 (1962). The classic exposition of the notion of descent is discussed by Firth and Schneider in the new edition of W.H.R. RIVERS, ***Kinship and Social Organization*** (1968); while other views on the same topic are found in D.M. SCHNEIDER and K. GOUGH (eds.), ***Matrilineal Kinship*** (1961); H.W. SCHEFFLER, "Ancestor Worship in Anthropology," *Curr. Anthrop.*, 7:541–551 (1966); and R.M. KEESING, "Shrines, Ancestors, and Cognatic Descent," ***Am. Anthrop.***, 72:991–1019 (1970). The significance of kin terminology is examined in R.M. KEESING, "On Quibblings over Squabblings of Siblings," *SWest.* J. *Anthrop.*, 25:207–227 (1969), and in many articles in the collections mentioned above. Marriage rules are discussed in CLAUDE LEVI-STRAUSS, *Les structures élémentaires de la parenté* (1967; Eng. tran., ***The Elementary Structures of Kinship***, 1969), and "The Future of Kinship Studies," ***Proceedings of the Royal Anthropological Institute 1965*** (1966); R. NEEDHAM, ***Structure and Sentiment*** (1962); and by HOMANS and SCHNEIDER in G.C. HOMANS, ***Sentiments and Activities*** (1962). Ancestor worship is dealt with in M. FORTES, "Some Reflections on Ancestor Worship in Africa," in M. FORTES *et al.*, ***African Systems of Thought*** (1965); and J.D. MCKNIGHT, "Extra-Descent Group Ancestor Cults in African Societies," ***Africa***, 37:1–20 (1967). Fictive kinship is examined in E.A. HAMMEL, ***Alternative Social Structures and Ritual Relations in the Balkans*** (1968). A good general account of the history of kinship studies is S. TAX, "From Lafitau to Radcliffe-Brown," in F. EGGAN, ***Social Anthropology of North American Tribes***, 2nd ed. (1955); additional information is contained in F. EGGAN, "Lewis H. Morgan in Kinship Perspective," in G.E. DOLE and R.L. CARNEIRO (eds.), ***Essays in the Science of Culture*** (1960).

(Jo.A.B.)

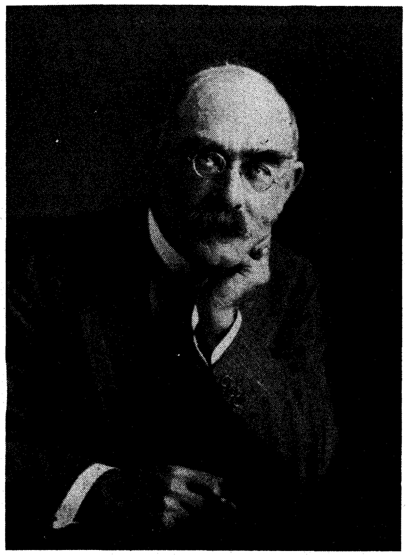
Kipling, Rudyard

Rudyard Kipling, English novelist, short story writer, and poet, is best remembered for his celebration of British imperialism and his tales and poems of British soldiers in India and Burma as well as his children's stories. He was extraordinarily popular in the late 19th and early 20th century, but following World War I his reputation as a serious writer suffered through his being widely viewed as a jingoistic imperialist. Latterly, however, he has again enjoyed high regard.

Joseph Rudyard Kipling was born in Bombay, India, on December 30, 1865. His father, John Lockwood Kipling, an artist and scholar who had considerable influence on his son's work, became curator of the Lahore museum and is described presiding over this "wonder house" in the first chapter of *Kim*, Rudyard's most famous novel. His mother was Alice Macdonald, two of whose sisters married the highly successful 19th-century painters Sir Edward Burne-Jones and Sir Edward Poynter, while a third married Alfred Baldwin and became the mother of Stanley Baldwin, later prime minister. These connections were of lifelong importance to Kipling.

Childhood

Much of his childhood was unhappy. Kipling was taken to England by his parents at the age of six and was left for five years at a foster home at Southsea, the horrors of



Kipling.
Elliott and Fry

which he described in the story "Baa Baa, Black Sheep" (1888). He then went on to the United Services College at Westward Ho, north Devon, a new, inexpensive, and inferior boarding school. It haunted Kipling for the rest of his life—but always as the glorious place celebrated in *Stalky & Co.* (1899) and related stories: an unruly paradise in which the highest goals of English education are met amid a tumult of teasing, bullying, and beating. The Stalky saga is one of Kipling's great imaginative achievements. Readers repelled by a strain of brutality—even of cruelty—in his writings should remember the sensitive and shortsighted boy who was brought to terms with the ethos of this deplorable establishment through the demands of self-preservation.

Kipling returned to India in 1882 and worked for seven years as a journalist. His parents, although not officially important, belonged to the best Anglo-Indian society, and Rudyard thus had opportunities for exploring the whole range of that life. All the while he had remained keenly observant of the thronging spectacle of native India, which had engaged the interest and affection of his earliest childhood. He was quickly filling the journals he worked for with prose sketches and light verse. He published *Departmental Ditties* in 1886, *Plain Tales From the Hills* in 1888, and between 1887 and 1889 brought out six paper-covered volumes of short stories. When he returned to England in 1889, his reputation had preceded him, and within a year he was acclaimed as one of the most brilliant prose writers of his time. His fame was redoubled upon the publication of *Barrack-Room Ballads* in 1892. Not since the English poet Lord Byron had such a reputation been achieved so rapidly. When the poet laureate Alfred Lord Tennyson died in 1892, it may be said that Kipling took his place in popular estimation.

In 1892 Kipling married Caroline Balestier, the sister of Wolcott Balestier, an American publisher and writer with whom he had collaborated in *The Naulahka* (1892), a facile and unsuccessful romance. The young couple moved to the United States and settled on Mrs. Kipling's property in Vermont, but their manners and attitudes were considered objectionable by their neighbours. Unable or unwilling to adjust to life in America, the Kiplings eventually returned to England. Ever after Kipling remained very aware that Americans were "foreigners," and he extended to them, as to the French, no more than a semi-exemption from his proposition that only "lesser breeds" are born beyond the English Channel.

Novels

During his years in America, however, he published *The Light that Failed* (1890), the story of a painter going blind and spurned by the woman he loved; *Captains Courageous* (1897), which, in spite of its sense of adventure, is often considered a poor novel because of the excessive descriptive writing; *Kim* (1901), which, although essentially a children's book, must be considered a classic; and The *Jungle Books* (1894 and 1895), stylistically superb and further proof that Kipling was best at telling a story but inconsistent in producing balanced, cohesive novels.

In 1902 Kipling bought a house at Burwash, Sussex, which remained his home until his death. Sussex was the background of much of his later writing—especially in *Puck* of *Pook's Hill* (1906) and *Rewards and Fairies* (1910), two volumes that, although devoted to simple dramatic presentations of English history, embodied some of his deepest intuitions. In 1907 he received the Nobel Prize for Literature. In South Africa, where he spent much time, he was given a house by Cecil Rhodes, the diamond magnate and South African statesman. This association fostered Kipling's imperialist persuasions, which were to grow stronger with the years. These convictions are not to be dismissed in a word; they were bound up with a genuine sense of a civilizing mission that required every Englishman, or, more broadly, every white man, to bring European culture to the heathen natives of the uncivilized world. Kipling's ideas were not in accord with much that was liberal in the thought of the age, and as he became older he was an increasingly isolated figure. When he died (in London, January 18, 1936), two days before King George V, he must have seemed to many a far less representative Englishman than his sovereign.

Kipling is said to have twice declined the Order of Merit, one of the highest distinctions that can be conferred upon a British subject, and also to have intimated, in 1895, that he would decline the poet laureateship, should it be offered to him.

Assessment

As a poet Kipling will scarcely rank high, although his rehabilitation has been attempted by so distinguished a critic as T.S. Eliot. His verse is indeed vigorous, and in dealing with the lives of soldiers and sailors in particular it breaks new ground. But balladry, music-hall song, and popular hymnology provide its unassuming basis; and even at its most serious—as in "Recessional" (1897) and similar pieces in which Kipling addressed himself to his fellow countrymen in times of crisis—the effect is rhetorical rather than imaginative. But it is otherwise with Kipling's prose. In the whole sweep of his adult storytelling, he displays a steadily developing art. While his later stories cannot exactly be called better than the earlier ones, they are as good—and they bring a subtler if less dazzling technical proficiency to the exploration of deeper though sometimes more perplexing themes. There is much in Kipling's later art to curtail its popular appeal. It is compressed and elliptical in manner and sombre in many of its themes. But that his reputation among informed critics should have declined steadily during his lifetime can scarcely be accounted for except in terms of political prejudice.

MAJOR WORKS

SHORT STORIES: *Plain Tales from the Hills* (1888), 40 stories, enlarged in later editions; *Soldiers Three* (1888); *Wee Willie Winkie* (1888), including "Baa Baa, Black Sheep"; later revised editions include also "Under the Deodars," "The Phantom Rickshaw," and "The Man Who Would Be King"; *Life's Handicap* (1891), including "The Man Who Was," "Without Benefit of Clergy," and "The Mark of the Beast"; later revised editions include also "The City of Dreadful Night" and "The Courting of Dinah Shadd"; *Many Inventions* (1893), including "My Lord the Elephant," "His Private Honour," and "Love o' Women"; later revised editions include also "The Finest Story in the World"; *The Jungle Book* (1894), *The Second Jungle Book* (1895), stories linked by poems for children; *The Day's Work* (1898), including "The Bridge-Builders," "The Ship that Found Herself," "The Tomb of His Ancestors," ".007," "William the Conqueror," "The Maltese Cat," and "The Brushwood Boy"; *Stalky & Co.* (1899), nine school stories, with prefatory poem; *The Complete Stalky & Co.* (1929), adds five stories and a concluding poem; *Just So Stories* (1902), animal fables linked by poems for young children; *Traffics and Discoveries* (1904), stories linked by poems, including "The Bones of Discipline," "Their Lawful Occasions," "They," and "Mrs. Bathurst"; *Puck of Pook's Hill* (1906), and *Rewards and Fairies* (1910), the two volumes combined in *All the Puck Stories* (1935), stories, with accompanying poems, on episodes in English history for children; *Actions and Reactions* (1909), stories linked by poems, including "An Habitation Enforced," "The Puzzler," and "Little Foxes"; *A Diversity of Creatures* (1917), stories linked by poems, including "The Village That Voted the Earth Was Flat" and "My Son's Wife"; *Debits and Credits* (1926), stories linked by poems, including "The United Idolaters," "The Janeites," "In the Interests of the Brethren," and "A Madonna of the Trenches"; *Limits and Renewals* (1932), stories linked by poems, including "The Woman in His Life," "The Church That Was at Antioch," "Beauty Spots," "Aunt Ellen," and "Uncovenanted Mercies."

NOVELS: *The Light That Failed* (1890), version with happy ending serialized in 12 chapters, version with unhappy ending as a book in 14 chapters, revised in 15 chapters 1891, later further revised; *Kim* (1901), for children.

VERSE: *Departmental Ditties and Other Verses* (1886, later enlarged); *Departmental Ditties, Barrack-Room Ballads, and Other Verses* (1890); enlarged edition of *Barrack-Room Ballads and Other Verses* (1892); *The Seven Seas* (1896); *Recessional and Other Poems* (1899); *Songs from Books* (1912, enlarged 1913), collects and revises prefatory and interspersed poems from prose volumes; *The Years Between* (1919).

OTHER WORKS: *From Sea to Sea*, 2 vol. (1899), sketches and travel pieces; *Letters of Travel* (1920), letters and sketches, 1892–1913; *Somethzng of Myself, for My Friends Known and Unknown* (posthumously 1936, dated 1937), unfinished.

BIBLIOGRAPHY. The standard bibliographical work is F.V. LIVINGSTONE, *Bibliography of the Works of Rudyard Kipling* (1927), together with its supplement (1938). The only substantial biography, CHARLES E. CARRINGTON, *Rudyard Kipling: His Life and Work* (1955), is judicious and sympathetic and benefits from a thorough knowledge of Kipling's background in three continents. Three of the best short critical discussions of Kipling's work are those of BONAMY DOBREE in *The Lamp and the Lute* (1929); EDMUND WILSON in *The Wound and the Bow* (1941); and T.S. ELIOT in *A Choice of Kipling's Verse* (1941). Among longer studies two are outstanding: J.M.S. TOMPKINS, *The Art of Rudyard Kipling* (1959); and BONAMY DOBREE, *Rudyard Kipling: Realist and Fabulist* (1967). The first of these, which deals mainly with the prose fiction, is at once enthusiastic and discriminating; the second is a penetrating examination of the philosophical basis of Kipling's work. In *Kipling's Mind and Art* (1964), ANDREW RUTHERFORD has brought together an excellent selection of essays by various hands. R. LANCELYN GREEN (ed.), *Kipling: The Critical Heritage* (1971), affords a valuable conspectus of the development of Kipling's reputation.

(J.I.M.S.)

Kirgiz Soviet Socialist Republic

Located in the heart of the Asian landmass, the Kirgiz (or Kirghiz) Soviet Socialist Republic (Kirgiz S.S.R., Kirgiziya, or Kirgizistan) has an area of 76,600 square miles (198,500 square kilometres) and was inhabited by an estimated 3,294,000 people in January 1975. Almost 44 percent of the population (1970) is composed of Kirgiz, a Turkic people, former nomads whose known history dates back at least to 200 B.C. Kirgiziya became an autonomous *oblast* under the Russian Soviet Federated Socialist Republic in 1924 and an autonomous soviet socialist republic in 1926. It was made a constituent republic only in 1936, with its capital at Frunze (until 1925 named Pishpek).

By the 1970s Kirgiziya had become transformed into a rapidly industrializing state, with the former nomads settled under a collectivized agricultural system. Though urbanization is accelerating, many of the old ways of life, nevertheless, survive.

To the north and northwest, Kirgiziya is bounded by the Kazakh S.S.R. On the southwest lies the Uzbek S.S.R., and on the south, beyond the Turkistan and Zaalay ranges, the Tadzhik S.S.R. On the southeast, the Kokshaal-Tau Range, part of the Tien Shan, forms the border with China. Most of Kirgiziya's borders run along mountain crests, and the few lowland areas—the Chu and Fergana valleys---occur near the places where the borders dip down to meet the plains and deserts that lie to the west and north. A deep hollow in the northeast holds Lake Issyk-Kul, one of the finest natural features in all of Central Asia.

For related information, see TIEN SHAN (MOUNTAINS); ISSYK-KUL, LAKE; INNER ASIA, HISTORY OF; RUSSIA AND THE SOVIET UNION, HISTORY OF; TURKISTAN, HISTORY OF; ALTAIC LANGUAGES; and CENTRAL ASIAN PEOPLES, ARTS OF.

THE LAND

Topography. Kirgiziya is above all a mountainous country. At its eastern extremity, on the Chinese border, rise Victory Peak (Pik Pobedy), the second highest in the entire Soviet Union, and Pik Khan-Tengri, the fourth highest. This region is the core of the massive Tien Shan system, which thrusts on eastward into China; from it, mountainous belts spread westward over most of Kirgiziya. On the republic's southern border appear the Kokshaal-Tau, Alay, and Zaalay ranges, the last named influenced geologically by the vast structural knot of the Pamirs lying to the south in the adjoining Tadzhik Soviet Socialist Republic.

Mountain peaks

In the southwest lies the Fergana Valley. A similar huge hollow lies in the northeast, comparatively close to Khan-Tengri; bounded by the westward-thrusting arms of the Kungey-Alatau and Terskey-Alatau ranges, it contains Lake Issyk-Kul, the clear and deep waters of which are backed by snow-clad peaks. The rugged mountain and basin structure of much of the country, and the high alpine plateau of the central and eastern regions, are separated from the Fergana Valley on the west by the Fergana Range. running southeast to northwest, which merges into the Chatkal Range. The last named is linked to the Issyk-Kul region by a final enclosing range, the Kirgiz.

Apart from the Fergana Valley, the only important lowland areas are the Chu and Talas valleys near the northern border of the republic, the former containing the capital, Frunze. These lowland areas, occupying 15 percent of the total area, are of immense significance for settlement.

Drainage. The crests of the ranges are mantled in perpetual ice and snow. The total area thus clad exceeds 2,500 square miles and includes the 37-mile-long southern branch of the Inylchek glacier, one of the world's greatest. Numerous short rivers, with many rapids, pour from the mountains into the inner basins. The Naryn River, running down into the Fergana Valley and continuing northwestward as a tributary of the Syrdarya, is of major significance, as is the Chu, running parallel to, or forming part of, the northern boundary with Kazakhstan.

Climate. Important influences on the climate are the remoteness of Kirgiziya from the oceans and the sharp change of elevation from neighbouring plains. The republic is surrounded by the Central Asian plains and deserts to the north, west, and southeast, making the contrast with the climate and landscape of its mountainous interior all the more striking. The lower parts of its fringing ranges lie in belts of high temperature and experience the hot, moisture-consuming winds from the deserts beyond. Conditions in the transitional zone above this hot desert region but below the cold high deserts vary greatly according to relief patterns: westward- and northward-facing slopes generally receive more moisture, which increases with altitude. Summer in the valleys is hot and dry, and the mean temperature in July is 82" F (28" C), while the average January temperature is —0.5" F (—18" C). Absolute temperatures have an even wider range, from a high of 109" F (43" C) to a low of —65" F (—54" C). Annual precipitation varies from 7.1 inches (180 millimetres) in the eastern Tien Shan to 30–40 inches (750–1,000 millimetres) in the Kirgiz and Fergana mountains. In the most populous valleys, it ranges from four to 20 inches.

Precipitation

Vegetation and animal life. Except at the highest altitudes, the elevated landscapes of Kirgiziya are characterized by flower-strewn alpine and subalpine pastures. Woodlands are found along the lower valleys and on the slopes of the north-facing ranges; occupying 3 to 4 percent of the total area of the country, they are coniferous and include the striking Tien Shan white spruce. Animal life in the forests is a varied mixture of European and Asian species that includes brown bear, wild pig, lynx, gray wolf, and ermine. The wooded ravines and the valleys of the mountainous steppe regions are the abode of the *arkhar—a* mountain sheep—mountain goats, deer, and snow leopards. Birds are plentiful in the mountains;

rock partridges are found on cliffs and around water sources, and jackdaws and the snow, black, and bearded varieties of vulture live in the alpine areas. By comparison, animal life in the desert regions is scanty: rodents—yellow gophers and jerboas—hares, and a large-eared hedgehog are typical; birds include the desert bullfinch and red starling, and among reptiles are the steppe turtle and some snakes.

The human imprint. Since the beginning of the Soviet period, the pattern of settlement has changed considerably, with large immigration and industrialization. Cities have grown remarkably, with 38 percent of the population clustered in 18 large cities and 47 other urban communities by the mid-1970s, compared with only 12 percent urbanization in the mid-1920s. The densest settlement is in the Chu and Fergana valleys. The capital, Frunze, a town of some 30,000 people in the 1920s, in 1970 registered a population of 430,618. Only 53,059 (12.3 percent) of these people were Kirgiz, however. More than 66 percent (284,676) of the city's residents were Russian, and Ukrainians made up 6.2 percent (26,901). Tatars, Uighurs, Uzbeks, and Jews made up smaller minorities of the Frunze populace.

Patterns of urbanization

Around the edges of the Fergana Valley are five towns ranging in population from 16,000 to 120,000. Clockwise from the north they are Tash-Kumyr, Kok-Yangak, Osh, Kyzyl-Kiya (all coal-mining towns except Osh, a silk-processing centre), and Sulyukta. Another city of similar size is Przhevalsk, at the eastern end of Lake Issyk-Kul.

The modern buildings and planned layout of the new cities contrast with rural life, which differs according to ethnic background. Russian and Ukrainian villages are scattered in the north, together with communities of Dungans, or Hui (Chinese Muslims). There are old-fashioned Uzbek settlements, known as *qishlaq*s, in the south. The Kirgiz kyshtaks, villages built during the Soviet period to speed the transition from a nomadic to a settled existence, are scattered everywhere, as are collective and state farms.

The development of a modern transportation network is breaking down much of the old isolation. Frunze was formerly linked with the Fergana region only by a railway that made a long detour over the lowlands to the west, through Kazakhstan and Uzbekistan, but since 1965 traffic has flowed along the direct Frunze–Osh highway that cuts through the mountains and halves the distance. Alpine terrain still presents a great impediment to the construction of railways, but a network of highways is gradually replacing the winding paths along which people travelled for centuries by horse or camel or on foot.

THE PEOPLE

Ethnic backgrounds. Kirgiziya's population is made up mainly of Kirgiz, Russians, Uzbeks, Ukrainians, and Germans (the last deported to Central Asia from the Soviet west in 1941). The 1,284,773 Kirgiz in 1970 constituted less than half of the 2,932,805 total, Russians more than 29 percent, Uzbeks 11.3 percent, Ukrainians 4.1 percent, and Germans 3.1 percent. Many smaller groups made up the remainder.

The early Kirgiz

The first Kirgiz tribes settled in the Tien Shan region in ancient times and had a long history as one of the great nomadic groups of Central Asia. During the Soviet period, the almost complete abandonment of nomadic life under compulsion, and the replacement of individual farming by state collectivization in agriculture, have caused great changes in Kirgiz society.

Demographic trends. The Kirgiz death rate has diminished to little more than one-sixth that of the early 20th century, while the birth rate has doubled, giving an average annual natural increase of 2.4 percent. Between 1959 and 1970 Kirgiz in their republic grew by 53.5 percent, an average of nearly 4.9 percent each year, including natural and other forms of increase.

In 1970, 85.5 percent of the Kirgiz remained rural dwellers, engaged almost wholly in agriculture; the growing urbanization and economic development in industry have largely bypassed them. At the same time, 66 percent of all Russians in Kirgiziya lived in urban areas, as did 51 percent of Ukrainians, 37.5 percent of Germans, and 36 percent of Uzbeks. These groups, along with Tatars, Uighurs, Jews, and smaller but also considerably urbanized minorities, occupied a disproportionate share of urban-industrial employment in Kirgiziya.

The population density averaged 38 per square mile at the census of 1970, but the interior mountainous regions are virtually uninhabited, whereas Frunze alone accounts for almost 15 percent of the population.

The Kirgiz in Kirgiziya rose from 40.5 percent in 1959 to 43.8 percent in 1970, but demographers doubt that this high rate of increase will be maintained. In the identical period, though the percentage of Russians in the republic's population declined slightly, from 30.2 to 29.2, the absolute numbers of Russians grew from 623,562 to 855,935, a gain of more than 37 percent. Kirgiziya differs from most other union republics in that fairly large numbers of Russian and other Slavic and Soviet western immigrants have entered the farming population and not merely added to the population of the urban areas.

THE ECONOMY

Kirgiziya was formerly wholly agricultural, and stock raising was prevalent. By the 1970s the republic was a source of nonferrous metals, notably antimony and mercury ores, and a producer of machinery, instruments, hydroelectric power, and the products of light and food industries. The mineral wealth of the mountains—lead, zinc, and other metals—is being discovered, Kirgiziyan coal reserves are estimated at about 3,000,000,000 tons, and deposits of petroleum and natural gas have been found in the Fergana region. Extensive irrigation makes possible the expansion of agriculture and represents part of a large capital investment in the republic. During the years of the pre-World War II five-year plans, this investment produced plants for the extraction of antimony and for the processing of agricultural products, and the first metal-manufacturing plants and modernized blast furnaces were introduced. Large sugar factories were built, and emphasis was given to the primary processing of cotton and other fibre crops. During World War II, manufacturing expanded, and in ensuing decades the industries producing machines, building materials, and electric power were developed.

Industry. In addition to ore extraction, coal mining continues to be of basic importance. About 4,400,000 tons were produced annually by the mid-1970s, and new sources in the north and east were being exploited, in addition to the traditional southern reserves. The new deposits include major reserves in the Karakichi region of the inner Tien Shan, where coal layers up to 250 feet (75 metres) thick outcrop onto the surface. Petroleum and natural gas deposits have been exploited in the south, and annual extraction rates in the mid-1970s reached 260,000 tons of oil. Plans call for coal, oil, and natural gas to serve as the base for a chemical industry.

Food-processing and light industries, utilizing local agricultural raw materials, are also of importance in the republic. They include the production of meat, canned meat, fruits and vegetables; woollen, cotton, and silk fabrics; and leather goods. Cotton spinning and cleaning and enterprises producing worsted cloth, knitted garments, and leather and sheepskin coats are particularly significant.

Energy resources

The considerable energy resources of the republic are essential for industrial development. More than half the electric energy, some 4,400,000,000 kilowatt-hours annually, comes from waterpower. Completion of, and greater utilization of the power produced at, the Toktogul station, using the waters of the Naryn River, is one of the pressing industrial tasks.

Agriculture. Industrialization has stimulated the mechanization of Kirgiz agriculture, and many types of machines necessary to cope with the rugged terrain are now manufactured in the republic. Irrigation has been most important in increasing agricultural output; in the Fergana region, for example, where the republic's cotton is produced, the network of canals and reservoirs has been completely restructured.

The leading branches of agriculture are pasture-based stock raising and the cultivation of cotton, sugar beets, tobacco, and opium poppies. Some 80 percent of the total agricultural land is taken up by pastures and by fields used to grow hay. Under the collectivized, state-owned system, the always important stock raising of the Kirgiz is being linked with the cultivation of cereals and industrial crops, silage, and seed grasses. This process is encouraged on every kolkhoz (collective farm) and sovkhoz (state farm) but is most fully developed in the valleys on the republic's borders.

Livestock

Sheep and goats totalled almost 10,000,000 head by the mid-1970s, and milk and beef cattle—especially in the Chu Valley and the Issyk-Kul region—approached 1,000,000 head. Around Lake Issyk-Kul and in other mountain regions, the breeding of fine racehorses was being developed, as was the raising of swine, bees, and rabbits.

Horses are also important in the mountain kolkhozy; not only are they used as draft animals, but also their meat is eaten, and fermented mare's milk. known as koumiss, is the favourite nourishing beverage of the Kirgiz, also used as part of the treatment at mountain health resorts.

By the mid-1970s the total area under crops had reached 3,183,000 acres (1,288,000 hectares), of which about 44 percent was devoted to food grains, 6 percent to cotton, and 4 percent to sugar beets, while the remaining acreage was used for potatoes, vegetables, and fodder (not including hay).

The total cotton harvest approached 232,000 tons annually in the mid-1970s, and yield per acre was high. The annual production of sugar beets, mostly in the Chu Valley, was around 2,050,000 tons, providing work for a half dozen sugar factories and giving that area the Russian nickname *sakharnaya* ("sugar valley"). In the Naukat Valley in the south and also in the Talas Valley of the north, tobacco is cultivated. Horticulture and viticulture are developed in the Chu Valley and the Fergana area, with the mulberry trees of the latter supporting the raising of silkworms.

Hunting and fishing. The sparsely populated interior valleys contain almost inaccessible regions suitable for valuable fur-bearing species of animals. Kirgiz are enthusiastic hunters; indeed, five of the months in the Kirgiz calendar are named after game animals. Although modern firearms and traps have been introduced, there is still a preference for using the trained golden eagles to search out rabbits, foxes, and even wolves.

Lake Issyk-Kul, in spite of a paucity of plankton, is rich in fish. The catch—about 8,000 tons annually—is brought to the shoreline communities almost year round, the exception being in winter, when the fish descend to great depths. Artificial stocking of the lake has included the introduction of trout from Lake Sevan in Armenia, and fish-breeding farms are also used.

Transportation. Before 1924 the only railways in Kirgiziya were two narrow-gauge lines leading from the borders to the coal deposits of Kok-Yangak and Sulyukta. The construction of a line from Frunze across the Chu Valley and over the border to Lugovaya in Kazakhstan joined the north of the republic to the Turkistan–Siberian main railway line and, through it, to the south of the republic and the entire Soviet railway network. In 1948 a link extended the line up-valley from Frunze—it reached heights of 5,000 feet along the remote Boam Canyon—to Rybachye at the western tip of Lake Issyk-Kul. New southern lines reached the coal mines at Tash-Kumyr and Kyzyl-Kiya.

The road system

Highways are nevertheless being developed as the basic answer to the topographic problems confronting land transportation. One main route climbs from Frunze to Rybachye (with extensions along the north and south shores of Issyk-Kul), then swings south across difficult central terrain to Naryn, and then through the high Torugart Pass across the frontier with China and down to the city of Kashgar, in the Sinkiang Uighur Autonomous Region. The other major artery, the "route beyond the clouds," from Frunze to Osh, crosses the Kirgiz Alatau crest through a 10,500-foot (3,200-metre) tunnel. An important southern link is provided by the road joining Osh, via the Alay Pass, to the Pamir region of Tadzhikistan. An offshoot runs eastward through Irkeshtam to Kashgar. The total main and secondary road network in the country approached 12,800 miles (20,600 kilometres) by the mid-1970s, and about 2,523,000 ton-miles (3,684,000 ton-kilometres) of freight moved along the roads each year.

Cargo-carrying and passenger vessels ply Lake Issyk-Kul, whose waters, although surrounded by peaks clad with snow, remain unfrozen in winter. Very strong local winds, raising waves and waterspouts, are an ever-present hazard, and travellers, not unexpectedly, generally prefer to use the modern lakeside highways.

ADMINISTRATION AND SOCIAL CONDITIONS

The constitutional and political framework. The republican constitution dates from March 23, 1937. All significant state functions are subordinated to the central government in Moscow, the capital of both the Russian S.F.S.R. and the Soviet Union. Though sovereignty is claimed for the Kirgiz S.S.R. by Soviet authorities, in fact it exercises no independent control over cultural, economic, social, political, military, or foreign affairs affecting it.

The highest governmental body in the republic is the Supreme Soviet, selected from a single slate of nominees for a four-year term. The Supreme Soviet, which exercises only ceremonial functions, generates an executive and administrative Council of Ministers. Local soviets are selected every two years. The chief legal officer of the Kirgiz S.S.R., the procurator, is appointed by the central government in Moscow.

The Communist Party

The sole authorized political organization and centre of power in Kirgiziya is the Communist Party, a branch of the Communist Party of the Soviet Union (CPSU). Kirgiz constitute a disproportionately small part of the membership throughout the Soviet Union as well as within their own republic. In 1970 they made up 0.6 percent of the entire Soviet Union population but in 1972 only 0.3 percent of CPSU members. More important is the fact that, inside the Kirgiz S.S.R., there were 38,881 Kirgiz party members out of a total membership of 103,028, just 37.7 percent, compared with the 38,847 (also 37.7 percent) Russian members, whose share of the republic's population in 1970 was less than 30 percent. Ukrainians supplied 8.7 percent of Kirgiziya's party members from their 4.1 percent of its population; the Uzbeks, by contrast, provided only 6.1 percent, although they constituted more than 11 percent of the republic's population. These figures reveal both the Kirgiz disadvantage and a general disparity between the proportion of membership in this privileged organization allowed to the indigenous Central Asians as compared to the various immigrant groups.

The Komsomol (Young Communist League), closely tied to the CPSU, counts more than 350,000 youths enrolled in Kirgiziya, a large proportion of them from the ranks of outsiders rather than Kirgiz or other Central Asians.

Labour unions, which enroll about 981,000 members in the republic, exert little influence upon protecting the interests of workers in relations with management. In effect, they serve the government by leading campaigns for higher labour productivity and by supervising granting of vacations to working people as well as access to resorts, sanatoriums, and other facilities.

Education and other services. Between 1959 and 1970 the number of Kirgiz with college degrees rose from 21 to 56 per 1,000 persons in the republic over nine years of age. The ratio of men to women in this category was about three to one in 1970, an improvement from more than four to one in 1959. At the time of the Soviet census of 1970, however, 98.3 percent of Kirgiz over 60 years of age, 91.8 percent between the ages of 55 and 59, and 79.5 percent between 50 and 54 had had no secondary education.

By the mid-1970s, ordinary public schools enrolled more than 820,000 pupils of all nationalities; technical schools and similar institutions with specialized curricula admitted some 42,000. Nine institutions of higher learning enrolled almost 49,000 students from a mixture of ethnic groups. The Kirgiz State University and other principal institutions are located in Frunze, but Przhevalsk, Dzhalal-Abad, and Osh also have institutions of higher education.

Education is provided without direct tuition charge, the institutions being supported by turnover taxes, touching most working people, and levies upon factories and other firms.

Academy of Sciences

The Academy of Sciences, established as a branch of the central body in Moscow in 1954, has three departments and about a dozen research institutes, and there are 50 other scientific and other research establishments.

With the concentration of a growing immigrant population from outside the republic in the cities, housing development has had to undergo great expansion, as have hospitals, maternity and child welfare centres, and other public health bodies. By the 1970s the republic had a ratio of 2.3 doctors per 1,000 population. Medical services are tax-supported.

The mass communications media. Newspapers and magazines are published in both Russian and Kirgiz. Radiobroadcasting also is carried out in both languages, and most of the larger cities and all the districts have studios. There are republican television centres and studios in Frunze and regional studios in nearly all regional capitals. A network of relay lines connects the stations, enabling them to receive programs from Moscow, as well as those of the European Intervision and Eurovision networks.

Cultural life. Kirgiz cultural life has been greatly influenced by the rich oral literary tradition (including epic cycles and lyric poetry) of the region; by the development of a modern literary language and the change from the Arabic alphabet to Roman and finally to Cyrillic (with diacritical markings added) beginning in 1940; by imposition of Marxist-Stalinist-Leninist ideology in all fields of thought and the arts; and by persistent official attacks upon Islām (to which the Kirgiz were converted long ago) and the practice of religion.

A leading figure in drama and the novel today is the Kirgiz writer Chingiz Aytmatov, whose work, especially *Tales of the Mountains and Steppes*, which was widely circulated in English translation, won him a Lenin Prize in 1963. Although Aytmatov writes in both Kirgiz and Russian, the easy accessibility of his prose makes both originals and translations of the stories and longer writings extremely popular. His play, *The Ascent of Mt. Fuji*, written with the Kazakh playwright Kaltay Muhamedjanov, with its discussion of moral compromises made under Stalin, was a sensation when it first appeared on stage in Moscow in 1973 and has since been performed in English in Washington, D.C., and elsewhere. Kirgiz poets strive to accommodate their tradition to the new ideology and its literary guidelines known as Socialist Realism. Folk legacies are handed on by the bards, who recite the long Kirgiz verse epic *Manas*, expressing the fiercely independent outlook of the earlier Kirgiz.

In music, older traditions are represented in ensembles of the three-stringed *komuz*, plucked like a lute, and there is a Kirgiz Philharmonia. There are vigorous folk dance groups, and theatres perform Kirgiz and Russian plays as well as plays translated from other languages.

The Kirgiz motion-picture studio, which was established in 1942, produces feature films as well as newsreels, documentaries, and popular scientific films.

THE OUTLOOK

Under the Ninth Five-Year Plan, for 1971–75, industrial products were planned to increase by almost 50 percent, with manufacturing and light, hydroelectric, and mining industries particularly scheduled for expansion. Completion of the Toktogul hydroelectric station (first power generated in 1975) was to be followed by construction of the large Kurpsay project. Agriculture was also to be strengthened. An expansion of services and consumer goods should result in a rise in living standards.

Lively developments in literature, education, and publishing suggest the maturing of a sizable new Kirgiz intelligentsia. Gains, though slight, in Communist Party membership within Kirgiziya, together with the dramatic Kirgiz population growth, also presage further advancement for the Kirgiz people. For Kirgiz leaders, the primary problem will soon cease to be the modernization of their people and will become the taking on of greater responsibility for their own affairs. This promises to raise serious challenges when the highly trained Kirgiz younger generation cannot find suitable employment in the cities, where most jobs for skilled and educated persons are still filled by non-Kirgiz.

BIBLIOGRAPHY. For modern Kirgiziya, extensive reference lists appear in RICHARD A. PIERCE, *Soviet Central Asia: A Bibliography* (1966), as well as in the following publications, which also cover recent developments in the Kirgiz S.S.R. in detail: EDWARD ALLWORTH (ed.), *Central Asia: A Century of Russian Rule* (1967) and *The Nationality Question in Soviet Central Asia* (1973); and ZEV KATZ *et al.* (eds.), *Handbook of Major Soviet Nationalities (1975).*

(S.N.R./E.Al.)

Kirin

A province of the central (Northeast Manchuria) region in northeastern China, Kirin (Ji-lin in Pin-yin romanization) borders the Primorsky Kray (Maritime Province) of the Soviet Union to its east, North Korea to its southeast, the Chinese province of Liaoning to its south, Heilungkiang to its north, and the Inner Mongolian Autonomous Region to its west. It has an area of 72,200 square miles (187,000 square kilometres) or 2 percent of the national total, and a population registered in 1957 as 12,550,000 or 1.9 percent of the national total. The estimated 1970 population is 20,000,000. The capital (since 1954) of the province is Ch'ang-ch'un. Kirin was created a province in 1907, near the end of the Manchu dynasty, and was occupied by the Japanese Army in 1931, becoming a part of the puppet state of Manchukuo (1932–45). Just before Japan's surrender to the Allies on August 15, 1945, Soviet forces entered the region, dismantled key industrial installations, and removed them to the Soviet Union. Following the withdrawal of Soviet troops, Chinese Nationalists moved in, but by 1948 they had been driven out by Chinese Communist forces.

The landscape. *Relief and drainage.* The province may be divided into three parts: the eastern mountains, the western plains, and a transitional zone of rolling hills between them. Elevation decreases from the highlands in the southeast toward the Manchurian Plain in the northwestern part of the province. The mountains of eastern Kirin take the form of parallel ranges with the Cathaysian or Sinian trend, and are separated by broad valleys. The most famous of the ranges is the Ch'ang-pai Shan ("permanent white mountain") close to the Korean border. Its snow-covered summit is 9,000 feet (2,744 metres) above sea level—the highest peak in northeastern China. The summit is formed by a volcanic crater occupied by a lake called the T'ien Ch'ih ("celestial pond"). The range is the source of three important rivers: the Sungari, the Yalu, and the Tumen.

Mountains

The Yalu and Tumen rivers flow in opposite directions along the Sino-Korean border. The Yalu runs southwest along the border to the Korea Bay; the Tumen down the Ch'ang-pai Shan northeastward to the Sea of Japan. The two rivers are of great strategic importance, guarding the land approaches to northeastern China from the Korean peninsula.

The Sungari River (Sung-hua Chiang) is the major stream of Kirin. It flows for almost 500 miles within the province, draining an area of over 30,000 square miles. Its upper course runs northwest in a series of rapids through heavily forested mountains before it enters the Sungari Reservoir, a man-made lake some 105 miles long and 210 square miles in area, having a storage capacity of 12,300,000,000 tons of water. It irrigates more than 167,000 acres of farmland, and generates 567,000 kilowatts of electricity. Emerging from the reservoir, the Sungari flows past the city of Kirin, situated at the head of navigation of Sungari River and at the geographical centre of the province. At Fu-yii it enters the Manchurian Plain and is shortly afterward joined by its chief tributary, the Nen Chiang, which is in fact larger than the Sungari. It then turns sharply east to run along the provincial boundary for a short distance before it leaves Kirin Province.

The middle section of the Manchurian Plain forms the northwestern part of the province, and constitutes three-eighths of its area. It has a rolling topography, with an average elevation of about 650 feet above sea level.

Temperature and precipitation

Climate. Kirin Province forms a transitional climatic zone between northern and southern Manchuria. The winter is cold and long, and rivers are frozen for about five months; the ice on the Sungari is thick enough to support mule carts. Ch'ang-ch'un, midway between northern and southern Manchuria at 708 feet above sea level, has mean temperatures of 1.6" F (—16.9" C) for January and 74.5" F (23.6" *C)* for July. It has a mean annual precipitation (rain and snow) of 24.5 inches, 83 percent of it during the five warm months from May to September. Precipitation increases southeastward to more than 40 inches in the Ch'ang-pai Shan area, but decreases westward; the Manchurian Plain has only 16 inches.

Soils. There are two main types of soil in the province: podzols in the eastern mountainous region and black earth in the western plains. The podzols occur in several forms and are of both high and low fertility. Central and western Kirin are the areas of the black earths of the Manchurian Plain. Of high fertility and adequate thickness (20–50 inches) and containing a high percentage of organic matter (10 percent and more), they form good arable land. The young alluvial soils along the Sungari and its tributaries also provide excellent land for cultivation. In western Kirin are saline soils, alkali soils, and high-alkali soils of low agricultural value.

Tree types

Vegetation and animal life. The natural vegetation is prairie grass in the western plains and mixed conifer and broadleaf deciduous forest in the eastern mountainous area. The vegetation in the eastern mountains includes tree species such as the Japanese red pine, Manchurian ash, fish-scale pine, larch, birch, oak, willow, elm, and the Manchurian walnut. In the deep mountain interior, virgin forest has been preserved. Tree types are distributed in distinct belts depending mainly on altitude: between 800 feet and 1,600 feet elevation is the deciduous broadleaf belt, mainly mountain willow and thumb; between 1,600 feet and 3,000 feet is found mixed coniferous and broadleaf forest; between 3,000 feet and 5,900 feet occurs coniferous forest; and from 5,900 feet to 6,900 feet is found mountain birch.

Many valuable wild animals and medicinal plants are found in the forested mountain areas. The Manchurian hare, valued for its fur, and some species of rodents such as the rat hamster and the eastern field vole are believed to be peculiar to the Manchurian forest. Among the birds, finches, the buteo hawk, the needle-footed owl, the black and white barrier, and certain species of flycatcher are typical. Among semiaquatic animals, the lungless newts are notable. Certain species of snakes, such as the Schrenk racer, found in the inhabited areas of the Northeast and Korea, live in a semidomesticated state and are used to eliminate harmful rodents in orchards and gardens. The European wild boar, the common hedgehog, the Asian red deer, the harvest mouse, and the field mouse are among the more common Eurasian species. Valuable pelts include fox, chipmunk, the light-coloured polecat, the Manchurian hare, and the sable. The sable population, however, has become very small.

The people. Population growth. The annual rate of growth of population for the period 1957–67 was 3.1 percent, and it was estimated in 1970 that the population of the province had reached 20,000,000, compared with the 11,290,000 of 1953. The rapid population increase is accounted for partly by a mass immigration of farmers and partly by natural increase during two decades of peace and order. There has been some birth control, and the death rate has been rapidly declining in China because of relatively easy improvements in public hygiene and medicine, and, above all, a general improvement in nutrition among the peasants.

Density. The average population density (1967–68) was 236 per square mile—higher than the national average of about 194 per square mile. The Kirin figure is thus slightly higher than that of France and Austria, both of which had densities of nearly 231 per square mile in 1966. A study by the Institute of Geography of the Chinese Academy of Sciences reported that in 1957 the density varied greatly in different sections of the province; some densely populated counties (hsien) in northeast Kirin reached 414 per square mile, and sparsely populated counties of the forested area in the southeast had as few as 26 persons per square mile. More than 30 percent of the population is urban.

Urbanization. Kirin Province is the most urbanized in China: in 1957 the urban population accounted for nearly 30 percent of the total, a rapid advance from the 22 percent of 1949. At the end of 1957, the largest city, Ch'ang-ch'un, had 975,000 inhabitants; and Kirin, second largest, 568,000. Three cities (Ssu-p'ing, Liao-yiian, and T'ung-hua) had 100,000–250,000 inhabitants; four had 50,000–100,000; 25 had 20,000–50,000; and 31 more had between 10,000 and 20,000 persons.

Ethnic groups. The population registration in 1957 revealed that 90 percent of the inhabitants were Chinese, 6.3 percent Korean, and 2.9 percent Manchus. Chinese predominate throughout the province, except in the Yen-pien Korean Autonomous District. In the Yen-pien District, which is contiguous to North Korea, Koreans, who introduced the cultivation of rice to Kirin, form 59 percent of the population. They came originally from Korea. Most of the Manchus live in the central part of the province: in Kirin, in Yung-chi Special District (*chuan-ch'ii*), and T'ung-hua Special District. A few Muslims (Hui) are distributed in the cities and towns of the province, and some Mongolians are to be found in the Pai-ch'eng area in northwestern Kirin.

Administration and social conditions. Kirin province is one of the 29 primary administrative divisions of the People's Republic of China. The province is divided into eight subprovince level areas consisting of five special districts (chuan-ch'ii), one autonomous district (*tzu-chih-chou*), and two municipalities (shih) under the direct control of the province. The two municipalities are Ch'ang-ch'un Shih, the largest city and the provincial capital, and Chi-lin Shih (Kirin), the second largest city and former capital of the province. The subprovincial administrative areas are further subdivided into 36 counties (hsien), seven municipalities (shih), and two autonomous counties (tzu-chih-hsien). The latter are Ch'ang-pai Korean Autonomous Hsien, in the southeast, and South Gorlos (Ch'ien-kuo-erh-lo-ssu) Mongol Autonomous Hsien, in the northwest.

The role of the production teams

The traditional subcounty territorial unit is the hsiang (rural district or civil township). The hsiang was retained in the 1954 Constitution of the Chinese People's Republic, but during the mid-1950s became obsolete with the organization of "higher co-ops," and was finally relegated to administrative limbo with the establishment of people's communes in 1958. Each commune is administered by an elected commune committee, but actually the real power lies in the commune party committee of the Chinese Communist Party. The commune is subdivided into production brigades, the branch committee of which is the grass-roots control body of the party apparatus. The production brigade owns the land as well as the means of production, and it is subdivided into production teams that form the lowest level administrative unit. These are often agglomerated rural villages, or *chieh-tao* (residential street grouping) within a town. The production team is the basic economic accounting unit of rural China. During the Great Proletarian Cultural Revolution

(1966–69), many local administrations, both party and government, were disrupted by Red Guards and replaced with revolutionary committees constituted by representatives of the People's Liberation Army, the "revolutionary mass organizations" and "state organs of the dictatorship of the proletariat." The People's Liberation Army assumed an increasingly important role in local administration after 1969. Rural education, social welfare, the care and support of the aged, orphans and widows, rural recreation, medical care, and rural economic development and modernization are in part financed from profits of labour-intensive rural industry organized under the commune system.

Economic organization. ***Manufacturing.*** The two largest cities, Ch'ang-ch'un and Kirin, serve as major administrative, cultural, educational, commercial, and manufacturing centres.

The oldest city, Kirin, has a history of merely 300 years; Ch'ang-ch'un was founded only in 1825; and most of the towns have a history of not more than 60 years. Rapid urban-industrial development has occurred only during the last three or four decades.

Ch'ang-ch'un has long been a food-processing centre for the rich agricultural region of the Sungari Valley, where wheat, rice, maize, soybeans, kaoliang, and millets are grown and livestock raised. Lumbering is also important. Under the People's Republic, the industrial base has been greatly expanded. The huge Ch'ang-ch'un No. 1 Motor Vehicle Plant was designed and equipped with the technical assistance of the Soviet Union to produce 30,000 "Liberation" trucks (six cylinders, 90 horsepower, four tons) a year. Because of the Sino–Soviet rift, the Soviets withdrew their assistance in mid-1960, but the 23,000 Chinese workers, by the end of 1960, were producing up to 100 trucks a day. A railway carriage factory was also designed by Soviet experts, but when the Soviets withdrew support the plant was left unfinished. The designed capacity was 50 cars a day, but the daily output in late 1960 was only five a day. Ch'ang-ch'un has the second largest film studio in China and produced 26 films and 12 documentaries in 1960.

Vehicle production

Transportation. The province is served by 1,676 miles (1957) of railroads, about 10 percent of China's railways. It has 236 miles of railroad per 100 square miles of area, the densest in China. Of the 47 municipalities and counties in the province, 38 have direct access to a rail line.

The Sungari is the main artery of the inland navigation network. Large steamers can sail up to the town of San-ch'a-ho, smaller steamers to Feng-man, and wooden vessels to Liang-chiang-k'ou. Its tributary the Hui-fa Ho and the Tumen River are both navigable by wooden vessels. The Yalu is navigable by steamers up to Yii-shu-lin-tzu, and by wooden vessels above that point.

Kirin Province is served by over 6,000 miles of motor roads. The highway network has four regional centres—at Ch'ang-ch'un, Kirin, Yen-chi, and T'ung-hua.

Cultural life. In 1959 to 1960 there were 35 institutions of higher learning in the province, with a total student enrollment of 32,000. Kirin University, Kirin Polytechnic University, and Kirin Normal University were located in Ch'ang-ch'un. Higher education institutions closed during the Cultural Revolution (1966 to 1969), but in 1970 the North-East People's University was one of those which had reopened.

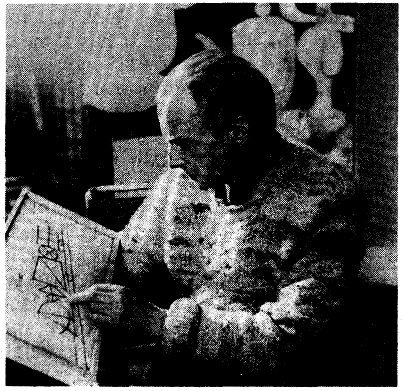
BIBLIOGRAPHY. The best regional economic monograph is SUN CHING-CHIH (ed.), ***Economic Geography of Northeast China*** (Eng. trans. 1962). Two of the best atlases on China are the C.I.A. publications: ***Communist China Map Folio*** (1967) and ***Communist China Administrative Atlas*** (1969). The best physical geography of China in a Western language is V.T. ZAYCHIKOV (ed.), ***The Physical Geography of China*** (Eng. trans. 1965).

(F.Hu.)

Klee, Paul

One of the most acute creators of forms in modern painting, Paul Klee has had a considerable influence on the art of the 20th century. This genius of modern art stands apart from all groups. If Expressionism can be considered as the "direct visualization of individual emotion," Klee perhaps can be said to belong to this movement. He was able to create a new language of colour, form, and space and to develop a private symbolism that communicates deeply. Klee was also an accomplished musician, and much of his system of visual form is built up like musical form. To Klee, "art was a symbol of creation, an example much as the earth exemplifies the cosmos." He proceeded to create his own reality, which does not represent the familiar reality but penetrates to a reality that is not known until one sees his paintings. The present state of things, he felt, is only a momentary and accidental arrangement, and he attempted to communicate a world as it might once have been or is in the process of becoming. If a work of art is able to be more than a pleasing decoration, if it is able to give insight and perhaps change lives, then it may be understood why Klee is one of the most admired artists of one of the most creative periods in the history of art.

Aufnahme Fotopress



Klee, 1939.

Early **life** and education. Klee was born on December 18, 1879, at Miinchenbuchsee, near Bern, capital of the Swiss Confederation. He was Swiss on the side of his mother, née Ida Maria Frick, whose family was from Basel; she had studied music at the Stuttgart conservatory. His father, Hans Klee, was a German, born in Thuringia in 1849, who had given up a singing career and contented himself with teaching music.

Klee's involvement with the musical and visual arts began early in his life. At the age of four, he sought refuge near his mother when the devils he was scrawling on scraps of paper seemed to him to come to life. In his Uncle Ernst's cafe he would daydream over the sinuous patterns in the marble tabletops, which suggested to him fantastic designs. The boy's imagination was also stimulated by the stories he heard from his grandmother Frick. At seven, he began to play the violin.

Childhood interest in art and music

As a student, Klee was considered attentive, but he early displayed an original turn of mind opposed to the established order. He described his boyhood education as "mountains immeasurably high but with no foot." As a youth he was particularly fond of Greek and drawing (12 notebooks filled with drawings are preserved in the Klee-Stiftung in Bern), but most of all he liked to play the violin. His musical talents were so developed that he was used as a substitute by the Bern symphony orchestra. The teen-age Klee also wrote precociously sophisticated poetry and short stories.

Upon graduating from secondary school in 1898, Klee decided to devote himself exclusively to painting. Choosing Munich over Paris as a place to study art, he attended an art preparatory school in the Bavarian capital until he entered the Munich Academy in 1900 and studied under Franz von Stuck, an academic painter known for his allegorical pictures. Stuck demanded of his pupils a strict knowledge of anatomy and preferred to make them concentrate most of their time on drawing rather than on painting. Repelled by his teacher's conformity, Klee left the academy in 1901. His enthusiasm for avant-garde German publications, such as the magazine ***Jugend*** and the satirical newspaper Simplicissimus, further influenced his departure from artistic and intellectual conventionalism.

During his art-student years he assiduously attended concerts, being especially attracted to the music of the German Romantic composer Richard Wagner and the works of the Austrian composers Richard Strauss and Mozart. It was at a musical evening in the winter of 1899–1900 that he first met Lily Stumpf, the daughter of a prominent Munich doctor; he married her in 1906.

in June of 1901 Klee left Munich, briefly returned to Bern, and then set out for Italy with his friend Hermann Haller, a painter who later became one of Switzerland's leading sculptors of the early 20th century. Visiting Milan, Genoa, Rome, Naples, and Pompeii, Klee was particularly drawn to the works of such great masters of the Italian Renaissance as Michelangelo, Botticelli, Tintoretto, and Pinturicchio.

Artistic development. He returned to Bern in May of 1902 and lived in the Miinchenbuchsee house of his childhood until his marriage. Artistically and intellectually, this was an important period in Klee's life. In art he decided he would like to begin everything again by himself, without taking account of the accomplishments of the past, although this attitude did not prevent him from studying anatomy and the handling of the mediums of the visual arts. He also continued to read voraciously.

In 1905, from May 31 to June 13, Klee visited Paris, where he saw an important retrospective exhibition of the works of the American artist James McNeill Whistler (1834–1903) and was overwhelmed by the art of the French Impressionist Pierre-Auguste Renoir (1841–1919). Returning to Bern, he collected into a portfolio 15 etchings he had executed and submitted them to the jury of the Neuen Miinchner Sezession exhibition. Ten of these were selected for the show, including "Jungfrau im Baum" ("Virgin in a Tree") and "Der Held mit dem Flügel" ("Hero with the Wing'). In these etchings, with their aggressive, fantastic figures, Klee reveals an almost menacing humour. He escaped the dullness and stifling atmosphere of Bern by this black humour and through his precisely detailed, grotesque imagery, exemplified in the etching "Zwei Manner, einander in hoherer Stellung vermutend, begegnen sich" ("Two Men Meet, Each Believing the Other to Be of a Higher Rank"). Working from nature as well as his imagination, Klee also executed drawings, watercolours, and paintings on glass. After having modelled some statuettes in Plasticine, Klee went to live in Munich in 1906.

Satirical
etchings

After marrying, the Klees lived in Schwabing, a suburb of Munich. They resided there until Klee was called to teach at the Bauhaus in the autumn of 1920. And it was in Schwabing that his son Felix was born on November 30, 1907. Lily, a pianist, largely supported the family by teaching music; Klee did not begin to sell his paintings until 1914. Although the Klees lived in relative poverty, Paul felt freer and more at home than in Bern, for he was able to choose between a wider variety of exhibitions and concerts, as well as encounter more stimulating contemporaries in the arts.

First
exhibits in
Switzerland
and
Germany

In 1911–12 Klee did the filiform pen-and-ink illustrations for an edition (1920) of Candide, a satirical work by the 18th-century French writer Voltaire. He also had his first one-man shows that year. They were held in Switzerland at the museums of Ziirich, Bern, Basel, and Winterthur. Having made the acquaintance of the German Expressionist painters Wassily Kandinsky and Franz Marc in 1912, Klee participated in the second Munich exhibition of the avant-garde Blaue Reiter (Blue Rider) group. He began to receive many invitations to exhibit his works from important private galleries such as the Sonderbund Gallery in Cologne and Galerie "Der Sturm" in Berlin.

During a 16-day stay in Paris in April 1912, the influential German collector Wilhelm Uhde introduced Klee to the leading Cubist painters working in France. Klee was especially taken with the colouristic works of Robert Delaunay, the originator of Orphic Cubism. He was also moved by the works of Vincent Van Gogh. The proto-Expressionist works of this Dutch painter disturbed Klee deeply, for he saw behind their intensity "a brain that suffers from the combustion of a star." The engravings of another late-19th-century proto-Expressionist, the Belgian artist James Ensor, also interested him at this time. Although showing a definite preference for the expressionism of northern European art, Klee placed Paul Cézanne above all recent painters and began to be impressed by his contemporary, the great French master of 20th-century painting Henri Matisse.

In April 1914 Klee undertook a trip to Tunis in the company of two other painters, Louis Moilliet and August Macke. The "three wise men of the West," as someone called them, were each profoundly influenced by this North African experience. While painting watercolours at Qayrawān (Tunisia), Klee experienced a revelation of the nature of colour. "Colour possesses me . . ." he wrote in his diary. "It will possess me always . . . . That is the meaning of this happy hour: colour and I are one." Although his Tunisian trip lasted only 12 days, it changed the course of his art from a descriptive image of natural reality to a less figurative art and to being, as Klee wrote, "abstract, with memories."

On August 1, 1914, World War I broke out. Klee's friend August Macke was killed September 26 on the Champagne front in France. Another of his friends, Franz Marc, one of the pioneers of German Expressionism, fell at Verdun on March 4, 1916. Called up for duty a few days after Marc's death, Klee served in the German army and was stationed at several locations, including the aviation depot at Schleissheim near Munich. where he painted and polished airplanes. It cannot be said that military life frightened or disturbed him greatly; he found it endurable, and its overall effect on him was that of a "fantastic dream." His superiors respected his reputation and permitted him to live in town and visit his family. He also obtained permission to write the preface to the catalog for an exposition of Franz Marc's paintings. In moments of leisure he painted in watercolours, sketched, and read. Demobilized after the defeat. he rented a studio in the Schlosschen Suresnes in Munich. The wealthy Bern collector Hermann Rupf requested some of his watercolours, for himself and for Kahnweiler, the Paris dealer and champion of the Cubists.

In 1919 he began to paint primarily in oil. His paint became increasingly thicker. Far from being a transcription of the tonalities of the external world, Klee's colour was invented, corresponding to the inner leaps of his spirit. In November 1920 the Bauhaus, a school of design founded by the German architect Walter Gropius on the principle that "to build is a biological and not an aesthetic phenomenon," invited Klee to join the staff as an instructor of painting. Klee accepted and left Munich for Weimar in January of the following year. At the Bauhaus, the general educational method was to emphasize the technical aspect of art. The ideal was not "art for art's sake" but a utilitarian art, an inspired craftsmanship, that would be applicable to decorative or monumental projects. Through the Bauhaus would be created "the new edifice . . . born of innumerable hands." One wonders how Klee, such an internal and poetic individual, was able to venture into this communal atmosphere. But he possessed a thorough knowledge of pictorial optics and a unique pedagogical gift. Klee taught in accordance with a method that he himself discovered and expounded in the Padagogisches Skizzenbuch (1925; Pedagogical Sketchbook).

Bauhaus
association

In the summer of 1924 Klee spent six weeks in Sicily—at Taormina, where he painted watercolours, and at Syracuse, which seemed to him an "abstract landscape." After continual attack by the press and public of Weimar, the Bauhaus was forced to close its doors in 1925. Encouraged by the German dramatist Gerhart Hauptmann, however, the Bauhaus moved its activities to Dessau, where, in 1926, it reopened in a building constructed by Gropius. In July the Klees moved to Dessau, where their neighbour was Kandinsky. Work in Klee's studio alternated with classes, music, lectures, and collaboration on the magazine Bauhaus.

On December 20, 1928, Klee embarked at Genoa for Egypt. Visiting Cairo, Luxor, and Aswan in a few days, this trip, like his earlier Tunisian journey, exerted a profound effect on him. Klee resigned his Bauhaus post in 1931 and accepted a position at the Diisseldorf Academy. He no longer had to teach, being able to support himself by the sale of his work, but he needed to satisfy his pedagogic nature.

Nazi harassment

Last years. Klee's existence in Germany was doomed when Adolf Hitler was appointed chancellor in January 1933. Attacked by the Nazis, Klee was treated as a barbarian and a corrupted foreigner. He left Germany, travelling to Italy, Switzerland, and France. Profoundly disgusted by the Hitler regime, Klee again took up residence in politically neutral Switzerland. Settling in Bern in 1933, he met his old friend Moilliet, who had become involved with creating stained-glass windows. In 1934 Klee had his first London exhibition; while in Bern and Basel, an important retrospective of his work was held in 1935. During the summer an attack of bronchitis followed by measles brought about pulmonary and cardiac complications. Ultimately, Klee's condition was diagnosed as scleroderma, a rare disease that results in a drying up of the mucous membranes and ends by striking the heart. Klee was initially unconcerned. He continued his work, took cures at Tarasp, and went to recuperate at Montana-sur-Sierre in the Swiss canton of Valais. Gradually, however, his hope of recovering diminished. In 1937 he became feverishly active and did not interrupt his work except for short walks. His palette took on brilliant strength, as seen in the pastel "Figur im Garten" ("Figure in the Garden"). In the spring, the French painter Georges Braque visited him, and in November, after having unveiled his "Guernica" in Paris, Pablo Picasso came to see Klee. In the same year, an exhibition of confiscated art, called Degenerate Art, opened in Munich. It included, among the many masterpieces of modern art condemned by the Nazis, 17 works by Paul Klee. On the occasion of Klee's 60th birthday, in 1939, he received telegrams and presents from everywhere, except official Germany.

In 1940 Klee wrote, "It is no accident that I have entered on the tragic path. Plenty of sheets of drawings have pointed the way and said: the time has come." Klee executed that same year a series of angels, which are closer to those of Islām, especially the ones described by Ḥāfeẓ, the 14th-century Persian poet, than to the angels of Christianity. He also painted "Tod und Feuer" ("Death and Fire"), his *Dies Irae*, which seems to foreshadow his approaching cremation. Klee began his final illness at the sanitarium of Orselina-Locarno on May 10. He was taken to the Saint Agnes clinic at Muralto-Locarno, where he died of a heart attack on June 29, 1940. On July 1 he was cremated at Lugano, and the urn was finally buried in September 1942 in the Schosshalden cemetery at Bern. The tombstone is inscribed with a passage from his diary: "I cannot be grasped in this world. For I am as much at home with the dead as with those beings who are not yet born."

Paul Klee is generally acclaimed one of the greatest visionaries of all time, an artist to whom the universe opened the secret of its metamorphosis. His work reveals a prodigious knowledge of pictorial illusion. Imbued with autohypnosis, Klee's art communicates a poetic spell created by a seemingly magical transfiguration of the real.

MAJOR WORKS

PAINTINGS: "Zoological Garden" (1918; Klee Foundation, Bern); "Villa R" (1919; Kunstmuseum, Basel); "Graduated Shades of Red-Green" (1921; Yale University Art Gallery); "A Young Lady's Adventure" (1922; Tate Gallery, London); "Dance, Monster, to My Soft Song!" (1922; Solomon R. Guggenheim Museum, New York); "Magic Theatre" (1923; Klee Foundation, Bern); "Around the Fish," (1926; Museum of Modern Art, New York); "Pastorale" (1927; Museum of Modern Art, New York); "Monument at the Edge of an Orchard" (1929; Klee Foundation, Bern); "The Mocker Mocked" (1930; Museum of Modern Art, New York); "Arab Song" (1932; Phillips Collection, Washington, D.C.); "Child Consecrated to Suffering" (1935; Alhright-Knox Art Gallery, Buffalo, New York); "Rich Harbour" (1938; Kunstmuseum, Basel); "Demonry" (1939; Klee Foundation, Bern); "Death and Fire" (1940; Klee Foundation, Bern).

DRAWINGS: (All but one at Klee Foundation, Bern): "Furniture Caricature" (1910); "War Destroying the Country" (1914; Kunstmuseum, Basel); "Before Birth" (1926); "Wandering Animals" (1928); "Abstract Script" (1931); "Please!" (1932); "Symptom to Be Recognized in Time" (1935); "Fear Erupting" (1939).

GRAPHICS: "Comedian II," etching (1904); "Destruction and Hope," lithograph (1916).

WRITINGS: *Pädagogisches Skizzenbuch* (1925; ***Pedagogical Sketchbook,*** 1953); ***Tagebiicher von Paul Klee, 1898–1918,*** ed. by Felix Klee (1957; ***The Diaries of Paul Klee, 1898–1918,*** 1964).

BIBLIOGRAPHY. WILL GROHMANN, ***Paul Klee*** (1954; Eng. trans. 1956), the fundamental work on Klee, which follows step-by-step the development of his life and work; LEOPOLD ZAHN, ***Paul Klee: Leben, Werk, Geist*** (1920), the first book written on Klee; WILHELM HAUSENSTEIN, ***Kairuan, oder eine Geschichte vom Maler Klee*** (1921), the German art historian demonstrates Klee's originality; RENE CREVEL, ***Paul Klee*** (1930), the Surrealist poet extols Klee's talent in this first publication about the artist in French; ALFRED H. BARR, JR., ***Paul Klee: Museum of Modern Art, New York, March 13–April 2, 1930,*** long preface to the catalog of 63 exhibited works (reprinted with revisions, 1941); KLEE-GESELLSCHAFT, BERN, ***Paul Klee,*** vol. 1, ***Dokumente und Bilder aus den Jahren 1896–1930*** (1949), with writings by Klee; WERNER HAFTMANN, ***Paul Klee, Wege bildnerischen Denkens*** (1950; Eng. trans., ***The Mind and Work of Paul Klee,*** 1954), a good analysis of the painter's mind and method; CAROLA GIEDION-WELCKER, ***Paul Klee*** (1952); GAULTIERI DI SAN LAZZARO, ***Klee, la vie et l'oeuvre*** (1957; Eng. trans., ***Klee: A Study of His Life and Work,*** 1957), copiously illustrated; PIERRE COURTHION, ***Klee*** (1953), a sketch of the artist's personality (in France); "Hommage à Paul Klee," ***XXème Siècle,*** 4:30–36 (1938); and ***Klee, un art prophétique*** (1967); HERBERT READ, ***Klee (1879–1940)*** (1948).

(P.Co.)

Knox, John

The foremost leader of the Scottish Reformation, John Knox set the austere moral tone of the Church of Scotland and shaped the democratic form of government it came to adopt in contrast to the rule by bishops that prevailed in England. Important political consequences resulted, both in his day and later, from his enunciation of the belief that man may rightfully resist rulers in order to obey God. In him, too, the Scottish Reformation found its historian.

Knox, engraving from *Icones*, by T. Beza, 1580.

By courtesy of the trustees of the British Museum; photograph, J. R. Freeman & Co. Ltd.

Early life. Almost nothing is known of Knox's life before 1540, the accounts given by his earlier biographers being mostly fanciful. He was born in the vicinity of Haddington, East Lothian; the traditional date of his birth (1505) has been abandoned, and a date around

1514 is now accepted. Of his parentage it is known only that his mother's name was Sinclair (Knox used the name "John Sinclair" as an incognito in times of danger); that his father's name was William; and that he and both Knox's grandfathers had fought, and two of them had died—perhaps at the Battle of Flodden against Henry VIII's troops. The family may have been farmers.

It is supposed that Knox trained for the priesthood under the scholar John Major, most probably at St. Andrews University. Knox did not take a master's degree, but he ended his training with a mind imbued with that delight in abstract thought and dialectical disputation, which, even in that age, was recognized all over Europe as typical of Scottish scholarship. He was in priest's orders by 1540 and was in 1543 known to be also practicing as an apostolic notary in the Haddington area, which would seem to indicate that he was in good standing with the ecclesiastical authorities.

Involvement in the Reformation. *Relationship with political Protestantism.* Two years later, however, Knox was in more equivocal company as tutor to the sons of two gentlemen of East Lothian deeply involved in the intrigues of political Protestantism. Under their protection, George Wishart, a Scottish Reformation leader who was to become an early martyr for the cause, began a preaching tour in the Lothians in December 1545. Knox was much in his company; and his complete conversion to the Reformed faith dates from his contact with Wishart, whose memory he cherished ever afterward. Wishart was burned for heresy in March 1546 by Cardinal David Beaton, archbishop of St. Andrews, who, rather than the weak governor, was the real ruler of Scotland. Wishart's execution began a chain of events that profoundly altered Knox's life. Three months later, Beaton was murdered by Protestant conspirators who fortified themselves in St. Andrews castle.

Conversion to Protestantism

Meantime, Knox, accompanied by his pupils, was moving from place to place to escape persecution and arrest. His desire was to go to Germany to study there at the Protestant seats of learning; but his employers sent word to him to take their sons to St. Andrews and continue their education under the protection of the castle. Thus, in April 1547, 11 months after the cardinal's murder, and against his own desire Knox arrived with his pupils in St. Andrews—still an unknown man. The three months he spent there transformed him, against his own predisposition, into the acknowledged spokesman and protagonist of the Reformation movement in Scotland. The Protestants in the castle had become involved in controversy with the university; several of them, becoming aware that a man of uncommon gifts had joined them, pressed upon Knox's conscience the duty of taking up "the public office and charge of preaching." Knox's inclination was for the quiet of the study and the schoolroom, not for the responsibilities and perils of the life of a preacher of a proscribed and persecuted faith. He resisted the call with tears, and only after great hesitation was he persuaded to preach in the town of St. Andrews a sermon that convinced friend and foe alike that the great spokesman of Scottish Protestantism had been found. This was the turning point of Knox's life;' from this time forward he regarded himself as called to preaching by God, and he was the more certain of the divine origin and compulsion of the call in that it ran counter to every inclination of his own.

At the end of June 1547, French assistance reached the governor of Scotland. The garrison of St. Andrews castle, bombarded from without and assailed by plague within, capitulated on terms that were not kept; Knox and others were carried off to slavery in the French galleys. English intervention secured his release 19 months later, though with permanently broken health.

Propagator of Protestantism in Scotland. In England the Protestant government of Edward VI was endeavouring to hurry clergy and people into the Reformation faster, if anything, than most of them were willing to go. For this program preachers and propagandists were urgently required; and since a return to a Scotland under Roman Catholic rule was impossible for Knox at this time, the English government promptly made him one of a select corps of licensed preachers and sent him north to propagate the Reformation in the turbulent garrison town of Berwick-upon-Tweed. He brought order to the town and established a congregation on Puritan lines, and there he met Marjorie Bowes, who was to become his wife. Early in 1551 he was given a new assignment in Newcastle and a little later was appointed to be one of the six royal chaplains whose duties included periodic residence at, and preaching before, the court as well as itinerant evangelism in areas where the regular clergy were lacking in Protestant zeal. The accession to power of the duke of Northumberland, who virtually ruled England from 1549 to 1553 during the minority of Edward VI, marked a governmental swing toward more extreme Protestantism; and Knox, now a leading figure of the Protestant camp, was pressed to accept the bishopric of Rochester. His refusal was motivated not by any doctrinaire objection to episcopacy but by his complete distrust of Northumberland's government and by a determination not be be bound to it by gifts or preferments. Later he refused the vicarage of Allhallows, London, mainly because it would have required his compliance with usages that he considered to be not yet sufficiently Reformed. He continued, under the patronage of the government, to exercise an itinerant ministry mainly, but not exclusively, in Buckinghamshire, Kent and London.

Appointment as a royal chaplain

Influence on the Church of England. In three respects Knox left his mark on the Church of England: He took part in the shaping of its articles. He secured the insertion into *The Book of Common Prayer* of the so-called black rubric that denies the corporal presence of Christ in the consecrated bread and wine used in Holy Communion and explains that kneeling at communion implies no adoration of the elements. And he was one of the chief foster fathers of English Puritanism, a reform movement started within the state church with a view to the more rigorous application of Reformation principles in doctrine and worship. In this connection, however, it should be noted that Knox regarded all sectarianism with horror. He could see no place for any religious body or sect apart from the national Protestant Church; and throughout his life he used his considerable influence in England to persuade the Puritans to remain within the English Church.

Escape to the Continent. On the accession of Mary Tudor, a Roman Catholic, to the throne in 1553, Knox was one of the last of the Protestant leaders to flee the country. He escaped to the Continent disturbed by the realization that the fate of "true religion" in England had turned on the religious opinions of one woman. He could see no security for the Reformation anywhere if the personal whim of a sovereign was permitted to settle the religion of a nation. Might it not be legitimate for Protestant subjects, in such circumstances, to resist—if necessary by force—the subversion of their religion by a Roman Catholic ruler? Knox laid this question before the Protestant leaders on the Continent. John Calvin, in Geneva, and Heinrich Bullinger, in Zürich, were prepared to counsel only patience and prayer, certainly not armed resistance. From the French Protestant theologian Theodore Beza, Knox may have received greater encouragement, for in 1554 Beza wrote his *Traité de l'authorité du Magistrat* in which he developed ideas significantly similar to those of Knox. Be that as it may, Knox formulated his fateful conclusion, later to be applied in Scotland, that God-fearing magistrates and nobility have both the right and the duty to resist, if necessary by force, a ruler who threatens the safety of true religion. Also in 1554 Knox published his *Faithful Admonition* to the Protestants who remained in England. Its extremism and intemperate language served to increase the sufferings of those to whom it was addressed; and, coming as it did from one who was in comparative safety, it alienated many in England from him.

Minister to English exiles

In the same year, on the insistence of Calvin, Knox became minister of a congregation of English refugees, mainly Puritan, in Frankfurt am Main; but he remained there for only a few months. He then became minister of the growing congregation of English exiles in Geneva, a pastorate that lasted until his final return to Scotland in 1559, but was interrupted at the outset by a visit (1555–

56) to Berwick and a nine-month sojourn in Scotland, in the course of which he married Marjorie Bowes. She died, having borne him two sons, in 1560.

In Edinburgh Knox was astounded by the progress made by the Reformed cause and by the eager reception given to him by all classes in the community. To the nobility, in visits to their country houses, he propounded his doctrine of "justifiable resistance" to Roman Catholic rulers who attacked the faith of Protestant subjects and urged them to withdraw from all the rites and ceremonies of the Roman Church and to band themselves together for the defense of Protestantism in case that should prove necessary. A peremptory summons from his congregation called him back to Geneva; but he left to the faithful in Scotland an important ***Letter of Wholesome Counsel*** (1556) enjoining not only private family worship but weekly meetings of believers for corporate Bible study and discussion. From these weekly meetings, Reformed congregations grew apace, and from the leaders of these congregations came the elders of the Reformed Church.

Geneva was to Knox a beloved city, "the most perfect school of Christ on earth since the days of the Apostles." There he spent the happiest years of his life, highly esteemed, in peace, and among kindred souls. From this period (1556–58) dates his elaborate and rather tedious treatise on ***Predestination*** as well as his ***first blast of the trumpet against the monstruous regiment*** [rule] ***of women,*** in which he states with uncurbed vehemence the common belief of his day that the exercise of authority by women is contrary to both natural law and revealed religion. The pamphlet was aimed at the three women who were holding the reins of government in England, France, and Scotland and were oppressing Protestantism; but, unfortunately for Knox, publication coincided with the accession in England of the Protestant Elizabeth I who indignantly and permanently debarred the rash author from her realm.

The religious crisis in Scotland

Recall to Scotland. In Scotland matters reached a crisis in the spring of 1559. Two years earlier the Protestant lords had signed a "band," or covenant, on Knox's advice, pledging themselves to foster and defend "The Congregation of the Lord" and its ministers (hence their name "Lords of the Congregation"). The queen regent, the French-born Mary of Guise, had deemed it politic to make concessions to them. But when hostilities between Spain and France ended early in 1559, opening the possibility of stronger French intervention in Scotland, the Queen Regent felt that the time had come to call a final halt to the expansion of Protestantism. To this end she summoned the Protestant preachers, as ringleaders of the growing Protestant insubordination, to appear before her on May 10, at Stirling. The Protestants replied by recalling Knox from Geneva, and the Protestant lords, lairds, and commoners mustered at Dundee. On May 4 Knox joined them and they advanced to Perth, where, after a vehement sermon by Knox, the friaries were sacked.

By the end of June, Edinburgh was temporarily in Protestant hands and Knox was preaching in St. Giles'; but the triumph was illusory and Knox knew it. The voluntary army of Protestants could not keep the field for more than a few weeks; the mercenary army of the Queen Regent could keep the field indefinitely and strike a crushing blow as Protestant strength declined. At this juncture Henry II of France died and power fell into the hands of the Guises, the brothers of the Queen Regent and uncles of the young queen of France—Mary, queen of Scots. Strong French intervention in Scotland was now assured in furtherance of the Guise plan to displace Queen Elizabeth of England and to unite France, Scotland, and England under Francis II, of France, and Mary. Thus a political issue of critical international importance cut athwart the religious issue in Scotland. A French victory in Scotland would place Elizabeth and England in peril. It therefore behooved England to make common cause with the Scottish Protestants. Knox lost no opportunity of driving this fact home to Elizabeth and her chief minister William Cecil (afterward Lord Burghley) though Cecil was already well aware of the peril and was held back only by the vacillation of his royal mistress. The autumn and winter of 1559 saw the Scottish Protestants in desperate plight. Only Knox's superhuman exertions and indomitable spirit kept the cause in being. In the blackest hour his great sermon at Stirling put fresh heart into the despairing Protestant leaders and staved off defeat at the hands of the government's French mercenaries. On Knox's resolution alone in these months hung the fate not only of Scottish Protestantism but of Elizabeth's England.

In the spring of 1560, Elizabeth at last consented to English action. In April 10,000 English troops joined the Scottish Protestants, the Queen Regent died in Edinburgh castle, and the disheartened French gave up the struggle. By treaty, French and English troops were then withdrawn, leaving the victorious Scottish Protestants to set their own house in order. The chances were heavily against a satisfactory settlement, for there was no strong central authority to take control. Queen Mary was a Roman Catholic and an absentee in France, and all her sympathies, religious and political, were with the defeated side. The Estates (Scottish Parliament) had never exercised much power; but now, meeting in August without royal authority, they proceeded to grapple with the religious issue. The Scots Confession (hurriedly prepared by Knox and three others), a document moderately Calvinist in tone but with greater emphasis on the sacraments than Calvin had advocated, was adopted as the faith of the nation. Papal jurisdiction was abolished, and the saying of mass prohibited, with death as the penalty for a third offense—though this penalty was inflicted in only one recorded instance.

Shaping the Reformed Church. It remained to set down the form of government and constitution of the Reformed Church. Knox, aided by a committee of distinguished churchmen, laid before the Estates the ***First Book of Discipline*** containing proposals for the constitution and finance of the Reformed Church. Worship was to be regulated by the ***Book of Common Order*** (also called Knox's Liturgy), a revision of the order used by Knox in Geneva. According to Knox's Liturgy, congregations were to be governed by elders elected annually by the people, and the elders were to aid the minister to maintain firm moral discipline among the people. Ministers were to be elected by the people but to be appointed only after rigorous examination of life and doctrine by their ministerial brethren. The ablest ministers were to be appointed superintendents of areas roughly corresponding to the old dioceses; they were to supervise the ministers and congregations in the area and were to be assisted by provincial synods of ministers and elders. Implied in the proposal is that final authority is to lie with a general assembly of the church containing both elders and ministers. This form of government is obviously not the "presbyterianism" of a later generation, which was based on government by a number of courts composed of clerical and lay elders, neither is it episcopacy, for the superintendents are not consecrated and do not constitute any kind of higher order but are in all points subject to the general assembly. But in the high place given to the laity, Knox's system contains the most essential element of later Presbyterianism.

The Book of Common Order

The ***Book of Discipline*** proceeds to outline a most elaborate educational scheme from parish school up to university, compulsory in its early stages for all and free to the poor, with equal opportunity of advancement for the son of peasant and of peer in the service of the church and commonwealth. It also lays plans for a much-needed scheme of systematic aid of the poor, a project dear to Knox's heart. Finally it urges that the endowments of the old church should be made available for the financing of these admittedly very costly schemes of the new church.

But the proposals thus outlined foundered on the rock of finance. The endowments of the old church were plunder in a poor land for the nobility, who had scant sympathy with Knox's "devout imaginings." The Estates refused to accept the ***Book of Discipline*** and the church was left to organize itself to the best of its ability. The Estates shelved the financial problem by the temporary expedient of granting to the remaining Roman Catholic clergy the life-rent of their benefices, provided they contributed to the maintenance of the Reformed Church out of their revenues. Knox was deeply embittered by the enforced abandonment of his schemes for education and poor relief and by the scant provision for the Reformed Church. Thus the financial question drove a wedge between Knox and the nobles who had been his friends and supporters, and between the church and the government. The rift was widened when the return of Queen Mary created a court party that resented the extremism of the ministers. Yet the poverty of the Reformed Church created a bond of sympathy between ministers and people, and governmental neglect or hostility compelled it to develop a sturdy autonomous life of its own that resisted later attempts to subject it to state control.

Relations with Queen **Mary**

Mary arrived in Scotland in 1561 already persuaded that Knox was to be her archenemy and that the country could not hold them both. Knox, who hoped at first that the young queen would prove pliable in Protestant hands, soon reached a similar conviction. The first three of his celebrated audiences with Mary were polite skirmishes; in the fourth, battle was joined in grim earnest. Hearing that Mary was contemplating mamage with Don Carlos of Spain, a match that would have had fatal consequences for the Scottish Reformation and probably for England as well, Knox sounded the Protestant alarm with all the vehemence at his command. Mary, enraged at this intervention by a heretic preacher and commoner in affairs of state, summoned Knox, berated him with hysterical fury, and charged him with treason; but the Privy Council refused to convict him. Knox filled Mary's cup of bitterness in 1564 by marrying, without the royal assent, Margaret Stewart, the 17-year-old daughter of the impoverished Lord Ochiltree, who was distantly related to the queen.

The hatred between Knox and Mary was now personal, deep, and vindictive. But Mary's sun was setting. In 1564 she dismissed her Protestant advisers and undertook the mismanagement of her own affairs. For a time the Reformed Church was in real danger, but in 1567 came Mary's ruin and abdication (Knox would have preferred her execution) and Knox's old friend James Stewart, earl of Moray, became regent. In him the Reformed Church would have found a powerful patron; and Knox, believing that the long battle of the Reformation was ended, was contemplating a peaceful retirement to Geneva when Moray was murdered and the country plunged into a struggle between the supporters of the queen and the regency. Knox was anxiously involved in the turmoil, but his strength was overtaxed and he suffered a paralytic stroke. When Edinburgh became a battleground between the factions in 1571, the leaders on both sides insisted on his removal to safety in St. Andrews, from where he returned in 1572 to die. When the news of the St. Bartholomew's Day Massacre of French Protestants reached Scotland, Knox dragged himself to his pulpit in St. Giles' and in white heat drove home the lesson of that tragedy to his hearers. He stood one more time in the pulpit of St. Giles' to introduce his successor. On November 24, 1572, he died in Edinburgh.

Assessment. Knox was and is a controversial figure, and his work and influence will always be differently assessed by men of differing religious and political views. Certainly his conviction that the Reformation was God's cause and must triumph, a conviction he had a remarkable power of impressing upon other minds, was the rock upon which the Reformed Church in Scotland was built. His power as a preacher lay in his capacity to fuse reason with emotion and to be a passionate logician in the pulpit, convincing mind and heart with equal force. Intolerant he undoubtedly was, but not more so than most religious leaders of his day; and his Calvinism was a good deal more moderate than that of a later age. He may justly be blamed for his "rude vehemence," his acrid judgments, and sometimes mischievous utterances; yet he was more temperate in action than in speech, and his private letters reveal an unexpected tenderness. His singleminded, lifelong, and incorruptible devotion to what he believed to be his duty, in an age when such qualities were rare indeed, must command respect. There is ample historical testimony that his moral life was in keeping with his rigorous creed. Knox's most considerable literary work is his ***History of the Reformation in Scotland.***

BIBLIOGRAPHY

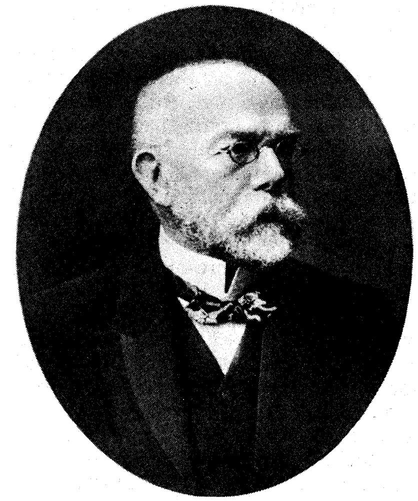
Works: D. LAING (ed.), ***The Works of John Knox,*** 6 vol. (1846–64), a work of distinguished scholarship containing the collected writings of Knox; W. CROFT DICKINSON (ed.), ***John Knox's History of the Reformation in Scotland*** (1949), Knox's own ***History*** in modern English, fully annotated; indispensable for the scholar and enjoyable for the layman.

Biographies: J.G. RIDLEY, ***John Knox*** (1968), a well-documented, modern biography incorporating the results of the most recent research, with a complete bibliography; E. PERCY, ***John Knox*** (1937), a sensitive and penetrating study of Knox and his times, but lacking in documentation; G. MACGREGOR, ***The Thundering Scot*** (1958); and E. WHITLEY, ***Plain*** Mr. ***Knox*** (1960), more popular recent studies for the general reader; J.S. MCEWEN, ***The Faith of John Knox*** (1962), a study of Knox's religious views and their relation to those of the other reformers.

(J.S.McE.)

Koch, Robert

Robert Koch, one of the pioneers of the science of bacteriology, was born on December 11, 1843, at Klausthal, near Hanover, Germany. He attended the local secondary school and the nearby University of Gottingen, where he studied medicine, graduating in 1866. He then became a physician in various provincial towns. After serving briefly as a field surgeon during the Franco-Prussian War of 1870–71, he became district surgeon in Wollstein, in what was then Germany, where he built a small laboratory. Equipped with a microscope, a microtome (an instrument for cutting thin slices of tissue), and a homemade incubator, he began his study of algae, switching later to pathogenic organisms.



Koch.
BY courtesy of the Nobelstiftelsen, Stockholm

One of Koch's teachers at Gottingen had been the anatomist and histologist Friedrich Gustav Jacob Henle, who, in 1840, published the theory that infectious diseases were caused by living microscopic organisms. In 1850 the French parasitologist Casimir-Joseph Davaine, who was among the first to observe organisms in the blood of persons suffering certain diseases, reported the transmission of anthrax by the inoculation of healthy sheep with the blood of animals dying of the disease and the finding of microscopic rod-shaped bodies in the blood of the dead sheep. Inspired by the work of the French microbiologist Louis Pasteur, Davaine in 1863 showed that it was highly probable that, because the sheep did not become diseased in the absence of these rodlike bodies, anthrax was due to the presence of such organisms in the blood. The natural history of the disease was, nevertheless, far from complete.

It was at this point that Koch began. He cultivated the anthrax organisms in suitable media on microscope

Discovery of the life cycle of anthrax

slides, demonstrated their growth into long filaments, and discovered the formation within them of oval, translucent bodies—dormant spores. Koch found that the dried spores could remain viable for years, even under exposed conditions. This finding explained the recurrence of the disease in pastures long unused for grazing, for the dormant spores could, under the right conditions, develop into the rod-shaped bacilli that cause anthrax. The anthrax life cycle, which Koch had discovered, was announced and illustrated at Breslau in 1876, on the invitation of Ferdinand Cohn, an eminent botanist. Julius Cohnheim, a famous pathologist, was deeply impressed by Koch's presentation. "It leaves nothing more to be proved," he said. "I regard it as the greatest discovery ever made with bacteria and I believe that this is not the last time that this young Robert Koch will surprise and shame us by the brilliance of his investigations." Cohn, whose discovery of spores was published earlier in 1876, was also very much impressed and generously helped to prepare the engraving for Koch's epochal paper, which he also published. One of Cohn's pupils, Joseph Schroeter, found that chromogenic (colour-forming) bacteria would grow on such solid substrates as potato, coagulated egg white, meat, and bread and that these colonies were capable of forming new colonies of the same colour, consisting of organisms of the same type. This was the starting point of Koch's pure-culture techniques, which he worked out a few years later. That a disease organism might be cultured outside the body was a concept introduced by Louis Pasteur, but the pure-culture techniques for doing so were perfected by Koch, whose precise and ingenious experiments demonstrated the complete life cycle of an important organism. The anthrax work afforded for the first time convincing proof of the definite causal relation of a particular bacillus to a particular disease.

In 1877 Koch published an important paper on the investigation, preservation, and photographing of bacteria. His work was illustrated by superb photomicrographs. In this paper he described his method of preparing thin layers of bacteria on glass slides and fixing them by gentle heat. Koch also invented the apparatus and the procedure for the very useful hanging-drop technique, whereby micro-organisms could be cultured in a drop of nutrient solution on the underside of a glass slide. In 1878 Koch summarized his experiments on the etiology of wound infection. By inoculating animals with material from various sources, he produced six types of infection, each due to a specific micro-organism. He then transferred these infections by inoculation through several kinds of animals, reproducing the original six types. In this study, he also observed differences in pathogenicity for different species of hosts and demonstrated that the animal body is an excellent apparatus for the pure cultivation of bacteria.

Discovery of tubercle bacillus

Koch, now recognized as a scientific investigator of the first rank, obtained a position in Berlin in the German Health Office, where he set up a laboratory in bacteriology. With his collaborators, he devised new research methods. To obtain a pure culture outside the body, Koch mixed the organisms in melted gelatin; then, after solidification and growth, he placed portions of pure colonies into separate tubes of broth or other media. Koch also concentrated his efforts on the study of tuberculosis, with the aim of isolating its cause. Although it was known that tuberculosis was due to an infective agent, the organism had not yet been isolated and identified. By modifying the method of staining, the bacillus was discovered and its presence established in preparations of tuberculous material. A fresh difficulty arose when, for some time, it proved impossible to grow the organism in pure culture. But, eventually, Koch succeeded in isolating the organism on a succession of media. Its etiologic role was now established. On March 24, 1882, Koch announced before the Physiological Society of Berlin that he had isolated and grown the tubercle bacillus, which he believed to be the cause of all forms of tuberculosis.

Meanwhile, Koch's work was interrupted by the appearance of cholera in Egypt and the danger of its transmission to Europe. As member of a German government commission, Koch went to Egypt to investigate the disease; although he soon had reason to suspect a particular comma-shaped bacillus as the specific cause of cholera, he was frustrated by the cessation of the epidemic. Nevertheless, he discovered the cause of amoebic dysentery and the bacilli of two varieties of Egyptian conjunctivitis. Proceeding to India, where cholera is endemic, he completed his task, discovering the cholera organism and its transmission via drinking water, food, and clothing.

Resuming his studies of tuberculosis, Koch investigated what effect an injection of dead bacilli would have on a person who subsequently received a dose of living ones; he concluded that the local reaction produced might prove the means by which the disease could not only be diagnosed but, in the early stages, perhaps even cured. In his studies, he used as the active agent a sterile liquid produced from cultures of the bacillus. But this liquid (tuberculin, 1890) proved disappointing as a curative agent, and, consequently, its importance as a means of detecting a present or past tubercular state was not immediately recognized.

Additional work on tuberculosis came later in Koch's career, but, after the seeming debacle of tuberculin, he was occupied (1891–99) with a great variety of investigations into diseases of man and animals—studies of leprosy, rinderpest, bubonic plague, surra, Texas fever, and malaria. The spread of malaria was still a mystery; Koch had nearly satisfied himself that it was transmitted by mosquitoes when the British bacteriologist Ronald Ross published his findings pointing to the same conclusion.

In 1901 Koch reported work done on the pathogenicity of the human tubercle bacillus for domestic animals. He believed that infection of human beings by bovine tuberculosis is so rare that it is not necessary to take any measures against it. This conclusion was rejected by commissions of inquiry in Europe and America in an extensive and important work that was stimulated by Koch's theory, even though it proved to be partly wrong. But, in most cases, the successful measures of prophylaxis were those firmly laid down by Koch.

Not an eloquent speaker, Koch was, nevertheless, by example, demonstration, and precept, one of the most effective of teachers, and his numerous pupils—from the entire Western world and the Orient—were the creators of the new era of bacteriology. His work on trypanosomes was of direct use to the eminent German bacteriologist Paul Ehrlich; this is only one example of Koch's instigation of epochal work both within and beyond his own immediate sphere. His discoveries and his technical innovations were matched by his fundamental concepts of the etiology of disease. He was awarded the Nobel Prize for Physiology or Medicine in 1905 for his tuberculosis research. Koch died on May 27, 1910, in Baden-Baden, but, long before his death, his place in the history of science was universally recognized.

BIBLIOGRAPHY. BERNHARD MOELLERS, ***Robert Koch: Persönlichkeit und Lebenswerk, 1843–1910*** (1950), still the most comprehensive biographical and analytical account of Koch's life and work; RICHARD BOCHALLI, ***Robert Koch: Der Schopfer der modernen Bakteriologie*** (1954), a short but very good general review; ARTHUR IGNATIUS, ***Robert Koch: Leben und Forschung*** (1965), a recent survey that adds relatively little new material, but provides an excellent conspectus; LESTER S. KING, "Dr. Koch's Postulates," ***J. Hist. Med.,*** 7:350–361 (1952), a reconsideration of the famous postulates, suggesting that they are often misunderstood; "Robert Koch, 1843–1910," ***Proc.*** R. ***Soc.***, Series *B,* 83:xviii-xxiv (1911), a brief, clear, and expert obituary notice.

(L.G.S.)

Koko Nor (Lake)

Koko Nor (Ch'ing Hai) is the largest drainless mountain lake of Central Asia ***(nor [nuur]*** and ***hai,*** "lake" in Mongolian and Chinese, respectively). It is situated at an altitude of 10,515 feet (3,205 metres) in the Nan Shan (Nan Mountain) system of China in a depression at approximately 37° N, 100" E. Mongolians occupy the northern shores and Tibetans the southern shores. There are some sizable settlements (Ch'iang-hsi-k'ou, Ta-la-ma-ho) lying on the Hsi-ning to Lhasa road, close to the southern shore of the lake. On the northern shore lies the small settlement of Kang-ch'a.

Physical characteristics. The length of the lake approaches 65 miles, the width 40 miles; the surface of the lake is approximately 2,300 square miles (about 6,000 square kilometres) in area in years when the water level is high, and about 1,600 square miles (4,200 square kilometres) in area when the water level is low. The greatest known depth is 123 feet. The water is azure in colour, the name deriving from the Mongolian term for sky blue.

Koko Nor's origin

The Koko Nor depression originated between the Pliocene and the Pleistocene epochs (some 2,500,000 years ago). The lake that formed in the depression originally drained into the Ma-ch'u Ho, but the rise of the basin's mountainous frame made this impossible. Melting waters from ancient glaciers thus accumulated and formed a big, deep lake in the late Pleistocene (at least 10,000 years ago). At this time the ancient lake was nearly one-third larger than its present size, and its depth extended to almost 160 feet. The subsequent exhaustion of the glaciers brought about a reduction in the level of the Koko Nor to its present position.

The depression in which the lake is set is a plain more then 10,500 feet high. A rolling, hilly relief, with many low mountains, prevails in the north. In the south the depression is bordered by the Nan Koko Nor Shan (South Koko Nor Mountains), which form a narrow chain connecting with distinct peaks. These are continually snowcapped and go to the eastern edge of the lake. Further to the east the ridge drops sharply, becoming flat and opening out.

The depression is covered with mainly red and gray sandstone and light-gray and claylike limestones. Traces of human activity are present in mountain loesses.

The shores of the Koko Nor open out gently; the delta of the Pu-k'o Ho (Pu-k'o River) empties into the western part of the lake. Along the shores woods spread in terraces, the highest of which extends up over the lake to a height of 160 feet. On the eastern shore there are many small, isolated lakes and a rising, wooded shore area. Numerous sandy islands dot the lake; the largest has a length of 5,410 feet and a width of more than 1,000 feet. Bottom deposits are formed mainly by silts—black, yellow, and pale yellow. In places sand can be found, but close to the shore there are pebbles. The mineral content of the water changes greatly from year to year, but sodium chloride is always present. The water is brackish and unsuitable for drinking.

Twenty-three rivers and streams empty into the Koko Nor, the biggest being the Pu-k'o Ho. They flow fastest in summer, causing a rise in the lake level.

Precipitation

Climate. The Koko Nor basin is distinguished by a comparatively dry climate. During the winter, snowstorms rage through the first half of March. The amount of snow that settles, however, is not great. Most precipitation (more than 70 percent) occurs in July and August, partly by thunderstorm, partly by cloudburst. On the southwest shore of the lake and on the slopes of the Nan Koko Nor Shan there is an annual precipitation of ten to 12 inches, while on the northern shore 14 to 16 inches falls annually, increasing up to 20 inches in the mountains to the north of the depression. During the summer water in the lake warms to 64°–68° F (18°–20° C). From November to March the lake freezes over, the ice becoming as much as two feet thick.

Flora and fauna. Adjacent to the lake are luxuriant multigrain steppe grasses, providing one of the best grazing areas in the Nan Shan. The principal vegetation is of wormwood (absinthe) and derris. Nettles, hollyhocks, and asters are common, and numerous other plants are found. In the mountains there are fir forests.

Fish are found in the lake, mainly of the carp family. There are few large mammals in the neighbouring area, as the territory is populated. The kulankiang (wild ass) and the Przhevalsky horse are found. The oaran-kukuyaman (blue sheep) lives in the mountains. Wolves are also found. The waterfront and the adjacent slopes are inhabited by a large variety of birds, including skylarks, grouse, sandpipers, cormorants, falcons, eagles, gray geese, and a few types of duck and gull.

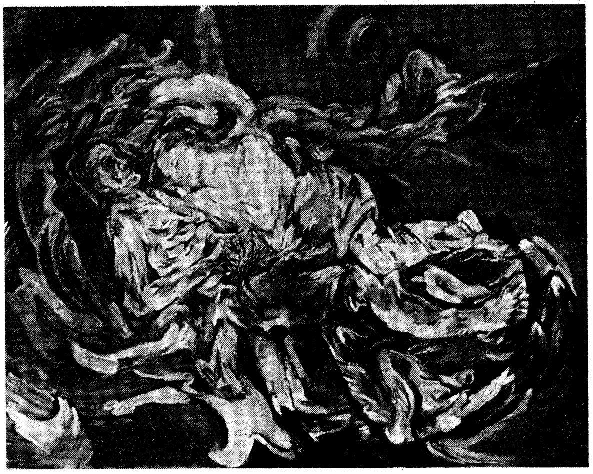
The chief occupation of the local population is nomadic cattle raising and tending of sheep, horses, and camels.

(N.T.K.)

Kokoschka, Oskar

Oskar Kokoschka, Austrian painter, humanist, and poet, has pursued a complex and productive career impelled by the dual concerns of art and humanitarianism. Most famous as an Expressionist artist, he has also written numerous plays. Both his art and his writing have reflected his political concern for the victims of war and dictatorship in the first half of the 20th century, and his outspoken political beliefs have forced him to spend much of his life in exile.

By courtesy of the Oeffentliche Kunstsammlung Basel, Switzerland



"The Tempest," self-portrait with Alma Mahler by Oskar Kokoschka, oil on canvas, 1914. In the Kunstmuseum, Basel, Switzerland. 1.81 X 2.21 m.

Oskar Kokoschka was born March 1, 1886, in the small Austrian town of Pöchlarn. When he was three, his father lost everything in a financial crash. The family was forced to move to Vienna, where his father worked as a travelling salesman, and his mother raised the children on limited means. Tragedy entered Kokoschka's life early, when his eldest brother died in 1891. Kokoschka attended elementary and high school in Vienna and received his first artistic impressions from the stained glass and Baroque frescoes of the Piarist church where he sang in the choir.

When he was 18, Kokoschka won a scholarship to the Kunstgewerbeschule (School of Arts and Crafts) in Vienna. Soon he became an assistant teacher, giving lessons at night and studying during the day. By 1907 he had also become a member of the Wiener Werkstatte (Vienna Crafts Studio), which supplied him with commissions until 1909. At the Kunstgewerbeschule he learned drawing, lithography, bookbinding, and other crafts. He was profoundly dissatisfied with the school, however, because it was devoted entirely to the decorative arts and completely omitted from its curriculum the study of the human figure. The Wiener Werkstatte, too, supported work only in the restrictive field of the decorative arts. From the beginning Kokoschka's primary artistic interest was the human figure; this interest is perhaps rooted in the deep concern for humanity that transcended even his concern for art. He tried to find practical means to pursue this interest. In his night classes he introduced the thin, muscular children of acrobats as models for his pupils, teaching the latter to make quick sketches—an innovation completely opposed to the aims of the school. He used the human figure as a decorative motif in the postcards, bookbindings, and bookplates he designed for Wiener Werkstatte commissions. Still, his real desire was to paint monumental pictures of people. He taught himself to paint in oils and executed some canvases; but economic necessity forced him to spend most of his time with decorative work, and the general artistic milieu in which he found himself continued in its failure to support his creative aspirations.

Friendship with Adolf Loos and early style

In 1908 he met the prominent Viennese architect Adolf Loos, who, having been impressed by one of Kokoschka's early paintings, took an active interest in the young artist. Like Kokoschka, Loos rejected the prevailing decorative ideal, and he enthusiastically launched Kokoschka's artistic career by introducing him to sympathetic artists, securing him commissions for paintings, and providing him with much-needed spiritual inspiration and support. During this early period Kokoschka painted mostly landscapes, developing a technique of vibrant, fluid lines and colours expressive of mood that formed the basis for all of his subsequent paintings. At first glance Kokoschka's landscapes seem to follow the principles of the Impressionist school because of their bright colours, ephemeral delineation of shapes, with tangled, multicoloured lines, and preoccupation with light. His vision, however, was different from that of the Impressionists, who sought, albeit in a revolutionary way, to represent only what strikes the eye. Kokoschka sought to express through his colours the inner sensibility of the observer viewing a scene. This aim is exemplified in one of his earliest paintings, "Dent du Midi," a snowscape in which the colours are warm, reflecting the response of the observer to the scene, rather than cool, evoking the actual light that must have emanated from the snow.

Humanist philosophy in Expressionist plays

At about this time Kokoschka began his career as a writer, composing several plays that heralded the new Expressionist theatre and expressed his compassionately humanist philosophy. The most important of them was Morder *Hoffnung* der Frauen (1907; "Murder Hope of Women"), a play that expressed his sensitivity to the moral crises of modern life and that was outspoken in condemning the political injustices of contemporary European society. He said in 1933 that in it he

> contrasted the callousness of our male society with my basic conception of man as mortal and woman as immortal; in the modern world it is only the murderer who wishes to reverse this state of affairs.

After a year in Berlin, where his first collective show was held, Kokoschka returned to Vienna in 1911, resuming his teaching at the Kunstgewerbeschule. At this point he used only models in motion for his classes. He exhibited paintings and drawings at Der Sturm gallery, along with works of the Russian Wassily Kandinsky, the Swiss Paul Klee, and the German Franz Marc. Soon public reaction to Kokoschka's plays caused such a scandal that he was dismissed from his teaching post.

In 1911 he met Alina Mahler, seven years his senior and the widow of the Austrian composer Gustav Mahler. He fell in love with her, and for three years they pursued a tempestuous affair that Kokoschka much later described as "the most unquiet time of my life." He succeeded in ending his involvement with her only with the outbreak of World War I and his enlistment in the Austrian Army.

Evolution of Kokoschka's portrait style

Shortly before his return to Vienna, Kokoschka had begun to paint more portraits, which show an extremely sensitive preoccupation with the character of the subjects and an increasing concern with expressing this character through colour. The earlier of these portraits make use of delicate, agitated lines describing figures painted in relatively naturalistic colours; solid colours are varied with multihued highlights, and certain features and gestures characteristic of the sitters or expressive of their psychology are exaggerated. Later, these portraits, still concentrating on hands and faces, are painted with increasingly broader strokes of more varied colour and heavier outlines, which, however, are broken and no longer solidly enclose forms. Kokoschka's most important painting of this period ("The Tempest") shows the artist and Alma Mahler resting together in a huge cockleshell in the midst of a raging sea. In this virtually monochromatic composition, all the forms are described by large, loose strokes of colour, and the direction of the strokes seems to cause the entire composition to swirl and spin. In all of these paintings, as with the landscapes, the involvement of the artist with the subject is an essential element, and this element continued to be the basis of his art throughout Kokoschka's life. In 1962 Kokoschka was to say,

> Painting . . . isn't based on three dimensions, but on four. The fourth dimension is a projection of myself. . . . The other three dimensions are based on the vision of both eyes . . . ; the fourth dimension is based on the essential nature of vision, which is creative.

Political disillusionment after World War I

Kokoschka saw active duty in the war for only a short time. In 1916 he was severely wounded and was taken to a military hospital in Vienna, then to one in Dresden. While recovering in Dresden he wrote, produced, designed, and staged three plays. In Orpheus und Eurydike (1918; "Orpheus and Eurydice") he expressed the terror he had experienced after being wounded. This play was adapted as an opera in 1926 by the German composer Ernst Krenek. The war and the takeover of the Russian Revolution by the Bolshevik regime disillusioned Kokoschka, as it did many intellectuals who had identified revolution with humanitarianism. He began to see revolution as a purely destructive force, and in 1920 he wrote a "Dresden Manifesto," which denounced all militancy in politics for its lack of human concern. Political and humanitarian themes disappeared for several years from his writing and art. The next ten years he taught, primarily as a professor of fine arts at the Dresdener Akademie (1919–23), and travelled in Europe, North Africa, and the Middle East.

Revival of humanitarian themes

In 1931 Kokoschka returned again to Vienna, where he completed his first political commission since the war, a joyful painting of children playing at an orphanage established by the Socialist city council. This painting was meant as a protest against the reactionary policies of the current Austrian chancellor. In 1934 he moved to Prague, where he met Olda Palkovska, his future wife. In Prague he was commissioned to do a portrait of the president of the Czech Republic, the philosopher Tomáš Masaryk. During the sittings he discussed with the aged statesman the philosophy of the 17th-century Moravian theologian John Amos Comenius, whose humanitarian views Kokoschka had admired from his youth. Kokoschka placed Comenius in the background of Masaryk's portrait, creating an allegory of the humanistic spirit from past to present. Comenius also became the subject of another play (*Comenius*, begun 1935).

In 1937 all of Kokoschka's works in Germany were removed by the Nazis from museums and collections as "degenerate art." This act outraged Kokoschka less for his own sake than because it boded ill for the future of culture and humanity. The same year, a great Kokoschka exhibition was held in Vienna, but Kokoschka was not encouraged. After the Munich Agreement between the English prime minister Neville Chamberlain and Adolf Hitler in 1938, he fled to London with Olda Palkovska.

Kokoschka's financial situation was so desperate in London that he was forced to paint mainly in watercolour. He began, however, a number of large canvases, entitled "The Red Egg," "Anschluss-Alice in Wonderland," "Loreley," "Marianne-Maquis," and "What We Are Fighting For," which were an anti-Fascist manifesto. They express his distress at the sufferings of humanity and are free from narrow ideological considerations; the series is an indictment of all the powers, not just the Fascist, that had caused suffering in World War II. In 1942 Kokoschka also painted a portrait of the Russian ambassador to London, Ivan Maysky, which contains a subtle warning against Soviet imperialism. The style of this painting is loose and expressive but calmer and more solid than that of his second Viennese period. He donated the fee for the portrait to the Red Cross for the care of German and Russian soldiers wounded at Stalingrad.

Public recognition and material success

After the war, beginning with a large but still basically unappreciated exhibition in Vienna, there began a series of shows of Kokoschka's works all over Europe and the United States. Kokoschka was financially secure for the first time. He continued to paint portraits and landscapes, and in 1950 he began his first great mythological composition, the three paintings of the "Prometheus Saga."

In 1953 Kokoschka established a seminar called Schule des Sehens (School of Seeing) at the Internationalen Sommerakademie für bildende Kunst (International Summer Academy for Visual Arts) in Salzburg. He continued to paint, completing in 1954 a second mythological trilogy, "Thermopylae." In the 1950s he designed tapestries and theatrical scenery and worked increasingly in lithography. He also continued his political art; his two moving posters of 1937 and 1945 protesting the effects of the Spanish Civil War and World War II on the children of Europe were followed in 1956 by a poster for Hungarian relief showing a stricken mother and a dead child.

Kokoschka's painting in the 1960s is perhaps best characterized by his "Herodotos" (1960–63), a luminously painted picture of the Greek historian inspired by visionary historical figures that appear in the space above his head. Kokoschka's late style is calmer, brighter, and more monumental than that of his early works. Although some critics have missed in the late paintings the agitation and surface intensity of his early works, the art historian Hans Maria Wingler expresses a true understanding of this most mature development of Kokoschka's style:

> What will determine the verdict of history is that the promise of his brilliant opening notes was fulfilled in the mature works of his old age, in which Kokoschka is drawing the balance of all his insight and experience. His power of transforming the inner substance of his mind into pictorial terms has grown with the years.

MAJOR WORKS

"Dent du Midi" (1909, private collection, Ziirich); "Still Life with Tortoise and Hyacinth" (1909; Österreichische Galerie, Vienna); "Adolf Loos" (1909; Staatliche Museen Preussischer Kulturbesitz, Berlin); "Portrait of Dr. Tietze and His Wife" (1909; Museum of Modern Art, New York); "Double Portrait (Oskar Kokoschka and Alma Mahler)" (1912; Edgar Horstman Collection, Hamburg); "The Tempest" ("Die Windsbraut"; 1914; Kunstmuseum, Basel, Switzerland); "Knight Errant" ("Der Irrende Ritter"; 1915; Solomon R. Guggenheim Museum, New York); "The Power of Music" (1919; Stedelijk van Abbemuseum, Eindhoven, The Netherlands); "Dresden, Augustus Bridge with Steamer" (1923; Stedelijk van Abbemuseum, Eindhoven); "London Bridge View of the Thames" (1926; Albright-Knox Art Gallery, Buffalo); "Tigon" (1926; Museum of Modern Art, New York); "Adele Astaire (Lady Cavendish)" (1926; Kunsthaus, Zurich); "Venice. Santa Maria Della Salute II" (1927; private collection, Paris); "Courmayeur" (1927; Phillips Collection, Washington, D.C.); "Self-Portrait with Cap" (1932; private collection, Epsom, England); "Tomaš G. Masaryk" (1935–36; Museum of Art, Carnegie Institute, Pittsburgh); "The Red Egg" (1940–41; Národní Galerie, Prague); "Anschluss-Alice in Wonderland" (1942; private collection, London); "Ambassador Maysky" (1942; Tate Gallery, London); "Loreley" (1942; private collection, Villeneuve); "Marianne-Maquis" (1943; private collection, Villeneuve); "What We Are Fighting For" (1943; private collection, Ziirich); "Cathleen, Countess of Drogheda" (1944–47; private collection, London); "Prometheus Saga" (1950; Count Antoine Seilern Collection, London); "View of Hamburg Harbour" (1951; Museum of Modern Art, New York); "Galatea" (1953; private collection, London); "Thermopylae" (1954; Universitat, Hamburg); "Sir Stanley Unwin" (1959; George Allen and Unwin Ltd. Collection, London); "Double Portrait (Oskar Kokoschka and His Wife Olda)" (1963–64; private collection, Villeneuve, France).

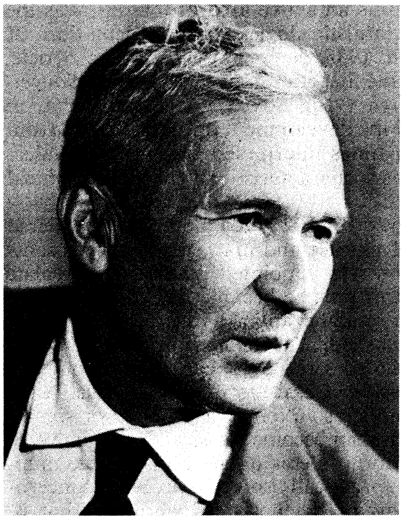
BIBLIOGRAPHY. Biographical and critical studies particularly recommended include: E. HOFFMANN, ***Kokoschka: Life and Work*** (1947); HANS M. WINGLER, ***Oskar Kokoschka, das Werk des Malers*** (1956; Eng. trans., ***Oskar Kokoschka: The Work of the Painter,*** 1958); BERNHARD BULTMANN, ***Oskar Kokoschka*** (1959; Eng. trans., 1961); JOSEF P. HODIN, ***Oskar Kokoschka: The Artist and His Time*** (1966); LUDWIG GOLDSCHEIDER, ***Kokdschka,*** written in collaboration with the artist, 3rd ed. (1967); REGINA BRANDT, ***Figurationen und Kompositionen in Dramen Oskar Kokoschkas*** (1967); and HEINZ SPIELMANN, ***Oskar Kokoschka: Handzeichnungen, Druckgraphik, Tapisserie*** 1965–1970 (1970). The artist's writings have been collected in ***Kokoschka Schriften 1907–1955*** (1956). See also the ARTS COUNCIL OF BRITAIN, ***Kokoschka: A Retrospective Exhibition of Paintings, Drawings, Lithographs, Stage Designs, and Books,*** 2nd ed. (1962).

(L.Go.)

Kolmogorov, A.N.

Andrey Nikolayevich Kolmogorov, the 20th century's most influential Soviet mathematician, contributed original general ideas in a number of branches of modern mathematics and, as a teacher, guided many young men to a mathematical maturity equal to his own. Raised to a high rank in the Academy of Sciences of his own country, he has been honoured in many other countries as well. His original contributions to abstract theories that describe chance include axioms that put probability theory on a nonintuitive logical basis. He is among the founders of the theory of stochastic processes (that is, infinite collections of random variables), particularly those called Markov processes, and his work has been honoured with equations that bear his name. Topology (the study of geometric structure that endures under continuous transformation), the representation of functions by Fourier series (or periodic components), and the theory of finite automata, which is at the confluence of logic and the theory of computing machines, have all been influenced by Kolmogorov. As a teacher of advanced mathematics as well as a leader in the state program of mathematics in the U.S.S.R. elementary schools, Kolmogorov has been known for his lectures and writing.

Novosti Press Agency



Kolmogorov, 1986.

Kolmogorov was born on April 25, 1903, in the central Russian town of Tambov. His father was an agronomist; his mother was from a family that adhered to progressive ideas. The young Andrey was raised by the sisters of his mother, after she died in childbirth.

Early career. At the age of 17 Kolmogorov graduated from the secondary school in Tambov and was enrolled in the Moscow State University. In the early period of his study in the university, he showed keen interest in Russian history and actively participated in a seminar conducted by Professor Sergey V. Bakhrushin, as a result of which he wrote a scientific paper on the registration of real estate in the ancient Novgorod republic in the 15th and 16th centuries. At this time Kolmogorov's interest embraced the ancient Russian arts, and, in fact, his interest in such art has been retained throughout his life. The lectures on mathematics that he attended during his first course at the university were delivered by professors Nikolay N. Luzin, Aleksey Vlasov, and Pavel S. Uryson, all among his country's most eminent teachers. Kolmogorov's active participation in Professor Vyacheslav V. Stepanov's seminar on trigonometrical series was especially important for his later career.

Set theory

In the autumn of 1921, when Kolmogorov was in the second course, he worked on a number of complex problems in the theory of trigonometrical series and operations on sets. Thus began his first highly productive period of research, which was closely tied with the basic directions of the Moscow mathematical school of that time, namely, with ideas in the theory of functions of real variables. In the spring of 1922 he completed a study in the theory of operation of sets (later published in 1928). In this period he also proved a theorem that gives a broad generalization to **a** well-known theorem, which was first formulated by a Soviet mathematician named Mikhail Souslin, concerning the necessary and sufficient conditions for a set to be of a certain type called a "Borel" set. Kolmogorov's result greatly influenced the research of a number of prominent mathematicians who later worked on this subject.

In June of 1922, Kolmogorov succeeded in attaining a result that was an outstanding sensation of that period: he constructed an integrable function whose Fourier series diverged at every point. His result, together with later work done in 1966 and 1967 by the Swedish mathematician Lennart Carleson and the U.S. mathematician Richard Hunt, settled the problem of convergence of Fourier series.

This period provided numerous publications on problems of differentiation, integration, and measurable sets —in every general problem Kolmogorov made considerable contributions.

Later Kolmogorov expanded his sphere of interest into problems of mathematical logic and the foundations of mathematics. One of the first results in this research, which he continued as an active area of interest, was the identification of "intuistic logics" with calculation of problems. Thus, in 1938, he published a large article, "Mathematics," in the first edition of the *Bolshaya Sovyetskaya Entsiklopediya* ("Big Soviet Encyclopaedia"), in which he described the development of mathematics from ancient to modern times interpreted in terms of dialectical materialism, as is characteristic of this philosophical attitude.

Law of large numbers

Probability theory. In 1925 Kolmogorov graduated from the faculty of physics and mathematics of Moscow State University and was appointed a research associate of the faculty. At this time he became interested in probability theory. In collaboration with another Soviet mathematician, Aleksandr Y. Khinchin, he studied problems of convergence of series composed of random elements. There followed a cascade of publications, each giving either a complete solution of a problem or significant progress, which then became classical. Thus, in 1928, he made advances in problems of the law of large numbers and in 1929 he practically solved a problem concerning the so-called law of the iterated logarithm (that is, the law of large deviations from the average in a prolonged sequence of statistically independent trials). He also augmented his significant study of conditions for the strong law of large numbers, the particular case of this work being a theorem generalized by him a little later: if random variables are mutually independent and have the same distribution, then their arithmetic mean, extended over *n* terms, converges with a probability of one to their common mathematical expectation.

Of great importance was a paper "General Theory of Measure and Probability Theory" published in 1929, which gave the first description of an axiomatic construction of probability theory based on measure theory. Ideas of this kind, aside from the earlier use of measure theory in the probabilistic researches of U.S. mathematician Norbert Wiener (*q.v.*), previously were put forward by Émile Borel, a French mathematician, and Anton Lomnicki, a Polish mathematician. The complete axiomatics of probability theory, as accepted in the latter half of the 20th century, were first given in the aforementioned publication. In 1933 Kolmogorov expanded the paper into the monograph *Grundbegriffe der Wahrscheinlichkeitsrechnung*, now well-known among mathematicians. In 1950 the monograph was translated into English under the title *Foundations of The Theory of Probability.*

Stochastic processes and topology. In 1931 Kolmogorov was elected a professor of Moscow State University; two years later he was appointed a director of the Institute of Mathematics of the university. This marked the second period of his scientific activities, which lasted about a quarter of a century and was characterized by broad conceptions of his studies. It is notable, first, that he elaborated the principles of stochastic process theory without aftereffect in the monograph *Analytical Methods of Probability Theory*. Other contributions concerned aspects of functional analysis—a comparatively recent branch of mathematics in which the traditional techniques of algebra and calculus are applied to entire collections of functions, which are mappings from one space to another, or equivalently ordered pairs of numbers. He also made contributions to studies of topology and turbulent flow.

Markov processes

Application of probability theory. In the study of probability theory, which is considered his basic speciality, Kolmogorov not only gave generalizations to a number of fundamental facts but also established new directions in research that drastically changed probability theory and, to a considerable extent, widened the possibilities of its application in numerous and different ways. In the publication concerning analytical methods in probability theory mentioned above, he laid down a theory of continuous processes without aftereffect, or, as it is called at present, Markov processes. He formulated two systems of equations in partial derivatives that bear his name; they have since become famous and are consistent with transition probabilities controlling the process. The elaboration of the theory was an answer to problems in physics and the theories of Brownian motion and diffusion. The work marked a new period in the development of probability theory and its application in physics, chemistry, civil engineering, and biology. Many prominent mathematicians worked in this direction; among them were U.S. mathematicians J.L. Doob and William Feller.

This fundamental study in probability was succeeded by a series of works that either continued the development of established ideas, introduced particular details of the theory, or expanded its application to classical problems of probability theory, physics, and chemistry. Of fundamental significance for limiting theorems concerning sums of independent random values were two papers on homogeneous stochastic processes without aftereffects. A fundamental study of Markov chains contained many new ideas in the theory of dynamic systems and served as a starting point for numerous works of other mathematicians.

Of considerable importance were Kolmogorov's investigations in problems of random stationary processes, the theory of which he related to Hilbert space geometry and the applications of which were numerous, including consequences for local isotropic turbulent flow. These investigations are widely known and much valued by mathematicians throughout the world. Wiener, the founder of cybernetics (the theory of control and communication as applied to animals and machines), had independently investigated many aspects of stationary processes, particularly those identified with statistical prediction, and he pointed out that the investigations by Kolmogorov were relevant to the statistical theory of information upon which the science of cybernetics in part depended.

Operator theory

In topology, Kolmogorov, simultaneously with U.S. mathematician James Alexander but independently of him, put forward certain considerations relating to an operator called the nabla operator and applied these to complexes and later to any topological space. Related algebraic structures called nabla groups turned out to be very effective and provided suitable instrumentation in the study of numerous problems of topology, including continuous maps. On this foundation, Kolmogorov put forward a conception of a homological ring, an important idea in topology. His formulation of the duality law (on equivalent representations of certain mathematical properties) was very important for the period of 1935–36. In this context the duality law concerned closed sets situated in a topological space that is locally bicompact and fully regular.

This period was also marked by works in geometry "On Topological Group Formulation of Geometry" and "On Formulation of Projective Geometry," as well as work in the areas of functional analysis and the optimum approximation of functions.

As a professor of the university and the director of its Institute of Mathematics, Kolmogorov always paid great attention to training young mathematicians in different branches of mathematics. Among those who were influenced by his thinking were Izrail M. Gelfand (functional analysis), Anatoly I. Maltsev (mathematical logics), Boris V. Gnedenko (probability theory), Grigory Bavli (probability theory), Ivan Verchenko (theory of real variable functions), Sergey M. Nikolsky (theory of real variable functions), and others who later became well-known mathematicians and were elected members of the Academy of Sciences and corresponding academy members in union republics and foreign countries.

In 1939 Kolmogorov was elected an academician of the Academy of Sciences of the U.S.S.R. and, a little later, an academician-secretary of the department of physical and mathematical sciences of the academy for a term of three years. He then guided the activities of the department, which was established to unite and direct scientific work of the most prominent researchers in this field.

In 1940, jointly with Khinchin, Kolmogorov was awarded the U.S.S.R. State Prize for investigations in the theory of stochastic processes.

During World War II Kolmogorov conducted comprehensive research in theories of firing and turbulent flow in what he considered to be his patriotic contribution to the struggle of the Soviet people against Nazi Germany.

Postwar period. In the mid-1950s, the third period of Kolmogorov's activities began. He became interested in information theory, dynamic-system theory, interconnections of information theory with function theory, classical mechanics, the 13th Hilbert Problem (representation of functions of a large number of variables by functions of a lesser number of arguments), finite automata, and complex theory. In all these topics Kolmogorov derived fundamental results. In this third period, as before, Kolmogorov was surrounded by numerous young talented mathematicians.

Contributions to education

Throughout the latter part of his life, Kolmogorov paid great attention to problems of mathematical education of schoolchildren. He was appointed chairman of the Commission for Mathematical Education under the Presidium of the Academy of Sciences of the U.S.S.R., he delivered lectures in a boarding school for talented young mathematicians gathered from all parts of the country, he wrote a number of introductory textbooks for general education, and he published numerous papers in pedagogy for parents and teachers. Under his leadership a new state program of training in mathematics in the Soviet general school was developed and introduced. In accordance with his proposal, an optional mathematics course was introduced in the ninth and tenth forms of general schools in the Soviet Union to permit fuller development of the scientific interests of schoolchildren.

The merits of Kolmogorov are praised both in the Soviet Union and abroad. He has been awarded the honourable title of Hero of Socialist Labour, the gold medal "Hammer and Sickle," the four highest Orders of Lenin, and the Order of Red Banner of Labour. Besides being the U.S.S.R. State Prize winner, he also won the Lenin Prize. Many foreign academies of sciences have elected him to membership: Paris Academy of Sciences, Royal Society of London, U.S. National Academy of Sciences, Polish Academy of Sciences, Romanian Academy of Sciences, and others. Kolmogorov is an honorary member of many mathematical societies and the American Meteorological Society, and holds an honorary doctorate of the Paris, Stockholm, and Warsaw universities.

(B.V.G.)

Konoe Fumimaro

Prince Konoe Fumimaro was a political leader during one of the most eventful periods in Japanese history—the mid-1930s to the end of World War II. As prime minister in 1937–39 and 1940–41, he tried unsuccessfully to restrict the power of the military forces and to keep Japan's war with China from widening into a world conflict.

Early life

Konoe was born October 12, 1891, to the foremost of the five families from among which regents (*sesshō*) and Imperial advisers (*kampaka*) were chosen. His father, Atsumaro, was a prominent statesman.

Konoe was an intelligent and sensitive child, but he tended to be nervous and timid, and he developed a cynical attitude toward society. He was educated with the intellectual elite of the nation, first attending Tokyo Imperial University and then Kyōto Imperial University, from which, during Japan's participation in World War I, he was graduated by the law faculty.

Konoe Fumimaro.
By courtesy of the International Society for Educational Information Tokyo, Inc.

As a student he became intimate with Western philosophy, literature, sociology, and social and economic problems, particularly the problems of the poor. One of his major advisers at Kyōto University later became one of Japan's leading Marxist economists. During this period, Konoe translated into Japanese Oscar Wilde's essay "The Soul of Man Under Socialism" (1891), which was published in an intellectual journal. Its sale was prohibited by the government because it was judged dangerous to the public order.

Unlike most politicians in Japan before World War II, Konoe did not begin his political career either through civil-service examination or through membership in a political party. Instead, he entered (*c.* 1920) under the tutelage of the *genrō* (elder statesman) Saionji Kimmochi, a perceptive advocate of parliamentary government and a member of the court aristocracy.

After the war Konoe was an attendant of the Japanese delegation to the Paris Peace Conference. At this time he wrote an article rejecting the Anglo-American principle of pacifism as a convenient cover for maintaining the status quo by nations which, unlike Japan, already possessed extensive spheres of influence. He also vigorously criticized as impractical the principle of an international league as a means of abolishing economic imperialism and racial prejudice.

Konoe entered the upper house of the Diet by virtue of his rank as prince. He advocated reform of that house and of the peerage system, opposed Fascism, and called for reform of the army general staff to prevent army interference in foreign affairs. His goal was the expansion of parliamentary politics. From the outbreak of war with China in Manchuria (1931), Konoe felt that a complete political renovation of the Japanese militaristic power structure, carried out by political leaders, was required in order to realize the nation's destiny. After serving as vice president of the upper house, Konoe was appointed president in 1933. He declined to form a Cabinet in 1936, when on Saionji's recommendation, he was asked to by the Emperor. But in June 1937 he agreed to form a non-party Cabinet, which he hoped would gain the support of the nation. He decided to adopt the most reasonable of the army's demands while controlling its more reckless elements. He declared that he sought to realize social and international righteousness and to alleviate internal friction and discord.

Prime minister

In July Chinese and Japanese forces clashed near Peking, and the two countries soon were engaged in full-scale, although undeclared, war. Konoe made various unsuccessful efforts to end the conflict, and in January 1939 his Cabinet fell. He was appointed head of the Privy Council and was given a Cabinet post in the Hiranuma Kiichirō Cabinet.

Konoe's first Cabinet had been plagued by the separation of state affairs and the army's right of supreme command. He believed that to restrain the army and to settle the war required a government based on political power derived from a national organization. In June 1940 he resigned as head of the Privy Council, planning to develop such a mass national movement, but, before his plans were fully developed, he was persuaded to form his second Cabinet. His plan for a mass, popular organization was finally realized later that year with the formation of the Imperial Rule Assistance Association.

That September Japan concluded a military alliance—the Tripartite Pact—with Germany and Italy. Meanwhile, amid steadily deteriorating relations with Great Britain and the United States, Konoe tried to prevent a widening of the Sino-Japanese conflict and thus avert American participation; in April 1941 he concluded a nonaggression pact with the Soviet Union.

Faced with a further deterioration of Japanese–American relations and an American encirclement of Japan, Konoe concluded that a solution to the Sino-Japanese conflict could be reached only through American mediation. Thus, from April 1941, he devoted all his energy to Japanese–American negotiations, hoping to enter into discussions with the U.S. president, Franklin D. Roosevelt. In July, the third Konoe Cabinet was formed in order to eliminate Foreign Minister Matsuoka Yōsuke, who opposed these negotiations. But in October Konoe resigned over differences with the army minister Tōjō Hideki.

With the widening of the war following the Japanese attack on Pearl Harbor in December 1941, Konoe came under military surveillance and was forced to leave the centre of politics. In 1944 he cooperated with other leading political figures to bring about the collapse of the Tōjō Cabinet. After the war, in 1945, he became deputy minister of national affairs in the Higashikuni Cabinet. In January 1946 he was served with an arrest warrant by the occupation army on suspicion of being a war criminal. On January 16, the day he was to report, he committed suicide by drinking poison.

BIBLIOGRAPHY. TEIJI YABE, *Konoe Furnimaro,* 2 vol. (1958), is an excellent and detailed biography (in Japanese). YOSEIITAKE OKA, *Konoe Fumimaro: Unmei no Seijika* (1972), provides a unique account of Konoe's personality and character in the historical context of his age.

(Sh.H.)

Koppen, Wladimir Peter

Wladimir Peter Koppen, Russian-born German meteorologist and climatologist, best known for his systematic method of identifying and mapping the world's climates, played a major role in the advancement of climatology and meteorology for more than 70 years. His achievements, practical and theoretical, profoundly influenced the development of atmospheric science.

Koppen was born September 25, 1846, in St. Petersburg (now Leningrad), and remained in Russia until he was 20. His grandfather was one of the German physicians invited to Russia by the empress Catherine the Great to improve sanitation in the provinces. He later became personal physician to the tsar. His father, Peter von Koyyen (1793–1864). worked at the Academy in St. Petersburg as geographer, statistician, and historian. In gratitude for his services to Russian culture, Tsar Alexander II (reigned 1855–81) appointed him Academician, highest academic rank in Russia. He also granted him in 1858 a seaside estate called Karabakh on the south coast of Crimea.

His father's scholastic success and versatility inspired Koppen at an early age to apply his own intellect and perception to the varied environment of the Crimea. The complex geography of the low mountain ranges along the Black Sea coast provided the setting for his first explorations. While attending secondary school at Simferopol (1858–64), some 30 miles north of Karabakh, where the coastal ranges yield to extensive plains, he frequently travelled the mountain route inland from the sea. The floral richness and climatic variety of the region, he later emphasized, first awakened his lasting interest in the geography of the plant world and its relationship to climate.

Early interest in geography

In 1864 Koppen began study at the University of St. Petersburg, specializing in botany. Koppen returned to Karabakh many times, and the environmental changes he saw between the dark northern forests and the subtropical shores of the Crimea broadened his geographical perspectives.

In 1867 Koppen transferred to the University of Heidelberg, completed his doctoral dissertation on the relation of plant growth to temperature, and received his degree in 1870. A mark of Koppen's extraordinary integrity was his insistence on travelling for his final examinations from Heidelberg, where the faculty might have been prejudiced in his favour, to the University of Leipzig to assure the impartiality of his examiners. Following the Franco-Prussian War (1870–71), in which he served in the ambulance corps, Koppen returned to St. Petersburg as assistant at the Central Physical Observatory. Three years later he accepted a position with the German Naval Observatory at Hamburg as head of the newly established division of weather telegraphy, storm warning systems, and marine meteorology. In 1879 he was given the new title of meteorologist of the observatory, and in 1884 he produced a world map of temperature belts, ranging from polar to tropical latitude, each distinguished by the number of months having temperatures above or below certain mean values.

Climatic classification system

A climax in geographical climatology was reached in 1900 when Koppen introduced his mathematical system of climatic classification. Each of five major climate types was assigned a mathematical value according to temperature and rainfall. Since then, many of the systems introduced by other scholars have been based on Koppen's work.

Koppen retired from his position at the Hamburg observatory in 1919 and moved to Graz, Austria, in 1924. In 1927 he undertook, with Rudolph Geiger, the editorship of a five-volume *Handbuch der Klimatologie*, which was nearly completed when he died on June 22,1940.

Throughout his distinguished career Koppen retained his intellectual flexibility. Well-informed on a broad range of subjects, he was keenly receptive to new ideas and methods, especially those offered by youthful scientists, who found in him a patient and constructive listener. Although he was not widely travelled, he knew a great deal about the world, and he saw his work and his nonprofessional interests in full global perspective. The deep concern he felt for his fellow man was evident in the time and energy he devoted to problems of land-use reform, school reform, improved nutrition for the underprivileged, alcoholism, and calendar reform. In the cause of world peace he strongly advocated widespread use of Esperanto, which he spoke as fluently as he did German and Russian. Between 1868 and 1939 he produced more than 500 publications, some of which he translated into Esperanto.

Work as social reformer

Koppen's fondness for children was well known. He was a founder of the Eimsbütteler Boys Home at Hamburg, where he was a frequent and regular worker. He also accepted into his family, which included his wife and their five children, a nephew and niece whose father had died. When a group of Russian students fled to Germany, he arranged housing for them in Hamburg and later assisted some of them to reach America. These selfless acts called for considerable sacrifice, for his means were limited.

Koppen was a small, dignified man. He was modest: he eschewed his inherited right to use "von" before his name, he rarely referred to his many honours, and he preferred to travel by third-class railway carriage. Köppen was one of the last scholars of an era when an erudite man could attain competence in, and make significant contributions to, many branches of natural science. Prominent among scholars of his own time, he helped to, pave the way for the scientific specialists of the 20th century.

BIBLIOGRAPHY. No English language study of Koppen **has** been published. In German, a scholarly biography by Koppen's daughter, ELSE **WEGENER-KOPPEN**, *Wladimir Koppen,* ein Gelekrtenleben (1955), is available.

(G.R.R.)

Köprülü Family

The Kopriilu family was a famous family of statesmen considered to have been responsible for preventing the Ottoman Empire from decline in the second half of the 17th century. The founder of the family, Kopriilu Mehmed Pasha, became grand vizier (chief minister) in 1656 when the Ottoman Empire was in a critical situation. Venetians were then threatening Istanbul after their capture of Tenedos (Turkish Bozca Ada) and Lemnos (summer 1656). Cut off from supplies, the city was in distress. Rising prices and continuous financial crisis had put the troops and the populace in the capital in a rebellious mood. The palace clique with the mother of the Sultan at its head was prepared to relinquish its power to a strong man. When Kopriilii was offered the grand vizierate, he asked the Sultan's mother, who was the real ruler behind the Sultan, to promise solemnly never to interfere or let anybody interfere with the grand vizier's absolute authority in the government. The pledge was solemnly reiterated by the Sultan himself at the ceremony of appointment (September 14, 1656).

Kopriilii Mehmed Pasha. For more than half a century power in the Ottoman state had been determined as a result of struggles and compromises between rival groups. Grand viziers who were supposed to represent the Sultan's absolute power had virtually become dependent on the Palace and the Janissary corps, or provincial forces, with the result that there was a lack of authority and complete disorder in the administration. Faced with crisis, the Palace chose Kopriilu Mehmed, an old vizier in retirement, recommended to the Sultan's mother by a clique as the wisest and most experienced man available. A product of the peculiar Ottoman institution of trained palace pages, he came from Rojnik, a village near Berat in Albania. After his services and training in the palace, he became a governor general in the provinces of Trabzon (Trebizond; 1644), Egri (Eger) (1647), Karaman (1648), and Anadolu (1650), and sat as vizier in the imperial council for only a week in 1652, and then, dismissed, he retired to Koprii, his father-in-law's seat, a small town in northern Anatolia, thence his nickname Kopriilii, of Koprii.

Well informed in the most complicated Ottoman politics, Kopriilu, once grand vizier, began by placing his supporters at the key positions and mercilessly repressing opponents and rivals. When challenged by the *sipahi*s (the Ottoman cavalry), who wanted to put another man, a former leader of the Anatolian mercenaries, in power, Kopriilii secured the support of Janissaries (the core of the Ottoman standing army) and *shaykh al-islām*, head of the *ʿulamāʾ* (scholars trained in Muslim religion and law), and thus nipped the rebellion in the bud. His second test came when he organized and led an expedition against the Venetians the next summer. His success in repelling the Venetian navy in the Dardanelles (July 19, 1657) and the subsequent recovery of Tenedos (September 4) and Lemnos (November 15) won him the prestige he needed to establish his authority. During this expedition, he was unusually severe against the Janissaries and others who had neglected their duties. Enjoying the absolute confidence of the Palace, Köprülü tried to re-establish central authority over the vassal princes beyond the Danube and in the provinces of Anadolu, Syria, and Egypt. An expedition against George II (Gyorgy Rákóczi), prince of Transylvania, resulted in Rákóczi's replacement by a new prince loyal to the court (1658) and later in the annexation of the provinces of Yanova (Jeno) (August 1, 1660) and Várad (August 27, 1660). But the Ottoman expansion in Transylvania opened a period of rivalry and wars between the Ottomans and Habsburgs for the next four decades.

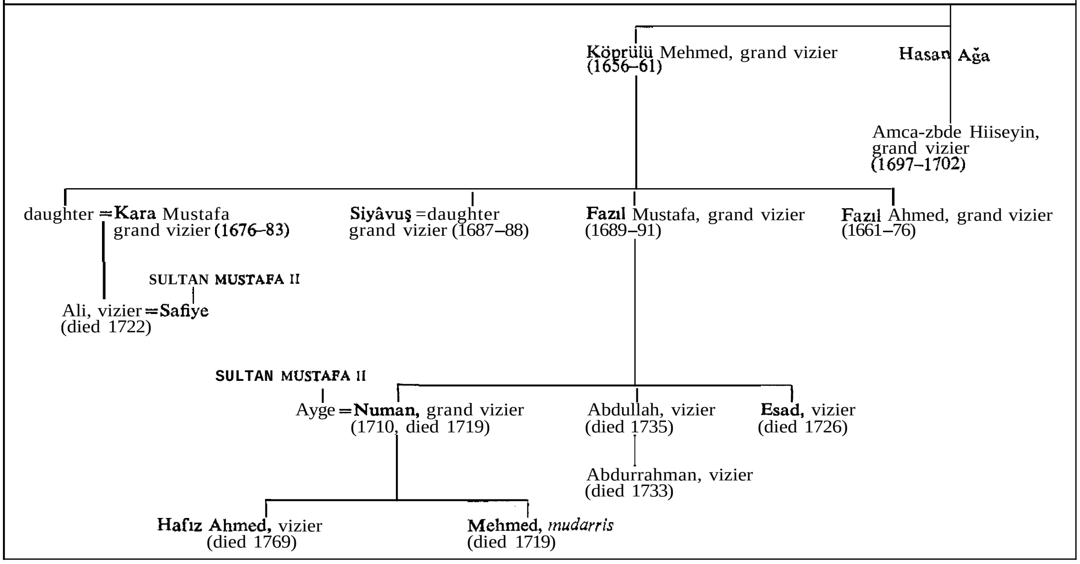
Military victories

As a protest against Köprülü's unusually despotic government the principal governors general in Anatolia and Syria had not joined the imperial army in Transylvania. In the fall of 1658, while Kopriilu was in the field, they rose up and marched toward the capital. Abaza Hasan, then Pasha of Aleppo and chief of the rebels, held his power as the leader of the sekbans, unruly mercenary troops in Anatolia. The skillful tactics of the old vizier eventually made his rivals powerless and all the rebel pashas were executed (February 1659). In the summer of the same year Kopriilu sent a general inspector to Anatolia with authority to exclude from the state registers all those from among the *reaya* (non-Muslim taxpayers) who claimed to belong to the military class. This measure, aimed primarily. at the suppression of the sekbans, reestablished central authority in the region. When Koprulu died in Adrianople on October 31, 1661, he had succeeded in restoring the central authority in the Ottoman Empire in accordance with its age-old principles of government. He had won the confidence and respect of the young sultan and his mother by his loyalty to the interests of the state. The Sultan followed his advice in appointing as grand vizier his son Fazıl Ahmed, a young man of 26, who had served as his deputy in his last days.

Fazıl Ahmed. The eldest son of Kopriilu Mehmed, Fazıl Ahmed was born at Köprü in 1635. His father had chosen for him a career in the learned professions. At 16 he was made professor but entered the civil service when his father became grand vizier. He was made governor general of Erzurum (1659) and then Damascus (1660) before he was called as deputy grand vizier during an illness of his father. Ahmed became grand vizier on November 1, 1661 and proved to be as energetic and skillful as his father in asserting his authority. When, during his first campaign against the Austrians in 1663, he learned of a plot against himself fomented by Şâmî-zâde Mehmed, chief of the secretaries and member of the clique who brought Kopriilii Mehmed to power, he did not hesitate to have him executed. Against those who belittled him and struggled against him he always had the unswerving support of the Sultan. Whenever Ahmed was away from the capital, Kara Mustafa, husband of his sister, was left there as his deputy. In the battlefield he also had the close cooperation of Gurcii Mehmed, Kaplan Mustafa, his brother-in-law, and other able generals. The prestige of the empire that his father had re-established was so great as to bring under his command during his campaigns in central Europe the auxiliary forces of the vassal principalities—Transylvania, Moldavia, Walachia, and the Crimean Khanate. He also followed his father's reforms aiming at ieducing the number of the Janissaries and *sipahi*s and also made them more efficient by restoring discipline among them. Ahmed also placed emphasis on the age-old Ottoman policy of protection of the reaya, which was possible only under a strong central government. Though his failure at the battle of St. Gotthard (August 1, 1664) tarnished his earlier success against the Austrians—the capture of Érsekújvar in September 1663—he was able to make a treaty favourable to the Ottomans at Vasvár (August 10, 1664). The Sultan received him in Edirne with great favours. Ahmed won unparalleled prestige when he conquered Candia (modern Iráklion) in Crete (September 27, 1669) after a siege of 28 months and thus terminated a long and dangerous war with Venice. An expedition against Poland in 1672 in which the Sultan himself took part was a great achievement for Kopriilii. Under the Treaty of Buczacz (October 18, 1672) Podolia was annexed to the empire and the Ukraine was surrendered to the Cossacks under Ottoman suzerainty. Poland was also to pay a yearly tribute of 22,000 gold pieces. But the Poles under John Sobieski rejected these onerous conditions and took up a counteroffensive, which caused the war to continue four more years until the Treaty of 26-rawno, October 17, 1676. Russian intervention in the Ukraine led Ahmed to invade the region (summer 1674). Exhausted and for some time ill as a result of long expe-

Fazıl Ahmed's appointment as grand vizier

The Kopriilii Family


Köprülü Mehmed, grand vizier (1656–61)
Hasan Ağa
Amca-zbde Hiiseyin, grand vizier (1697–1702)
daughter = Kara Mustafa grand vizier (1676–83)
Siyâvuş = daughter grand vizier (1687–88)
Fazıl Mustafa, grand vizier (1689–91)
Fazıl Ahmed, grand vizier (1661–76)
SULTAN MUSTAFA II
Ali, vizier = Safiye (died 1722)
SULTAN MUSTAFA II
Ayge = Numan, grand vizier (1710, died 1719)
Abdullah, vizier (died 1735)
Esad, vizier (died 1726)
Abdurrahman, vizier (died 1733)
Hafız Ahmed, vizier (died 1769)
Mehmed, mudarris (died 1719)

ditions and excessive use of wine, Ahmed died on November 3, 1676. Contemporary Ottoman sources found in him all the qualities requisite for an Oriental statesman —wide knowledge, wisdom, justice, and generosity. He also distinguished himself in Islāmic law and Persian literature. He left a cash fortune of over 300,000 gold pieces. Like other pashas he maintained a private force of *sekban*s (1,500 in number).

Kara Mustafa. After Ahmed the grand vizierate was given to Kara Mustafa, his brother-in-law. Ottoman power then at its zenith was used by this ambitious man to establish once and for all Ottoman sovereignty in the Ukraine by the war against Russia, 1678–1681, and central Europe by the siege of Vienna, 1683. His failure outside Vienna caused the collapse of the whole edifice the first two Kopriiliis had erected. The Austrian Emperor, Venice, and Poland made a Holy League (1684) against the Ottomans, in which Russia joined two years later.

Decline of the family

Consequent to the defeat at Vienna the Köprülü family lost its influence. Fazıl Mustafa, younger brother of Ahmed (born in 1637 at Koprii), vizier since 1680, had to resign. But later when his brother-in-law Siyâvuş became grand vizier he was made second vizier (October 2, 1687), and they both played a major role in deposing Mehmed IV. But soon rebels turned against them, and Fazıl Mustafa saved his life only with the protection of the Sultan. In 1689 when the Austrian army advanced in the Balkans Mustafa was called to the grand vizierate. In the campaign of 1690 he liberated Nish and Belgrade from occupation; he was killed fighting an imperial army under Louis of Baden at Slankamen on August 19, 1691; Mustafa was mortally shot while rushing to support his right wing.

Other members of thefamily. It fell to Kopriilii Mehmed's nephew Amca-zâde Hiiseyin Pasha, grand vizier between September 13, 1697, and September 29, 1702, to conclude the peace treaty with the allies at Carlowitz (January 26, 1699). Another grand vizier of the Kopriilii family was Numan Pasha, eldest son of Fazıl Mustafa. Becoming grand vizier on June 16, 1710, he was soon dismissed on August 17 because of his timid policy toward Russia and his inexperience in the political manoeuvres. His brother Abdullah Pasha, vizier in 1700 and commander in chief of an army sent against Iran in 1734, was killed in the battlefield (June 14, 1735). The third son of Fazıl Mustafa, Esad, vizier on February 19, 1718, was not good at administration and pursued a literary career. He died in Retimo in September 1726.

The fourth generation of the Kopriiliis still occupied important posts as viziers, governors, and commanders. Numan's son Hafız Ahmed was made vizier on April 20, 1731, and died as the governor of Egypt in 1769. Abdullah's eldest son Abdurrahman became vizier and governor of Van and later of Trabzon (Trebizond; died in 1733). Kara Mustafa's son Ali Pasha was a prominent statesman and general between 1701 and 1720.

Assessment

Most of the Kopriiliis lived on the rich religious endowments that the first two Kopriiliis had made. From the first Kopriilii on they were all interested in careers as *ʿulamāʾ*. The family was said to have God's aid and infallibility. One of them, Numan, was even venerated at his death as a saint. Public attention turned to them at critical times and exerted a strong pressure on the palace to bring them back to power. Also family ties and solidarity played an important role in bringing them to power and maintaining them there. One of the slaves of Kopriilii Mehmed, Siyâvuş, became grand vizier and prepared the way for Mehmed's grandson to grand vizierate. The Kopriiliis followed as a tradition a policy of abolishing abuses in tax collection and had a reputation of treating the common people with justice.

BIBLIOGRAPHY. JOSEPH VON HAMMER-PURGSTALL, ***Geschichte des osmanischen Reiches,*** vol. 6–7 (1830–31, reprinted 1963), useful information from the Ottoman chronicles; J.W. ZINKEISEN, ***Geschichte des osmanischen Reiches in Europa,*** vol. 4–6 (1856–59), supplements Hammer–Purgstall from Western sources; EDWARD S. CREASY, ***History of the Ottoman Turks,*** new and rev. ed., ch. 15–17 (1878, reprinted 1961), a summary of Hammer-Purgstall; EKKEHARD EICKHOFF, ***Venedig, Wien und die Osmanen, Umbruch in Siidosteuropa,*** *1645–1700* (1970), a comprehensive account of the Köprülüs' foreign policy.

(H.I.)

Korea, History of

The Korean people share a common racial origin with other peoples of North Asia, and the Korean language belongs to the Altaic language family of the region. There was a close relationship between Korean culture and that of neighbouring peoples in the Neolithic Period and the Bronze Age. For example, Korean combware pottery, widely used in the Neolithic Period, is commonly found in the region; Korean bronze daggers, belt hooks, and multi-knobbed mirrors also display the traits of bronze tools unearthed in the area. In this manner, the lineage and formation of early Korean culture can be roughly traced. One branch of the cultural ancestry went as far as to the Shantung Peninsula and formed an element of ancient Chinese culture; another went southward to Japan to form the main current of early Japanese culture.

Because of this background, the early Korean people were in constant confrontation with the Chinese. China was antagonistic toward the northern peoples, with whom the Koreans were associated, and naturally assumed an aggressive attitude toward the Koreans. Thus, Korea's formation and development were nurtured in its struggles against China.

After the unification of the Korean peninsula by Silla, Korea maintained friendly relations with China out of the need to deter invasions from the northern nomadic peoples. Later, Manchuria and Mongolia came under the successive control of the northern nomadic peoples and served as a base for invasion of Korea and China. To cope with this threat, Korea and China felt the need for a military alliance.

An even more important reason for friendly relations between Korea and China was cultural homogeneity. After the Iron Age, both China and Korea developed agricultural economies and thus shared many cultural similarities. China's culture was more advanced than Korea's, however, and Korea absorbed much from China—*e.g.,* written Chinese characters, laws and decrees, Confucianism, and painting. Nevertheless, Korea had its own language and invented its own alphabet; it developed its own way of life and adapted and improved upon all its borrowed culture to fit its own needs.

By the late 19th century the influence of the West began to be felt in East Asia. This was chiefly because Japan, the less advanced nation in the region, took the lead after opening its doors to the West. Stimulated by this, Korea tried to carry out reforms, doing away with traditional systems. But Korea fell prey to Japanese imperialism, which took advantage of reform movements in Korea for its own benefit. On liberation from Japanese rule in 1945, Korea was thrown into the vortex of world politics.

THE DAWN OF HISTORY

The Stone Age. Stone artifacts of the Paleolithic Period were unearthed at Kulp'o-ri in North Hamgydng Province (Hamgydng-pukto) and at Sŏkch'ang-ni in South Ch'ungch'ŏng Province (Ch'ungch'ŏng-namdo). Of 13 stratified Paleolithic sites, each cultural stratum produced chipped-stone tools of different shapes. Radioactive-carbon dating indicates the Paleolithic provenance of Sŏkch'ang-ni. Inhabited sites with round fireplaces were discovered there along with carved pebbles that show the way of life in the Paleolithic Period.

Neolithic life

The Neolithic Period was well established by 3000 BC, and a major characteristic was the use of combware pottery, chiefly found at seashore and river-basin sites, where inhabited places and shell mounds were also discovered. Stone spears and flint arrowheads have also been found, as well as bone hooks and stone weights, used for fishing. Remains of the Late Neolithic Period include stone plows and sickles, which indicate the beginning of farming. People lived in dugouts, mostly shallow round or rectangular hollows with fireplaces in the centre and possibly covered with thatched roofs. These shelters were huddled together in groups. The size of such villages is yet to be determined, but legends indicate the family members might have lived together, forming clan communities.

The use of metals and the emergence of tribal states. Bronzeware was probably first used about the 8th century BC, though some scholars think that it predates the 10th century. As the Bronze Age started, the design of pottery changed to undecorated earthenware. The uncovering of such pottery indicates that Bronze Age Korean people lived on hillsides, in dugouts raised slightly above ground. Half-moon-shaped stone knives and grooved stone axes show that rice farming was practiced, and bronze daggers and bronze arrowheads indicate wars of conquest. Dolmens, used as tombs, which were discovered in south Manchuria and the Korean peninsula, show the boundary of ancient Korean culture. Because only important persons were buried in dolmens, their number and location indicate that many small Bronze Age tribal states were probably established by powerful men.

The most advanced state was ancient Chosdn, established in the Taedong-gang (Taedong River) Basin. According to myth, the son of Heaven, Hwanung, descended to Earth and married a bear-turned-girl, who bore a son, Tangun, the founder of Chosŏn. Perhaps Tangun, who called himself a grandson of Heaven, ruled a tribal state in which rituals and politics were not separated. Chosŏn developed into a league of tribes in the area of the Taedong and Liao rivers (*c.* 4th century BC). Around this time ironware came to be used. Iron plows and sickles indicate the use of animals in farming and more efficient harvesting methods. Wooden houses were built on the ground, and ondol, a floor-heating system, was developed. Iron weapons were manufactured. The appearance of horse equipment and coaches indicates that horses and chariots were employed in wars.

Wiman (Wei Man in Chinese), said to have defected from China, became ruler of Chosŏn about 194 BC. More likely, he was indigenous to Chosŏn. Wiman's Chosŏn was overthrown by the Han Empire of China and replaced by four Chinese colonies in 108 BC.

DEVELOPMENT OF ANCIENT STATES

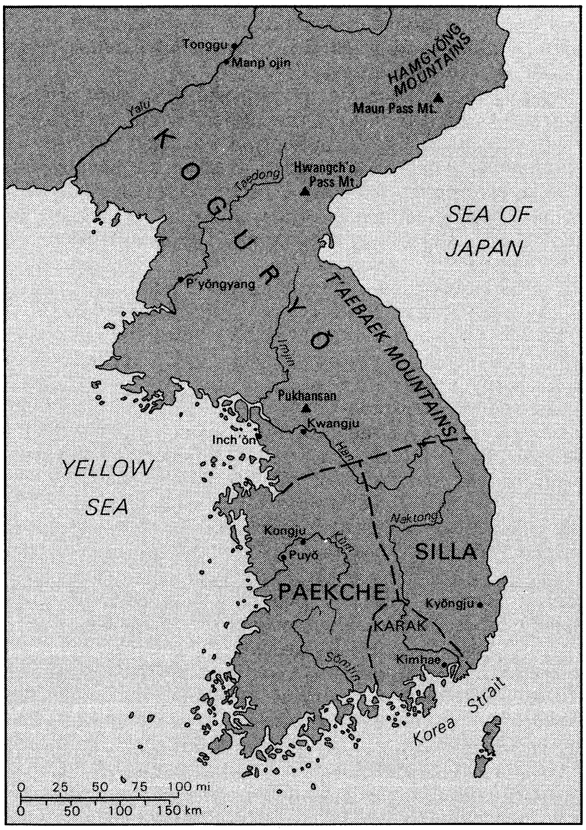
The Three Kingdoms. Apart from Chosŏn, the region of Korea developed into tribal states. To the north, Puyŏ developed in the Sungari River Basin in Manchuria. Chin. south of the Han-gang (Han River), was split into three — Mahan, Chinhan, and Pyŏnhan. These states were leagues, tribal federations centred on a leading state. The tribal league states stretched across a wide area from the Sungari Basin in Manchuria to the southern Korean peninsula. They evolved into three conflicting kingdoms—Koguryŏ, Paekche, and Silla. According to myths, Koguryŏ was founded by Kojumong in 37 BC, Paekche by Onjo in 18 BC, and Silla by Pak Hyŏkkŏse in 57 BC. The actual development of a state, however, was begun for Koguryd by King T'aejo (reigned AD 53–146?), for Paekche by King Koi (reigned 234–286), and for Silla by King Naemul (reigned 356–402).

Common characteristics of the Three Kingdoms

The Three Kingdoms shared several common characteristics. They evolved into statehood through frequent wars of expansion. Centralized military systems were organized, and training institutions (kydndang in Koguryŏ, hwarangdo in Silla) were developed. The power of the king in each state was strengthened. Royal hereditary systems were established.

Another common trait was the appearance of central aristocrats, tribal chiefs who moved to the capital. These

From E. Reischauer and J. Fairbank, East *Asia*, the *Great* Tradition, Copyright © 1958 and 1960 by Edwin O. Reischauer and John K. Fairbank; Houghton Mifflin Company



Korea during the Three Kingdoms period (*c.* AD 400)

aristocrats were divided into several social classes and gained certain privileges as they advanced socially and politically. The typical class system was Silla's Kolp'um, or bone-rank system, in which the families of rulers customarily monopolized political power. Silla had a state conference, Hwabaek, composed of men of *chin'gol,* or "true bone" (royal or formerly royal origin), which made important state decisions.

The states all experienced a centralization of power. The nations were divided into administrative units—the largest called *pu* in Koguryd, *pang* in Paekche, *chu* in Silla—that controlled many castles. The central government sent officials to these provincial units, where they saw to it that the people, as royal subjects, provided taxes and corvée labour.

The Three Kingdoms developed highly advanced cultures. Each compiled its history, apparently to consolidate the authority of the state. Also noteworthy was the introduction of Buddhism, characterized at the time as a nationalistic religion praying for the protection and welfare of the state.

Unified Silla. With the support of China, Silla conquered and subjugated Paekche in 660 and Koguryd in 668. Not until 676 did Silla drive out the Chinese and gain complete control of the Korean peninsula. But the surviving Koguryd people in northern Manchuria established Parhae, under the leadership of Dae Cho-ydng, which soon came into direct confrontation with Silla. This period could well be called an age of opposing southern and northern states, but it is customary to place the primary focus on Silla because little is known about Parhae (though it built a highly civilized state that the Chinese referred to as Haedong-sdngguk, the Prosperous Country of the East). After Parhae's collapse its territory was controlled by the northern nomadic peoples and was thus not part of Korean history.

Unified Silla saw the flowering of absolute monarchy, which reduced almost to nothing the influence of the Hwabaek. A central administrative body called Chipsabu was established to execute royal decrees. Aristocrats were now granted salaries and given land, but the latter was to revert to the state after tenure, thus reducing direct control of land and of the governed by the aristocracy. Monarchs built extravagant palaces and annexes and royal tombs. The nation was divided into administrative units: state, county, and ward. Five provincial capitals prospered as cultural centres.

Among the aristocracy, Avatamsaka Buddhism provided the ideological backing for autocratic monarchy. The underprivileged general public was attracted most to the *Sukhāvatī-vyūha-sūtra,* teaching bliss in the next world. The flowering of Buddhism produced many beautiful temples and great works of art, the most remarkable of which were Pulguk-sa, Sokkuram (a grotto shrine), and the bell at Pongdok-sa.

Confucianism prospered among low-echelon aristocrats, who used it as a foothold for political advancement. The National School, Kukhak, was established, and a sort of civil service examination system, called *toksŏ samp'um* kwa, was installed.

Disputes over succession in the 8th century

Emergence of provincial magnates. Frequent conflicts and rebellions over succession took place among the Silla aristocracy in the late 8th century. Aristocrats eventually restored the authority of the Hwabaek, overthrowing royal despotism. Even lower-ranking aristocrats demanded the abolition of restrictions imposed by the strict status system. New powerful families appeared in many provinces and grew because of the weakening of central control. Provincial military fortresses were established to repel Chinese pirates. The most active was the Ch'dnghae fortress led by Chang Po-go, who almost monopolized trade with China and Japan and had a private army of 10,000. Silla settlements in Chinese coastal cities in the Shantung Peninsula were also engaged in trade. Also powerful were the village rulers, who became castle lords by reinforcing control over military, administrative, and economic affairs. Many farmers, taxed both by the central government and castle lords, chose to become drifters or robbers, often staging rebellions.

Largely as a result of these trends, two provincial leaders, Kydnhwdn and Kungye, established the Later Paekche (892) and Later Koguryd, also called Majin or T'aebong (901), kingdoms, which, along with Silla, are commonly referred to as the Later Three Kingdoms. In this period Sdn (Zen) Buddhism was most popular, emphasizing the importance of realizing, through contemplation, the inborn Buddha nature of the individual.

KORYO

Social structure and culture. Koryd was founded in 918 at Songak (Kaesŏng) by Wang Kŏn, who in 936 established a unified kingdom in the Korean peninsula. Wang Kŏn went to great lengths to absorb the people of the overthrown states, even accepting the survivors of Parhae, which had been destroyed by the Khitan (Liao). Proudly declaring itself the successor of Koguryd, Koryd launched active campaigns to recover lost territory, clashing frequently with the Khitan in the north. Koryd eventually expanded its territory to the Yalu River (Korean Amnok-kang).

The Koryd ruling class consisted largely of provincial castle lords and former Silla aristocrats. The rulers regarded their family lineage highly. Marriage into a powerful family, especially a family of royal blood, was an important means for maintaining and elevating social and political status. Sons of a family above the fifth of nine official grades received official posts without going through civil service examinations.

The central government was run by the two supreme organs—Samsong, the highest administrative body, and Chungchuwon, the secretariat to the king. These two formed the Supreme Council of State. Koryd thus practiced aristocratic council-centred politics. Central aristocrats were granted land which was later to be returned to the state. Officials above the fifth grade were given land for permanent possession. Even the land supposed to be returned was actually handed down for generations because the grantees' sons usually became officials. Aristocrats expanded their land by reclaiming, purchasing, or seizing by force, and land became the primary source of wealth.

Aristocrats believed in Buddhism as a religion, for spiritual fulfillment and personal happiness; and in Confucianism, for its political precepts and ethical principles. The same was true of the government, which built grand Buddhist temples, such as Hŭngwang-sa, to observe rituals and pray for the prosperity of the nation, but which also set up a school named Kukchagam to teach Confucianism.

The military coup of 1170

Military rule. Civilian officials constituted the core of the ruling class; the military was generally discriminated against. Indeed, the supreme commander for military affairs was a civilian. Military officials were not eligible for the second grade of the official hierarchy and were excluded from the Supreme Council. Even in the same official grade, military men received less land than did their civilian counterparts. This discrimination eventually led to a military coup d'etat in 1170. The revolutionaries massacred a large number of civilian officials and gained complete control of government. Struggles for hegemony soon occurred among the leaders and were ended by Gen. Ch'oe Ch'ung-hdn, who established a military regime of his own that lasted about 60 years.

The monarch remained as a figurehead, deprived of political power, which was in the hands of the Ch'oe families. The Ch'oes set up a private army for personal security and a new public military organization for national security, but the latter also served, in effect, as their private army. They also established a body of civilian officials to take care of the state personnel administration, thus controlling both military and civilian leaders.

Buddhism was suppressed and retreated to remote mountain areas, where it formed a new Sŏn sect called Choge-jong, which became the main current of Korean Buddhism. The underprivileged farmers, stimulated by a general political atmosphere in which subordinates rose against superiors, staged rebellions across the country over a period of 30 years. The upheavals were at first natural and spontaneous protests against unfair oppression, but they developed into organized campaigns for liberation and for the seizure of power. The rebellions, eventually brought under control through appeasement and the use of naked force, were nevertheless instrumental in improving the lot of the underprivileged.

In 1231 the Mongols invaded Koryd, and the Ch'oe regime resisted for about 30 years. Even farmers and servants stood up bravely, and the Mongols, who had conquered most of Eurasia, could not take Koryd by force. As the exploitation of farmers by the Ch'oe grew more severe, however, the people became estranged, and the regime was finally overthrown by civilian leaders, who concluded a peace treaty with the invaders.

Social change in later Koryŏ. After the peace treaty Koryd was subject to some political interference from the Mongols but retained its political and cultural independence. Koryd went to some lengths to show its national and cultural superiority over the invaders by turning out highly advanced poems and national history books.

This period saw the remarkable development of huge farms, run by powerful aristocrats, which were scattered across the country. The landowners lived in the capital and sent private vassals or servants to collect taxes from the tenants, the common people who farmed the land. These tenants were often required to pay taxes to more than one owner because it became popular for landholders to share ownership. Tenants were also subject to forced labour and military duty for the state. Many farmers chose to become servants, thus getting protection from the aristocrats and avoiding the state duties. Some aristocrats also caught drifters and illegally made them servants. Thus, the number of servants increased, but these servants were actually on a level with tenants. They were not slaves in the Western sense but were more like serfs. The increase in the number of landholders and servants reduced the source of state tax revenue and the number of people to be mobilized in war.

Through civil-service examinations, the central government recruited a new bureaucratic force consisting of ***sadaebu*** (scholar officials), who generally had small farms in their native towns under their own management. These men were usually not satisfied with Buddhism and mere interpretations of the classics, and they adopted Neo-Confucianism, which brought a metaphysical approach to the understanding of the universe. Because of the financial straits of the government the new officials could not receive land commensurate to their rank; thus, their demand for land reform was strong. Eventually, with the support of Gen. Yi Sdng-gye, they seized power and established a new land-distribution system, under which land was granted according to the rank of office. These reforms ended the Koryd dynasty in 1392 and established the Yi dynasty. (K.-b.L.)

End of the Koryd dynasty

THE YI DYNASTY

The establishment of a Confucian state. The Yi dynasty was established and the region under its control named Chosdn with permission of the emperor of China. The Yi dynasty, with 26 monarchs, ruled until the Japanese annexation of Korea in 1910. Hanyang (now Seoul) was made the capital. The Confucian ethical system was adopted officially and replaced Buddhism, which had become corrupt. Many Confucian institutions of learning were set up, and Neo-Confucian scholars gained government posts through civil-service examinations.

The early Yi dynasty flourished intellectually and culturally, especially in the reign of Sejong the Great, the fourth monarch. With the technique of movable-type printing, developed in Korea in 1234, many publications in such fields as medicine, astronomy, geography, history, and agriculture were produced. In 1420 a royal academy called Chiphydnjdn was established, and many promising young scholars engaged in study and research there. In 1443 the Korean phonetic alphabet, Han'gŭl, was completed under Sejong's direction.

In the reign of Sejo, the seventh monarch, a powerful centralized and civilian-oriented government structure emerged. Laws were codified. The highest administrative body was the Supreme State Council. The country was divided into eight administrative provinces, and all officials were appointed by the central government.

Late in the 15th century, Korean scholars made original contributions to the theoretical refinement of Confucianism. In the mid-16th century many of these scholars were recruited to government positions. Idealistic in orientation, they criticized the bureaucratic establishment and recommended drastic measures toward the realization of Confucian ideals. But relentless counterattacks and pressures forced most of the scholars to quit their posts. They set up private academies of Confucian learning, called *sŏwŏn*. These academies produced many eminent scholars, including Yi Hwang (T'oegye) and Yi Yi (Yulgok), whose distinct theories of the universe evolved into mutually antagonistic schools.

Foreign invasions. In 1592 Toyotomi Hideyoshi, the military ruler of Japan, sent a large expeditionary force to Korea in an alleged attempt to invade China. Korean land forces suffered a series of defeats, but Korean naval forces, led by Adm. Yi Sun-shin, secured full control of the sea. Yi won the greatest naval victory in Korean history over the Japanese troops off the southern coast. People of almost all ranks, even Buddhist priests, volunteered to fight the Japanese. Contingents came from Ming China to help Korea. After about a year the Japanese were forced to withdraw. In 1597 Toyotomi launched another invasion, but after his sudden death, in 1598, the Japanese again withdrew. The war left most of Korea in waste. Palaces, government buildings, and private houses were burned; cultural treasures were lost or destroyed. Some scholars and artisans were taken to Japan, where they were required to teach the methods of Korea's more advanced technology.

Korea's greatest naval victory

In the early 16th century, nomadic Manchu violated the borders of both Ming China and Korea. Ming and Korean punitive attacks on Manchu strongholds in 1619 were beaten back, and in 1627 Manchu nomads overran the northern regions of Korea. Only after Korea had agreed to recognize their claim to "brotherhood did the Manchu pull out of the occupied territory. In 1637 the Manchu captured Seoul and wrested an unconditional surrender from the king. The Manchu then overthrew the Ming and established the Ch'ing dynasty. Thus, the tribute that Korea had paid to the Ming now had to be switched to the Ch'ing.

Silhak and popular culture. A series of significant changes in Korea began in the mid-17th century and made a great impact on virtually every sector of Korean society in the 18th century. In agriculture, rice transplantation became popular. Irrigation systems were improved. Advances in farming brought dramatic boosts in agricultural produce and raised the standard of living for farmers. With the cultivation of such special crops as tobacco and ginseng, commerce and trade developed apace. The government started minting coins and collecting farm rents in cash. Markets were held in many places across the country. Particularly active were merchants from Kaesdng, who had a nationwide network that put every fair in the country within their sphere of influence.

In the realm of scholarship, attention shifted from traditional theorizing to matters of relevance to the needs of the society and nation. Such scholars are often referred to as the Silhak, or Practical Learning school. Silhak scholars fell into four major groups. One group advocated comprehensive administrative reform, calling upon the government to rationalize the systems of civil-service examination, education, taxation, and land administration. Another group stressed the need to foster commerce, industry, and technology. A third conducted critical examinations of the Confucian Classics. The fourth group focused on the study of the history, geography, and language of Korea. Their research and publications provided a basis for study for the generations that followed.

Comparable new trends appeared in arts and letters. Mass-oriented literary and artistic works came into fashion, a great change from the tradition of catering exclusively to the upper classes. The new works were not only written in the easy-to-read Han'gŭl but also gave frank expression to popular discontent. Singing dramas or traditional Korean operas, most of them adapted from popular novels, were also popular with the masses. Many artists specialized in pictures of blacksmiths at work, farmers in the field, traditional wrestling matches, and in landscapes of the countryside. Pottery with a simple blue and white glaze was produced in large quantities for mass consumption.

Arrival of Europeans

The introduction of Catholicism. Europeans began to arrive in East Asia in the mid-17th century. In 1656 a Dutch merchant ship went aground off the southern shore of Cheju-do (Cheju Island). The 36 survivors were taken to Seoul for detention. About 13 years later Hendrik Hamel and seven others escaped and returned home. Hamel wrote an account of his experiences—the first book on Korea published in the West.

Along with the European merchants in East Asia came Catholic priests. Korea's first contact with Christianity was through missionaries in China. Korean envoys to China in the 16th century brought back with them a world atlas and scientific instruments made by the priests, as well as literature on science and Christianity. Some Silhak scholars had converted to Catholicism by the late 18th century, even before missionaries reached Korea. Most of the early converts were aristocratic scholars. A number of commoners were later attracted to the Catholic Church, finding a hope of salvation in the Christian doctrine of equality of all men before God and a new source of joy in the Christian belief in life after death. Catholicism spread from Seoul to the provinces slowly but steadily.

The incompatibility of Catholicism with Confucianism posed a serious problem. The two sides could not compromise on the great importance Confucianism attached to ancestor worship, which Catholicism rejected as pure idolatry. The government began to suppress the Catholic Church in the belief that it defied the existing sacrosanct mores of Confucianism. During persecutions in 1801, 1839, and 1866, scholar converts were either put to death or forced to turn renegade; foreign missionaries were ferreted out and beheaded. But rank-and-file Catholics rallied around the church and kept it alive. In 1831 the Holy See set up a Korean parish. French priests smuggled themselves into the country and engaged in clandestine activities.

The advent of Silhak, mass-oriented arts, and Catholicism in the 17th and 18th centuries is an indication of a modern Korea in the making. But in the 19th century, young princes came to the throne, and power-hungry maternal relatives seized power and plunged the government into a state of collapse. One peasant uprising followed another in the provinces, and the whole nation was seething with popular discontent and resentment.

Many farmers sought refuge in religion. A new religion founded (*c.* 1860) by Ch'oe Che-u, a fallen aristocrat scholar, advocated sweeping social reform; it had much in common with traditional animism and it appealed to the farmers. This religion was called Tonghak, or Eastern Learning, as a counterpoise to Sdhak, or Western Learning—*i.e.*, Catholicism.

CONTACT WITH WORLD POWERS

The opening of the door. King Kojong was too young to rule when he ascended the throne in 1864, and his father, Taewŏn-gun, was appointed regent. Taewŏn-gun set out to restore all the powers of the monarchy and pursued a policy of national isolation to devote government efforts to mending its own fences. He put into force bold political reform measures, such as liberal civil service recruitments and the abolition of a great number of private Confucian academies.

During his regency Western men-of-war and merchant vessels came in search of trade and friendship, but the regent refused them. Korean soldiers and civilians burned and sank the merchant ship USS "General Sherman" near P'yŏngyang in revenge for acts of plunder committed by the crew. Korean forces repulsed two attacks by French warships in 1866. In 1871 a United States flotilla came to retaliate for the "General Sherman" but was beaten back. Such incidents strengthened Taewdn-gun's resolve to keep the country's doors closed.

Treaty with Japan

Japan made repeated futile attempts to establish diplomatic relations with Korea. The Japanese military thereupon raised an outcry for a war of conquest on Korea. Meanwhile, Taewdn-gun came under widespread criticism for the enormous financial burden he had imposed on the people. Popular resentment forced him to step down as regent in 1873. Relatives of Queen Min took over the helm of state and initiated policies opposed to Taewŏn-gun's. Japan, which had been watching every development in Korea, now dispatched a fleet and pressured Korea to sign a treaty of trade and friendship. The ports of Pusan, Wŏnsan, and Inch'dn were opened to the Japanese in 1876.

The growing Japanese presence in Korea was disturbing to the Ch'ing rulers of China. When conservative soldiers tried to restore Taewdn-gun, the Ch'ing used it as an excuse for stationing forces in Korea. Thus began a period of open Ch'ing interference in Korean affairs. The Ch'ing forced Korea to sign a treaty of trade that heavily favoured Chinese merchants. Korea signed a treaty of trade and friendship with the United States through the offices of China. Similar treaties with Great Britain, Germany, Russia, and France followed, and resident foreign missions were established in Seoul.

Once the doors were opened, a modernization movement was begun. Students and officials were sent to Japan and China. Western-style schools and newspapers were founded. The government, however, could not push ahead with a consistent policy of modernization, for the King was feebleminded and the ruling classes were hopelessly divided into radicals and moderates.

In a coup of 1884 the Radicals seized power and drew up a bold blueprint for reform. But a Ch'ing contingent moved in and overthrew their three-day-old regime. This led to the signing of the Li-Itō Convention, designed to guarantee a Sino-Japanese balance of power in the Korean peninsula.

The Tonghak Revolt and government reform. Government expenditures greatly increased, largely because of appropriations for machinery imports and government reorganization, and the financial picture was aggravated by obligations to pay reparations to foreign governments. Heavier tax levies were imposed on farmers, who provided the bulk of government revenue. The import of such necessities as cotton textiles upset the traditional self-sufficiency of the farming community. Usurious loans by Japanese rice dealers helped reduce the peasantry to abject poverty. Angry farmers turned increasingly to Tonghak.

Despite ruthless government persecution, Tonghak took deep root in the peasantry. Its followers staged large-scale demonstrations calling for an end to injustice. Negative official response precipitated the Tonghak Revolt (1894), in which the Tonghak followers and the peasantry put up a united front for popular liberation. Government troops armed with Western weapons suffered ignominious defeats in southern Korea, weakening the government's military grip on the country. Foreign intervention seemed the last resort open to the rulers. Ch'ing troops soon moved in at the request of the government. Then Japan dispatched its large units uninvited, and the two alien powers were in sharp and sudden confrontation.

The rebels laid down their arms voluntarily to defuse the threat. But a war broke out in 1894; Japan emerged victorious, and the two powers signed the Treaty of Shimonoseki, recognizing Japanese hegemony in Korea.

Japanese hegemony

The Japanese dictated to the Korean government a wide range of reforms. Korea set up a council to plan and initiate reforms and issued pertinent decrees. Western-style institutions and a cabinet were formed. Civil-service examinations were discontinued. Such social practices as class discrimination were abolished. But public reaction to the reform policy was unfavourable. The government realized that old practices and institutions die hard and that reform takes more than mere decrees and imitation of things Western.

International power struggle and Korea's resistance. Japan's supremacy in Korea and its subsequent seizure of the Liaotung Peninsula in Manchuria was more than Russia, with its long-cherished dream of southward expansion, could tolerate. With German and French support, Russia pressured Japan to return the peninsula to China. At the same time, encouraged by Russia, the Korean government showed signs of drifting toward an anti-Japanese course. The Japanese government, however, promptly engineered the assassination of Queen Min (October 1895), the suspected mastermind behind the anti-Japanese attitude. Fearing for his own life, the King took refuge in the Russian Legation, granting such concessions as mining and lumbering rights to Russia and other powers.

Popular movements for the restoration of Korean sovereignty arose under the leadership of such figures as Sŏ Chae-p'il (Philip Jaisohn). After many years of exile Sŏ organized, in 1896, a political group called the Tongnip Hyŏphoe (Independence Association); he also published a daily newspaper named *Tongnip Sinmun* (The Independent) as a medium for awakening the populace to the importance of sovereignty and civil rights. On the urging of the Tongnip Hyŏphoe, the King returned to his palace and declared himself emperor and the Korean empire equal to other nations.

The Boxer Rebellion in China led to a Russian invasion of Manchuria and to the Russo-Japanese War (1904–05). The Korean government at first declared neutrality. But under Japanese pressure, it signed an agreement allowing Japan to use much of the country for operations against the Russians.

Japan won the war, and the resulting Treaty of Portsmouth (September 1905), signed through the mediation of the United States, recognized Japan's undisputed supremacy in Korea. Its hand thus strengthened, Japan forced the Korean emperor into signing a treaty that made Korea a Japanese protectorate (December 1905).

Although the Korean emperor sent a secret emissary to the international peace conference held at The Hague in 1906 to urge the big powers to intercede with Japan on behalf of Korea, the mission failed, serving only to infuriate Japan. Under Japanese coercion, the Emperor then abdicated in favour of his son, Emperor Sunjong. The Korean Imperial Army was disbanded, and in 1910 Japan annexed Korea.

Japanese annexation of Korea

A Korean Army, led by deposed officials and Confucian scholars, had arisen against the Japanese in the southern provinces following the 1905 treaty. For five years the militiamen effectively harassed the Japanese occupation forces, especially in 1908 and 1909. With the annexation, however, they were driven into Manchuria. Large numbers of Koreans emigrated to Manchuria, Shanghai, and Hawaii around this time.

KOREA UNDER JAPANESE RULE

Military control. Japan set up a government general in Seoul, with the governor generalship filled by generals or admirals appointed by the Japanese emperor. Koreans were deprived of freedom of assembly, association, the press, and speech. Many private schools were closed because they did not meet the standards arbitrarily set by the government general. The colonial authorities used their own school system as a tool for assimilating Korea to Japan, placing abnormally heavy stress on the Japanese language and excluding such subjects as the language and history of Korea. The Japanese built a nationwide transportation and communications network and established a new monetary and financial system. They also encouraged Japanese business activities in Korea while barring Koreans from similar operations.

The government general promulgated a land-survey ordinance that forced landowners to report the size and area of their land. By failing to report, many farmers were deprived of their land. Land and forestry owned jointly by a village or a clan were also appropriated by the Japanese because no single individual could lay claim to them. Much of the land acquired by the government general was sold cheaply to Japanese. Many of the dispossessed took to the woods and became brand tillers; others emigrated to Manchuria and Japan in search of jobs (the majority of Korean residents in Japan today are their descendants).

The March 1st Movement. A turning point in Korea's resistance movement came on March 1, 1919, when nationwide anti-colonial rallies were staged. The former emperor Kojong, the supreme symbol of independence, had died a few days before, bringing mourners from all parts of the country into Seoul. A declaration of independence was read at a rally in Seoul on March 1. Waves of students and citizens took to the streets, calling for independence. This movement took the form of peaceful demonstrations appealing to the conscience of the Japanese. An estimated 2,000,000 persons took part. The authorities responded with brutal repression, unleashing their gendarmerie and army and navy units, and thus ending the demonstrations. They arrested some 47,000 Koreans, of whom about 10,500 were indicted, and killed and wounded nearly 23,000.

In April, independence leaders, including Syngman Rhee, An Ch'ang-ho, and Kim Ku, formed a Korean provisional government in Shanghai. It brought together all Korean exiles and established an efficient liaison with leaders inside Korea. Japan realized that its iron rule had to be replaced with more sophisticated methods. The gendarmerie gave way to an ordinary constabulary force, and some freedom of the press was granted. But the oppressive and exploitative Japanese colonial policy remained ruthless as ever, though with less conspicuous methods.

Korean provisional government in exile

Taking advantage of a wartime business boom, Japan took leaps forward as a capitalist country. Korea became not only a market for Japanese goods but also a fertile and untapped market for capital investment. Meanwhile, industrial development in Japan was achieved at the sacrifice of agricultural production, creating a chronic shortage of rice. The government general enforced projects for the improvement of rice production throughout Korea. Many farmers were ordered to turn their dry fields into paddies. The program was temporarily discontinued during the worldwide depression in the early 1930s. But it was soon resumed to meet the increased needs of the Japanese military when Japan went to war against China, beginning in 1931. Most Koreans were forced to subsist on low-quality cereals imported from Manchuria instead of their own rice.

The end of Japanese rule. Of the several dailies and magazines founded shortly after the March 1st Movement, the newspapers *Dong-a Zlbo* and *Chosun Zlbo* spoke the loudest for the Korean people and inspired them with the ideals of patriotism and democracy. In the academic community, scholars conducted studies of Korean culture and tradition. Many novels and poems were written in colloquial Korean.

A major anti-Japanese mass rally was held in Seoul in 1926, occasioned by the funeral of Sunjong. A nationwide student uprising originated in Kwangju on November 11, 1929, calling an end to Japanese discrimination. These and other resistance movements were led by the New Cadre, composed of a whole spectrum of Korean intellectuals.

In 1931 the Japanese imposed militaristic rule once again. After the outbreak of the Sino-Japanese War (1937) and of World War II (1941), Japan tried to obliterate Korea as a nation, forcing Koreans to worship at *Shintō* temples and even to adopt Japanese names and banning academic societies devoted to Korean studies as well as newspapers and magazines published in Korean. The Japanese needed manpower to replenish the dwindling ranks of their military and labour forces. They drafted hundreds of thousands of able-bodied Koreans to fight for Japan and to work in mines, factories, and military bases.

When Shanghai fell to the Japanese, the Korean provisional government moved to Chungking. It soon formed a liberation army of Korean independence fighters scattered all over China and declared war on Japan in 1942. The small liberation army fought with the Allied forces in China until the Japanese surrender in 1945, which ended 36 years of Japanese rule over Korea. (K.-r.L.)

DIVISION OF KOREA

The Cairo Declaration, issued on December 1, 1943, by the United States, Great Britain, and China, pledged independence for Korea "in due course." The vague phrase aroused the Korean provisional government in Chungking (southwest China) to request interpretation from the United States. Their request, however, received no answer.

At the Yalta Conference held in February 1945, Pres. Franklin D. Roosevelt proposed to Joseph Stalin of the U.S.S.R. a four-power trusteeship for Korea between the U.S., Great Britain, the U.S.S.R., and the Republic of China. Stalin generally agreed to Roosevelt's offer, but they did not reach any formal agreement on the future state of Korea, and after the Yalta meeting there was a growing uneasiness between the Anglo-American Allies and the U.S.S.R.

Initial concord on Korea

Throughout the Potsdam Conference in July 1945, U.S. military leaders insisted on encouraging Soviet entry into the war against Japan. The Soviet leaders asked the U.S. about invading Korea, and the U.S. declined to do so on the grounds that such an expedition would not be practicable until after a successful landing had taken place on the Japanese mainland. The ensuing Potsdam Declaration included the statement that "the terms of the Cairo Declaration," which promised Korea its independence, "shall be carried out. . . ." In the terms of its entry into the war against Japan on August 8, the U.S.S.R. subscribed to support the independence of Korea. On the following day Soviet troops went into action in Manchuria and landed on the northern tip of Korea.

The General Order No. 1, drafted on August 11 by the United States for Japanese surrender terms in Korea, provided that Japanese forces north of the 38th parallel of latitude were to surrender to the Soviet commanders while those south of that line were to surrender to the U.S. commander. Stalin did not raise objections to the contents of the order, and on September 8 American troops arrived in southern Korea almost a month after the first Soviet entry. On the following day the United States received the Japanese surrender in Seoul. There were now two zones, for the Soviets had already begun to seal off the 38th parallel.

The historic decision to divide the peninsula has aroused speculation on several counts. Some believe that the division of Korea was made simply for the sake of military expedience in receiving the Japanese surrender. Others tend to think that the division was politically motivated to prevent the Soviet forces from occupying the whole of Korea. Considering the fact that American policy toward Korea during World War II had aimed to prevent any single power's domination of Korea, it may be reasonably concluded that the reason for the division was to stop the Soviet advance south of the 38th parallel.

The South. The end of Japanese rule caused political confusion among Koreans in both zones. In South Korea various political parties sprang up. Although they were roughly divided into rightists, leftists, and middle of the roaders, they had a common goal; the immediate attainment of self-government. As early as August 16, some Koreans organized a Committee for the Preparation of Korean Independence. On September 6 the People's Republic of Korea was proclaimed by delegates attending a national assembly called by the committee. The republic was headed by Woon-hyung Lyuh, who was closely associated with the leftists. But the U.S. military government, under Lt. Gen. John R. Hodge, the commanding general of the United States armed forces in Korea, refused to recognize the republic, asserting that the military government was the "only government" in Korea, as stipulated in General Order No. 1. The U.S. policy in Korea was to establish a trusteeship that would supersede both the U.S. and the Soviet occupation forces in Korea. The exiled Korean provisional government, on returning, declared itself a political party, not a government.

In late December the Council of Foreign Ministers (representing the United States, the Soviet Union, and Great Britain) met in Moscow and decided to create a four-power trusteeship of up to five years. On receiving the news, Koreans reacted violently. On February 14, 1946, to soothe the discontent, the military government created the Representative Democratic Council, as an advisory body to the military government. This body was composed of Koreans and had as its chairman Syngman Rhee, former president of the Korean government in exile.

In October the military government created an Interim Legislative Assembly, half of whose members were elected by the people and half appointed by the military government. The assembly was empowered to enact ordinances on domestic affairs but was subject to the veto of the military government. The anti-trusteeship feeling came to a climax a few months later, when the assembly condemned trusteeship in Korea. The military government thereupon declared this condemnation as "illegal."

National resentment at treatment by the trustee powers

The North. Unlike the U.S. forces in the South, the Soviet Army marched into the North accompanied by an army of Korean Communists. By placing the latter in key positions of power, the U.S.S.R. easily set up a Communist-controlled government in the North. On August 25 the People's Executive Committee of South Hamgydng Province was created by the South Hamgydng Province Communist Council and other nationalists. The Soviet authorities recognized the committee's administrative power in the province, thus setting a pattern for the committee's role throughout the North Korean provinces. In this way the Soviet Union placed the North under its control without actually establishing a military government. On October 19 the people organized the Bureau of Five Provinces Administration, a central governing body; but this was replaced on February 9, 1946, by a Provisional People's Committee for North Korea. This new agency adopted the political structure of the Soviet Union.

Kim 11-sung, who returned in the uniform of a major of the Red Army, was introduced to the people as a national hero on October 14, 1945. Shortly after his public appearance, Kim was elected as the first secretary of the North Korean Central Bureau of the Communist Party. In February 1946 the Provisional People's Committee for North Korea was organized with Kim as its chairman and took over the existing central administrative bureaus. A year later, in February 1947, a legislative body was established under the name of the Supreme People's Assembly, and with the strong support of the Soviet occupation authorities Kim began to consolidate his political power. The North Korean Workers Party had been created on August 23, 1946. By December of that year its membership had reached 600,000.

***The two* republics.** After the simple discussion on trusteeship between Stalin and Roosevelt at the Yalta meeting, it was not until May 1945, at a meeting with the Chinese Foreign Minister in Moscow, that Stalin confirmed his agreement to set up a four-power trusteeship. The ensuing Moscow Conference, held in late December, created a four-power trusteeship and established a Joint U.S.–U.S.S.R. Commission of the rival U.S. and Soviet military commands in Korea for the settlement of the question of a unified Korea. When the Joint Commission convened in Seoul from March to May 1946, the Soviet delegates demanded that those Korean political groups that had opposed trusteeship be excluded from consultation. The United States refused, and on this rock foundered all attempts by the commission to prepare for the unification of Korea. The commission met again from May to August 1947, but it achieved nothing for the creation of a unified Korea.

The United States presented the entire matter of Korean unification to the United Nations in September 1946. The United Nations General Assembly adopted a resolution, proposed by the United States, rejecting the Soviet position. The resolution called for general elections in Korea under the observation of a UN Temporary Commission on Korea, those elected to make up a National Assembly, establish a government, and arrange with the occupying powers for the withdrawal of their troops from Korea. But the U.S.S.R. barred the Temporary Commission from entering North Korea. The South, however, held elections under the supervision of the Temporary Commission on May 10, 1948. The National Assembly convened on May 31 and elected Rhee as its speaker. Shortly afterward a constitution was adopted, and Rhee was elected president on July 20. Finally, on August 15, the Republic of Korea was inaugurated, and the military government came to an end. On December 12 the UN General Assembly declared that the republic was the only lawful government in Korea.

Construction of the People's Republic

Meanwhile, on November 18, 1947, the Supreme People's Assembly of North Korea set up a committee to draft a North Korean constitution. The committee adopted a new constitution in April 1948, and on August 25 elections for members of the Supreme People's Assembly were held with a single list of candidates. On September 3 the constitution of the Democratic People's Republic of Korea was ratified by the Supreme People's Assembly, which was holding its first meeting in P'yŏngyang. Kim 11-sung was appointed premier of the People's Republic, and on September 9 the Democratic People's Republic of Korea was proclaimed. The U.S.S.R. recognized the People's Republic as the only lawful government in Korea on October 12.

THE KOREAN WAR

South Korea began to organize a police constabulary reserve in 1946. On December 14, 1948, the Department of National Defense was established. By June 1950, when the war broke out, South Korea had a 98,000-man force equipped only with small arms, which was barely enough to deal with internal revolt and border attacks. The United States occupation forces completely withdrew from Korea by June 29, 1949, leaving behind them a force of about 500 men as a U.S. Military Advisory Group to the Republic of Korea to train the South Korean armed forces. On June 7, 1949, in a message to the United States Congress calculating the security measures necessary for the protection of the Pacific from Communist domination, Pres. Harry S. Truman declared that Korea had become a testing ground in the ideological conflict between Communism and democracy. On that ground, he asserted, U.S. aid should be granted with a long range program. On October 6, 1949, the United States granted South Korea $10,200,000 for military aid and $110,000,000 for economic aid for the fiscal year 1950, the first year of a contemplated three-year program. In addition, the U.S. Congress approved $10,970,000 for military aid on March 15, 1950. The military equipment committed under the U.S. military-assistance program was still en route, however, when North Korean troops invaded the South. South Korea was unprepared to resist the total invasion from the North.

Early in 1946 the Soviet authorities in North Korea had organized a 20,000-man constabulary and army units, and in August 1946 the North Korean Army was established, its title being changed to the Korean People's Army in February 1948. The Soviet occupation forces left North Korea in December 1948, leaving behind 150 advisors for each army division for training purposes. On March 17, 1949, the U.S.S.R. concluded a reciprocal-aid agreement with North Korea in which it agreed to furnish heavy military equipment; and by June 1950 North Korean forces numbered 135,000, including a tank brigade. As early as 1946, the Soviets were sending thousands of Koreans to the U.S.S.R. for specialized training, and during 1949–50 the People's Republic of China transferred about 12,000 Korean troops from its army to the North Korean forces. The North Korean forces were thus far superior to the forces of South Korea in training and equipment.

Cost of the war

The North Korean troops launched a full-scale invasion of South Korea on June 25, 1950. The war concluded with an armistice on July 27, 1953, having lasted for three years and one month and having accounted for about 4,000,000 casualties, including civilians. South Korean casualties were some 1,313,000 (1,000,000 civilians); Communist casualties were estimated at about 2,500,000 (including 1,000,000 civilians). The United States lost 33,629 dead in action, South Korea 47,000, and the UN forces 3,194; but the estimated losses of the People's Republic of China in action were 900,000 men and of North Korea 520,000. During the war, 43 percent of Korea's industrial facilities was destroyed and 33 percent of its homes devastated.

UN intervention. On June 26 (June 25, New York time) the UN Security Council approved a resolution describing the invasion of South Korea as a "breach of the peace and an action of aggression" and called upon the members to render every assistance in restoring peace. The Soviet Union was unable to impose a veto because its delegate had been boycotting the meetings to protest the fact that the People's Republic of China had no seat in the United Nations. On June 27 President Truman issued the order for United States air and naval forces to resist Communist aggression in Korea, and that afternoon the Security Council of the United Nations ratified Truman's decision to send air and sea aid to Korea, calling upon the UN members to render such assistance to Korea as might be necessary to restore peace. But Seoul, the South Korean capital, fell on June 28, and most of the South Korean Army was destroyed. On June 30 Truman ordered United States ground forces in Japan into Korea; the first U.S. troops reached the battlefield on July 4. The UN approved the creation of a unified command in Korea, and Gen. Douglas MacArthur was appointed commander. Sixteen member nations sent armed contingents, but the United States furnished the great bulk of the air units, naval forces, supplies, and money.

The North Koreans continued to advance recklessly despite the presence of U.S. troops in the field. In early August the UN retreat came to an end in a defense perimeter along the Naktong-gang (Naktong River) line (forming a semicircle in southeast Korea). South Korea was now almost overrun by the North Koreans, except for the small beachhead around Pusan-hang (Pusan Harbor; in Korea's extreme southeast). On September 15 MacArthur counter-attacked, catching the Communists on the flank by an amphibious attack on Inch'bn (on the coast west of Seoul). They were trapped and either surrendered or fled in panic. By October 1 the UN forces were back at the 38th parallel. On September 27 the U.S. Joint Chiefs of Staff ordered MacArthur to destroy the North Korean armed forces, and two days later Truman authorized him to advance into North Korea. On October 7 the UN General Assembly approved the resolution to permit entry into North Korea and created a UN Commission for the Unification and Rehabilitation of Korea. On October 20 the UN forces entered P'yŏngyang, the North Korean capital, and on October 26 reached the Manchurian border at the Yalu River.

Chinese intervention. The Chinese Communists, who had moved troops along the Yalu after the Inch'ŏn landing, in November entered Korea in overwhelming numbers. By the end of 1952, 1,200,000 Chinese were engaged in the war under the command of P'eng Te-huai. They forced the UN forces to retreat in disorder, and Seoul was re-evacuated on January 4, 1951. But around P'ybngtaek (about 30 miles south of Seoul) the Chinese were halted, and in February the UN General Assembly formally condemned the People's Republic of China as an aggressor. The UN counter-offensive began in late January. By March 31 the UN forces had again reached the 38th parallel. MacArthur now publicly advocated an extension of the war to China because of the Chinese intervention, but this advocacy was regarded as a challenge to the United States president's conduct of foreign policy. Consequently, on April 11, Truman dismissed MacArthur from all of his commands, and Gen. Matthew B. Ridgway took his place. From then until the armistice, the UN forces fought a holding action along the 38th parallel; indeed, in many places the UN forces were slightly north of the line.

Armistice and aid. The Soviet delegate to the United Nations proposed a discussion of a cease-fire and an ar-

Negotiating difficulties

mistice on June 23, 1951, and on July 10 negotiations began between the United Nations and the Communist commanders at Kaesdng, later resumed at P'anmunjdm (both about 30 miles northwest of Seoul; the former in North Korea, the latter in South Korea). Many issues stood between the two negotiators, the first being the Chinese demand that all foreign troops be withdrawn from Korea in the face of the steadfast refusal of the United States to withdraw all UN troops from South Korea. The second issue was the boundary: the Communists demanded the restoration of the 38th parallel, but the United States insisted on the existing battle line. The third and most important issue was that of prisoners. The UN forces held 171,000 prisoners, 50,000 of them unwilling to return to their Communist countries. The Communists, not to lose face, were determined to have all prisoners back. On this matter the negotiations were deadlocked and did not resume until after the death of Stalin in March 1953. The United States administration under Dwight D. Eisenhower was inaugurated in early 1953 and, deeply concerned with balancing the budget, was determined to end the impasse even if this involved resumption of hostilities. On the other hand, the war weariness of the Communists was increasing. In April the first 6,670 Communists and the 684 UN personnel were exchanged at P'anmunjdm. Soon the prospects for armistice negotiations seemed to improve. The Communists agreed to hand over to a neutral commission the UN-held prisoners of war who did not wish to be repatriated, But Syngman Rhee opposed any term that would leave Korea divided and demanded that the military offensive be resumed. On June 18 Rhee suddenly released 27,000 North Korean anti-Communist prisoners in defiance of the United Nations, whereupon the Communists broke off negotiations. On July 20 negotiations were resumed. Rhee gave in and agreed to support the armistice even though he would not sign it. In return the United States promised to extend economic aid and conclude a mutual-security pact to protect South Korea against further aggression.

Armistice signed

The armistice was signed on July 27, 1953. The United Nations had won most of its demands. The military line became the boundary between North and South Korea, and commissions were established to enforce the cease-fire regulations. A Neutral Nations Commission for Repatriation was entrusted with the repatriation of prisoners, 21,809 of whom—among them 7,582 Korean and 14,227 Chinese—chose to stay in South Korea or go to Taiwan.

The U.S. Army provided Korea with $181,200,000 during the occupation period of 1946–48. This money, which was provided under the assistance programs for occupied areas, was spent mainly on preventing hunger and disease. For the period of 1949–52, the U.S. provided $485,600,000 for economic aid and $12,500,000 for military aid. After the war the UN Korean Reconstruction Agency (UNKRA) was established to carry out economic aid to South Korea, 34 member and five nonmember states contributing $148,500,000. The UNKRA came to an end in 1958, but meanwhile UN Emergency Relief also contributed $474,400,000 and other international voluntary agencies $85,000,000. Most of the UN contributions were provided by the United States, and total U.S. aid to South Korea from 1946 to 1978 exceeded $5,893,900,000.

POSTWAR KOREA

South Korea. The political experience of the people of the Republic of Korea since 1948 represents a wide range of variation. They have lived under 12 years of Syngrnan Rhee's authoritarian rule, tempered by the rise of a vocal opposition coalition under semicompetitive conditions; the collapse of this system after student demonstrations that brought the nation close to civil war in April 1960; nine months of multiparty liberalism under the cabinet system of Prime Minister Chang Mydn that permitted competition among conservative and emergent socialist parties; and, since May 1961, a presidential system dominated by the armed forces, which had maintained a tight and effective rule over Korea.

The First Republic. The first Korean republic, established in 1948, adopted a presidential system and elected Syngman Rhee as the first president. He was re-elected in August 1952 while the nation was still at war. Even before the outbreak of the Korean War there had been a serious conflict between Rhee and the opposition-dominated National Assembly that elected him in 1948. The dispute involved a constitutional amendment bill that the opposition introduced in an attempt to defeat Rhee by replacing the presidential system with a parliamentary cabinet system. The bill was defeated, but the dispute was carried down to Pusan, the wartime provisional capital, where the National Assembly was reconvened.

Early disagreement over president's role

When the opposition introduced another amendment bill in favour of a parliamentary system, Rhee counteracted by pushing through an amendment bill that provided for the popular election of the president. Later, in 1954, Rhee succeeded in forcing the National Assembly, then dominated by the ruling party, to pass a constitutional amendment bill providing a life-term presidency for himself. On the basis of the revised constitution, Rhee was able to run successfully for his third term of office in May 1956: Rhee's election for the fourth time, in March 1960, was preceded by a period of tension and violence, followed by student demonstrations resulting in many casualties. Rhee resigned under pressure and fled to exile in Hawaii, where he died in 1965 at the age of 90.

The Second Republic. The second Korean republic, which adopted a parliamentary system in place of the presidential system, lasted only nine months before it was overthrown by a military coup in May 1961. With a figurehead president elected by both houses of the legislature, power was shifted to the office of Prime Minister Chang Mydn, who was elected by the lower house by a narrow margin of 10 votes.

In spite of some strenuous efforts to initiate reforms in a society already blemished by social and economic ills accumulated over a long period of time, the Chang regime was unsuited to deal with the explosive situation created as an aftermath of a violent political change. To make the situation worse, factionalism came to prevail in political life. As the ultimate source of authority now shifted to the office of the prime minister, constant efforts were made by conservative and more moderate factions to coalesce with a group of independents, either to keep or upset the majority within the legislature. Even before the Chang regime launched a full program of economic reform, the Democratic elite was seriously weakened by these factional struggles within its ranks.

Military junta of 1961

The military coup. With the military seizure of political power, postliberation Korean politics entered a new phase characterized by inflow of military officers into the government. The military junta that took over the government on May 16, 1961, dissolved the National Assembly and imposed a strict prohibition on political activity. The nation was then placed under martial law, and the Supreme Council for National Reconstruction (SCNR), headed by Maj. Gen. Park Chung Hee, took the reins and began to establish a new system of national government.

In November 1962 the SCNR made public a constitutional amendment bill that provided for a strong president and a weak, single-chamber National Assembly. The bill was approved by a national referendum held one month later. Then, in February 1963, Chairman Park of the SCNR issued a statement that he would not take part in the civilian government to be formed later in the year if civilian political leaders would uphold a nine-point "political stabilization proposal." On February 27, 1963, Chairman Park made a public pledge not to take part in the new government, for 53 political and military leaders had sworn to support his stabilization proposal.

In March, however, following bitter turbulence in the ruling junta itself and a chaotic situation created by the proliferation of minor political parties, Chairman Park proposed that military rule be extended for four years. The proposal met vigorous opposition from civilian political leaders, but some 160 military commanders, most of general officer rank, submitted a resolution to Park supporting the extension. Under considerable pressure, Park again changed course and announced, on April 8, 1963, a plan for holding elections toward the end of the year. In late May, Park was named presidential candidate of the Democratic Republican Party.

Election of Park Chung Hee

The Third Republic. The election for the president of the Third Republic took place on October 15, 1963. Park, by a narrow margin, defeated the opposition candidate, Yun Po-sun, former president of the Second Republic, who had remained in office (more or less as a figurehead) at the request of the junta to provide constitutional continuity for the military regime. When political activity was permitted to resume, Yun took the initiative in mustering opposition groups and became the presidential candidate of the Civil Rule Party. In May 1967 Park was re-elected to his second term of office, and the Democratic Republican Party won a large majority of the seats in the National Assembly. Members of the opposition New Democratic Party, whose candidate, Yun, had been defeated for the second time, claimed election fraud and refused to take their seats in the National Assembly until some months later.

During his second term, President Park was confronted with the constitutional provision that limited the president to two consecutive four-year terms. After considerable political turmoil and demonstrations by the opposition and the students, Democratic Republican Party members in the legislature passed an amendment that provided for the incumbent president to become eligible for three consecutive four-year terms. The amendment was approved by a national referendum in October 1969. In the presidential elections held on April 27, 1971, Park defeated his opponent, Kim Dae-jung of the New Democratic Party; in elections for the eighth National Assembly, however, the New Democrats made impressive gains, securing 89 seats as compared with the 113 seats of the ruling Democratic Republicans.

The Yushin system. In December 1971, shortly after his re-election to a third term of office, Park declared a state of national emergency on the basis of the "dangerous uncertainties of the international situation," and ten months later, in October 1972, he suspended the constitution and dissolved the legislature. A new constitution, which extended the power and permitted the re-election of the president for an unlimited number of six-year terms, was promulgated in December.

National Conference for Unification

The institutional framework of the Yushin (Revitalization–Reform) system departed radically from the Third Republic. The National Conference for Unification was created as "a national organization based on the collective will of the people as a whole to pursue peaceful unification of the fatherland." The conference is a body of not less than 2,000 nor more than 5,000 members who are directly elected by the voters for a six-year term. The president of the republic is the chairman of the conference, and the conference is charged with the election of the president. Other functions of the conference include approving a list of one-third of the members of the National Assembly, who are appointed by the president for *a* three-year term (other members of the legislature are elected directly by the voters for a six-year term) and approving constitutional amendments proposed by the National Assembly. Under this arrangement, Park was elected without opposition on December 23, 1972, and was re-elected for another term in December 1978.

Under the Yushin system, the Park government continued to achieve a high rate of economic growth. Despite the balance-of-payments difficulties of 1974 and 1975, a consequence of the petroleum crisis of 1973 and a much less favourable international environment than expected, growth in gross national product (GNP) averaged 11.2 percent a year during the third Five-Year Economic Plan (1972–76). Although the net growth in world trade in manufactured goods was sluggish during 1973–76, Korea's export volume more than doubled. At the same time, Koreans experienced great success in obtaining construction contracts in the Middle East. The government projected that the GNP would increase by 9.2 percent a year during the fourth Five-Year Economic Plan (1977–81). Whether South Korea could sustain such a growth rate was contingent on the growth of the world economy; on high rates of growth of exports through export diversification and structural changes in the Korean economy; and on continued government policies attaching high priority to a wider distribution of growth benefits to ensure a stable social order.

The Assassination of President Park. In October 1979, Korea faced a major national crisis that culminated in the death of President Park Chung Hee. According to the government, he was shot by Kim Jae-kyu, director of the Korean CIA on the evening of October 26, 1979, at the CIA dining lodge in Seoul. Five others, all presidential security force members, were also fatally shot. It was the first time in the history of the republic that its president had died in such an abrupt and tragic manner. After the death of the president, Prime Minister Choi Kyu-hah became acting president under Article 48 of the Yushin constitution. Emergency martial law was immediately proclaimed throughout the country, except for Cheju Island, and the army chief of staff, Gen. Chung Seung-hwa, was appointed martial law commander.

The death of President Park left the nation with a number of questions concerning its future, the most serious of all being the matter of effecting a peaceful transition of government. The Yushin consitution states that the National Conference of Unification must elect a successor within three months to fill the remainder of the presidential six-year term, which had begun in December 1978. The opposition, however, had demanded that the Park regime return to a liberal parliamentary democracy system and thereby had closed all avenues for the opposition to compete in the election for the presidency. As Korea moved into the 1980s, basic questions were how the government and opposition would resolve their conflicting positions on succession and whether Korea would develop a more openly democratic political system.

Rise of Kim 11-sung

North Korea. In 1945 Kim 11-sung, the supreme commander of the Korean People's Army, arrived in P'yŏngyang, the capital of North Korea, with the Soviet contingents and was pushed by the Soviet authorities to the forefront of North Korean politics. In 1948, when the Democratic People's Republic of Korea was established, he became the first premier of the North Korean Communist regime and in 1949 he became chairman of the Korean Workers Party.

After 1956, as the Sino-Soviet conflict was intensified, Kim was compelled to shift his positions vis-a-vis the two great powers, Moscow and Peking, no less than three times: from pro-Soviet to neutral, from neutral to pro-Chinese, and from pro-Chinese to independent. The division between the two great powers was also reflected in conflicts within the North Korean leadership. The pro-Chinese group, known as the Yenan faction, was purged by Kim during 1956–58. At about the same time, ha eliminated a pro-Soviet faction from the Central Committee of the Korean Workers Party.

In 1966, after a visit to P'ydngyang by the Soviet prime minister, Aleksey N. Kosygin, Kim announced what became known as the independent party line in North Korea, in which he stressed the principles of "complete equality, sovereignty, mutual respect, and noninterference among the Communist and Workers' Parties." From this statement, the KWP ideologists worked out four principles: "*juche* [autonomy, or identity] in ideology," "independence in politics," "self-sustenance in economy," and "self-defense in national defense."

In the late 1960s the Kim 11-sung regime carried out a program of strengthening the military forces, turning the country into a fortress with a standing army of about 500,000 by the late 1970s and much modem equipment. There was also a strong militia, with a total number estimated at 1,250,000 in the mid-1970s.

North Korea's emphasis on strengthening its military forces proceeded in parallel with its continued stress on the construction of a self-reliant economy. With aid from the Soviet Union, the People's Republic of China, and the east European countries, North Korea implemented a series of economic development plans and made significant gains. When the Soviet Union, under Nikita S. Khrushchev, suspended its aid to North Korea, the Seven-Year Plan (1961–67) of the North Korean regime was seriously affected, as indicated by the extension of the plan for another three years.

North Korea responded favourably to the appeal of the South Korean Red Cross in August 1971 to reunite the families separated by the division of the country in 1945. The Red Cross held several talks in Seoul and P'ybngyang alternately during 1972. While the talks were in progress, on July 4, 1972, the South and North Korean agreement on the principles of Korean reunification was made public by Yi Hu-rak, then the director of the KCIA, who in May 1972 made a secret visit to P'ybngyang, where he met Kim 11-sung and signed the agreement with his North Korean counterpart, Kim Yong-ju, a younger brother of Kim 11-sung. Yi's visit to P'ybngyang was reciprocated by a visit of Park Sung-chul, then vice premier of North Korea, to Seoul on May 29, 1972.

The joint communique of July 4 provided for the establishment of a South–North Korea Coordinating Committee to discuss and implement the means necessary for reuniting the divided families. After holding the last session on March 14, 1975, North Korea unilaterally boycotted the Coordinating Committee and cut off the hot line installed between Seoul and P'ybngyang in 1972. The Red Cross talks continued intermittently until December 1977, but no agreement was reached by the delegates on reuniting the families. The gulf dividing the two parts of Korea is so wide that the representatives of the two governments in the South–North Korea Coordinating Committee failed to achieve even agreement on the agenda to be discussed. South–North Korean relations remain hostile and highly unstable, and the Korean peninsula remains a threat to peace and stability in East Asia. (B.-h.H.)

BIBLIOGRAPHY

Early Korea: PYONG-DO YI, ***Hankuksa Taegwan*** (1948), a compendium of Korean history, revised several times; TAKASHI HATADA, ***A History of Korea*** (Eng. trans. 1969), a socioeconomic study; KI-BAIK LEE, ***Hankuksa Sillon*** (1967), contains a unique periodization based on changes of ruling powers; WOO-KEUN HAN, ***The History of Korea*** (Eng. trans. 1971), an outline with emphasis on the modern period; CHIN-TAE SON, ***Hankuk Minjoksa Kaeron*** (1948), a social history covering the period up to the fall of the Silla dynasty; CHAE-WON KIM and PYONG-DO YI, ***Hankuksa Kodaepyon*** (1959), a history of the ancient period; PYONG-DO YI, ***Hankuksa Chungsepyon*** (1961). a history of medieval Korea; WON-YONG KIM, ***Hankuk Kogohak Kaeron*** (1966), valuable as an introductory work on Korean archaeology; and KI-JUN CHO, ***Hankuk Kyongjesa*** (1962), an economic history of Korea.

Yi dynasty and Japanese rule: SANG-BAEK YI, and SUN-GUN YI, ***Hankuksa,*** 4 vol. (1961–65), a history of the Yi dynasty, with primary focus on political changes; KWAN-U CHON, "Hankuk Silhak Sasangsa," in ***Hankuk Munhwasa Daege,*** vol. 6 (1970), deals with the so-called practical thought that was highly developed in 18th-century Korea and with the nature of the several schools of thought and their developments; SANG-KI KIM, ***Tonghak Kwa Tonghaknan*** (1947), deals with Tonghak, or Eastern learning, developed in the early 1860s in reaction to Christian influence; HONG-YOL YU, ***Hankuk Chonju-Kyohoesa*** (1962), a history of the Catholic Church in Korea; L. GEORGE PAIK, ***The History of Protestant Missions in Korea, 1832–1910*** (1929, reprinted 1970); HOMER B. HULBERT, ***The Passing of Korea*** (1906, reprinted 1969), gives a good picture of how Korea was deprived of its sovereignty toward the end of the 19th century amid the conflict of big powers in Korea; ANDREW J. GRAD, ***Modern Korea*** (1944; reprinted 1979), a critique of Japanese imperial rule in Korea.

Korea since 1945: CARL BERGER, ***The Korea Knot: A Military-Political History*** (1964), a concise but excellent study of the military and political history of Korea since the end of World War II; SOON SUNG CHO, ***Korea in World Politics, 1940–1950: An Evaluation of American Responsibility*** (1967), an analysis of events in Korea during the turbulent decade from World War II to the start of the Korean War—a significant contribution to the understanding of the twist and turn of events in post-liberation Korean politics; GREGORY HENDERSON, ***Korea: The Politics of the Vortex*** (1968), contains rich historical data on both traditional and contemporary Korean society and an interesting comparison of China, Korea, and Japan, written by a former U.S. foreign service officer who spent a lengthy period in Korea; CHONG-SIK LEE, ***The Politics of Korean Nationalism*** (1963), a comprehensive study of Korean nationalism and the nationalist movement during the Japanese colonial rule that also treats the historical background of the Sino-Japanese War, the repercussion of the war on the Korean government, and the process by which the country lost its independence; GLENN D. PAIGE, ***The Korean Decision, June 24–30, 1950*** (1968), a detailed account of the United States decision to intervene in Korea in 1950, a significant contribution to the literature in foreign policy-making theories; and ***The Korean People's Democratic Republic*** (1966), an excellent analysis of the history of the North Korean Communist system, describing changes in economic, political, and ideological aspects; DAVID REES, ***Korea: The Limited War*** (1964), one of the best books dealing with the outbreak of the war; HAHN-BEEN LEE, ***Korea: Time, Change and Administration*** (1968), an imaginative and skillful application of time perception to administrative behaviour under conditions of rapid social change in Korea throughout the postliberation period; SUNG-JOO HAN, ***The Failure of Democracy in South Korea*** (1974), a study of the causes of the collapse of Chang Myŏn's liberal democratic government in May 1961.

(K.-r.L./K.-b.L./B.-h.H.)

Korea, North

The Democratic People's Republic of Korea (Chosbn Minjujuŭi In'min Konghwaguk) occupies the northern section of the Korean peninsula, which juts out from the Asian mainland between the Sea of Japan (East Sea) and the Yellow Sea. The country is bordered by China and the Soviet Union to the north and by the Republic of Korea (South Korea) to the south.

North Korea has an area of 46,800 square miles (121,200 square kilometres), occupying about 55 percent of the peninsula. In the late 1970s, it had an estimated population of about 17,072,000. The national capital, P'ybngyang, is a major industrial and transport centre near the west coast,

North Korea was created in 1948 as a result of the post-World War II military occupation by the United States in the south of the peninsula and by the Soviet Union in the north. In close alliance with China and the Soviet Union, the North Korean government has endeavoured to transform the prewar agricultural country into a self-sufficient industrial nation. Economic programs have been successful, but the industrialization has been at the expense of agriculture and of personal freedom. The government, run by the Korean Workers Party, has attempted to achieve strict production quotas with the use of forced labour, socialization of the entire economy, and ideological indoctrination.

Foreign relations are oriented toward the Communist countries and those non-Communist nations such as Japan that are active in the Afro-Asian solidarity movement. Close economic and political ties have been maintained with the Soviet Union and China, while a hostile attitude has been sustained toward the United States and, generally, South Korea. (For an associated physical feature, see YALU RIVER. For a detailed discussion of the capital city, see P'YONGYANG. For history, see KOREA, HISTORY OF.)

The Kaema Plateau and **Nangnim-sanmaek**

The* landscape.** ***Relief, drainage, and soils. Mountains and valleys characterize most of the country. The Kaema Plateau in the northeast has an average elevation of 3,300 feet above sea level and forms the topographic roof of the entire Korean peninsula. Paektu-san (9,003 feet [2,744 metres]), the highest mountain in North Korea, rises at the northern edge of this plateau; it is an extinct volcano topped by a large crater lake. The Nangnim-sanmaek (Nangnim Mountain Range) runs from north to south through the middle of the country, forming a divide between the eastern and western slopes of the peninsula. The Kangnam, Myohyang, Ŏnjin, and Mybrak mountains, with their roots in the Nangnim-sanmaek, extend parallel to each other toward the southwest. Large river valley plains have developed between the western mountains; they merge along the narrow, irregular coastal plain on the west coast. The Hamgydng mountains, extending from the Nangnim-sanmaek to the northeast, form a steep slope between the Kaema Plateau and the Sea of Japan. The T'aebaek-sanmaek of South Korea extends into North Korea; one peak, Kŭmgang-san (5,373 feet [1,638 metres]), is famous for its scenic beauty.

The longest river of North Korea is the Yalu, known as the Amnok-kang. It rises at Paektu-san (Paektu Mountain) and flows southwest for 501 miles to its mouth on Korea Bay. The Tumen River also begins at Paektu-san but runs northeast for 324 miles to the Sea of Japan. There are no large streams along the east coast except for the Tumen River, and all the significant rivers, such as the Yalu, Ch'dng-ch'dn, Taedong, Chaerydng, and the Yesdng, drain to the Yellow Sea. The relatively large valley plains of the western rivers are major agricultural regions.

More than 60 percent of the soils are locally derived from the weathering of granitic rocks or various kinds of schists (crystalline rocks). The soils are generally brownish in colour, abundant in sandy materials, and low in fertility. Well-developed reddish-brown soils that are derived from limestone are found in North Hwanghae Province (Hwanghae-pukto) and the southern part of South P'ybngan Province (P'ybngan-namdo). The Kaema Plateau shows development of podzolic soils (ash-gray forest soil) as a result of its cold climate and coniferous forest cover. Although most of the soils are infertile and lacking in organic content, the valley plains have relatively rich alluvial soils.

Climate. North Korea has a generally cool continental climate. The winter season from December to March is long and cold; mean temperatures in January range between 21" F (—6" C) in the south and —8" F (—22" C) in the northern interior. The summer, from June to September, is hot, with mean July temperatures above 68° F (20" C) in most places. Accordingly, the annual range of temperatures is large—about 54° F (30" C) at P'ydngyang and about 77° F (43" C) at Chung-gangjin, where the lowest temperature in the Korean peninsula, —46.5" F (—43.6" C), has been recorded. Because of ocean currents and the mountain ranges bordering the narrow coastal lowlands, winter temperatures on the east coast are about 4° or 5° F (2° or 3" C) higher than those of the west coast.

Most of the country receives around 40 inches of precipitation annually. The northern inland plateau, however, receives 24 inches and the lower reaches of the Taedong-gang (Taedong River) Valley 32 inches. The upper Ch'ŏng-ch'ŏn-gang area averages between 48 and 52 inches yearly. Approximately 70 percent of the annual precipitation falls in the four months from June to September; this heavy concentration of rainfall is related to the humid summer monsoon from the Pacific. Less than 5 percent of the total precipitation occurs in winter. There are about 200 frost-free days along the coast and fewer than 150 inland.

Vegetation and animal life. Vegetation on the Kaema Plateau, especially around Paektu-san, is composed of coniferous trees, such as the Siberian fir, spruce, pine, and Korean cedar. The western lowlands were originally covered by temperate mixed forests with many types of plants, but continuous deforestation has resulted in only remote patches of the original forests. Most of the lowlands are now cultivated, except for some of the hills that are covered with small pine groves mixed with oaks, chestnuts, and elders. Along streams that are subject to flooding, or where the ground is too stony for cultivation, reeds, sedges, wild mulberry trees, and Italian poplars are found. Common river fish include carps and eels.

Because of deforestation, the population of deer, mountain antelopes, goats, tigers, leopards, and panthers has greatly decreased and is restricted to the remote forests. In the plains, however, it is still possible to see wild pigeons, herons, cranes (which nest near human habitations), and many migratory water fowl, which alight in the rice fields.

Historic provincial divisions

Traditional regions. Close examination reveals numerous distinct regions, each with a different natural environment and historical background. Of the eight Korean provinces of the Yi dynasty (1392–1910), North Korea contains the three provinces of P'ybngan, Hwanghae, and Hamgyŏng and the northern parts of Kangwŏn and Kydnggi provinces. Each province was not only a political unit but also had characteristics of a cultural region in terms of dialect, customs, and a way of life. North Korea may also be divided into the two larger traditional regions of Kwanso to the west and Kwanbuk to the east, roughly divided by the Nangnim-sanmaek while Kwanso comprises the provinces of North P'ybngan (P'yŏngan-pukto), South P'ybngan (P'ydngan-namdo), North Hwanghae (Hwanghae-pukto), South Hwanghae (Hwanghae-namdo), and Chagang, while Kwanbuck includes North Hamgybng, South Hamgybng, and Yanggang provinces.

Human settlement. Urbanization has increased rapidly since 1953; by 1967, 48 percent of the population were urban. Most of the rural population inhabits the eastern and western coastal lowlands and river valley plains. The inland areas of Chagang and Yanggang provinces are sparsely settled because of the lack of arable land and the cold climate, which is not suitable for rice culture. Villages in the lowlands and valley plains are usually clustered together on the southern foot of hills, which offer protection against the cold northwestern winter wind. Scattered fire fields are tilled by a small number of shifting cultivators in the Kaema Plateau, especially in Yanggang Province. The Upper Yalu and Tumen river valleys contain settlements associated with lumbering, and fishing villages are numerous along the coast, especially on the east.

Urban settlement

Cities that developed during the Japanese occupation (1910–45) were largely associated with the exploitation of natural resources, industry, and transportation. P'yŏngyang is the hub of the national railway system, Namp'o is a western port, Songnim contains an iron-ore refinery, and Sinŭiju contains factories for the production of electrical equipment, chemicals, textiles, and consumer goods. Mining cities include Aoji, Chaerybng, Iwdn, Kilchu, Musan, and Pukchin. The Communist regime's heavy emphasis on manufacturing resulted in the continuous expansion of the early industrial centres and caused a population flow into the urban areas from the countryside. Most of the cities were destroyed during the Korean War (1950-53) and have since been rebuilt. There are no high-rise buildings, but seven- and eight-story buildings line the main streets of P'ydngyang. Workers are expected to live in apartments rather than individual homes, and housing projects are supported almost solely by the government. Heating systems in the apartments and urban water supplies are inadequate. The streets are strangely empty of pedestrians, as the Koreans have few leisure hours.

People and population. *Population groups.* Because they are a nation with a long history, Koreans believe that they belong to a single racial stock. Physical and cultural characteristics vary only slightly from one region

North Korea, Area and Population

	area		population
	sq mi	sq km	1968 estimate
Provinces (*do* or *to*)			
Chagang-do	6,200	16,050	780,000
Hamgyŏng-namdo*	7,350	19,000	1,315,000
Hamgyŏng-pukto*	6,050	15,700	1,110,000
Hwanghae-namdo	2,900	7,500	1,340,000
Hwanghae-pukto	3,300	8,550	1,060,000
Kangwŏn-do	4,100	10,600	1,030,000
P'yŏngan-namdo	4,700	12,200	2,250,000
P'yŏngan-pukto	4,600	11,900	1,760,000
Yanggang-do	5,400	14,000	435,000
Special cities (*t'ŭkpyŏlsi*)			
Ch'ŏngjin-si†	750	1,900	385,000
Hamhŭng-si†	300	800	530,000
P'yŏngyang-si	700	1,800	1,275,000
Special district (*chigu*)			
Kaesŏng-chigu	450	1,200	289,000
Total North Korea	46,800	121,200	13,559,000

*The prefixes "nam" and "puk" mean south and north, respectively. †According to some sources, these are second-order and not first-order subdivisions. Although their status is not definite, it seems to meet the same criteria as does P'yŏngyang-si; *i.e.,* they are large urban centres.
Source: U.S. Department of State.

to another. Koreans all have typical Mongoloid physical features; their average stature is a little shorter than that of the northern Chinese and slightly taller than that of the Japanese. All Koreans speak the Korean language. which is related to Japanese and contains Chinese loan words. The Korean script. known in North Korea as Chosdn muntcha (and in South Korea as Han'gŭl), is composed of phonetic symbols for the ten vowels and 14 consonants. In North Korea Chosdn muntcha has been used exclusively without Chinese characters in newspapers and other publications since 1945. About 80.000 Chinese lived in the cities before 1945. but their present population figure is unknown.

The way of life and the value system of Koreans are based fundamentally on Confucian thought. To a lesser extent. Buddhism is also important. Roman Catholic and Protestant beliefs were introduced in the 17th and 19th centuries; Sdn-ch'on and P'ydngyang were major centres of Christian activities. World War II brought repression of Christianity. and all foreign missionaries were expelled from the country.

The monotheistic religion of Ch'dndogyo ("Society of the Heavenly Way"; originally known as Tonghak) was founded by the Confucian teacher Ch'oe Che-u in 1860. A combination of Buddhism. Confucianism. and Christianity. Ch'dndogyo played a leading role in the independence movement of 1919. Shamanism—the religious belief in gods. demons. and ancestral spirits responsive to a

MAP INDEX

Political subdivisions

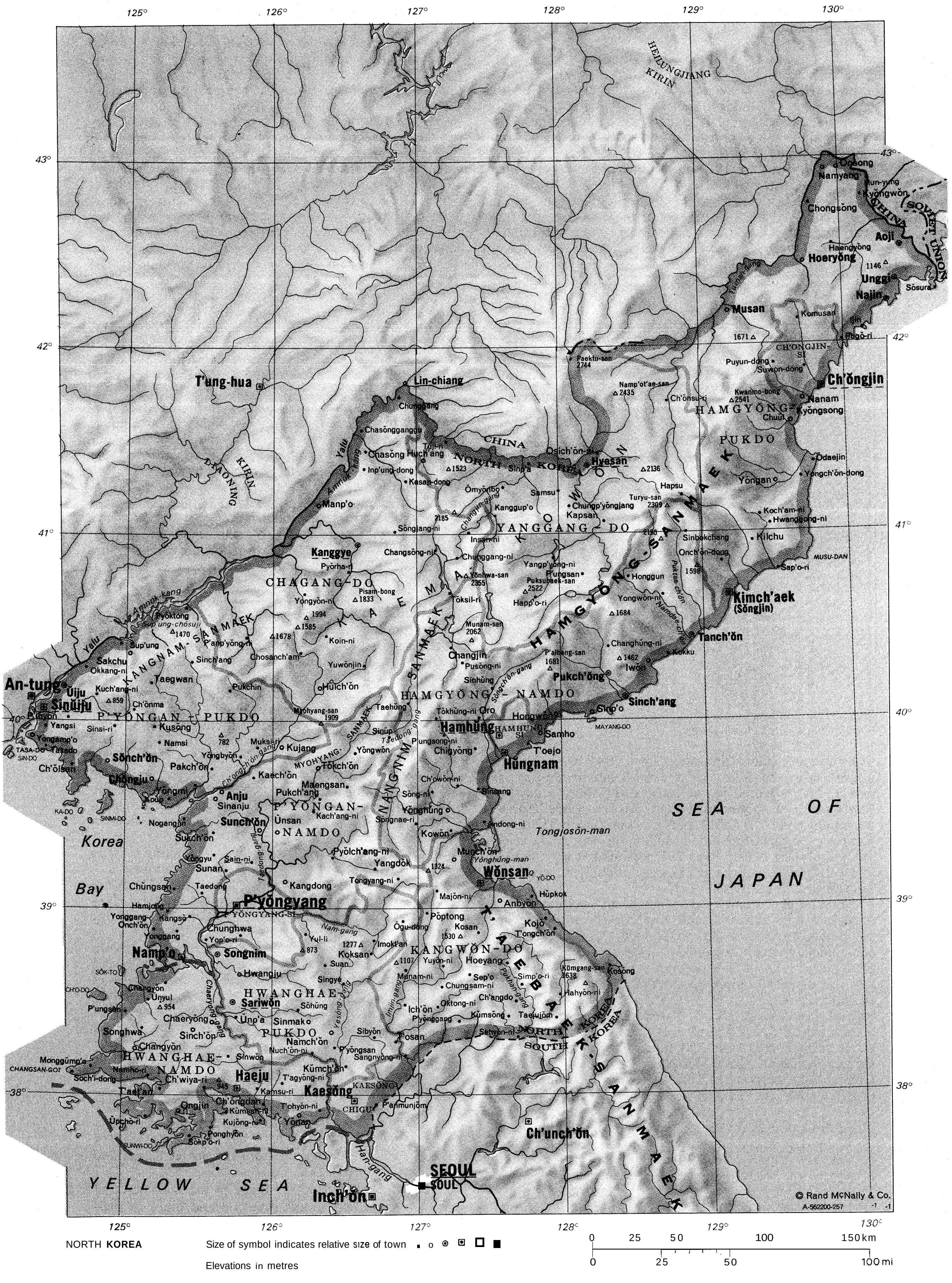
Chagang-do	40·50n	126·30e
Ch'ongjin-si	41·55n	129·40e
Hamgyŏng-namdo	40·00n	127·30e
Hamgyŏng-pukdo	41·45n	129·50e
Hamhung-si	39·55n	127·40e
Hwanghae-namdo	38·15n	125·30e
Hwanghae-pukdo	38·30n	126·25e
Kaesŏng-chigu	38·00n	126·38e
Kangwŏn-do	38·45n	127·35e
P'Yŏngan-namdo	39·20n	126·00e
P'yŏngan-pukdo	40·10n	125·20e
P'yŏngyang-si	39·05n	125·50e
Yanggang-do	41·15n	128·00e

Cities and towns

Anbyŏn	39·03n	127·32e
Andong-ni	39·28n	127·27e
Anju	39·36n	125·40e
Aoji	42.31 n	130·23e
Chaerybng	38·24n	125·36e
Ch'angdo	38·30n	127·40e
Changhŭng-ni	40·24n	128·19e
Changjin	40·23n	127·15e
Changsbng-ni	40·58n	127·04e
Changyŏn	38·37n	125·16e
Chasŏng	41·27n	126·37e
Chigyŏng	39·51n	127·26e
Ch'ŏlsan	39·46n	124·40e
Ch'ŏngdan	37·58n	125·56e
Ch'ŏngjin	41·47n	129·50e
Ch'ŏngju	39.41n	125·13e
Chongsŏng	42·46n	129·48e
Ch'ŏnma	40·03n	125·01e
Ch'bnsu-ri	41·43n	128·45e
Ch'osan	40·50n	125·47e
Chosanch'am	40·22n	126·11e
Ch'owbn-ni	39·40n	127·17e
Chunggang	41·45n	126·52e
Chunggang-ni	40·52n	127·20e
Chunghwa	38·52n	125·47e
Chungsam-ni	38·34n	127·09e
Chŭngsan	39·06n	125·22e
Chuŭl	41·33n	129·34e
Ch'wiya-ri	38·03n	125·32e
Haeju	38·02n	125·42e
Haengyŏng	42·32n	129·56e
Hahyŏn-ni	38·33n	127·57e
Hamhŭng	39·54n	127·32e
Harnjong	38·59n	125·20e
Happ'o-ri	40·40n	127·49e
Hapsu	41·13n	128·51e
Hoerybng	42·27n	129·44e
Hoeyang	38·43n	127·36e
Honggun	40·46n	128·27e
Hongwŏn	40·02n	127·57e
Huch'ang	41·25n	127·03e
Hŭich'on	40·10n	126·17e
Hŭngnam	39·50n	127·38e
Hun-yung	42·53n	130·12e
Hŭpkok	39·04n	127·44e
Hwanggong-ni	40·03n	129·27e
Hwangju	38·42n	125·46e
Hyesan	41·23n	128·12e
Ich'on	38·30n	126·50e
Ijin	42·05n	130·08e
Imokt'an	38·50n	126·41e
Inp'ung-dong	41·25n	126·34e
Iwŏn	40·19n	128·39e
Kach'ang-ni	38·24n	126·11e
Kaech'bn	39·42n	125·53e
Kaesbng	38·00n	126·30e
Kamsu-ri	38·03n	125·54e
Kangdong	39·09n	126·05e
Kanggup'o	41·07n	127·31e
Kanggye	40·58n	126·34e
Kangsŏ	38·58n	125·26e
Kapsan	41·04n	128·19e
Kasan-dong	41·18n	126·55e
Kilchu	40·58n	129·20e
Kimch'aek	40·41n	129·12e
Koch'am-ni	41·06n	129·23e
Koin-ni	40·28n	126·22e
Kojŏ	38·57n	127·51e
Kokku	40·22n	128·44e
Koksan	38·46n	126·40e
Komusan	42·08n	129·41e
Kosan	38·52n	127·24e
Kosŏng	38·40n	129·19e
Koŭp	39·36n	124·28e
Kowbn	39·26n	127·14e
Kuch'ang-ni	40·09n	124·46e
Kujang	39·52n	126·01e
Kujŏng-ni	37·53n	125·54e
Kŭmch'ŏn	38·10n	126·29e
Kŭmsan-ni	37·55n	125·41e
Kŭmsŏng	38·25n	127·36e
Kusŏng	39·59n	125·15e
Kyŏngsŏng	41·35n	129·36e
Kyŏngwŏn	42·48n	130·09e
Maengsan	39·40n	126·30e
Majŏn-ni	39·06n	127·07e
Manp'o	41·09n	126·17e
Monggŭmp'o	38·09n	124·47e
Muksi-ri	39·52n	125·54e
Munam-ni	38·41n	126·54e
Munch'ŏn	39·16n	127·15e
Musan	42·14n	129·13e
Najin	42·15n	130·18e
Narnch'bn	38·19n	126·23e
Namho-ri	38·07n	125·10e
Namp'o	38·45n	125·23e
Namsi	39·54n	124·36e
Namyang	42·57n	129·53e
Nanam	41·43n	129·41e
Nogangjin	39·30n	125·23e
Nuch'ŏn-ni	38·14n	126·16e
Ŏdaejin	41·34n	129·40e
Ogu-dong	38·57n	126·56e
Okkang-ni	40·18n	124·42e
Oktong-ni	38·27n	127·07e
Ŏmyŏnbo	41·16n	127·36e
Onch'ŏn-dong	40·51n	129·07e
Ongjin	37·57n	125·21e
Onsŏng	42·57n	129·52e
Oro	40·02n	127·26e
Osich'ŏn-ni	41·25n	128·16e
Pakch'bn	39·45n	125·35e
P'anmunjŏm	37·57n	126·40e
P'anp'yŏng-ni	40·28n	125·49e
P'ihyŏn	40·01n	124·37e
Ponghyŏn	37·49n	125·36e
Pŏptong	38·59n	127·05e
Pugŏ-ri	42·01n	129·59e
Pukch'ang	39·36n	126·17e
Pukchin	40·10n	125·43e
Pukch'bng	40·15n	128·20e
P'ungsan	38·28n	125·01e
P'ungsong-ni	39·56n	127·11e
Pusŏng-ni	40·19n	127·19e
Puyun-dong	41·55n	129·30e
Pyŏktong	40·35n	125·20e
Pyŏlch'ang-ni	39·17n	126·26e
P'yŏnggang	38·26n	127·16e
P'yŏngsan	38·19n	126·23e
P'yŏngyang	39·01n	125·45e
Pybrha-ri	40·48n	126·32e
Sain-ni	39·15n	125·51e
Sakchu	40·23n	125·01e
Samho	39·56n	127·53e
Samsu	41·19n	127·59e
Sangnyŏng-ni	38·14n	126·54e
Sap'o-ri	40·49n	129·31e
Sariwŏn	38·31n	125·44e
Sehyŏn-ni	38·20n	127·41e
Sep'o	38·39n	127·22e
Sibybn	38·19n	126·41e
Simp'o-ri	38·36n	127·41e
Sinanju	39·36n	125·36e
Sinbokchanu	41·01n	128·54e
Sinch'ang	40·07n	128·28e
Sinch'ang	40·19n	125·27e
Sinch'bn	38·22n	125·28e
Singye	38·36n	126·30e
Sinhŭng	40·11n	127·34e
Sinmak	38·25n	126·14e
Sinp'a	41·24n	127·46e
Sinp'o	40·03n	128·12e
Sinsang	39·38n	127·25e
Sinsi-ri	39·59n	124·58e
Sinŭiju	40·05n	124·24e
Sinŭp	39·54n	126·47e
Sinwŏn	38·14n	125·47e
Soch'i-dong	38·06n	124·45e
Sŏhŭng	38·27n	126·10e
Sŏkp'o-ri	37·46n	125·27e
Sŏnch'ŏn	39·48n	124·55e
Songhwa	38·21n	125·08e
Sŏngjang-ni	41·02n	126·50e
Sonjin, see Kimch'aek		
Sŏngnae-ri	39·28n	126·59e
Sŏng-ni	39·38n	127·06e
Songnim	38·44n	125·38e
Sŏsura	42·16n	130·37e
Suan	38·42n	126·22e
Sukch'ŏn	39·24n	125·38e
Sunan	39·13n	125·41e
Sunch'ŏn	39·26n	125·54e
Sup'ung	40·27n	124·57e
Suwon-dong	41·54n	129·43e
Taedong	39·05n	125·32e
Taegwan	40·13n	125·12e
Taehŭng	40·06n	126·56e
Taejujŏm	38·24n	127·58e
T'aet'an	38·04n	125·18e
T'agyŏng-ni	38·04n	126·03e
Tanch'ŏn	40·27n	128·54e
Tasado	39·50n	124·27e
T'oejo-dong	39·54n	127·46e
Tohyŏn-ni	39·53n	124·52e
Tŏkch'ŏn	39·46n	126·19e
Tŏkhŭng-ni	40·02n	127·08e
Tŏksil-ri	40·44n	127·12e
T'ongchŏn	38·54n	127·54e
Tongyang-ni	39·08n	126·52e
T'osan	38·18n	126·43e
Tuji-ri	41·31n	127·12e
Ŭiju	40·12n	124·32e
Ŭnch'ŏn	38·34n	125·26e
Unggi	42·20n	130·24e
Unp'a	38·26n	125·45e
Ŭnsan	39·25n	126·01e
Ŭnyul	38·30n	125·10e
Ŭpchŏ-ri	37·53n	125·09e
Wŏnsan	39·09n	127·25e
Yangsi	39·59n	124·28e
Yangp'yŏng-ni	40·53n	127·58e
Yangdbk	39·13n	126·39e
Yŏnan	37·55n	126·10e
Yongamp'o	39·55n	124·24e
Ybngan	41·15n	129·30e
Yŏngbyŏn	39·49n	125·48e
Yongch'ŏn-dong	41·18n	129·40e
Yonggang	38·52n	125·21e
Yonggang-Onch'ŏn	38·53n	125·14e
Yŏnghŭng	39·32n	127·13e
Yŏngmi	39·39n	125·31e
Yŏngwŏn	39·49n	126·32e
Yongwon-ni	40·41n	128·42e
Yongyŏn-ni	40·44n	126·09e
Ybngyu	39·18n	125·34e
Yop'o-ri	38·50n	125·33e
Yul-li	38·52n	126·13e
Yuwbnjin	40·18n	126·37e
Yuyŏn-ni	38·42n	127·10e

Physical features and points of interest

Amnok-kang, see Yaiu		
Chaeryong-gang. river	38·39n	125·28e
Changjin-gang, river	41·24n	127·45e
Changsan-got, cape	38·08n	124·39e
Ch'o-do, island	38·32n	124·50e
Ch'ŏngch'ŏn-gang. river	39·35n	125·28e
Hamgyŏng-sanrnaek	41·50n	128·30e
Imjin-gang, *river*	37·47n	126·40e
Japan. Sea of	39·30n	129·30e
Ka-do, *island*	39·33n	124·40e
Kaema-kowŏn, plateau	40·40n	127·00e
Kangnam-sanmaek. mountains	40·30n	125·30e
Korea Bay	39·15n	125·00e
Kŭmgang-san, mountain	38·35n	128·10e
Kwanmo-bong, mountain	41·42n	129·13e
Mayang-do, *island*	40·00n	128·12e
Munam-san, mountain	40·22n	127·08e
Musu-dan, point	40·50n	129·43e
Myohyang-san, mountain	40·02n	126·17e
Myohyang-sanrnaek. mountains	40·30n	127·00e
Namdae-ch'ŏn, river	40·46n	129·22e
Nam-gang, river	39·03n	125·52e
Namp'ot'ae-san, mountain	41·44n	128·24e
Nangnim-sanmaek, mountains	40·30n	127·00e
Paektu-san, mountain	42·00n	128·03e
P'albong-san, mountain	40·16n	127·57e
Pisam-bong, mountain	40·41n	126·34e
Pukhan-gang, river	38·20n	127·48e
Puksubaek-san, mountain	40·42n	127·44e
Puktae-ch'ŏn, river	40·28n	129·00e
Sin-do, island	39·48n	124·14e
Sinmi-do, *island*	39·33n	124·53e
Sŏjosŏn-man, *bay*	39·20n	124·50e
Sŏk-to, *island*	38·40n	125·00e
Sŏng-ch'ŏn-gang, *river*	39·48n	127·35e
Sunwi-do, island	37·44n	125·15e
Sup'ung-chŏsuji, reservoir	40·30n	125·05e
T'aebaek-sanmaek. mountains	38·35n	127·45e
Taedong-gang, *river*	38·42n	125·15e
Tasa-do, island	39·47n	124·25e
Tongjosŏn-man, bay	39·30n	128·00e
Tuman-gang, river	42·18n	130·41e
Turyu-san, mountain	41·10n	128·47e
Yaiu (Amnok-kang), river	39·55n	124·22e
Yellow Sea	37·30n	125·00e
Yesŏng-gang, river	37·53n	126·24e
Yŏ-do, island	39·14n	127·38e
Yŏnghŭng-man, bay	39·15n	127·30e
Yŏnhwa-san, mountain	40·46n	127·23e


125°
126°
127°
128°
129°
130°
43°
42°
41°
40°
39°
38°
HEILUNGJIANG
KIRIN
LIAONING
CHINA
SOVIET UNION
NORTH KOREA
SOUTH KOREA
T'ung-hua
Lin-chiang
Chunggang
Chasŏnggangg
Chasŏng
Huch'ang
Inp'ung-dong
Kasan-dong
Manp'o
Yalu
Amnok-kang
Paektu-san 2744
Namp'ot'ae-san 2435
Hyesan
Sinp'a
Samsu
Kanggup'o
Ŏmyŏnbo
Chungp'yŏngjang
Kapsan
Hapsu
Turyu-san 2309
Musan
Hoeryŏng
Onsong
Namyang
Kyŏngwŏn
Chongsŏng
Aoji
Unggi
Najin
Sŏsura
Komusan
Pugŏ-ri
Ch'ŏngjin
CH'ONGJIN-SI
Puyun-dong
Suwon-dong
Kwanmo-bong 2541
Nanam
Kyŏngsong
Chuŭl
HAMGYŎNG-PUKDO
Ŏdaejin
Yŏngch'ŏn-dong
Yŏngan
Koch'am-ni
Hwanggong-ni
Kilchu
MUSU-DAN
Sap'o-ri
Kimch'aek
(Sŏngjin)
Tanch'ŏn
Kokku
Iwŏn
Pukch'ŏng
Sinch'ang
Sinp'o
YANGGANG-DO
KAEMA KOWŎN
HAMGYŎNG-SANMAEK
Sinbokchang
Onch'ŏn-dong
Honggun
Yongwŏn-ni
Changhŭng-ni
P'ungsan
Puksubaek-san 2522
Happ'o-ri
Yangp'yong-ni
Insan-ni
Chunggang-ni
Yŏnhwa-san 2355
Tŏksil-ni
Munam-san 2062
Changjin
Pusŏng-ni
Sinhŭng
P'albong-san 1681
HAMGYŎNG-NAMDO
Hongwŏn
Samho
Hamhŭng
HAMHUNG-SI
T'oejo
Hŭngnam
MAYANG-DO
Oro
Tŏkhŭng-ni
Chigyŏng
Pungsong-ni
Sinsang
Ch'owŏn-ni
Yŏnghŭng
Kowŏn
Sŏngnae-ri
Chŏngp'yŏng
Kŭmya
Tongjosŏn-man
Yŏnghŭng-man
Munch'ŏn
Wŏnsan
YŎ-DO
Hŭpkok
Anbyŏn
Kojo
T'ongch'ŏn
Kŭmgang-san 1638
Kosŏng
Hahyŏn-ni
SEA OF JAPAN
Kanggye
CHAGANG-DO
Pyŏrha-ri
Changsŏng-ni
Sŏngjang-ni
Pisam-bong 1833
Yŏngyŏn-ni
Koin-ni
Yuwŏnjin
Hŭich'ŏn
Chosanch'am
KANGNAM-SANMAEK
Sup'ung-chŏsuji
P'yoktong
Sup'ung
Sakchu
Okkang-ni
Taegwan
Pukchin
An-tung
Uiju
Sinŭiju
Kuch'ang-ni
Ch'ŏnma
P'YŎNGAN-PUKDO
Yangsi
Sinsi-ri
Kusŏng
Namsi
Yŏngbyŏn
Myohyang-san 1909
Taehŭng
Kujang
Hyangsan
Yŏngwŏn
MYOHYANG-SANMAEK
Tŏkch'ŏn
Maengsan
Kaech'ŏn
Anju
Sinanju
Sunch'ŏn
P'YŎNGAN-NAMDO
Ŭnsan
Kach'ang-ni
Pyŏlch'ang-ni
Yangdŏk
Sŏnch'ŏn
Pakch'ŏn
Chŏngju
Kwaksan
Yŏngmi
TASA-DO
Tasado
SIN-DO
Ch'ŏlsan
KA-DO
SINMI-DO
Nogang-ni
Sukch'ŏn
Korea Bay
Yongyu
Sain-ni
Sunan
Chŭngsan
Taedong
P'yŏngyang
YŎNGYANG-SI
Kangdong
Tongyang-ni
Hamjong
Yŏnggang-Onch'ŏn
Kangsŏ
Chunghwa
Yop'o-ri
Namp'o
Songnim
SŎK-TO
CH'O-DO
Hwangju
Changyŏn
Ŭnyul
P'ungsan
Sariwŏn
HWANGHAE-PUKDO
Chaeryŏng
Ŭnp'a
Sinmak
Sohung
Sinch'ŏn
Songhwa
Changyŏn
HWANGHAE-NAMDO
Monggŭmp'o
CHANGSAN-GOT
Soch'i-dong
Ch'wiya-ri
Haeju
T'aet'an
Ongjin
Ch'ŏngdan
Kŭmch'ŏn
Kaesŏng
KAESŎNG
Panmunjŏm
Yŏnan
Ŭpch'ŏ-ri
Kujŏng-ni
Ponghyŏn
Sokp'o-ri
SUNWI-DO
CHIGU
Majŏn-ni
P'ŏptong
Kosan
Ŏgu-dong
Yul-li
Imokt'an
Koksan
Suan
KANGWŎN-DO
Yuyŏn-ni
Hoeyang
Singye
Munam-ni
Sep'o
Simp'o-ri
Chungsam-ni
Ch'ŏrwŏn
Kŭmsŏng
Ich'ŏn
P'yŏnggang
Sehyŏn-ni
Taejujŏn
Ch'ŏnam
Namch'ŏn
P'yŏngsan
Sangnyŏng-ni
Kŭmch'ŏn
T'agyŏng-ni
Kamsu-ri
T'ohyŏn-ni
NANGNIM-SANMAEK
T'AEBAEK-SANMAEK
Ch'unch'ŏn
SEOUL
SŎUL
Inch'ŏn
Han-gang
YELLOW SEA
© Rand McNally & Co.
A-562200-257
Size of symbol indicates relative size of town
Elevations in metres
0 25 50 100 150 km
0 25 50 100 mi

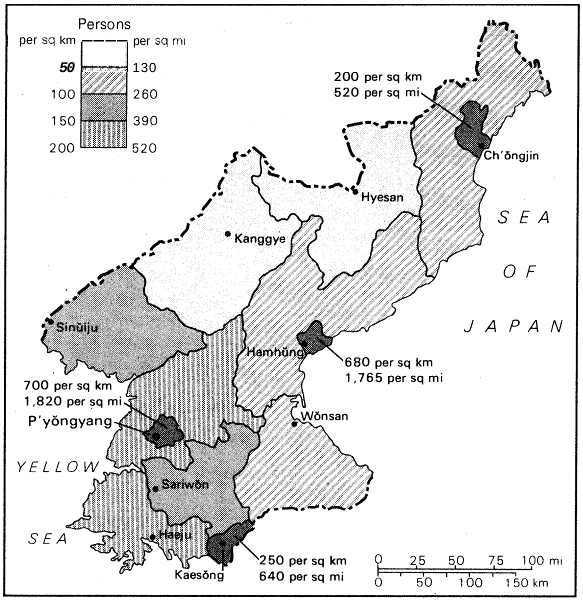
NORTH KOREA

priest, or shaman--existed in Korea before the introduction of Buddhism and Confucianism and still prevails in rural villages.

Religious repression

The Communist regime has constitutionally confirmed freedom of religion but does not practice it, allegedly for fear that it will weaken the Communist Party and the government. Ch'ŏndogyo, however, is used as a means of propaganda. After the Korean War, churches and Buddhist temples were confiscated and looted, and many were converted to other purposes.

Demography. In 1970 the United Nations estimated that during the period 1965–70 the North Korean birth rate was 39 births per 1,000 population, the death rate 11 per 1,000, and the rate of natural increase 28 per 1,000. The total population was estimated at about 14,280,000 in 1971. The movement of population northward during the 1930s and 1940s was reversed after the Korean War, and North Korea lost about 2,000,000 people to its southern neighbour.



Population density of North Korea.

The stress on industrialization since 1945 has promoted migration to the cities, and the farm labour shortage is severe. North Korea is not concerned with overpopulation because an abundant labour force is needed for its unusually high goals of economic achievement. There is a campaign to repatriate Koreans living in Japan. The average population density was 305 persons per square mile in 1971. The coastlines are heavily populated, while the interior only sparsely so.

The national economy. The means of production are socialized, and priorities and emphases in economic development are set by the government. The economy is self-sufficient except for industrial needs such as fuel and machinery. Like other Communist countries, North Korea places special emphasis on capital goods rather than consumer goods.

Natural resources. North Korea contains about 80 to 90 percent of all known mineral deposits on the peninsula. It is estimated that 200 minerals are of economic value, including gold, tungsten, graphite, magnesite (magnesium carbonate), barite (barium sulfate), and molybdenum (a metallic element used in hardening steel).

Major iron and coal deposits

Iron-ore reserves are estimated at about 2,400,000,000 tons; the deposits at Musan, North Hamgydng Province are of low quality, while those of North Hwanghae Province and South P'ybngan Province are of high grade. Rich deposits of anthracite (hard coal) occur along the Taedong-gang, not far from P'yŏngyang, and there are small amounts of lignite (brown coal) at Aoji and Anju. Petroleum was reportedly discovered in the 1950s, but there is no evidence of production.

The northern interior contains large forest reserves of larch, spruce, and pine trees. Most of the coastal slopes have been excessively deforested, however, and reforestation programs stress economic forestry. Hydroelectric-power resources were developed highly during the Japanese regime along the Yalu River and its upper tributaries, such as the Changjin, Pujdn, and Hŏch'ŏn rivers. Power production is based mainly on hydroelectricity, but thermal electricity is becoming important because of the increasing demands of industrialization and the deficiency of hydroelectric power during the dry season.

Sources of national income. Since World War II, North Korea has changed from an agricultural to an industrial nation. In 1946, 59 percent of the gross national product (GNP) was derived from agriculture and 23 percent from industry. As of 1965, agriculture contributed only 22 percent to the GNP compared to industry's 78 percent.

Farm cooperatives

Despite the disproportionately small agricultural contribution due to the labour shortage and low productivity, there has been an increase in cultivated land, irrigation projects, chemical fertilizer supplies, and mechanization through the government program of socialization. By 1958 all farms were incorporated into more than 3,000 cooperatives; each cooperative is comprised of about 300 families on about 1,000 acres. The farm units are controlled by management committees, which issue orders to the work teams, set the type and amount of seed and fertilizer to be used, and establish production quotas. Produce is delivered to the government, which controls distribution through state stores. Each farmer is paid for his labour in money or in kind and is allowed to keep chickens, bees, fruit trees, and a garden.

There are also state and provincial model farms for research and development. The workers are paid in money and are allowed a garden. Livestock husbandry is concentrated on the state farms because the land is little suited to grazing and few feed grains are grown.

The main food crops are grains, such as rice, corn (maize), millet, barley, and wheat. Although production has increased since the 1950s, grain must still be imported. Sweet potatoes, soybeans, and fruit trees are raised extensively. Industrial crops include tobacco, cotton, flax, and rape (an herb grown for its oilseeds).

Forestry has declined since World War II and contributed about 3 percent of total industrial output in 1963. The sea is the main source of protein for North Koreans, and the government has continually expanded commercial fishing since the 1950s. Most fishing activity centres on the east coast and on Tasa-do (Tasa Island) in the mouth of the Yalu River. The annual marine catch includes anchovy, mackerel, pollack, tuna, crustaceans, and seaweed.

Mining is a state enterprise directed by a manager who is appointed by the government. The most emphasis is given to the extraction of coal and iron ore. There are more than 100 small coal mines, an iron mine at Sunhŭng, a nickel mine at Puyun, and a copper mine at Mandŏk.

The three most important industries are concerned with the production of iron and steel, centred at Songnim and Ch'ŏngjin; of industrial and agricultural machinery at Kangsŏn, near P'ydngyang; and of textiles, centred at P'yŏngyang and Hamhŭng. Other industrial products include chemicals, armaments, vehicles, cement, glass, ceramics, and some consumer goods.

Industrial development is related to the country's large supply of electric power. Production has increased steadily, although it has not kept pace with industry, and electricity is sold to China. Most power is provided by hydroelectric facilities, such as those at Sup'ung, Purydng, Kŭm'gang, and Kanggye. There has been expansion of thermal facilities through aid from the Soviet Union and China.

Banking

The North Korean Central Bank is the sole bank of issue. It receives all national revenues and precious metals and provides government agencies with working capital. The Industrial Bank has branches in every farm cooperative; it administers the government insurance system, operates savings accounts for government and individuals, and is the only bank to grant loans. The Foreign Trade Bank handles all foreign transactions and currencies and is supervised by the Central Bank.

More than 80 percent of foreign trade is conducted with the Soviet Union and China. Since the 1960s, trade has been permitted with non-Communist countries, including Japan, France, Australia, West Germany. Hong Kong, The Netherlands, and the United Kingdom. Imports mainly consist of machinery (including machine tools and precision instruments), fuel and related oil, chemical or rubber products; exports are pig iron, magnesia products, iron ore, and nonmetallic minerals.

Management of ***the*** economy. The farm cooperative is not only an economic unit but also the basic unit of technical, ideological, and political control. A Cooperative Farm Management Committee was established in 1961 in each county (*kun*) in order to tighten control over the individual units. Since grain is marketed only through state-operated stores, the government controls production, pricing, and distribution of grain. The industrial sector is organized into state-owned enterprises and production cooperatives. the latter being confined largely to handicrafts, marine processing, and other small-scale operations. In 1963 state enterprises accounted for 91 percent of the gross industrial product, and cooperatives contributed the remainder.

Through the three national economic plans promulgated since 1954, the government has given high priority to manufactures—especially the chemical and heavy industries—at the expense of agriculture. In order to increase the low productivity of labour, the state adopted an independent accounting system and a mass-mobilization measure known as the Ch'bllima ("Flying Horse") campaign.

The nation's railways

Transportation. The 6,500 miles of railroads, the principal means of transportation, in 1963 carried 93 percent of all freight. It is reported that 21 percent of the railroads are electrified. The basic railway pattern runs in a north–south direction roughly parallel to the coasts with branch lines to the river valleys. Because of the high mountains, there is only one east–west railway line between P'ybngyang and Wbnsan. The Kybng–Ui line on the west coast runs from Kaesbng near the South Korean border to Sinŭiju on the Chinese border, connecting the major cities. From this major line a branch from P'yŏngyang southwestward to Namp'o connects centres of machine building and foundries. The Manp'o line, which runs northward from P'ybngyang to Manp'o on the Yalu River, connects the western interior to China's Northeast (formerly Manchuria). The Wŏlla line is the major railway on the east coast; it runs from Wdnsan northward to Najin, and continues to Namyang on the Chinese border. Several branch lines serve the inland areas and mining centres.

Highway transportation is not as important as railroads because few motor vehicles are available. Major roads parallel the rail lines, and there are few east–west roads. Most of the 3,500 miles of roads are not paved.

River transportation plays an important role in transporting agricultural products and minerals. The most important rivers utilized for freight transportation are the Yalu, Taedong, and Chaerybng. Namp'o—the entry port to P'ydngyang—Haeju, and Tasado are the major ports on the west coast, as are Wbnsan, Ch'bngjin, and Najin in the east.

Air services are controlled by the air force. Flights are maintained between the major cities, and international services connect P'ybngyang with Peking and Moscow. Sunan Airport, ten miles north of P'ybngyang, serves as the international airport; domestic airports are located at Hamhŭng, Ch'bngjin, and Wbnsan.

Administration and social conditions. The government. Constitutionally, the 45-member Cabinet, Supreme Court, and Supreme Procurator (an agency that maintains surveillance over all citizens) are responsible to the Supreme People's Assembly, which is the highest organ of state power. The actual source of authority, however. derives from the extraconstitutional political body of the Korean Workers Party headed by Kim 11-sung, who has been the head of the party and the government since 1949. The government is highly centralized and totalitarian in nature and was often officially described as a transmission body of the party.

There are nine provinces (do or to); three special cities (si) of P'ybngyang, Ch'bngjin, and Hamhŭng; and one special region (chigu) of Kaesŏng. These are further subdivided into cities, counties, and ri, the smallest administrative unit. Local people's assemblies elect the members of their people's committees, which execute administration duties and make local economic plans and budgets with the approval of higher authorities.

Control of political activity

There are a number of political parties and social organizations who serve to support the Korean Workers Party. All political activities are sponsored by the party or require its sanction and must closely follow the party line and policies. Elections provide a means whereby assent is registered for the policy and program of the party; they do not allow freedom of expression. There is seldom more than one candidate on the ballot for each constituency, and the electoral system is completely controlled by the party.

The judicial system consists of the courts and the procuracy. They are independent of each other, and actually the procurator's office functions as the fourth branch of the government. The courts consist of the Supreme Court, whose judges are elected for three-year terms by the Supreme People's Assembly, and the provincial and people's courts, whose members are elected by local people's assemblies. Judges are usually party members or are controlled by the party.

The unusually large armed forces total about 401,000 men, about 3 percent of the population. Since 1966 there has been an emphasis on military preparedness, and economic plans have been altered to support military expenditures. In 1971 the military expenses constituted 30 percent of the total budget. The army is the largest force; there is also a 30,000-man air force and an 11,000-man navy. Both men and women are subject to conscription for three to four years service. The paramilitary militia is composed of about 1,250,000 members.

The social sector. Education is directly controlled by the party and serves as a process of indoctrination in Communist ideology and a means to supply skilled workers, technicians, and scientists to meet the government's economic goals. **All** students are required to engage in productive labour along with their studies, which emphasize science and technology. In 1967 education was made compulsory for those between the ages of seven and 16. The system is comprised of four years of primary, five years of middle, and two years of high school. There are also four to six years of university training. Technical schools are usually located near factories.

Health and welfare services

Medical care is free, and there is at least one clinic in each village. There are no recent statistics on the number of medical personnel, but available evidence indicates a significant shortage of doctors and medicine. Medical benefits are provided by social insurance for workers who are temporarily or permanently disabled and women during pregnancy and childbirth. There are also funeral benefits-and old-age pensions. Homes for the aged in each province have been operative under the Ministry of Labour since 1964.

Reconstruction of houses after the Korean War was given high priority, and dwellings are said to have improved considerably. Rural mud-walled, thatched-roofed huts have been replaced by brick buildings with tile or slate roofs. Urban housing is classified into five groups that range from one room and a half-sized kitchen to individual houses with gardens.

The national police and secret agencies under Ministry of Social Security control people's movements and social activities even down to the household level. Because of the high priority for industrialization and defense, the provision of consumer goods and social services is inadequate. It is believed, however, that the material economy and the lot of the peasant have improved since World War II, although it is still necessary to ration such daily goods as food and clothing.

Cultural life and institutions. The compound religious thoughts of Shamanism, Buddhism, and Confucianism have deep roots in Korean culture. Although the country has received continuous streams of foreign cultural influence mainly from China, Koreans have kept their identity and maintained and developed their unique language and customs Westernization, begun in the late 19th century, was in harmony with Korean tradition and slowly transformed the culture without much conflict until the 1940s. After World War II, the occupying Soviets did not recognize the Korean traditional family system or Confucian philosophy; age-old lineage records were burned, and the kinship system was broken. Through education, people were molded to fit the pattern of party idealism, and private life and individual freedom became extremely limited.

The state of the arts

Development plans since the Korean War have demanded almost superhuman patience and labour from the North Koreans. As a result, the people have had to lead an austere existence. The standard of living has improved, but leisure and cultural activities have continued to be regimented and gerred toward organized group activities, such as rallys and museum tours. The government strongly believes in nationalism and is concerned with the maintenance and advancement of the traditional fine arts and other cultural features. The selection of cultural items is based on Communist ideology, and writers and artists attempt to enhance class consciousness and propagate the superiority and independence of Korean culture. Literary works often deal with the personality cult of Premier Kim 11-sung and anti-imperialism.

All of the writers, artists, dancers, and musicians are assigned to government institutions, such as the National Theatre for the Arts, National Orchestra, and National Dancing Theatre in P'ydngyang and to provincial organizations of music, ballet, and drama. Museums have been well sponsored by the government, and many archaeological sites have been excavated to promote the growth of a strong nationalistic feeling. There are more than a dozen museums, including the Korean Revolutionary Museum and the Korean Fine Arts Museum. Archaeological sites are located in Nangnang district of P'ydngyang and at Kungsan, near Yonggang.

Of the approximately 30 daily newspapers, the Korean Workers Party Central Committee's *Nodong sinmun*, the government's Minju Chosdn, and the General Federation of Korean Trade Union's *Nodongja sinmun* have the largest circulations. The Korean Central News Agency controls the dissemination of information, and all papers are strictly censored. The government considers radio and television to be important mass media, and they play a great role in ideological education. Radio broadcasts reach every corner of the country; most people in the rural areas listen to wire receivers given to every village by the government. It was estimated that by the early 1970s almost all North Korean households had access to radio broadcasts as a result of a government project to link household loudspeakers to village receivers. Television broadcasting in North Korea has been made available to all parts of the country, and the number of television sets, both imported and domestically produced, has increased annually.

Prospects for the future. North Korea's future continues to depend upon the correction of economic imbalances and the maintenance of political stability. Unification with South Korea awaits the easing of continued military and political tension. Direct bilateral talks, which began in the early 1970s, have continued but have produced few results.

BIBLIOGRAPHY. SHANNON MCCUNE, ***Korea's Heritage: A Regional and Social Geography*** (1956), a work that provides a general description of the country and of its people. NENA VREELAND ***et al., Area Handbook for North Korea***, 2nd ed. (1975), one of the most comprehensive and objective modern studies of all aspects of the country. ROBERT A. SCALAPINO (ed.), ***North Korea*** Today (1963), a comprehensive collection of essays on various subjects. RYU HUN, Study ***of North Korea*** (1966), a detailed study that is based on North Korean materials and sources.

(C.Le.)

Korea, South

Korea is a peninsula thrusting from the northeast Asian mainland in a southerly direction for about 600 miles (965 kilometres). It reaches southeast to within approximately 120 miles of Honshu, the principal island of Japan, and to within approximately the same distance of the Shantung Peninsula of China. Elongated and irregular in shape, the Korean peninsula separates the Yellow Sea and the Sea of Japan (called the East Sea in Korea).

Because of its strategic location between China, Japan, and the Soviet Far East, Korea has long suffered from the inroads of aggressive neighbours. The peninsula was partitioned along the 38th parallel in 1945, after World War II, by the military occupation of the United States in the south and the Soviet Union in the north. In 1948, by a United Nations resolution, the Republic of Korea (South Korea) was formed in the south, and in the same year the Democratic People's Republic of Korea (North Korea) was formed in the north. As a result of the Korean War, which began in June 1950, following an invasion of South Korea by troops from North Korea, a demilitarized zone was established in 1953. It runs roughly from the mouth of the Han-gang (Han River) on the west to a little south of the town of Kosdng on the east coast and is about 150 miles in length. The present effective administrative area of South Korea is about 45 percent of undivided Korea: 38,221 square miles (98,993 square kilometres). South Korea is divided administratively into nine provinces (do or *to)* and two special cities (*t'ŭkpyŏlsi*), Seoul (Sŏul) and Pusan. It maintains diplomatic and trade relations with more than 100 countries. South Korea has not been admitted to the United Nations but has joined almost all of the United Nations agencies. The country has made considerable economic progress since the 1950s, with the government pursuing ambitious industrialization programs with special emphasis on mobilization of domestic capital and the promotion of export industries.

The landscape. Physical geography. Korea is largely mountainous, with small valleys and narrow coastal plains. The Taebaek-sanmaek (Taebaek Mountains), forming the backbone of the peninsula, run in a north–south direction along the eastern coastline. From them extend several lower mountain ranges oriented in a northeast–southwest direction. Principal rivers (gang), such as the Han, Kum, and Naktong, all have their sources in the Taebaek-sanmaek, and they flow between the ranges. The mountains are not very high, none exceeding 10,000 feet (3,050 metres) above sea level. The highest peak in South Korea, Halla-san on Cheju-do (Cheju Island) is 6,398 feet (1,950 metres) above sea level. There are comparatively extensive lowlands along the lower parts of the Han, Kum, and Naktong rivers. The eastern coastline is relatively straight, whereas the west has an extremely complicated ria (*i.e.*, creek-indented) coastline with many islands. The Yellow Sea and the complex coastline causes one of the highest tidal ranges in the world—about 30 feet maximum at Inch'dn., the entry port for Seoul.

Geological foundation

Geologically, the Korean peninsula consists in large part of Precambrian rocks (more than 570,000,000 years old), such as granite and gneiss. There are two volcanic islands, Cheju and Ullŭng, and a small-scale lava plateau in Kangwdn Province (Kangwdn-do). Most soils derive from granite and gneiss. Sandy and brown-coloured soils are common, and they are generally well leached and have little humus content. Podzolic soils are found in the highlands owing to the cold of the long winter season.

Climate. Because of continental influences the climate of Korea is characterized by a cold winter and a hot summer. The annual range of temperature is greater in the north and in interior regions than in the south and along the coast. The average monthly temperature in January drops below freezing except along the southern coast, and the July average monthly temperature rises to about 78° F (25" C). The average monthly temperature in January at Seoul is about 23" F (—5" C), and about 78° F (25" C) in August; at Pusan, on the southeast coast, the average January temperature is 35° F (2° C), and the August average is about 78" F (25° C), as in Seoul. The annual rainfall varies from about 40 to 55 inches (1,016 to 1,397 millimetres). Taegu, the driest area, has 38 inches, and Pusan, one of the wettest areas, receives 55 inches of annual rainfall. About 70 percent of the annual rainfall is received during the summer monsoons and shifting polar fronts. Occasionally, late summer typhoons cause heavy showers and storms along the southern coast. The frost-free season varies from 170 to 226 days.

Vegetation and animal life. The long, hot, humid summer is favourable for the development of extensive forests, which cover about two-thirds of the total land area. Because of fuel needs in the long cold winter, however, and the high population pressure, the original forest has almost disappeared. Except for subtropical broadleaf forests in a narrow belt along the southern coast, most areas contain broadleaf and coniferous trees. Wild-animal life is similar to that of northern China and the Manchurian region. Tigers, leopards, lynxes, and bears, formerly abundant, have almost disappeared, even in remote areas.

Patterns of settlement. The pace of urbanization since 1960 has caused decreases in rural population, not only proportionally but also absolutely, and by 1980 it had declined to about 43 percent of the total population. Agriculture is the most important occupation in rural areas, and rural settlements are thus found close to arable lands—mainly in river valleys and coastal lowlands. Agglomerated villages are common, ranging from a few houses to several hundred. Villages are frequently located along the foothills facing toward the south, backed by hills that give protection from the severe northwestern winter monsoon winds. Fields are divided into tiny plots and are cultivated by manual labour and animal power. Two types are found: rice fields and upland fields. Rice fields are usually irrigated, while upland fields are unirrigated dry fields, in which barley, wheat, soybeans, and millet are grown. Along the coastline small clustered fishing villages are found. Although the fishing population is not a large portion of the rural population, fishing is an important industry for export and for obtaining protein foods. Logging, mainly of coniferous trees, is limited to the mountain areas of Kangwdn and Kydngsang provinces. Settlements in mountain areas are usually scattered, in contrast to the lowlands. Logging is usually practiced during the off-season for agriculture.

Urban expansion

Rapid expansion of urban areas in the past decade, especially the expansion of Seoul and Pusan, has resulted in considerable changes in urban landscapes. Before 1960 there were very few high-rise buildings; even in Seoul most structures were lower than 10 stories. By the early 1970s, however, buildings of more than 20 stories had become common in the city. Because of this rapid growth, city services, such as water, transportation, and sewage systems, have not met the increasing needs. The difficulties in improving these services is exacerbated by the fact that old and new buildings are still located side by side; modern high-rise buildings stand next to traditional one-story Korean houses, even in the business district.

The people. ***Ethnic background.*** The Korean people may originally have had links with the people of Central Asia, the Baikal region, Mongolia, and the coastal areas of the Yellow Sea. Tools of Paleolithic type and other artifacts found in Sokch'ang-ni, near Kongju, are quite similar to those of the Baikal and Mongolian areas. The physical characteristics of Koreans show Mongolian racial traits, such as dark straight hair, straight noses, high cheek bones, and the Mongolian eyelid fold. The population is quite homogenous, with only about 30,000 foreigners, most of them urban Chinese, in South Korea.

Religion. There is little uniformity of religious belief in Korea, which is confusing to outsiders. In a typical Korean family, the women may adhere to the Buddhist religion, while the men may be followers of the Confucian ethical system. The teachings of Confucius, including ancestor worship, are probably the most important basic belief and are especially strong in the rural areas. Deep-rooted Buddhism is more popular with women, and old Buddhist temples are part of the typical rural and urban scene. Christianity is relatively new in Korea, although it claims a large number of devoted adherents and has had a profound effect on the modernization of Korean society. In 1980 there were about 8,000,000 Christians, of whom about 1,300,000 were Roman Catholics. An eclectic religion, Ch'dndogyo—a combination of Buddhism, Confucianism, Christianity, and even Taoism—spread widely in the latter part of the 19th century. About 1,200,000 adherents of Ch'dndogyo were reported in 1980. Shamanism still remains strong in the minds of rural people, especially among women. Ch'angga Hakhoe (Value Creation Learning Society), introduced from Japan in 1963, has become popular in the low-income areas of major cities; it is a militant society of followers of the 14th-century Japanese Buddhist Nichiren, organized in Tokyo as the Sōka-gakkai in the early 1930s.

Demography. Since 1960 birth and mortality rates have decreased rapidly. In 1960 the birth rate was about 42 and the mortality rate about 13 per 1,000; by 1980 the birth rate had dropped to about 23 and the mortality rate to less than 7 per 1,000. The rate of natural increase, 2.7 percent in the early 1960s, dropped to an average of 1.8 percent by 1980. The decrease in the birth rate has been caused chiefly by a national campaign for family planning conducted and emphasized since 1965. It has also reflected the increasing number of educated people in South Korea. The country, however, continues to struggle to reduce its population growth rate.

South Korea, Area and Population

	area		population*	
	sq mi	sq km	1970 census	1980 census
Provinces (do or *to*)				
Cheju-do	705	1,830	365,000	463,000
Chŏlla-namdo	4,698	12,168	4,006,000	3,779,000
Chŏlla-pukto	3,106	8,045	2,435,000	2,288,000
Ch'ungch'ŏng-namdo	3,394	8,791	2,861,000	2,956,000
Ch'ungch'ŏng-pukto	2,870	7,433	1,482,000	1,424,000
Kangwŏn-do	6,522	16,893	1,867,000	1,792,000
Kyŏnggi-do	4,274	11,069	3,358,000	4,935,000
Kyŏngsang-namdo	4,575	11,848	3,119,000	3,323,000
Kyŏngsang-pukto	7,676	19,881	4,560,000	4,963,000
Special cities *(t'ŭkpyŏlsi)*				
Pusan-si	167	433	1,881,000	3,160,000
Sŏul-t'ŭkpyŏlsi	234	607	5,536,000	8,367,000
Total South Korea	38,221	98,9931	31,469,0001.	37,449,000†

*De jure. †Figures do not add to total given because of rounding.
Source: Official government figures.

Migration patterns

Before World War II, Koreans migrated to two major regions: Manchuria and Japan. People migrating into Manchuria were mainly from northern Korea, while those who went to Japan were mostly from southern Korea. It is estimated that in 1945 about 2,000,000 Koreans lived in Manchuria and Siberia and about the same number in Japan. About one-half of the Koreans in Japan returned to South Korea just after 1945. The most important migration, however, was the north-to-south movement of people after World War II, especially the movement that occurred during and after the Korean War. About 2,000,000 people migrated to South Korea from the North during that period, settling largely in the major cities.

More than one-half of the South Korean population lives in 30 cities and the country's two metropolises, Seoul and Pusan. Seoul, the capital, contains about one-fifth of the total population, which in 1980 was about 37,449,000.

The economy. The South Korean economy has progressed rapidly in the manufacturing and mining sectors under its economic development plans. Export-oriented manufacturing, which received strong government support, was particularly successful, and the country's balance-of-payments has steadily improved, although a payments deficit continued to exist into the early 1980s.

Natural resources. Natural resources in South Korea are meagre. The leading resources are coal, iron ore, graphite, gold and silver, tungsten, lead, and zinc, comprising almost two-thirds of the total value of mineral resources. Iron ore is exported mainly to Japan, and tungsten is exported to the United States. Both are pro-

MAP INDEX

Political subdivisions

Cheju-do	33·20n	126·30e
Chŏlla-namdo	34·45n	127·00e
Chŏlla-pukto	35·45n	127·15e
Ch'ungch'dng-namdo	36·30n	127·00e
Ch'ungch'ŏng-pukto	36·45n	128·00e
Kangwŏn-do	37·45n	128·15e
Kyŏnggi-do	37·30n	127·15e
Kybnasang-narndo	35·15n	128·30e
Kyŏngsang-pukto	36·15n	128·45e
Pusan (Pusan-si)	35·10n	129·05e
Seoul (Sŏul-t'ŭkpyolsi)	37·34n	127·00e

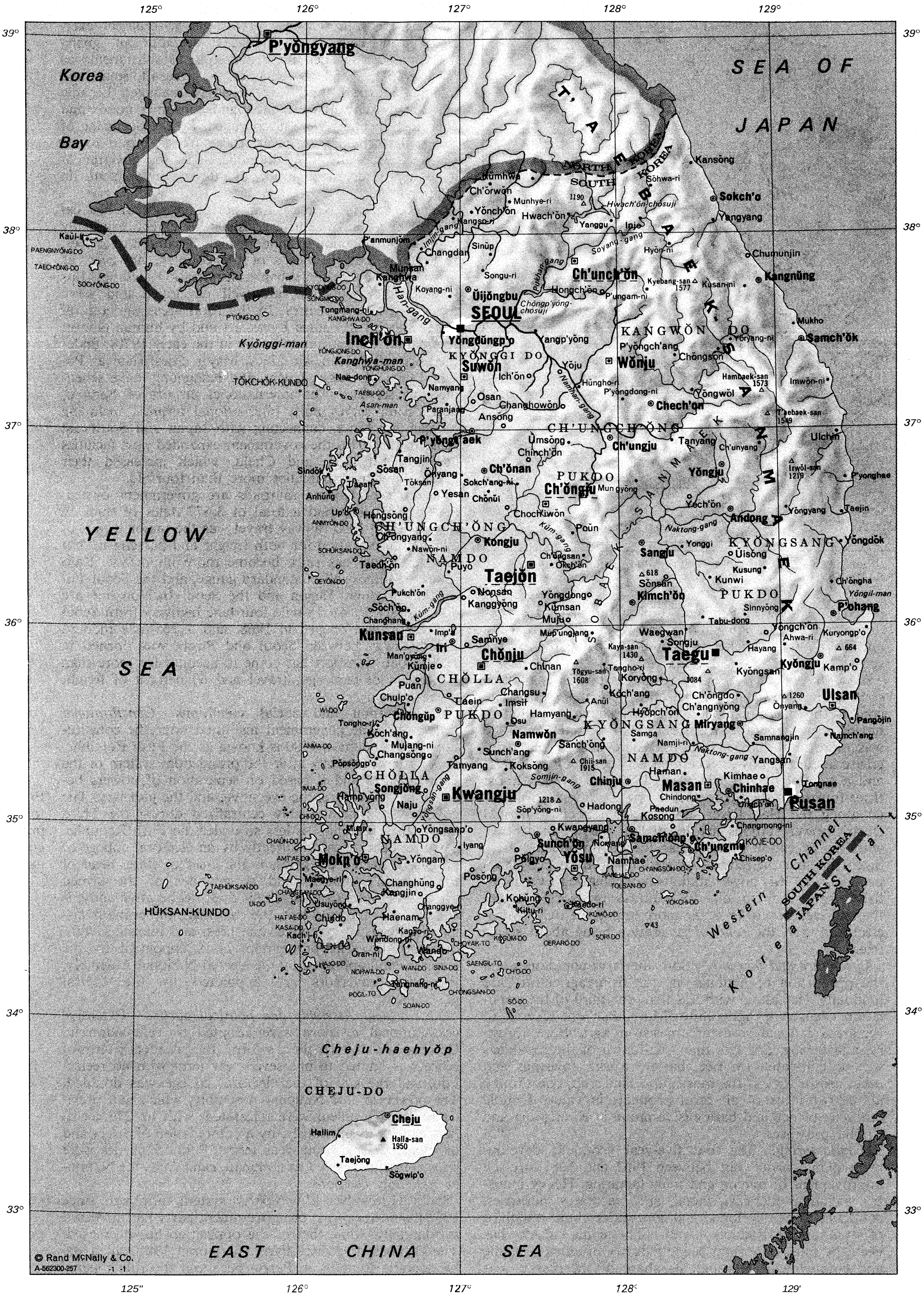
The name of a political subdivision if not shown on the map is the same as that of its capital city.

Cities and towns

Ahwa-ri	35·54n	129·02e
Andong	36·35n	128·44e
Anhŭng	36·41n	126·10e
Ansŏng	37·02n	127·16e
Anŭi	35·38n	127·48e
Changdan	37·56n	126·45e
Changgye-ri	34·33n	126·49e
Changhang	36·01n	126·40e
Changhowbn	37·08n	127·39e
Changhŭng	34·41n	126·52e
Changmong-ni	34·58n	128·41e
Ch'angnyŏng	35·33n	128·29e
Changsŏng	35·20n	126·49e
Changsu	35·40n	127·32e
Cheju	33·31n	126·32e
Chinan	35·48n	127·25e
Chinch'ŏn	36·52n	127·26e
Chindo	34·28n	126·15e
Chindong	35·07n	128·29e
Chinhae	35·09n	128·40e
Chinju	35·l l n	128·05e
Chisep'o	34·50n	128·42e
Choch'iwdn	36·37n	127·18e
Ch'bnan	36·48n	127·09e
Ch'unch'bn	37·52n	127·43e
Ch'ŏngdo	35·38n	128·43e
Ch'ŏngha	36·13n	129·20e
Ch'bngju	36·39n	127·31e
Ch'ŏngsan	36·22n	127·46e
Chbngshn	37·22n	128·38e
Chŏngŭp	35·36n	126·51e
Ch'ŏngyang	36·27n	126·48e
Chŏnju	35·49n	127·08e
Chŏnŭi	36·42n	127·11e
Chbrwbn	38·16n	127·12e
Chuip'o	35·37n	126·40e
Churnunjin	37·54n	128·49e
Ch'ungju	36·58n	127·58e
Ch'ungrnu	34·51n	128·25e
Ch'unyang	36·56n	128·54e
Hadong	35·05n	127·44e
Haenam	34·34n	126·35e
Hallim	33·24n	126·16e
Harnan	35·15n	128·24e
Hamp'yŏng	35·05n	126·30e
Harnyang	35·32n	127·42e
Hayang	35·55n	128·47e
Hongch'bn	37·42n	127·52e
Hongsŏng	36·36n	126·39e
Hŭngho-ri	37·14n	127·44e
Hwach'bn	38·06n	127·41e
Hyŏn-ni	37·57n	128·20e
Hydpch'bn	35·35n	128·08e
Ich'ŏn	37·17n	127·27e
Irnp'a	35·59n	126·49e
Imsil	35·37n	127·15e
Imwŏn-ni	37·15n	129·20e
Inch'ŏn	37·28n	126·38e
Inje	38·05n	128·09e
Iri	35·56n	126·57e
Iyang	34·53n	127·00e
Kach'i-ri	34·27n	126·08e
Kaedo-ri	34·35n	127·39e
Kagyo-ri	34·25n	126·48e
Karnp'o	35·48n	129·29e
Kanggyŏng	36·10n	127·00e
Kanghwa	37·45n	126·30e
Kangjin	34·39n	126·45e
Kangnŭng	37·45n	128·54e
Kangsb-ri	38·06n	126·58e
Kansŏng	38·22n	128·29e
Kaŭl-li	37·58n	124·37e
Kiitu-ri	34·35n	127·20e
Kirnch'ŏn	36·07n	128·05e
Kimhae	35·14n	128·52e
Kŏch'ang	35·26n	126·42e
Kŏch'ang	35·41n	127·55e
Kohŭng	34·37n	127·16e
Koksŏng	35·17n	127·17e
Kongju	36·27n	127·07e
Korybng	35·44n	128·15e
Kosbng	34·58n	128·18e
Koyang-ni	37·42n	126·56e
Kŭmje	35·48n	126·52e
Kŭmsan	36·07n	127·30e
Kunsan	35·58n	126·41e
Kunwi	36·15n	128·34e
Kuryongp'o	35·59n	129·32e
Kusan-ni	37·43n	128·49e
Kusung	36·18n	128·55e
Kwangju	35·09n	126·54e
Kwangyang	34·59n	127·34e
Kyŏngju	35·51n	129·14e
Kydngsan	35·48n	128·43e
Maegye-ri	34·45n	126·18e
Man'gyŏng	35·52n	126·48e
Masan	3511n	128·32e
Miryang	35.31n	128·44e
Mokp'o	34·48n	126·22e
Muan	34·58n	126·26e
Mujang-ni	35·26n	126·32e
Muju	36·02n	127·40e
Mukho	37·33n	129·06e
Munhye-ri	38·10n	127·19e
Mun'gyŏng	36·44n	128·07e
Munsan	37.51n	126·48e
Mup'ungjang	35·58n	127·49e
Nae-dong	37·16n	126·27e
Naju	35·03n	126·43e
Narnch'ang	35·26n	129·16e
Narnhae	34·50n	127·54e
Namji-ri	35·23n	128·29e
Narnwbn	35·25n	127·21e
Namyang	37·14n	126·44e
Nawŏn-ni	36·25n	126·40e
Nonsan	36·12n	127·05e
Noryang	34·56n	127·52e
Okch'ŏn	36·20n	127·34e
Ŏnyang	35·34n	129·07e
Ŏnyang	36·47n	127·00e
Ŏran-ni	34·22n	126·29e
Osan	37·11n	127·04e
Osu	35·31n	127·18e
Paedun	35·03n	128·21e
Pangŏjin	35·29n	129·26e
Paranjang	37·08n	126·55e
P'ohang	36·03n	129·20e
Pŏlgyo	34·52n	127·21e
Pŏpsŏngp'o	35·20n	126·27e
Posŏng	34·47n	127·04e
Poŭn	36·29n	127·43e
Puan	35·45n	126·44e
Pukch'ŏn	36·13n	126·45e
P'ungam-ni	37·43n	128·11e
Pusan	35·06n	129·03e
Puyŏ	36·18n	126·54e
P'yŏngch'ang	37·23n	128·22e
P'yŏngdong-ni	37·10n	128·02e
P'yonghae	36·46n	129·28e
P'yŏngt'aek	37·00n	127·05e
Samch'ŏk	37·27n	129·10e
Samch'ŏnp'o	34·57n	128·03e
Samga	35·25n	128·05e
Samnangjin	35·23n	128·50e
Samnye	35·55n	127·05e
Sanch'ŏng	35·26n	127·54e
Sangju	36·26n	128·09e
Seoul (Sŏul)	37·33n	126·58e
Sindŏk	36·47n	126·10e
Sinnyŏng	36·04n	128·46e
Sinŭp	37·54n	127·12e
Sŏch'ŏn	36·05n	126·41e
Sŏgwip'o	33·13n	126·34e
Sŏhwa-ri	38·15n	128·13e
Sokch'ang-ni	36·44n	127·17e
Sŏkch'o	38·12n	128·36e
Songjŏng	35·10n	126·46e
Sŏngju	35·55n	128·16e
Songu-ri	37·49n	127·09e
Sŏnsan	36·16n	128·17e
Sŏp'yŏng-ni	35·01n	127·24e
Sŏsan	36·47n	126·26e
Sŏul, see Seoul		
Sunch'ang	35·23n	127·07e
Sunch'ŏn	34·57n	127·28e
Suwŏn	37·17n	127·01e
Tabu-dong	36·03n	128·31e
T'aean	36·46n	126·16e
Taech'ŏn	36·22n	126·34e
Taegu	35·52n	128·35e
T'aein	35·40n	126·55e
Taejin	36·34n	129·24e
Taejŏn	36·20n	127·26e
Taejŏng	33·15n	126·17e
Tamyang	35·21n	126·58e
Tangjin	36·54n	126·37e
Tangnang-ni	34·12n	126·52e
Tanyang	36·57n	128·21e
Tŏksan	36·43n	126·39e
Tongho-ri	35·31n	126·29e
Tongho-ri	35·49n	127·54e
Tongmang-ni	37·37n	126·26e
Tongnae	35·12n	129·05e
Ŭijŏngbu	37·44n	127·03e
Ŭisŏng	36·22n	128·41e
Ulchin	36·59n	129·23e
Ulsan	35·34n	129·19e
Ŭmsŏng	36·56n	127·41e
Ungch'hn	35·07n	128·44e
Up'o	36·36n	126·22e
Usuydng	34·35n	126·18e
Waegwan	35·58n	128·24e
Wando	34·18n	126·47e
Wŏndong-ni	34·23n	126·40e
Wŏnju	37·22n	127·58e
Yanggu	38·06n	127·59e
Yangp'yŏng	37·30n	127·29e
Yangsan	35·21n	129·03e
Yangyang	38·04n	128·36e
Yech'dn	36·40n	128·26e
Yesan	36·41n	126·50e
Ybju	37·18n	127·37e
Yŏngam	34·48n	126·40e
Yŏngch'ŏn	35·58n	128·56e
Yŏngch'ŏn	38·07n	127·05e
Yŏngdŏk	36·26n	129·23e
Yŏngdong	36·10n	127·48e
Yŏngdŭngp'o	37·31n	126·54e
Yonggi	36·24n	128·24e
Yŏngju	36·50n	128·37e
Yŏngsanp'o	34·58n	126·44e
Yŏngwŏl	37·12n	128·28e
Yŏngyang	36·40n	129·07e
Ybryang-ni	37·30n	128·43e
Yhsu	34·46n	127·44e

Physical features and points of interest

Amt'ae-do, island	34·50n	126·07e
Anma-do. island	35·21n	126·02e
Anmyŏn-do, island	36·30n	126·22e
Asan-man, bay	36·56n	126·51e
Changsan-do, island	34·38n	126·10e
Ch'angsŏn-do, island	34·52n	128·01e
Chaŭn-do, island	34·53n	126·03e
Cheju-do, island	33·20n	126·30e
Cheju-haehyŏp, *strait*	33·40n	126·28e
Chi-do, island	35·04n	126·13e
Chii-san, mounfain	35·20n	127·44e
Chin-do, island	34·25n	126·15e
Ch'o-do, island	34·14n	127·15e
Chŏngp'yŏng-chbsuji. lake	37·40n	127·30e
Ch'ŏngsan-do, island	34.11n	126·54e
Choyak-to, island	34·22n	126·54e
East China Sea	32·45n	126·30e
Hajo-do, island	34·17n	126·03e
Halla-san, mountain	33·22n	126·32e
Hambaek-san, mountain	37·09n	128·55e
Han-gang, *river*	37·45n	126·11e
Hat'ae-do, island	34·32n	126·03e
Hŭksan-kundo, islands	34·30n	125·20e
Hwach'ŏn-chbsuji. lake	38·07n	127·52e
Imja-do, island	35·05n	126·05e
Imjin-gang, river	37·47n	126·40e
Irwŏl-san, mountain	36·50n	129·06e
Japan. Sea of	38·30n	129·30e
Kanghwa-do, island	37·40n	126·27e
Kanghwa-man, bay	37·20n	126·35e
Kasa-do, *island*	34·28n	126·03e
Kŏgŭm-do, island	34·27n	127·11e
Kŏje-do, island	34·52n	128·37e
Korea Strait	34·00n	129·00e
Kŭm-gang, river	36·00n	126·40e
Kŭmŏ-do, island	34.31n	127·45e
Kyebang-san, mountain	37·43n	128·29e
Kyodong-do, island	37·45n	126·16e
Kyŏnggi-man, bav	37·25n	126·00e
Naktong-gang, river	35·07n	128·57e
Namhae-do, island	34·48n	127·57e
Namhan-gang, *river*	37·31n	127·18e
Nohwa-do, island	34·12n	126·35e
Oeraro-do, island	34·27n	127·30e
Oeyŏn-do, island	36·14n	126·05e
Paengnyŏng-do, island	37·57n	124·40e
Pogil-to, island	34·09n	126·33e
Pukhan-gang, river	3731n	127·18e
P'yŏng-do, *island*	37·35n	125·43e
Saengil-to, island	34·19n	126·59e
Sinji-do, island	240 n	126·50e
Soan-do, island	34·09n	126·39e
Sobaek-sanrnaek. mountains	36·00n	128·00e
Soch'ŏng-do, island	37·46n	124·45e
Sŏ-do, island	34·02n	127·18e
Sohŭksan-do, island	34·04n	125·07e
Somjin-gang, river	34·58n	127·46e
Songmo-do, island	37·42n	126·18e
Sori-do, island	246 n	127·48e
Soyang-gang, river	37·52n	127·40e
T'aebaek-san, mountain	37·06n	128·55e
T'aebaek-sanmaek. mountains	37·20n	128·50e
Taebu-do, island	37·15n	126·35e
Taech'ŏng-do, island	37·49n	124·43e
Taehŭksan-do, island	34·40n	125·25e
Tŏgyu-san, mountain	36·48n	127·23e
Tŏkchŏk-kundo, islands	37·14n	126·07e
Tolsan-do, island	34·38n	127·45e
Ui-do, island	34·37n	125·51e
Wan-do, island	34·21n	126·42e
Western Channel	34·40n	129·00e
Wi-do, island	35·36n	126·17e
Yellow Sea	36·00n	125·00e
Yokchi-do, island	34·38n	128·15e
Yŏngil-man, *bay*	36·02n	129·26e
Yŏngjong-do, island	37·30n	126·31e
Yŏngsan-gang, river	34·54n	126·32e



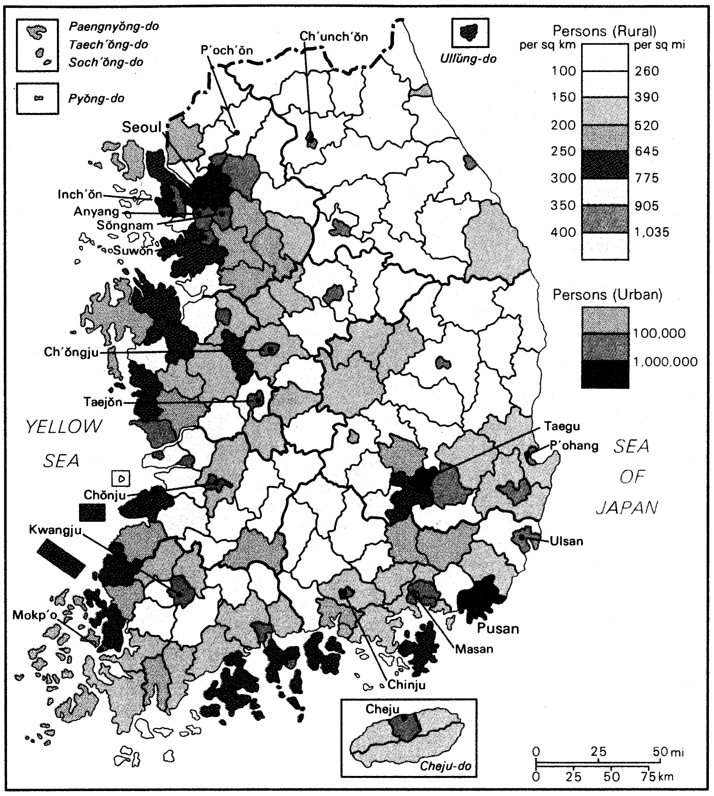
SOUTH **KOREA**

Size of symbol indicates relative size of town ▪ ○ ⊙ ⊡ □ ■

Elevations in metres

0 25 50 100 150 km

0 25 50 100 mi



Urban population distribution and rural population density of South Korea

duced in Kangwŏn Province. Gold and silver are produced mostly in Ch'ungch'dng and Kydngsang provinces, and a large-scale smelter is located at Changhang, on the southwest coast near Kunsan.

Energy production. Energy resources consist mainly of coal, petroleum, and hydroelectric potential. Anthracite coal is the major exploited energy resource, production having increased rapidly since 1960. Production in 1980 was more than 20,000,000 tons. There was a fivefold increase in electric power generated between 1961 and 1971, and a twofold increase between 1972 and 1980. Almost three-fourths of the hydroelectric power stations are located along the Han River, not far from Seoul. Thermal electric-power stations are located in the urban centres, with the exception of one near the coal-producing area. Since the first oil refinery started to produce petroleum products in 1964, power stations have changed over gradually from coal to oil. In contrast to North Korea, thermal electric power is far more important than hydroelectric power, the ratio of generation being about six to one.

Agriculture and forestry. Both the farm population and the proportion of national income from agriculture are gradually decreasing, with less than one-third of the population engaged in agriculture. More than one-seventh of the gross national product comes from agriculture, forestry, and fishing, and less than one-fourth of the republic's area is cultivated for rice, barley, wheat, soybeans and other crops. Rice is the most important crop, constituting about two-fifths of all farm products in value. Double cropping of rice and barley is common in Kybngsang and Chŏlla provinces.

Development of industry

Manufacturing. The four five-year economic development plans carried out between 1962 and 1981 stressed improvement of mining and manufacturing. Heavy industry, including chemical, metal, and machinery industries, continues to be developed and constitutes about one-third of all industrial products in value. The textile industry has been the most important single industry in terms of value and employment, although its share of the market has decreased. In 1980 textile exports accounted for almost one-third of all manufactured goods exported. South Korea's principal export markets are the United States, Japan, and Southeast Asian countries.

During the four five-year plans, the government exercised strong controls on industrial development, giving most support to such large industrial establishments as fertilizer plants and oil refineries. As a result, small and middle industries that were privately managed became increasingly difficult to finance, and consumer goods and consumer spending were discouraged. By the beginning of the Fifth Republic, however, these trends began to be reversed, especially in credit policies, as the government increasingly divested itself of direct involvment in industry.

Transportation. Korea's transportation system has expanded and improved to a considerable extent, especially with the introduction of modern highway and air services. It has not met the needs of the country, however, as indicated by congested urban transportation facilities. Bus transportation networks are well developed and serve most of the rural centres. In the rural areas agricultural products are still hauled by oxcart and by human labour. Internal air transportation began in the early 1970s, under the aegis of Korean Air Lines. Major cities, such as Pusan, Taegu, Kwangju, Cheju, and Kangnŭng, have scheduled air services. The international airport of Kimp'o, near Seoul, serves most international airlines, including Korean Air Lines, connecting with Japan and Southeast Asia. In the 1970s the government expanded port facilities at Pusan, Inch'dn, and Cheju, which increased their freight-handling capabilities more than fourfold.

The bulk of Korean railroads are government owned, and in 1980 they had a total of 3,677 miles of track in service. Until 1960 railway travel was the major means of inland transportation for both freight and passengers, but road transport has since become more important. Railroads are almost all of standard gauge, and the Seoul–Pusan line through Taejdn and the Seoul–Inch'dn line are double-tracked. An 18-mile four-lane highway from Seoul to Inch'dn was opened in 1968, and a 265-mile modern superhighway between Seoul and Pusan was opened in 1970. Road transport has come to account for more than 90 percent of passenger travel and 60 percent of freight transport.

Administration and social conditions. ***Constitutional framework.*** The government instituted after the constitutional referendum in 1980 is known as the Fifth Republic. It is patterned mainly after the presidential system of the United States and is based on separation of powers between the legislature, the executive, and the judiciary. The government system, highly centralized during the Third and Fourth republics, is less so under the Fifth Republic. The president, chosen indirectly by an electoral college of popularly elected delegates, is the chief of state, head of the executive branch, and commander of the armed forces. The State Council, the highest deliberative body, is composed of the president, the chairman, the prime minister, the heads of executive ministries, and ministers without portfolio. The prime minister is appointed by the president and approved by the elected National Assembly. Provincial governors are appointed by the central government,

Provisions for presidential government

The National Assembly and political parties. Extensive constitutional revisions were adopted by referendum in October 1980. Under these reforms the president's term of office was limited to one seven-year term with no reelection, and the size of the electoral college was doubled. The powers of the National Assembly, which had previously been curtailed, were reinstated, with its 276 members chosen, as previously, by a combination of direct and indirect election to four-year terms. Political parties were reinstated and allowed to nominate candidates for National Assembly elections.

South Korea has a two-party system, although since 1972 the power of the pro-government party has increased substantially, while that of the opposition has decreased. The Democratic Justice Party (DJP; until 1981 called the Democratic Republican Party) has been the ruling party since its founding in 1963. The Democratic Korea Party (DKP; until 1981 called the New Democratic Party), founded as a coalition of opposition groups in 1967, has been the major opposition. There are also minor political parties.

The judiciary. The judicial branch comprises the Supreme Court, three appellate courts, 12 district courts, and a family court. The Supreme Court is empowered to interpret the constitution and all other state laws and to review the legality of government regulations and activities. The chief justice is appointed by the president with the consent of the National Assembly, upon recommendation of the Judge Recommendation Council.

Education. Six years of primary school education is compulsory, and by 1980 almost all children of school age were enrolled. About 92 percent of primary school graduates go on to three years of middle school, and a little more than 85 percent of middle-school graduates go on to high school or to technical schools. Some 46 percent of high-school graduates go to higher educational institutions. Before World War II there were only 19 college-level institutions in South Korea, but by 1980 the number had increased to 84. Admission to the colleges and universities is in most cases granted through competitive entrance examinations.

Health and welfare. The availability of medical services has increased, but medical facilities and the number of personnel are inadequate to meet the country's needs. Improvement has been hampered partly by a continuous migration of medical personnel to foreign countries: in 1970, more than 4,500 doctors were practicing in or under contract to foreign countries, mostly the United States and Africa. Within South Korea there were 22,000 doctors by 1980—about six for every 10,000 persons.

Government welfare activities

Government welfare activities are new and limited in range. The programs include care of disabled war veterans, homes for the aged and for homeless and disabled war widows and orphans, vocational training of women. and care of juvenile delinquents. Since the devastation of the Korean War, United Nations agencies, civilian and military agencies of the United States, and private volunteer agencies have played a significant role in improving living conditions in the south. The shortage of housing remains a problem, but it has been partially solved by central and local government-housing programs in metropolitan areas. The national police force is strongly developed in order to counteract Communist infiltration.

Living conditions in South Korea are relatively low, compared to conditions in developed nations, but since the Korean War they have improved steadily. From 1968 to 1979 per capita disposable income increased more than sevenfold. Public health and sanitation have been greatly improved, thus reducing epidemics. Life-expectancy rates rose from about 53 years in the late 1950s to about 66 years by 1980. Despite improvement of social conditions there is little sign of a narrowing of the gap between rural and urban and low- and high-income groups.

Cultural life and institutions. Shamanism, Buddhism, and the philosophy of Confucius constitute the most important background of modem Korean culture. Since World War II, especially after the Korean War in 1950, the modem trends have rapidly progressed. Traditional thought, however, still plays an important role under the surface. Korea belongs to the Chinese cultural realm, although Koreans have maintained a distinct cultural identity throughout their history. The National Museum maintains about 80,000 artifacts of Korean culture, including many national treasures, chiefly in the central museum in Seoul. Its four branches are located at Kybngju, Puyŏ, Kongju, and Chbnju.

Architecture. After the Three Kingdoms period (57 BC to AD 668), Korean culture was strongly influenced by the Chinese, although this influence was given a distinctive Korean stamp. Korean architecture shows Chinese influence, but it is adapted to local needs and environment, utilizing wood and granite, the most abundant building materials. Beautiful examples are found in old palaces, Buddhist temples, stone tombs, and Buddhist pagodas.

Painting and ceramics. One of the earliest examples of Korean painting is found in the mural paintings in the kings' tombs of Koguryŏ (37 BC–AD 688). The best known mural paintings are those in the Sangyong Tomb, at Yonggang, located in North Korea. Ceramic arts became highly developed, flourishing during the Koryŏ period (918–1392) and diffusing to Japan, and every province continues to produce its distinctive ceramic wares.

Dance and music. Folk dances survive, and folk music, accompanied by native musical instruments, is performed occasionally at ceremonies and festive occasions. The government has made an effort to preserve the traditional arts. The National Classical Music Institute (formerly the Prince Yi Conservatory), for example, plays an important role in the preservation of folk music. It has had its own training centre for national music since 1954 and has been graduating 30 musicians each year. The Korean National Symphony Orchestra and the Seoul Symphony Orchestra give concerts in Seoul and Pusan, and in 1970 these ensembles celebrated the 80th anniversary of the introduction of Western music to Korea.

The mass media. Since 1945 there has been a sharp rise in the number of newspapers and in radio and television broadcasting. There are 25 national daily newspapers, including two English-language papers, as well as 21 local newspapers. In 1980 there were 54 radio broadcasting stations and 5 FM stations. Television was introduced in 1956, and there are 26 stations, 21 of which are run by the government.

Prospects. The Republic of Korea has been in a transitional stage between a traditional and an industrial society. An economic structure centred on primary industry is gradually changing into one oriented toward the secondary and tertiary economic sectors. Industrialization has accelerated urbanization and created new problems, such as a rural–urban gap and an overconcentration of population in the large metropolitan areas of Seoul and Pusan. The economy continues to be weak because of heavy dependence on foreign capital and a shortage in its international balance of payments. The U.S. aid program, begun in 1945, was terminated in 1970, but Japan has maintained a flow of investment capital as well as aid to South Korea. Inflation and the price of raw materials, especially oil, also continues to affect the country's rate of economic growth and foreign exchange.

South Korea is faced with the problem of trying to reconcile its differences with the north, while maintaining its relations with its allies, notably the United States and Japan. During the 1970s and early 1980s there were several rounds of low-level talks between the two Koreas, but these ended in a return to mutual hostility. The unification of the two countries, however, remains a high priority.

BIBLIOGRAPHY. SHANNON MCCUNE, ***Korea's Heritage: A Regional and Social Geography*** (1956), and ***Korea: Land of Broken Calm*** (1966), provide a general description of Korea's geography, people, and culture. K.G. CLARE ***et al.***, ***Area Handbook for the Republic of Korea*** (1969), is a good source of general information on social, political, economic, and national security matters. TAE-HUNG HA, ***Guide to Korean Culture*** (1968), is a survey of the varied phases of Korean culture. See also the ***Korea Annual;*** and REPUBLIC OF KOREA, NATIONAL BUREAU OF STATISTICS, ECONOMIC PLANNING BOARD, ***Korea Statistical Handbook*** (annual), ***Korea Statistical Yearbook*** (annual), ***Monthly Statistics of Korea,*** and ***Social Indicators in Korea 1981,*** which include current facts and statistics on government, foreign relations, the economy, social affairs, education, and culture. REPUBLIC OF KOREA, OVERSEAS INFORMATION, MINISTRY OF CULTURE AND INFORMATION, ***A Handbook of Korea*** (1978), includes a detailed discussion of and extensive bibliography on the land, race, history, culture, arts, customs, government, foreign policy, and social developments.

(C.Le.)

Korean Language

Korean is spoken by an estimated 55,000,000 people on the Korean peninsula and its coastal islands, and many among the approximately 663,000 Korean residents in Japan still speak the language. As the language of a sizable population occupying a strategic position in east Asia between China and Japan, Korean is important both historically and culturally. A considerable body of belles lettres and other literature is written in it.

Origin and classification. The direct precursor of the modern Korean language was the language of the ancient

The Korean Alphabet (Han'gŭl)

	McCune-Reischauer romanization*			approximate equivalent sound in English
	initial	medial	final†	
Consonants				
ㄱ	k	g	k‡	rut, a*g*ain, ba*ck*
ㄲ	kk	kk	k§	s*k*y
	n	n	n	*n*ine
ㄷ	t	d	t‡	take, ideal, hot
ㄸ	tt	tt	t§	stand
ㄹ	r	r	l‖	water, *l*eap
ㅁ	m	m	m	*m*an
ㅂ	p	b	p‡	put, a*b*out, hop
ㅃ	pp	pp	p§	spend
ㅅ	s	s	t‡	so
ㅆ	ss	ss	t§	mess sergeant
ㅇ	(silent)	ng	ng¶	singer, si*ng*
ㅈ	ch	j	t‡	chin, adjust
ㅉ	tch	tch	t§	mea*t* chopper
ㅊ	ch'	ch'	t♀	achieve
ㅋ	k'	k'	k♀	account
ㅌ	t'	t'	t♀	a*tt*end
ㅍ	p'	p'	p♀	appear
ㅎ	h	h		home

	McCune-Reischauer romanization	approximate equivalent sound in English
Vowels and Diphthongs		
ㅏ	a	dot
ㅐ	ae	buck
ㅑ	yað	*ya*
ㅒ	yae	*ya*mmer
ㅓ	ŏ	*cu*t
ㅔ	e	s*e*t
ㅕ	yŏ	*you*ng
ㅖ	ye	*ye*t
ㅗ	o	law
ㅘ	wa□	*wa*nder
ㅙ	wae	*wa*x
ㅚ	oe	Ger. k*ö*nnen, or English wet
ㅛ	yo	*yaw*l
ㅜ	u	m*oo*n
ㅝ	wŏ	*wo*n
ㅞ	we	*we*t
ㅟ	wi	weep
ㅠ	yu	*you*
ㅡ	ŭ	book
ㅢ	ui	
ㅣ	i	n*ea*t

*The system of phonetic transcription of Korean generally used for scholarly purposes in the United States. †Only seven consonant sounds (*k, t, p, n, l, m, ng*) are pronounced at the end of a word or before a consonant. Others are phonetically "reduced" to one of these seven when they occur in these positions; consequently, different letters may be romanized in the same way. The same letter may be romanized differently in different phonetic contexts because of sound changes. ‡Lenis (lax or softer than the corresponding tense consonant), voiceless, and slightly aspirated at the beginning of a word and typically voiced, except the sound *s*, between voiced sounds: *s* is palatalized as in English "she" when it comes before *i*, or *y* sounds. At the end of a word or before a consonant the sounds *k, t, p* are unreleased. §Tense (fortis or "reinforced"), voiceless, and unaspirated. ‖In South Korean published writing this letter is not found at the beginning of words except in some recent loanwords; in North Korean published writing it is found at the beginning of both Korean and foreign words. This letter, represented by *r*, is pronounced like the *t* in "water" when it comes between vowels or semi-vowels and at the beginning of words; when it comes at the end of a word or before a consonant, it is pronounced like English initial l. There is also a long *l* sound, romanized *ll*, that occurs between vowels or semi-vowels; this is the result of sound assimilations involving l. There is no special letter in the alphabet for it. ¶The sound of this letter in medial position is to be distinguished from *n'g*, which is the combination of *n* + *g* sounds. ♀Voiceless, and strongly aspirated. ðThe *y* part of a diphthong is represented in the Korean orthography by an extra stroke added to the letter for a vowel. □The w part of a diphthong is represented in Korean orthography by the letter for the vowel *o* or the letter for the vowel u. ◇This diphthong has no close approximation in English.

kingdom of Silla (c. 57 BC–AD 935), the original territory of which was in the southeastern part of the peninsula. The language of Silla spread over most of the peninsula when, in the 7th century, that kingdom conquered and annexed the territories of its rival states, Paekche (c. 18 BC–AD 660) in the southwest and Koguryŏ (c. 37 BC–AD 668) in the north. It is believed that the people of Silla and Paekche spoke dialects of the same language, while the people of Koguryŏ spoke a different, though related, language.

Possible relationship to Altaic

The most fruitful hypothesis concerning the affinity of Korean to other languages has been that it belongs to the Altaic family, which has traditionally included the Manchu-Tungus, Mongol, and Turkic groups of languages. Like them, Korean is a language of the agglutinative type, and it shares with them many features of phonology and grammar. (An agglutinative language typically combines into a single word several components, each of which remains relatively distinct in form and meaning.) This theory is supported by a number of cognates (words of similar form that presumably developed from a single ancestral form) and some sets of sound correspondences. The relationship is, however, probably a remote one.

Some scholars have proposed that Korean and Japanese are related. Their grammatical structures are similar in many ways, but their phonologies (sound systems) differ greatly. It is still an open question whether the sound correspondences demonstrated so far rigorously support a genetic relationship.

A standard spoken and written language is taught in the schools and used as the norm throughout Korea. This form of Korean was defined in the 1930s and is based primarily on the speech of the middle class of Seoul, with certain peculiarities excluded. Regional dialects, which are still vigorous, can be grouped into six major divisions: central, northeastern, northwestern, southeastern, southwestern, and Cheju-do, off the southern coast. Generally referred to by the name of the provincial areas with which they have traditionally been associated, these regional dialects differ from one another most conspicuously in intonation (the pitch pattern of a sentence), vocabulary, and the form of some grammatical endings.

Ancient Korean records

Knowledge of the early stages of the language that existed before the introduction of the native Korean alphabet in the middle of the 15th century is still slight. The most extensive records of ancient Korean extant are 25 texts of songs, called *hyangga*, from the kingdom of Silla; these are quoted in works written in Chinese during the succeeding Koryŏ period (935–1392). In these writings Korean is written in Chinese characters used as phonetic and semantic symbols of Korean words.

There are no generally agreed upon dates for the stages of development of the language. A possible dating is as follows: Old Korean (before the 12th century), Middle Korean (beginning of the 12th century to the end of the 16th century), Modern Korean (from the beginning of the 17th century).

The most penetrating external influence exerted on the language has been that of Chinese, from which many words were borrowed over the centuries. This influence has been limited largely to vocabulary. In addition, until the end of the 19th century, the prevailing written language of serious Korean literature was Chinese (*hanmun*) or, later, a deliberately Koreanized version of it.

Standardization of the spoken and written language became an urgent problem toward the end of the 19th century, when Korea undertook large-scale reforms to meet the social needs occasioned by internal changes and by pressures from the West. Efforts by Korean scholars, educators, and writers to develop a standard spoken and written language culminated in the publication by the Chosŏnŏ hakhoe (Society for the Study of the Korean Language) in 1933 of its *Han'gŭl mach'umppŏp t'ongiran* ("A Proposal for Unifying the Orthography") and its P'yojunmal *moŭm* ("Collection of Standard Forms of Words") in 1936. These standards became official only in 1945, upon Korea's liberation from Japanese rule. Some divergencies in the spelling system and in word usage have since developed in the north and in the south.

Native Korean alphabet

The writing system. The native Korean alphabet was introduced in 1446, after centuries of the use of cumbersome methods (known as *idu*) to transcribe Korean with Chinese characters. The new set of 28 letters (not an adaptation of an existing alphabet) was designed by a group of scholars commissioned by Sejong (reigned 1419–50), the fourth king of the Yi dynasty (1392–1910). Although the alphabet was meant to be a "script for the people," as its original name *Hunminjŏngŭm* implies, it did not win acceptance as a respectable form of writing among the literati, or scholar-officials, who continued to write *han*mun. This alphabet, also called Ŏnmun (the common script), did, however, give rise to a body of premodern popular literature. Finally, in the beginning of the 20th century, the alphabet received general recognition as the national writing system. Today it is known as Han'gŭl in South Korea and as Chosŏn Muntcha in North Korea.

Two methods of writing Korean are in use today: the purely alphabetic method and the "mixed script" method, in which Sino-Korean words (Chinese loanwords) may be written in their original characters and read in their Korean pronunciation. In North Korea only the alphabetic method has been in use since 1949; in South Korea both methods are in use, though it has been the government policy to dispense with Chinese characters in gradual stages. Four of the original 28 letters have gone out of use. The accompanying alphabet chart includes combinations of letters that are units in alphabetizing. Letters are grouped into syllable blocks, which contain at least an initial consonant plus a vowel. If the phonetic syllable has no initial consonant, a special letter is used as a filler for the initial consonant space in the syllable block and is silent in that position.

Three-way consonant distinction

Linguistic characteristics. Phonology. The McCune-Reischauer system of transcription is used for the Korean in this article; see the alphabet chart for the phonetic values of the transcription symbols. In the phonology, the consonants show an unusual three-way distinction among a series of lenis (soft) consonants, p, t, k, *ch*, s; a fortis (hard), unaspirated series, pp, tt, kk, *tch*, *ss*; and a fortis, aspirated series, p', *t'*, k', *ch'*. (Aspirated consonants are pronounced with an audible release of air.) These consonants are all typically voiceless (*i.e.*, pronounced without vibration of the vocal cords), but the lenis consonants p, *t*, k, and *ch* are voiced (*i.e.*, pronounced with vibrating vocal cords) when they come between voiced sounds, thus becoming pronounced as b, d, *g*, and *j* (as in "jam").

The central characteristic of the Korean phonological system is that the consonants are more severely restricted as to position of occurrence within a word than within a morpheme (*i.e.*, within the stems and affixes of words). For example, the sound [s] cannot be pronounced at the end of a word, although some word stems do end in *s*; or such sequences as [s] + [k] cannot be pronounced in Korean, but they can result from adding a suffix beginning with k to a stem ending in s. (The square brackets indicate that included letters stand for sounds pronounced rather than the conventional spelling.) The language has an extensive system of sound changes to resolve these conflicts between combining forms and permissible pronunciations. The standard orthography spells words in terms of the morphemes (constituent units) that comprise them, as long as the sound changes are covered by a regular phonological rule. Thus, most morphemes preserve their orthographic identity, although they may not be pronounced as spelled; *e.g.*, the noun stem for "five" is always spelled tascis, whether in the form *tasŏs-i* "five (as subject)," pronounced [tasŏsi], or in the form *tasŏs* "five (as a word by itself)," pronounced [tasŏt], with a final [t]. Forms based on the verb stem pis- "to comb," are spelled pis-6, pis-ko, but are pronounced as [pisb] and [pikko] respectively. If sound changes are not covered by a phonological rule of the language, the principle is to spell words phonetically.

Grammar. Aside from interjections and primary (not derived) adverbs, a word in Korean typically consists of a stem plus an ending. Stems are of two types: (1) nouns and (2) verbs and adjectives. Nouns take endings that mark their syntactic role in the sentence—such as subject, topic (often equivalent to English "as for," as in "As for Korea, it's a peninsula."), object, genitive—or that express such meanings as "to, in, from." They are indifferent as to grammatical number and gender. Verbs and adjectives have much the same pattern of forms and enter into many of the same kinds of constructions. They may take one or a combination of two tense suffixes, roughly designated as past and future tense; the lack of a tense suffix indicates present tense or tenselessness. To the stem or to the stem plus the tense suffix is added an ending that concludes the sentence (called a sentence-conclusive form), or a conjunctive, nominal (noun), adnominal (*i.e.*, attributive), or adverbial form.

Basic clause structure

The basic structure of a clause is subject + predicate; in the predicate the main verb or adjective comes last. An example is [ab6jiga^1 dje^2 segŭmŭl^3 naesiGssiimnida^4] "Father^1 paid^4 his taxes^3 yesterday^2." This order applies to conjunctive clauses (*i.e.*, clauses similar to "but he left early" in "He came, but he left early.") as well as to independent clauses, and it holds, too, for independent clauses regardless of mood—declarative, interrogative, imperative, and so on; *e.g.*, [abdjiga^1 6je^2 segŭmŭl^3 naesiŏssŭmnikka^4?] "Did Father' pay^4 his taxes^3 yesterday^2?"

The mood of the Korean sentence is specified by an ending on the verb or adjective that concludes the sentence. This ending expresses a complex of meanings—not only whether the sentence is a statement, a question, or a command, but also whether the speaker is reporting something from personal observation or from hearsay, and whether the speaker is certain or hesitant, and so on. At the same time, this ending specifies one of several possible levels of address for the sentence, reflecting the social relationship, conventionally defined, between the speaker and the person spoken to. For example, "Where^2 is^3 the post office'?" is [uch'egugi^1 Gdi^2 iss*ŭmnikka*^{3}?] when spoken to a senior or a stranger, [uch'egugi bdi iss*ŏ*?] when asked of a friend, and [uch'egugi ŏdi inni?] when addressed to a child. The relationship between speaker and person spoken to is one of three dimensions of a highly developed system of honorifics (forms showing deference). A second dimension, occurring between the speaker and the person spoken about (*i.e.*, the subject of the verb or adjective), requires a stem marked by the element [(ŭ)si-] when speaking of a senior. Thus, [ab6jiga^1 sanpporŭl^2 ka*s*inda^3] "Father' is going^3 for a walk^2" contrasts with [ch'dlsuga sanpporŭl kanda] "Ch'ŏlsu (a boy's name) is going for a walk."

The third dimension concerns the speaker in speaking of himself before a senior.

Pronouns, especially those for 2nd person ("you"), are used sparingly. The pronoun subject or object may be left out and understood from the linguistic or nonlinguistic context.

Modifying elements precede that which is modified. The most characteristic way to modify nouns is by the use of adnominal (attributive) forms of verbs and adjectives or of whole clauses; examples are [choh*ŭn*^{1} nalssi^2] "weather' that is good'," or, "nice' weather^2"; [naega' tanid*ŏn*^{2} hakkyo^3] "the school^3 that I' attended^2"; [ka*l*^{1} ttae^2] "time^2 to go^1." Adnominal + noun constructions play a variety of important roles in Korean syntax; *e.g.*, in nominalization processes (processes by which nouns or nounlike expressions are formed from verbs or verbal expressions, such as English "unifying" from "unify").

In multiclause sentences the clauses are attached one to the other in linear fashion. The last clause in the sentence is the main clause, containing the verb or adjective with the sentence-conclusive ending. The conjunctive relation is expressed by the conjunctive ending on the main verb or adjective in the preceding clause; *e.g.*, [mari tomada chokkŭmssik tarŭ*jiman*^{1} hakkyoesŏnŭn p'yojunmarŭl karŭch'inda^2] "Speech varies a bit from pro to province, *but*^{1} in the schools they teach the standard language^2."

Vocabulary. Among the Sino-Korean words, which make up more than half of the vocabulary, there are many ordinary as well as learned words. Sino-Korean elements and now also Sino-Japanese ones are the primary resources for new technical terminology. Most Sino-Korean elements are used as nouns or nounlike units in Korean, regardless of their original part of speech. They can be made into verbs or adjectives by adding certain stems that mean "to do," "to be," "to become." This device is also a productive one for native Korean nouns.

BIBLIOGRAPHY. The McCune-Reischauer system of transcription used in this article is the one most widely used; it is described in detail in G.M. MCCUNE and E.O. REISCHAUER, "The Romanization of the Korean Language," ***Trans. Korea Brch R. Asiat. Soc.***, vol. 29 (1939), and in an abbreviated account in ***ibid.***, vol. 38 (1961). For the history and development of the Korean alphabet, see the English resumé in KI-MOON LEE, ***Kugo P'yogippop ui Yoksajok Yon'gu*** (1963). For Chinese characters in Korea, see HYONG-GYU KIM, "Chinese Characters and Korean Language," ***Korea Journal***, vol. 3 (1963). The introductory chapter, "A Short History of Korean Literature," in DOO SOO SUH, ***Korean Literary Reader*** (1965), points out relations between Korean writing systems and literature. Surveys and bibliographies of studies in Korean linguistics are found in FRED LUKOFF, "The Republic of Korea," in THOMAS A. SEBEOK (ed.) ***Current Trends in Linguistics***, vol. 2 (1967), and in KI-MOON LEE, "Linguistics," in SUNG-NYONG LEE (ed.), ***Korean Studies Today*** (1970). The basic western work in the field of Korean grammar, and one which has provided much of the inspiration and direction for studies in Korean historical and comparative linguistics, is G.J. RAMSTEDT, "A Korean Grammar," ***Mémoires de la Société Finno-Ougrienne***, vol. 82 (1939). SAMUEL E. MARTIN, ***Korean Morphophonemics*** (1954), is the most comprehensive technical description of sound alternations in Korean phonology. There is no general grammatical description of Korean in English available as yet. A.A. KHOLODOVICH, ***Ocherk grammatiki koreiskogo iazyka*** (1954), presents a description in Russian of modern Korean phonology, morphology, and syntax. RENE DUPONT and JOSEPH MILLOT, ***Gramnzaire core'ene*** (1965), assumes some familiarity with Korean writing and phonology on the part of the student but serves as a good reference work for modern Korean morphological forms and syntax, illustrated by many example sentences. BRUNO LEWIN, ***Morphologie des koreanischen Verbs*** (1970), contains a concise description of Korean verb forms and an alphabetically arranged list of the endings, with example sentences. The following textbooks constitute complete courses in the elementary spoken language and include explanations of phonology and grammar: FRED LUKOFF, ***Spoken Korean***, 2 vol. (1945–47); CHANG HAI PARK, ***An Intensive Course in Korean***, 2 vol. (1961–65); B. NAM PARK, ***Korean Basic Course***, 2 vol. (1968–69); SAMUEL E. MARTIN and YOUNG SOOK C. LEE, ***Beginning Korean*** (1969). The method of writing and reading ***Han'gul*** is explained and illustrated in the first chapter of ANTHONY V. VANDESANDE and FRANCIS Y.T. PARK, ***Myongdo's Korean '68***, pt. 1 and 2 (1967). A comprehensive dictionary of the spoken and written language compiled for western students of Korean is SAMUEL E. MARTIN, YANG HA LEE, and SUNG UN CHANG, ***A Korean-English Dictionary*** (1967), with the entries in *Han'gŭl* and their pronunciations recorded in transcription.

(F.L.)

Korean Religion

The religions of Korea have developed from native folk traditions as well as from sophisticated religions introduced from abroad.

Nature and types of religion in Korea. ***Folk traditions.*** Though the original residents of the Korean peninsula have not yet been definitely identified, their linguistic and religious trends share many characteristics with the southern Tungus branch of the Altaic-speaking peoples of northern Asia. Their primeval thought form was shamanistic, a belief system centring on the occult, psychic transformation and healing powers believed to be possessed by a particular religious personage called a shaman. According to shamanism, curses on hair, nails, or any parts of the body can doom the whole of the victim; and dreams and omens are viewed not only as signs but also as the causes of happenings. Such practices in shamanism point to the fact that these folk traditions had not yet developed a concept of the whole, which is differentiated into and counterposed with its component parts. (The concept of the whole is a philosophical view that enables a systematic structure—*e.g.*, social or religious—to be built.) They held to a system of appendages, as in the parts of the body, that allowed for a syncretism that was rich in vitality and affluency. Thus, shamanism has withstood the assault of historical changes and preserved many quaint customs of the folk culture.

Shamanistic characteristics

The great religious traditions. The structure of the sophisticated religious traditions in Korea, in comparison with the simplicity of the folk traditions, usually has had a dichotomized form. In the matter of truth and falsity, for example, Confucianism accepts truth as true and falsity as false, whereas Buddhism considers that neither truth nor falsity is a valid criterion for judgment. Taoism, however, claims that truth cannot exist without falsity and that they are two phases of one reality. Western Christianity, by placing its faith in Christ, the God-man or the Creator-creature, accepts paradox as the ultimate form of truth. Such types of religion, out of a dichotomized framework of reference, tend either to negate the existing social structure or, by affirming the status quo, to become the state religion.

History of religion in Korea. ***Prehistoric northeast Asian religion and early Korean primitive religion.*** Recent archaeological findings in the Kongju area in east central South Korea show clear evidence of human occupation of the Korean peninsula over 35,000 years ago.

By courtesy of the Chun Sung Woo Kan Song Museum, Seoul

Shaman in a séance, painting by Shin Yun-bok (1758–1820). In the Hyung-pil Collection, Chun Museum, Seoul.

From the 3rd millennium BC to historical time, Altaic people of northeast Asia moved into the peninsula and became the dominant ethnic group. Among the religious remains of the period are simple dolmens (prehistoric stone monuments), which are common in northeast Asia, though usually absent in China proper. By the 2nd millennium BC, a new wave of Tungus migrated into the country. Like other Altaics, their foundation myth asserted their shaman-king, such as Tangun, as the descendant of Heaven. The god of heaven—as one of three gods (*sam* sin), the other two being the god of earth and the god of ancestors, believed to be more important than the numerous other gods—was worshipped more than any other god. Tangun, a great shaman and the mythical founder of Korea in 2333 BC, established worship of the god of heaven, according to tradition, at a great altar located near Seoul. The Tungus revered the earth in a fertility cult incorporating a ritual pole with a symbolic goose on top. All impressive natural phenomena were accepted as either deities or the works of deities. The shaman propitiated the malice of evil spirits and invoked the protection of the benevolent ones. This form of faith remains strong among some Koreans in the 20th century.

Religion in the Three Kingdoms period

The spread of Chinese religions and Buddhism to Korea. The established religions of China began to penetrate into Korea during the period of the Three Kingdoms (*c.* AD 57–668): The Koguryŏ kingdom in the north, the Paekche kingdom in the southwest, and the Silla kingdom in all of Korea after 668. To match the formation of complex bureaucratic hierarchies, tribal worship and shamanism were supplemented by a sophisticated state cult, usually a blend of Buddhism, Taoism, and Confucianism, that flourished in both China and Korea during the Six Dynasties period of China (AD 222–589). Buddhism was officially accepted at the courts of Koguryŏ in 372, of Paekche in 384, and of Silla in 528. Under King Chinhŭng of Silla (540–576), the traditional drills of the younger members of the nobility, centring on mystical experience in the native cult, military skills, and comradeship for national leaders—called Hwarangdo, the chivalry of Silla—were fused with a messianic faith in Maitreya (the bodhisattva, or "buddha-to-be"). This fusion of a chivalric code and a messianic faith generated an unprecedented zeal to transform Korea into the Buddha land of the world and established the impetus for unification of Korea under Silla.

Under unified Silla (668–935) important Buddhist monks often went to study the *dharma* (Sanskrit: "law") and the many and various Buddhist scriptures in China and sometimes travelled on to India, the land of Buddhism's origin. On returning to Korea, these pilgrims brought back knowledge of various sects then popular in China. The monk Uisang, in 669, brought back many Buddhist scriptures and relics and established the Hwaom (Avatamsaka) sect. Wonhyo, perhaps the greatest monk, strove for a unified outlook in Buddhism and established the Pobsong (*dharmatā*) sect. The Hwaom and Pobsong sects were the most influential of the o-gyo, the "five doctrinal sects" that placed their emphases on the Buddhist scriptures, the other three being the Yolban (based on the *Nirvāṇa-sūtra*) sect established by Poduk, the Kyeyul (based on the *Vinaya Piṭaka*) sect founded by Chajang, and the Pobsang (based on the *dharma-lakṣaṇa*) sect of Chinpo. After becoming a court religion, Buddhism followed a form of geomancy (divination by means of line or landscape conformations) called p'ungsuchirisol (theory of wind, water, and earth) from Taoism in order to select temple sites vital for the national safety. Pulguk-sa and Sokkuram grotto on a high ridge near Kyŏngju, with a commanding view of the Eastern Sea, apparently were intended to protect the country from Japanese marauders. Another task of the period was to make Buddhism popular among the masses. Wonhyo tore down the barrier between the secular and the sacred, asserting that all creatures had inherent and eventual buddhahood; their differences, he proclaimed, were matters of degree in cultivation. All were entitled to the richer promise of worldly rewards and greater hope for coming Nirvāṇa (the state of bliss in which there would be an extinction of suffering).

The zenith and decline of Buddhism

In the Koryd dynastic period (935–1392), Buddhism, though it had been adulterated by Taoistic geomancy and the Yin-Yang theory of opposites (in which earthly and heavenly cosmic forces were seen in various forms of relationship and opposition), was at its height both at the court and among the people. Many temples were built with funds from the national treasury, and monks assumed significant social positions, so much so that in 1197 seven sons of the King became monks. By the 13th century, however, the political frustration under Mongol domination was sublimated in religious devotion; 81,137 blocks of the Korean Tripiṭaka were carved in the exile capital at Kanghwa island, completed in 1251 after 16 years of work, in an effort to win back Buddha's protection against the invaders. The corruption of the Buddhist monks in positions of influence—gained as a result of royal favour—however, brought about outcries of resentment, and, after the advent of the Yi dynasty in 1392, official sponsorship of Buddhism was removed, King Sejong in the 16th century eliminated all but 36 Buddhist temples and merged the existing sects into two group-: Gyo (or Kyo; a doctrinal sect) and the Sŏn (Zen); *i.e.*, Buddhist intuitive meditative sects. Official disfavour and even persecution brought Buddhism to a low state. In its struggle to exist, it often combined with the folk belief maintained in Taoism and shamanism, thus losing its distinctive intellectual characteristics. Yi Sŏng-gye, the founder of the Yi dynasty, had adopted Confucianism as the state religion and recruited government officers through examinations based on the Confucian Classics. During the 16th century, Neo-Confucianism, a reinterpretation of Confucianism based on the views of its founder, Chu Hsi, a Chinese scholar, rose to a position of great significance, encouraging the rise of many Confucian philosophers and scholars. Neo-Confucianism became the only acceptable interpretation of the Confucian Classics, and its teachings were propagated through a school system (Sung Gyun Kwang) ranging from the elementary to the university level. The schools often maintained their own shrines for worship of Confucius and his 72 disciples.

Western influence and the development of native religious movements. In 1592, when the Japanese under Hideyoshi invaded Korea, a Jesuit priest, Gregorio de Cespedes, came with them. His stay was very brief, however, and his influence negligible. By the 17th century, Nestorianism, an Eastern Christian doctrinal view considered heretical by the Western and Eastern Orthodox churches, had come to China and extended its influence even to Korea. Though Korean embassies to Peking often met with Roman Catholic missionaries during this period, it was not until the middle of the 18th century that some Korean tribute envoys visited a Roman Catholic church in Peking and inquired about Christianity and the new sciences. The leading Confucianists of the time, such as Lee Su-kwang and Chong Yak-yong, were attracted by the new science and high ethics of Christianity and initiated a new movement, the "School of Practical Learning," as a means of making up for the now barren Neo-Confucianism of Chu Hsi. A number of discontented lower officers embraced Roman Catholicism and organized a native congregation that invited Chinese and then French missionaries to celebrate mass and to hear confession. The Confucianists in power looked upon Christianity as subversive because it refused to offer cultic service to the ancestors; in 1791 two Christians, Yun Chi-chung and Kwan Sang-yon, were executed on these charges. In 1801 a petition to the Vatican from the Korean church for the dispatch of Western troops to open up missions in Korea was intercepted, and a series of persecutions forced Catholicism underground until 1866. In that same year (1866) Robert Thomas became the first Protestant martyr in Korea. With the establishment of diplomatic relations between Korea and the United States in 1882, U.S. Protestant missionaries gained a free access to Korea. Soon after the Japanese gained control of Korea after defeating Russia in 1905, a religious revival

Arrival of Protestantism and the rise of the new religions swept over the Protestant churches (1907) that strengthened them. Western medicine and education were introduced to the masses. The new education imbued them with a democratic spirit that eventually encouraged them to take an active part in the 1919 revolution. By 1907 there were already 989 Protestant churches and about 70,000 followers. Three years later Japan annexed Korea as part of its empire.

Before the Japanese annexation, an indigenous religious movement that was anti-Western and anti-Roman Catholic — Tonghak (Eastern Learning) — arose. Founded in 1860 by Ch'oe Che-u, who was executed four years later, it emphasized social action and the principle that God and man are one. Tonghak, a syncretism of Taoism, Confucianism, and Buddhism, caused a revolt in 1894 that precipitated the Sino-Japanese War; in 1905 it was renamed Ch'ŏndogyo, or the religion of the heavenly way.

Japanese influence and recent developments in Korean religion. During the period that Korea was under Japanese suzerainty (1910–45), the Japanese used religion as a means to further colonial policy. They first united the Buddhist Gyo and Sdn groups into one Cho-ke-jong in 1911 and bound the various denominations of Christianity into one Reformed Congregation of Churches (1942) and forced Christians to join in Shintō (a Japanese religion combining folk traditions and nationalism) worship. Nearly 1,000 Shintō shrines were erected by the Japanese, but Shinto gained very little support. These modifications in religious organizations and worship disappeared as soon as the nation obtained independence in 1945, after the defeat of Japan in World War II.

During the Japanese occupation, various sectarian movements arose, many of them similar to Ch'ŏndogyo. After the war, more than 170 new religious sects emerged; they were syncretistic, messianic (looking for a saviour and hoping to establish a kingdom of heaven on earth), and shamanistic. Appealing to the masses with their claims of healing and prophetic powers, they demonstrated a potential for social and political power. Among the Christian, Buddhist, and Confucian religions, however, schisms have lessened their effectiveness for social reform. Under the impact of the 1960 student revolution, however, a new wave of social concern has become evident in Christianity and the other religions.

Sources of religious traditions. *Religious texts.* Very few shamanistic texts exist. Early in the 20th century some of the chants used by shamans were collected by Japanese and Korean scholars. The Buddhist *Tripifaka Koreana,* a monumental collection of texts from the Koryd period, together with 4,745 carved blocks of the Silla period, are preserved in the Haein–sa.

Korea did not have its own collections of Taoist writings. Korean Taoists, however, usually used Ko Hung's *Pao-p'u-tzu,* as references. The Confucian scholar Lee Hwang's "Anthology on the Nature of Principle" is a representative collection that included Neo-Confucian works up to the 17th century.

Foundation myths

Mythological texts and traditions. The earliest foundation myth, as recorded in the 12th century, *Samguk sagi,* by the Confucian scholar Kim Pu-sik, states that Hwanin, a divine being, and his son Hwanung descended to earth, intending "to benefit humanity to the broadest extent." He ruled the earth with three minor spirits (wind, cloud, and rain) and 360 subordinate deities in charge of the daily life. During his rule, a bear and a tiger petitioned him to convert them into damsels. Hwanung prescribed a magical herb to be consumed while they were hibernating in a cave for 40 days. Only the bear persisted and became thus a woman. She conceived under a sandalwood tree a son called Tangun, or prince of sandalwood tree, the founder of Korea, who ruled over the tribes of Dragon, Horse, Deer, Crane, Eagle, and Egret and mediated to them the things of heaven.

Another group of foundation myths is concerned with oviparous (external egg) origins. The Silla dynasty claimed that their progenitor, King Bul-ku-nae, was hatched out of an egg left by a celestial white horse. This oviparous myth was of Chinese origin, whereas the earlier myth of a divine ancestor was Altaic. The Koryŏ and Yi dynasties did not fabricate any new foundation myths of their own. The Tangun myth has gained in popularity through the centuries and has become the accepted myth of the country.

Common characteristics and unique features of Korean religion. *Beliefs and doctrines.* Because all impressive phenomena of life were associated with deities, the pantheon of shamanism included deities of fertility, longevity, the household, the township, luck, protection, pestilence, and many other things and conditions. Sometimes deities from Buddhism, Taoism, and even Christianity were added to the shamanistic pantheon. This was because it was believed that a careless omission of any deity might endanger the fortune of an individual as well as society. Sometimes, according to shamanistic belief, one could be seized by the wandering ghost of a person who had died an unnatural death, and thus propitiation was required to send the ghost off to his next incarnation. In such a world in which spirits were of significance, the overall coherence and intelligibility of life was believed to be regulated by the shamans, who, through their propitiative and preventive offerings, enabled a person to sustain his life. Under such conditions, one's morals had to be diffused to meet the requirements of the various deities. Confucianism in Korea from the Koryŏ period on depended heavily on Chu Hsi's Neo-Confucianism. First, Chu Hsi supplemented aphoristic Confucian teachings with a metaphysical foundation borrowed from the *I Ching* ("Classic of Changes"), as well as Taoistic teaching on the Ultimate, or T'ai Chi, made up of the Yin and Yang complex. It was supposed to have permeated profusely all existences.

The Neo-Confucianism of Chu Hsi

Second, Chu Hsi developed a theory that the mind is the cosmic principle in a miniaturized form. Because everyone projects his psychological state to his family activities, social involvement, national welfare, and even to the boundless cosmos, one should attempt to make his nature good by means of gaining the harmony of and equilibrium between any two opposing yet complementary forces, *i.e.*, the Yin and the Yang; the sway of passion to either extreme would cause a person to be deprived.

Third, Chu Hsi believed his teaching could be preserved and carried on through instruction in the five Confucian Classics: *Shih Ching* ("Classic of Poetry"), *Shu Ching* ("Classic of History"), I *Ching* ("Classic of Changes"), *Ch'un Ch'iu* ("Spring and Autumn Annals"), and *Li chi* ("Collection of Rituals"). Finally, emphasis was placed on putting Confucian teachings into social action. Many cooperative ventures, for example, were set up to experiment with Confucian concepts of social propriety (*jen*) between man and man.

Korean Confucianists of the Yi dynasty appropriated Neo-Confucianism and tried to solve their crucial philosophical problems. Li Hwang extended the ethical implications of *jen* to include all-out involvement, reverence, and mindfulness. Neo-Confucianism was thus able to clarify the frequent misunderstanding of its moralistic formalism. Another scholar, Li Yi, refined the anthropological concepts of Neo-Confucianism. Chu Hsi had accepted the dichotomy of man as composed of a rational soul and bodily feeling. The vague term mind was used to connect the two. Li Yi resolved the soul–body relation as follows:

> Principle is principle; Existence is existence
> One at the same time two, but not two
> two at the same time one, but not one
> without change
> without division
> without separation
> without confusion

His achievement laid the Neo-Confucianist keystone, which was parallel to the doctrinal formulations of Western Christology (nature of Christ) at the Council of Chalcedon (AD 451) that were embodied in the Athanasian Creed saying:

Our Lord, Jesus Christ

Complete in Godhead	Complete in Manhood
truly god	truly man
One substance with Father	One substance with us

> One person two natures, not two persons
> distinction is not annulled by union
> without confusion
> without change
> without division
> without separation

In China, Taoism — like shamanism in Korea — was an inclusive receptacle that assimilated other religions and cultural elements. After the Silla dynasty, Taoism was accepted at the court in Korea together with Confucianism and Buddhism as an essential component of organized religion, "one leg of the religious tripod." It became the courtly religion for national safety. At the mass level its teachings and functions were assimilated by and almost superseded by native shamanism. Philosophical Taoism, under strong influences of Confucianism, assumed Zen Buddhistic form. Its teaching on naturalness through non-action helped to eliminate various artificial attempts for enlightenment.

The teachings of Wonhyo

When Buddhism reached Korea, it attempted to win over a people long imbued with shamanism. The various sects of Mahāyāna Buddhism of the north and south dynasties were successful in filling the needs and curiosities of the diverse religious groups of the Three Kingdom period. When Silla intended to unify Korea, Wonhyo propounded a teaching that attempted to harmonize the doctrinal trends of the period. He taught that, somewhat similar to fire on ice, the flame of all mundane desires would melt the frozen buddhahood, thus quenching reflexively man's drive for desire. All teachings, Wonhyo also believed, were not contradictory but complementary to each other, making up one whole truth. Likewise, all men are made of the same buddha nature. The distinction of the sacred from the secular was a matter of difference only in degree. Wonhyo's urging for the equality of all mankind overcame the current insistence on social rank determined by birth, or "bone ranks": unity of all men became the social goal of the time. It offered a religious ground for the long-awaited political unity of Korea.

The Zen Buddhist schools, however, reinterpreted primitive Buddhism by means of Taoist argumentative techniques. By means of neither/nor, or double negation, all complementary items or actions were denied of their validity.

> The things taught by the Tathāgata are, in their essential nature, neither conceivable nor scrutable neither existent nor non-existent neither phenomena nor non-phenomena" (Diamond *Sūtra*).

The above *sūtra* (aphoristic saying) means that neither I, the subject, nor its complement, the world, really exist; *e.g.,* neither the secular nor the sacred have real existence. The double negation, the Zen schools believed, would release man from self-made perplexity and unveil to him the essential nature "as is" (*tathāgata*).

Ritual practices and institutions. Korean shamans have been sensitive to aesthetic tastes; this has enabled them to refine the coarser aspects of the old Tungus culture and to contribute to various spheres of folk culture. In various stages of seances the shaman changed his cloak and magical tools. His five-colour rainbow topcoat became a standard of Korean fashion. The shaman's fan dances and sword dances established basic patterns for Korean dancers. Many of their chantings, both the delicate and the more powerful, became the bases of popular ballads.

Shamans usually did not stand above the people but rather with the people as attentive leaders and counsellors who often were not without humour. From time to time, however, they could indulge in greed for power and gains through sorcery. For these reasons they were suppressed by various dynasties, especially under the Confucian kings of the Yi dynasty.

Cultic practices were occasioned by calendric celebrations as well as by calamities, such as pestilence, famine, drought, or other misfortunes. The procedures in the cultic séances of shamans included purification through incantation, invocation, praise, presentation of the case, the spirits' response, entertainment with song and dance, and the parting. The ordinary comings and goings of daily life were reflected in the worship procedures. This may be the reason why, in parting, some gifts are given to the spirits for their trip back to the other world.

Buddhism, Taoism, and Confucianism, as institutional religions, usually had titular heads, appointed by the state, and clerics who had passed qualification examinations on their canonical writings. The birthday of the founder was celebrated — for Buddha, with offerings, chanting, and lanterns in commemoration of his enlightenment; and for Confucius, with music for cosmic harmony and a courtly offering to supplicate the deceased founder. In passage rites (*e.g.,* at birth or death), Buddhists offered for people in general a propitiatory incantation for a better reincarnation; for the Buddhist devotee, the rituals were performed for the purpose of aiding one to enter Nirvana. Confucians usually presented seasonal offerings to the deceased parents, but these were not, strictly speaking, worship services but rather extensions of this-worldly filiality to be deceased. Because of this custom, the worst curse for traditional Confucianists and those influenced by Confucianism is not to be "damned to hell" but to "die without descendant," so that the deceased has to suffer eternal thirst and hunger without offerings.

Ch'ŏndo-gyo and Tajong-gyo

Contemporary forms of indigenous Korean religions. The contemporary forms of the many indigenous Korean religions are well represented by Ch'ŏndogyo and Tajong-gyo. Followers of Ch'bndogyo chant the formula given by its founder, Ch'oe Che-u: "As one waits on the Lord in Heaven, all creatures would find peace." To them, the human heart in its original nature, as taught by Neo-Confucianists, was the perfect image of heaven. In that blissful state, the cosmic order was in peace. The mind, in gaining the original state, calls forth a desire to reform this disorderly earth into a heaven. A bowl of clear water is placed on an altar to reflect heaven to the bowing worshipper, and a spoonful of rice is set aside daily to be used for making firm the heaven on earth. Their social concerns often plunged the sect into armed rebellions against the government and caused them to participate in the 1919 revolution against Japanese suppression.

Tajong-gyo, or the Tangun cult, was formulated by Na Chul. It worships the Lord, the Light, or the Progenitor of the Heaven. The triune deity consists of Great Wisdom, Power, and Virtue, which is parallel to the mind, body, and breath of humanity. The union and harmony of the Heavenly Trinity with the trinity of humanity, they believe, will generate a redemptive power that will renew humanity and reform society. The trichotomy of man, his universe, and its pursuit of ultimate harmony in terms of Yin-Yang theory were derived mainly from Neo-Confucianism. The cult shifted the Confucian emphasis on the present to the future and is optimistic about the coming millennium that will change this chaotic world.

Korean religious symbolism. Korean shamanism has had very few symbols for a deity or for use in worship. A pair of upright figures at the entrance to a town represent the General of the Heaven and Guardian of the Earth. Most of the shamanistic symbols depict the often negative presences of the invisible ones (ghosts and spirits): tridents have been used to threaten off the plague ghost; and fans were swung to blow away, or attract, hidden spirits.

Koreans shared with other Asian Buddhists the symbols and iconographic elements—*e.g.,* images, lanterns, and pagoda — that depicted various themes about the Buddha, who, being enlightened, came to rest in Nirvāṇa. Developed out of such common characteristics, Paekche sculptures (4th–7th centuries) had a distinct style in which length was exaggerated over width, somewhat like the Spanish artist El Greco's figures or sculptures on western European medieval columns. They reflected the high religiosity of the period. The Zens channelled their creativity into paintings. To them, the expression of the movement or force seen in the object was more important than the objects themselves. The moment of the enlightenment of Bodhidarma, the founder of Zen (or

Ch'an, as it was called in China), was usually depicted with the saint having the glaring eyes of a trance state, having lost consciousness of self and the world. For Confucianists, any part of the body, because it is received from the parents by birth, is sacred. A devotion to the body directed them to be realistic in painting. Because the human heart was assumed to be the counterpart of the Ultimate of the cosmos, "the heart in the middle of the chest" became a constant reference point in ethics. "Put your hand on the chest" meant to swear by the undeceitful and unswaying part of the self.

Symbolism in Confucianism and Taoism

Taoism especially was the Oriental religion that set trends for the arts. The artist must grasp the primal unity that is to be fused with the painting. He should "see without looking" in order to catch the flow of the movement rather than the appearance of the object. Thus, the artist should feel the spirit in its natural state. In landscape painting, he should lead the viewers to the mountain, which melts into a mist and the mist into nothing—"formlessly fashioned, that exists before heaven and earth, without sound, without substance, depending on nothing." The mystical void beyond, according to this view, joins the artist and the viewer in an unnegotiated oneness.

Religion in contemporary Korea. After Japanese control was ended in 1945, the area of Korea north of the 38th parallel came under Communist domination, and freedom of religion was suppressed. In South Korea freedom of religion was guaranteed by the constitution, and a great revival of religions occurred. In the late 1960s there were nearly 7,500 Protestant churches with over 900,000 members; Roman Catholics listed over 1,150 churches and nearly 790,000 members. Also, in the late 1960s, there were over 2,700 Buddhist temples with nearly 940,000 adherents. The Christian churches, having come from the West, have been attempting to accommodate themselves to Korean traditions, but the Buddhists, having absorbed much of native folk traditions, have been attempting to regain their peculiarly Buddhist teachings and practices, as well as reforming the organizational structures of their sects.

Confucianism, in the early 1960s, maintained over 230 *hyanggyos,* or schools, and listed over 170,000 adherents. Though as a dominant cult Confucianism is on the decline, it still provides the moral foundation for Korean society. The new religions of Korea, noted earlier, have appealed to those on the lower levels of Korean society and claim the allegiance of millions. The folk religions imbued with shamanism continue to dominate the lives of millions.

BIBLIOGRAPHY. Most of the important works on this subject are written in Korean and have not been translated into English. These include HYUN SANG YUN, *Chosen Yuhak Sa* (1960), a history of Korean Confucianism; KIM TUK WHANG, *Hanguk Chongkyo Sa* (1963), a general history of religions in Korea; LEE NUNG WHA, *Chosen Bulkyo Tongsa* (1968), a history of Korean Buddhism; *Chosen Musokko* (1927), a study of Korean shamanism; and *Hanguk Dokyo Sa* (1959), a history of Taoism in Korea; YU HONG LYOL, *Hankuk Chonju Kyohoe-sa* (1962), a history of the Catholic Church in Korea; and YU TONG SIK, *Hankuk Jongkyo Wa Kitokkyo* (1965), on the Korean religions and Christianity. The following English-language studies are recommended: C.A. CLARK, *Religions of Old Korea* (1932), an introduction to Korean religions with Western religious criteria—there are frequent mistranslations of the Chinese text; L.G. PAIK, *The History of Protestant Missions in Korea, 1832–1910* (1929), a comprehensive survey of Korean missions to 1910; and HAHM PYONG CHOON, "Religion and Law in Korea," *Kroeber Anthropological Society Papers,* vol. 40 (1969). Information may also be found in HAN WOO-KEUN, *The History of Korea* (1970); and JOSEPH M. KITAGAWA, *Religions of the East* (1960).

(T.-d H.)

KoSciuszko, Tadeusz

A Polish Army officer and statesman, Tadeusz Kościuszko gained fame both for his role in the U.S. War of Independence and for his leadership of the national insurrection in his homeland.

Participation in U.S. War of Independence

Tadeusz Andrzej Bonawentura KoSciuszko was born on February 4, 1746, in Mereczowszczyzna in the Kingdom of Poland (now in the Belorussian Soviet Socialist Republic) to a family of noble origin. He was educated at the Piarist college in Lubieszów and the military academy in Warsaw, where he later served as an instructor. Kościuszko's outstanding abilities soon attracted the attention of King Stanisław II Augustus Poniatowski, who sent him to Paris for further study in military and civil architecture and in painting. Returning home in 1774, he taught drawing and mathematics to the daughters of a Cossack leader, Józef Sosnowski; he fell in love with Ludwika, one of the daughters, and tried unsuccessfully to elope with her. Facing the wrath of her father, he fled to France, and in 1776 he went to America where he joined the colonial forces fighting for independence from the British. That August he was transferred to the Pennsylvania Committee of Defense in Philadelphia, where he took part in planning fortifications to defend the residence of the Continental Congress against the British. For this work he was given the rank of engineer colonel. In spring 1777, he was assigned to the army of Gen. Horatio Gates at Ft. Ticonderoga, in northern New York. Beginning in July he became active in Gates's army, closing by fortifications all roads along the Hudson River and thus contributing to the capitulation of the British army under Gen. John Burgoyne at Saratoga on October 17. He spent the next two years fortifying West Point, New York, where in March 1780 he was appointed chief of the engineering corps. That summer, serving under Gen. Nathanael Greene in North Carolina, he twice rescued the army from enemy advances by directing the crossing of the Yadkin and Dan rivers. In the spring of 1781 in South Carolina he conducted the Battle of Ninety-six and then a lengthy blockade of Charlestown. At the end of the war he was given U.S. citizenship and was made a brigadier general in the U.S. Army.

By courtesy of the Polish Museum of America, Chicago

Kościuszko, lithograph by C. Motte, 1828.

In 1784 he returned to Poland. Because of his association with the radical Czartoryski family, then in opposition to the King, he could not secure an appointment in the Polish Army. For five years he lived in poverty on a small country estate, in debt, moreover, because of his exceptional deed of freeing his serfs from part of their villein service. With the advent of liberal reforms in Poland, in 1789 he returned to military service. Under the protection of his former love, Ludwika, now the wife of Prince Lubomirski, and with the support of local nobility, on October 12 he was granted the rank of general major. At that time the 44-year-old general fell in love with an 18-year-old girl of noble birth, but again he was unable to win the father's permission for marriage.

In 1792 the Russian Army of the empress Catherine II the Great invaded Poland to end Polish internal reforms designed to liberate the nation from the Russian influence. In the ensuing war Kościuszko rose to fame as a division commander during the bloody Battle of Dubienka

(July 18). For this he was raised to the rank of general lieutenant by the king Stanislaw II Augustus Poniatowski, and the new Revolutionary government in Paris granted him honorary French citizenship. But, when the Polish king, fearing defeat, defected from the liberal cause, Kościuszko prepared to resume fighting. In Russian-occupied Poland, however, the reactionary party assumed power, forcing liberal statesmen into exile in Saxony. Kościuszko, against the King's wishes, gave up his commission and joined the exiles.

From Saxony, in January 1793, he was delegated to Paris to seek support for the Polish cause, first from the Girondists and then the Jacobins, pledging in return radical internal reforms in Poland and military diversion against Prussia and Austria, then at war with Revolutionary France. When he returned to Saxony in August, he faced new demands for starting an uprising in Poland in view of favourable indications there. Kościuszko agreed to command the national forces and went secretly to a place near Cracow (Krakow); but, finding preparations inadequate, he delayed the uprising and then went abroad again. His decision proved unwise because time allowed the enemy to undermine the conspiracy through widespread arrests and reduction of the army. Those left in the underground started the uprising on March 12, 1794. On their request Koiciuszko arrived in Cracow on March 24 and, amid an enormous assembly of people, solemnly swore an act of national uprising against all three occupying powers—Russia, Prussia, and Austria. Undertaking all political responsibility and military leadership, he set up an insurgent administration and military force. To do so he had to compensate for the quality of the enemy army with the quantity of his own. Therefore, he introduced conscription to military service, enlarged existing units by incorporating recruits into them, and developed new formations. Having no war industry, his forces could not be equipped with conventional firearms; hence, he armed his peasant recruits with pikes and traditional war scythes.

Leadership of the Polish uprising

After a smashing victory at Racławice (April 4), won by the scythe-bearing formations under his personal command, special new battle tactics were developed based on columns of men attacking on the run and backed by artillery fire. To win more army volunteers ,from the peasant masses, he issued the Manifesto of Polaniec, on May 7, suspending serfdom and reducing in half the existing villein service. This met with some resistance of the nobility. Defeats forced Kościuszko to retreat to his last stronghold, Warsaw. The defense of this city, besieged by Prussian and Russian armies for about two months, remains Koiciuszko's greatest military success, both as strategist and engineer. He managed to use the city population to build earthworks and to defend the city alongside the regular army. In critical moments he himself led the charge with fixed bayonets. Next, he stirred up an uprising in the occupied province of Wielkopolska, at the rear of the besieging armies, forcing the Prussian king Frederick William 11 to retreat. But Russian reinforcements retaliated quickly. Koiciuszko was unable to concentrate adequate forces in time, and he suffered his greatest defeat at Maciejowice, where he was wounded and taken prisoner. Without its leader, the uprising collapsed, and the Third Partition of Poland ended the existence of the country.

Koiciuszko, imprisoned in the Peter-Paul Fortress in St. Petersburg, was slowly returning to health when, upon the death of Empress Catherine II, her son, the emperor Paul I, granted him freedom in 1796. Despite a crippling illness, Kościuszko returned to the United States. On August 18, 1797, he arrived in Philadelphia, greeted enthusiastically by the people but held in suspicion by the incumbent Federalists. In the United States he led an active life in social circles and entered into a long-lasting friendship with Thomas Jefferson, who was then vice president; but, after receiving news of fresh possibilities to promote Poland's cause in France, he secretly left the United States on May 5, 1798. Before that he had appropriated some of his estate for setting free his Negro slaves and for educating them.

His return to France was a disappointment. A true republican, he distrusted Napoleon, and he refused to command Polish legions in a fight only for France, promising cooperation only if the liberation of his country would also be involved in the plans.

Hence, Kościuszko retired from public life and took up residence in Berville, near Fontainebleau. In 1806 Napoleon tried to gain Kościuszko's aid in the event of war with Russia. But Kościuszko again demanded political commitments, and Napoleon secured other, more agreeable Polish supporters for his plans. Kościuszko remained in exile when his country was rebuilt as the Grand Duchy of Warsaw (1807). Consequently, after Napoleon's fall in 1814, the Russian emperor Alexander I sought Kościuszko's help in bargaining for Poland's territories. After the Congress of Vienna and the formation of a new Polish realm under Russian rule in 1815, the emperor Alexander was anxious to be on good terms with Kościuszko and to have him return home. Kościuszko, however, again proposed unrealistic conditions, demanding social reforms and boundaries of the country reaching the Dvina and Dnepr rivers. When no answer came, he went to Solothurn, Switzerland, and freed all his serfs in Poland from villein service.

Kościuszko died at Solothurn on October 15, 1817. His remains were carried to Krakow and were buried in 1819 among the kings' tombs in the cathedral. The people, reviving an ancient custom, raised a huge mound to his memory near the city.

BIBLIOGRAPHY. TADEUSZ KORZON, *Kos'ciuszko, biografia z dokumentów wysnuta* (1906), a well-documented biography; A.M. SKALKOWSKI, *Kościuszko w świetle nowszych badań* (1924), a hypercritical account, although based on more recent research than Korzon; ST. HERBST, "T. Kościuszko," in *Polski słownik biograficzny,* vol. 14 (1969), an exhaustive biographical article; MIECISLAUS HAIMAN, *Kościuszko in the American Revolution* (1943) and *Kościuszko, Leader in Exile* (1946), two studies based on U.S. sources.

(S.He.)

Kossuth, Lajos

Kossuth was the mid-19th-century leader who inspired Hungary's struggle for independence from Austria. Though his brief period of power in the revolutionary years of 1848 and 1849 was ended by Russian armies, he remained a world-famous symbol of revolutionary nationalism.

By courtesy of the trustees of the British Museum: photograph, J.R. Freeman & Co. *Ltd.*

Kossuth, lithograph, 1856.

Kossuth was born at Monok, in northeastern Hungary, on September 19, 1802. His father came of Slovak, his mother of local German stock. The family was noble and of ancient creation but not wealthy, and Kossuth's father earned his living as an attorney for local landowning families. The Kossuths were Lutherans, and young Lajos studied at the Protestant academy of Sárospatak. After applying unsuccessfully for a post in government service, he found employment in his native county of Zemplén as agent to one of his father's clients, Countess Etelka Andrássy, with whom he formed an attachment. He did notable work during the great cholera epidemic of 1831 but found his life narrow and frustrating; he was also suffering, as he would all his life, from financial embarrassment. In 1832 his employer had him sent to the national Diet in Pozsony (now Bratislava) as "substitute delegate" for one of her relatives.

Political journalism

At this "long Diet" the new generation of Hungary's reformers was mounting its first full-scale offensive against the absolutist and obscurantist system under which Hungary was then ruled from Vienna, and in its excited atmosphere Kossuth developed his political and social philosophy. This was one of advanced radicalism. There was no postulate of the European liberalism of the day that he did not burn to see realized in Hungary—no abuse or injustice there left unremedied. But liberty meant for him, above all else, national liberty; and he felt passionately that until Hungary enjoyed de facto the internal freedom to which its laws entitled it, no social or economic progress was possible. The first battle, therefore, must be the political one. Sanguine and impulsive, he was blind to the dangers involved in too strong a challenge to Vienna.

Kossuth's mandate did not entitle him to participate in the Diet's debates, but he found a way of voicing his views. At that time the Diet's proceedings were not published, and Kossuth hit on the idea of issuing letters describing them. These reports, which were not verbatim records but colourful impressions barely distinguishable from political pamphlets, were copied by hand by enthusiastic young helpers and circulated throughout Hungary. Brilliantly written, they were widely and avidly read; and when the Diet ended in 1836, the county assembly of Pest invited him to write a similar series on its proceedings. Now, however, he was no longer protected by parliamentary immunity, and on May 4, 1837, he was arrested and, after 18 months' detention, sentenced to three years' imprisonment for subversion.

Released under an amnesty in 1840, Kossuth found himself a popular hero. The proprietor of a biweekly journal, the *Pesti Hirlap,* made him its editor. His articles were written in a fluent and beguiling style and gained him innumerable devotees at the same time that they alarmed the Austrian authorities, the Hungarian conservatives, and even Hungary's moderate reformers. He also antagonized the Croats and non-Magyars of Hungary by his chauvinistic insistence on the supremacy of its Magyar element. In 1844 his publisher dismissed him and he was refused permission to start a journal of his own. Metternich offered him journalistic employment in the service of the government, but this he refused. His next enterprise, inspired by the writings of the German economist and industrial promoter Friedrich List, was to found a society for promoting Hungarian industry, with the ultimate objective of achieving greater economic independence. This program proved a fiasco but afforded him a platform for continued agitation.

Role in the revolution of 1848

In 1847 the county of Pest elected Kossuth to represent it in the next Diet, in which he assumed leadership of the "national opposition," which had agreed on an extensive program of political and social reform. The reformers made a little progress in subsidiary fields, but deadlock had been reached on the central issue of political control when the news of the revolution in Paris (February 1848) gave Kossuth his opportunity. On March 3, in a speech of extraordinary power—for his tongue was as magical as his pen—he demanded the removal of the dead hand of Viennese absolutism as the only way to safeguard the liberties of Hungary and of all the peoples of the monarchy. He practically dictated to the Diet an address to the crown, embodying the reformers' program. When news of the revolution in Vienna reached the Diet on March 14, Kossuth expanded the address, and, as a member of the deputation that carried it to Vienna the next day, saw it accepted by the panic-stricken court.

Count Lajos Batthyány, the new Hungarian prime minister, allotted Kossuth the portfolio of finance in his government, a choice that proved dangerous, for the ultimate control of finance proved, with that of the defense services, to be precisely the chief bone of contention between Hungary and Vienna. Kossuth was soon at loggerheads with the new Ministry of Finance in Vienna—meanwhile, he had made himself the life and soul of the more extreme nationalist movement in Hungary, often to the embarrassment of his fellow ministers, who were striving to prevent a breach with Vienna. Kossuth often acted without consulting them or even in defiance of agreed decisions, appealing over their heads to the public in a journal edited and mainly written by himself. Yet they dared not dismiss him and could not even dispense with his services, for his nationwide popularity was their greatest asset.

It was Kossuth who, so far as any Hungarian did so, precipitated the final clash by persuading the Diet, in July, to tie the dispatch of Hungarian troops to Italy to political conditions obviously unacceptable to Vienna, at the same time calling for a big national force to defend Hungary against the danger he declared, not without reason, to be threatening it from the Croats and Serbs. When, in September, the Austrian-inspired Croat army invaded Hungary and Batthyány resigned, Kossuth became head of the committee of national defense appointed by the Diet as provisional authority. He was now virtual dictator of Hungary. The next months brought out all of his greatness and his weaknesses: his magnetism and his courage, his intolerance and his lack of realism, his wanton provocation of insuperable difficulties and his genius at overcoming them. No one but Kossuth could have given his people the heart to face the overwhelming odds against them; but he increased those odds by his intransigence and aggravated difficulties by his jealousy and suspicion of his best general, Artúr Görgey, and by his meddling in military affairs. The refusal of the Diet to recognize the abdication of the Austrian emperor Ferdinand I (December 2) was his work, as was the Diet's declaration of April 14, 1849, proclaiming the dethronement of "the perjured House of Habsburg-Lorraine." The Diet then elected Kossuth himself "governor" of Hungary, but when, after the arrival of the Russian armies, even he had to recognize the hopelessness of the situation, he resigned this post to Gorgey (August 11) and took refuge in Turkey.

Kossuth in exile

The Western powers put pressure on the Sultan to refuse Austria's and Russia's demand for his extradition, and Kossuth spent two years interned in Kiitahya in Anatolia. The United States government invited him to visit America and sent a frigate. He stopped in England on the way, where he addressed a series of mass meetings, speaking in English, which he had learned from the Bible and Shakespeare during his confinement. He was received with unprecedented popular ovations, and his reception in the U.S. was equally favourable, but in neither country could he obtain official support for Hungary's cause. He then settled in London. In correspondence with his followers at home, he endeavoured to keep alive in them the spirit of resistance. At the persuasion of the Italian patriot Mazzini, with whom he became intimate, he joined his revolutionary committee. Having moderated his views on the question of nationalities, he discussed with various circles, including the Moldavian and Serbian courts, plans—never to be realized and perhaps never quite realistic—for uniting Hungary, Croatia, Serbia, and Romania in a Danubian federation.

In 1859, when war between Austria and France was imminent, the French emperor Napolen III invited him in a personal interview to organize revolt in Hungary on the outbreak of war. Kossuth agreed, subject to certain safeguards and conditions. Military preparations were concerted with France and Piedmont, and Kossuth's own eloquence helped to deter Britain from intervening against France; but the plans collapsed when, in July, Napoleon concluded an armistice with Francis Joseph 1, leaving the Hungarians to their fate.

The promising international conjuncture never recurred; and in the following years Kossuth, living abroad in Turin, had to watch Hungary, guided by Ferenc Deák, move toward reconciliation with the monarch. He did so with bitterness in his heart and on the eve of the conclusion of the Austro-Hungarian Ausgleich, or Compromise of 1867, published an open letter calling down woe upon the measure and its author. This "Cassandra letter" stirred the opponents of the Compromise but could not prevent its adoption and subsequent maintenance. He spent his last years in loneliness, near-poverty, and increasing infirmity, sadly aware that Hungary's new leaders rejected the tenets to which he remained unalterably attached. He died in Turin on March 20, 1894. His body was brought back to Hungary and interred there amid nationwide mourning.

In 1841 Kossuth had married Terézia Meszlényi, who died in 1863. Their son, Ferenc Kossuth (1841–1914), was for a time president of the Hungarian Party of Independence.

After his death, Kossuth remained a popular idol in Hungary, his name a symbol of the aspiration for independence. His legend grew with the years and was further cultivated after 1945, when Hungary had lost much of the independence for which Kossuth struggled.

BIBLIOGRAPHY. There is no adequate full-length biography of Kossuth in any language. He himself produced one volume in book form, published in English in 1880 as ***Memories*** of My ***Exile.*** It mainly concerns his activities in 1859–61, and contains valuable material on his interviews with Napoleon III, his dealings with the Italian statesman Cavour, and his correspondence with the Balkan courts in connection with his plans for a Danubian federation. There are several collections of Kossuth's articles and speeches, the most recent and most complete being *Összes **Munkai,*** 15 vol. (1948–61). Many articles have been written about Kossuth, including the two-volume *Emlékkönyv* (1952) and the articles listed in the bibliography of *Magyarország **Tortknete,*** 2 vol. (1967).

(C.A.M.)

Kropotkin, Peter

Russian revolutionary and geographer, Peter Kropotkin was, from the late 1870s until his death in 1921, the foremost leader and theorist of the anarchist movement. Although he achieved renown in a number of different fields, ranging from geography and zoology to sociology and history, he shunned material success for the life of a revolutionist. Addressing meetings, founding periodicals, writing books and articles, and spreading the doctrines of his creed, he did more than any other figure to further the libertarian cause in Europe and around the world.

Brown Brothers

Kropotkin.

Kropotkin was born in Moscow on December 9, 1842, the son of Prince Aleksey Petrovich Kropotkin, and was educated in the exclusive Corps of Pages in St. Petersburg (Leningrad). For a year he served as an aide to Tsar Alexander II and, from 1862 to 1867, as an army officer in Siberia, where, apart from his military duties, he studied animal life and engaged in geographical exploration. On the basis of his observations he elaborated a theory of the structural lines of mountain ranges that revised the cartography of eastern Asia. He also contributed to knowledge of the glaciation of Asia and Europe during the Ice Age.

Conversion to anarchism

Kropotkin's findings won him immediate recognition and opened the way to a distinguished scientific career. But in 1871 he refused the secretaryship of the Russian Geographical Society and, renouncing his aristocratic heritage, dedicated his life to the cause of social justice. During his Siberian service he had already begun his conversion to anarchism—the theory that all forms of government should be abolished—and in 1872 a visit to the Swiss watchmakers of the Jura Mountains, whose voluntary associations of mutual support won his admiration, confirmed him in his libertarian beliefs. On his return to Russia he joined a revolutionary group that disseminated propaganda among the workers and peasants of St. Petersburg and Moscow. Caught in a police dragnet, he was imprisoned in 1874 but made a sensational escape two years later, fleeing to western Europe, where his name soon became revered in radical circles. The next few years he spent mostly in Switzerland until he was expelled at the demand of the Russian government after the assassination of Tsar Alexander II by revolutionaries in 1881. He moved to France but was arrested and locked up for three years on trumped-up charges of sedition. Released in 1886, he settled in England, where he remained for the next 30 years, until the Revolution of 1917 allowed him to return to his native country.

During his long exile Kropotkin wrote a series of influential books—the most important were ***Paroles d'un révolté*** ("Words of a Rebel," 1885), ***In Russian and French Prisons*** (1887), ***Fields, Factories and Workshops*** (1899), ***Memoirs of a Revolutionist*** (1899), ***Mutual Aid*** (1902), ***Russian Literature*** (1905), and ***The Great French Revolution** 1789–1793* (1909)—in which he set forth his libertarian philosophy. His aim, as he often remarked, was to put anarchism on a scientific basis. In ***Mutual Aid,*** which is widely regarded as his masterpiece, he argued that, despite the Darwinist concept of the survival of the fittest, cooperation rather than conflict is the chief factor in the evolution of species. Providing abundant examples, he showed that sociability is a dominant feature at every level of the animal world. Among humans, too, he found that mutual aid has been the rule rather than the exception. He traced the evolution of voluntary cooperation from the primitive tribe, peasant village, and medieval commune to a variety of modern associations—trade unions, learned societies, the Red Cross—that have continued to practice mutual support despite the rise of the coercive bureaucratic state. The trend of modern history, he believed, was pointing back toward decentralized, nonpolitical, cooperative societies in which men could develop their creative faculties without interference from rulers, priests, or soldiers.

Kropotkin's "anarchist communism"

In his theory of "anarchist communism," according to which private property and unequal incomes would give place to the free distribution of goods and services, Kropotkin took a major step in the evolution of anarchist economic thought. For the principle of wages he substituted the principle of needs. Each person would be the judge of his own requirements, taking from the common storehouse whatever he deemed necessary, whether or not he contributed a share of the labor. Kropotkin envisioned a society in which men would do both manual and mental work, both in industry and in agriculture. Members of each cooperative community would work from their 20s to their 40s, four or five hours a day sufficing for a comfortable life, and the division of labour would yield to a variety of pleasant jobs, resulting in the sort of integrated, organic existence that had prevailed in the medieval city.

To prepare men for this happier life, Kropotkin pinned his hopes on the education of the young. To achieve an integrated society he called for an "integral education" that would cultivate both mental and manual skills. Due emphasis was to be placed on the humanities and on the basic principles of mathematics and science, but, instead of being taught from books alone, children were to receive an active outdoor education and learn by doing and observing at first hand, a recommendation that has been widely endorsed by modern educational theorists. Drawing on his own experience of prison life, Kropotkin also advocated a thorough modification of the penal system. Prisons, he said, were "schools of crime" that, far from reforming the offender, subjected him to brutalizing punishments and hardened him in his criminal ways. In the future anarchist world, founded on mutual aid, antisocial behaviour would be dealt with not by laws and prisons but by human understanding and the moral pressure of the community.

Kropotkin combined the qualities of a scientist and moralist with those of a revolutionary organizer and propagandist. For all his mild benevolence, he condoned the use of violence in the struggle for freedom and equality, and during his early years as an anarchist militant he was among the most vigorous exponents of "propaganda by the deed"—acts of insurrection to supplement oral and written propaganda to awaken the rebellious instincts of the people. He was the principal founder of both the English and Russian anarchist movements and exerted a strong influence on the movements in France, Belgium, and Switzerland. But he alienated many of his comrades by supporting the Allied powers during World War I. His action, though prompted by the fear that German authoritarianism might prove fatal to social progress, violated the antimilitarist tradition and touched off bitter polemics that nearly destroyed the movement for which he had laboured nearly half a century.

Return to Russia

Events, however, took a brighter turn with the outbreak of the Russian Revolution. Kropotkin, now in his 75th year, hastened to return to his homeland. When he arrived in Petrograd (now Leningrad) in June 1917, after 40 years in exile, he was greeted warmly and was offered the post of minister of education in the provisional government, which he brusquely declined. Yet his hopes for a libertarian future were never brighter, for 1917 saw the spontaneous appearance of communes and soviets—soldiers' and workers' councils—that he felt might form the basis of a stateless society. With the Bolshevik seizure of power, however, his enthusiasm turned to bitter disappointment. "This buries the revolution," he remarked to a friend. The Bolsheviks, he said, have shown how the revolution was *not* to be made—that is, by authoritarian rather than libertarian methods. His last years were devoted chiefly to a history of ethics, which he was never to finish. He died at the village of Dmitrov near Moscow on February 8, 1921. His funeral, attended by tens of thousands of admirers, was the last occasion when the black flag of anarchism was paraded through the Russian capital.

Kropotkin's life exemplified the high ethical standard and the combination of mental with manual work that he preached throughout his writings. He displayed none of the egotism, duplicity, or lust for power that marred the image of so many other revolutionaries. Because of this he was admired not only by his own comrades but by many for whom the label of anarchist meant little more than the dagger and the bomb. The French writer Romain Rolland said that Kropotkin lived what Tolstoy only advocated, and Oscar Wilde called him one of the two really happy men he had known.

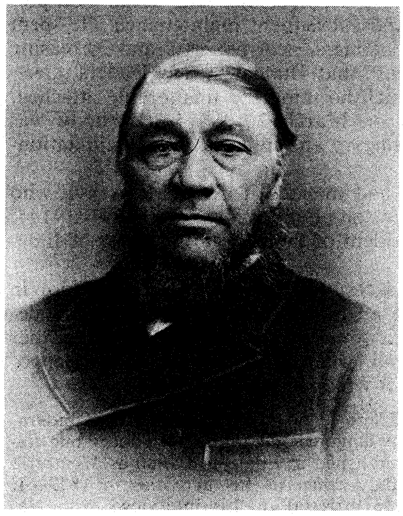
BIBLIOGRAPHY. For Kropotkin's life, see his classic autobiography, ***Memoirs of a Revolutionist*** (1899). The best biography in English is ***The Anarchist Prince*** by GEORGE WOODCOCK and IVAN AVAKUMOVIC (1950). For Kropotkin's place in the Russian anarchist movement, see PAUL AVRICH, ***The Russian Anarchists (1967).***

(P.A.)

Kruger, Paul

Farmer, soldier, and statesman, Stephanus Johannes Paulus (Paul) Kruger, during the last quarter of the 19th century, made his mark in South African history as the builder of the Afrikaner nation. He became famous as the defender of the independence of the Transvaal, or South African Republic, during the period when British imperialism in Africa was at its zenith.

Culver Pictures



Kruger.

Kruger was born in the Cradock district of the British Cape Colony on October 10, 1825. His parents were respectable farmers of Dutch descent on the northern outskirts of the colony. He had little formal education but was able to express himself clearly in writing. Of more importance was the religious instruction he received from his parents according to the strict tenets of Dutch Calvinism. When he was ten, his family took part in the general emigration of frontier farmers who sought to found an independent political existence in the northern interior. As a young boy, he was strongly influenced by the stirring events of the period when the emigrants had to struggle against the warlike tribes surrounding them and to establish an orderly government of their own.

Early political career

While still in his teens, Kruger played a part in public life as a local field-cornet, a post in which civil and military duties were combined. In January 1852 he was present when the Transvaal leader, Andries Pretorius, concluded the Sand River Convention with representatives of Great Britain, by which the independence of the Afrikaners (Boers) north of the Vaal River was recognized. He took part in 1855–56 as member of a commission that drew up the constitution of the new republic. During the civil disturbances of 1861–64, he played a prominent part as commandant general in unifying and pacifying the country in support of constitutional authority.

With the British annexation of the Transvaal in 1877, he became the recognized champion of his people in the struggle to regain independence. With that purpose in mind, he visited England in 1877 and 1878, and, when he failed to persuade the Disraeli government to undo the annexation, he helped organize a movement of passive resistance to British administration in the Transvaal. In 1880 he pinned his hopes to the promises of Gladstone, the Liberal leader. Disappointed when the new Liberal government failed to live up to his expectations, he succeeded in gaining the sympathy and political support of the Cape Colony against the British attempt to force South Africa, including the Transvaal, into a general federation. In December he led his people into active opposition, and, after a series of military victories that culminated in that of Majuba Hill (February 27, 1881), with great diplomatic skill, he succeeded in negotiating peace based on a limited independence. In 1883 he was elected president of the restored republic, and he held that office until 1902, when the Boers at last submitted to British authority.

Meanwhile, in 1883 he again visited England and, after protracted negotiations, concluded a new convention (that of London, February 27, 1884) with Britain, which rectified the western border and removed any reference to

British suzerainty over the Transvaal. On his return he found his republic embroiled with the Cape colonial authorities over control of the area along the western border, which was considered by Cecil Rhodes, the Cape statesman, to be the "Suez Canal" to the territory north of the Limpopo. In 1885 Kruger was forced to accede to British demands to withdraw from the area in question and to agree to a British protectorate over Bechuanaland. At the same time, the way was opened for the future expansion of the Cape Colony to the north.

Gold rush in the Transvaal

Kruger's greatest problem began in 1886 with the discovery of gold in the Witwatersrand area, where a new metropolis, Johannesburg, arose, some 40 miles south of the tiny republican capital, Pretoria. Large numbers of "outlanders" flocked to the Transvaal and established a cosmopolitan, mainly English community in the midst of a rural Boer society. Kruger saw this as a threat to the separate national identity of his people, "God's people" as he called them, and in 1890 he severely restricted the franchise to men resident at least 14 years. At the same time, he called into being a separate Volksraad (legislative body), in order to represent mining interests, but this did not satisfy the demands of the outlanders. The mining magnates of Johannesburg criticized Kruger's economic and railway policy, which was aimed at promoting the independence of the Transvaal but which resulted at the same time in raising the cost of production of gold. They complained of high railway tariffs, which Kruger's concessionaires, The Netherlands South Africa Railway Company, imposed in order to protect their railroad linking Johannesburg with Delagoa Bay. For political reasons, Kruger had to support this railway against the cutthroat competition of the Cape railways, which he was unable to exploit to his country's advantage.

Rhodes, the Cape premier, who had extensive gold interests and much political influence, hoped to achieve a united British South Africa. He supported the Rand capitalists and the outlander movement against Kruger's regime. When he failed to persuade Kruger to join a South African customs union, he decided to bring matters to a head. Kruger, after all, was the one obstacle that prevented Rhodes from realizing his dream of empire. By 1895 Kruger was aware that trouble was brewing in Johannesburg and that, behind the scenes of the internal conflict within the Transvaal, a larger issue was at stake, that of British supremacy as against republican independence. He felt that the matter of extension of the franchise to the newcomers was merely being used as a cat's-paw to further the schemes of Rhodes.

Prelude to the Boer War

Ever since 1890 he also had to contend with growing opposition of a section of his own people; but when Rhodes, with the full knowledge of Joseph Chamberlain, the British colonial secretary, sponsored an ill-fated raid, the Jameson Raid against the republic at the end of 1895, Kruger handled the affair so successfully that his prestige soared again. In the presidential election in May 1898, he received almost unanimous support. While Rhodes was forced into the background, British imperial interests now came to the front. The Colonial Secretary took up the cudgels on behalf of the outlanders and, in 1897, sent Sir Alfred Milner to South Africa as governor of the Cape Colony and high commissioner. Supported by Chamberlain, Milner began to force the issue and demanded that the residential qualification for voters in the Transvaal should be lowered to five years. In May 1899 a conference took place in Bloemfontein, the Free State capital, between Kruger and Milner. Although no agreement was reached, Kruger decided on a seven-year residential qualification. Milner refused the offer, tension increased, and Britain prepared an ultimatum. Both sides prepared for war, which was precipitated by Kruger when, on October 9, 1899, he presented his own ultimatum, demanding the withdrawal of British troops from the border.

War broke out two days later, and, notwithstanding initial Boer successes, British invading armies occupied the two Boer capitals. Kruger was forced to retreat with the last Boer army along the Delagoa Bay railway. Being too old to keep up with the ensuing guerrilla struggle, he was delegated to Europe, where he lived in Holland to the end of the war in May 1902. He died at Clarens, in Switzerland, on July 14, 1904, and his body found a temporary resting place at The Hague. He was finally buried at Pretoria on December 16, 1904.

BIBLIOGRAPHY. D.W. KRUGER, ***Paul Kruger,*** 2 vol. (1961–63), is the most recent biography, based on original research but popularly written. J.P. LA GRANGE LOMBARD, ***Paul Kruger, die Volksman*** (1925), highlights certain aspects of Kruger's life only; while M. NATHAN, *Paul Kruger,* 4th ed. (1944), has written a popular account based on published sources. C.T. GORDON, ***The Growth of Boer Opposition to Kruger, 1890–95*** (1970), is a recent scholarly work, based on original research and stressing important aspects of Kruger's policy. J.S. MARAIS, ***The Fall of Kruger's Republic*** (1961), is a most important contribution to knowledge of British policy vis-a-vis Kruger and his republic up to the eve of the Anglo-Boer War, 1899–1902. ***Lebenserinnerungen des Prasidenten Paul Kruger*** (1902; Eng. trans., ***The Memoirs of Paul Kruger,*** 1902), is useful for his early life, but of less value later and, notwithstanding the title, for the most part not memoirs but ordinary history written by his secretaries.

(D.W.K.)

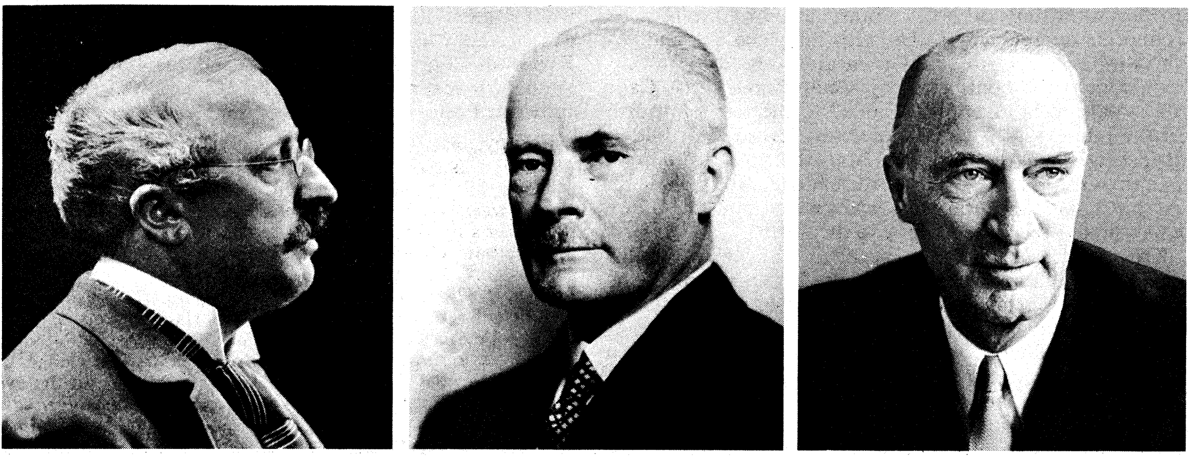
Krupp Family

From 1587 to 1968 members of the Krupp dynasty, the world's largest manufacturers of munitions, dominated the German city of Essen, the hub of the industrial Ruhr area (Ruhrgebiet), as surely as the rulers of any medieval fief. When the drums of German conquest rolled in 1870, 1914, and 1939, it was Krupp factories that provided first Prussia and then the German Empire with field guns, shells, tanks, battleship armour, and flotillas of submarines—always at immense profit to the House of Krupp. Until 1918 a suite in Villa Hiigel, the 300-room Krupp Castle in Essen, was always reserved for the reigning emperor. Later the dynasty became deeply committed to Adolf Hitler, and Alfried Krupp von Bohlen und Halbach (1907–67), the last of the line, was convicted of major war crimes at Nuremberg. In 1951 he was pardoned, and before his death he had become the most powerful industrialist in Europe.

Arndt Krupp (died 1624), the first member of the family to settle in Essen, arrived before a plague epidemic and bought large tracts of land from natives fleeing from it. Although Arndt and the four generations of Krupps who succeeded him grew wealthy, the family's rise to international significance did not begin until Friedrich Krupp (1787–1826) founded the dynasty's cast-steel factory ***(Gusstahlfabrik)*** in 1811. He never entirely matched the achievement of the British, who had mastered the art of casting steel, but his son Alfred (1812–1887) succeeded brilliantly. Known variously as "Alfred the Great" and "the Cannon King," Alfred perfected his technique by manufacturing rails and seamless-steel railroad wheels. Then he turned to guns.

Archiv fur Kunst und Geschichte

Alfred Krupp, portrait by Julius Grün, *c.* 1880.



(Left) Friedrich Alfred Krupp. (Centre) Gustav Krupp von Bohlen und Halbach. (Right) Alfried Krupp von Bohlen und Halbach, c. 1957.
Archiv fur Kunst und Geschichte

"The Cannon King"

Alfred Krupp. An eccentric genius and an insomniac, Alfred lay awake nights scribbling letters and plans for new instruments of death. If not completely paranoid, he was certainly suspicious of everyone around him. Although he worked at forges, he was terrified of fire and built his castle of glass and stone, without a stick of wood. At one point he tried to persuade his workmen to wear uniforms—he himself would have been arrayed as a field marshal. Because he believed that the odour of horse manure was inspiring, he built his study over his stable, with ducts to waft the scent upward. Yet his workmen were fanatically loyal to him. He was the first industrialist to introduce sick pay, a free hospital for his men and their families, pensions, and homes for retired workers.

In 1851 his cast-steel cannon was the sensation of London's Crystal Palace Exhibition, although his name did not become a household word until the Franco-Prussian War of 1870–71. In that conflict, the Prussians largely owed their triumph to Krupp's field guns, whose accuracy and range easily outperformed Napoleon III's bronze artillery. Despite Krupp's production of many peaceful items, his weapons had caught the public fancy. Almost overnight Krupp guns became a status symbol for 19th-century nations.

Alfred Krupp was in many ways the founder of modern warfare. At the time of his death he had armed 46 nations. As much as any other single individual, he had set the stage for the great holocaust that would begin in 1914, and, in recognition of his achievements in war weapons, the grateful governments of his day had awarded him 44 military medals, stars, and crosses, which included multiple honours from Spain, Belgium, Sweden, Romania, Austria, Japan, Russia, Turkey, and Brazil.

The "defensive and offensive weapons seesaw"

Friedrich Krupp. Whereas Alfred had been a blunt man, his son, Friedrich Alfred Krupp (1854–1902)—whom everyone called "Fritz"—was sly, and serpentine. Nevertheless, he shared his father's uncanny business sense and remarkable gifts for management. Probably the most profitable of his strategems was what the amused Emperor William II called the "Schutz- und *Trutzwaffen Schaukeln*," the "defensive and offensive weapons seesaw." It worked this way. Having perfected nickel-steel armour, Fritz advertised it in every chancellery. Armies and navies invested in it. Then he unveiled chrome-steel shells that would pierce the nickel steel; armies and navies invested again. Next he appeared with a high-carbon armour plate that would resist the new shells. But just when every general and admiral thought he had equipped his forces with invincible shields, Fritz once more produced good news for advocates of offensive warfare. It turned out that the improved plate could be pierced by "capped shot," with very expensive explosive noses. The governments of the world kept digging deep into their exchequers; altogether, 30 of them were caught in Fritz's arms game.

In seven years Fritz's fortune tripled, but tragedy stalked him. Although the father of two daughters, he had become an overt homosexual. Incontrovertible evidence of this, including photographs, fell into the hands of the Italian police and was soon published by German newspapers. Distraught and disgraced, Fritz killed himself in 1902, leaving as his heiress his elder daughter Bertha Krupp (1886-1957), then a teen-ager. It was unthinkable for the Reich's most martial industry to be run by a woman, so the Emperor himself conducted a search for an acceptable husband for Bertha.

Gustav Krupp. He chose Gustav von Bohlen und Halbach (1870–1950), a stiff Prussian diplomat. Efficiency was Gustav's religion; one of his oddest hobbies was reading train timetables, looking for typographical errors. With his domed forehead, set mouth, and quick mechanical gestures, he was a parody of Prussian rigidity. The Emperor himself gave the bride away, and, as a surprise for the newlyweds, he had the groom's name changed to Gustav Krupp von Bohlen und Halbach. The couple was thus granted the privilege of passing on not only the family fortune but also the Krupp name to its eldest son.

World War I armaments

In World War I Gustav Krupp, as he was known, provided many memorable contributions to Germany's arsenal. One, named in honour of his wife, was the 98-ton Big Bertha howitzer that shelled Liège and Verdun. Others included the great cannon that bombarded Paris from a range of 82 miles and Germany's submarines, which were built at the family's Kiel shipyards. Because the Germans lost, the war was, on the whole, bad business for Krupp, but not a total loss. Before the war, in 1902, Vickers, Ltd., a British manufacturer of artillery shells, had leased a Krupp fuze patent. Under the agreement, every Vickers shell was stamped KPz (Krupp patent fuze), eventually to be redeemed at 1*s*. 3d. per shell. After the war, Vickers paid off in a settlement based on German artillery casualties, which placed Krupp in the awkward position of having profited from Germany's war dead.

With this money, and with subsidies from the government of the Weimar Republic, Gustav began the secret rearming of Germany within a year of the Armistice. In his words, he was determined that Krupp should be ready "again to work for the German armed forces at the appointed hour without loss of time or experience." Submarine pens were furtively built in Holland; new cannon were covertly perfected in Sweden. Krupp helped finance the Nazi "terror election" of 1933, tightening Hitler's grip on the reins of government, and, as president of the Reichsverband der Deutschen Industrie—Germany's equivalent of the U.S. Chamber of Commerce--expelled all Jewish industrialists and became one of the country's most ardent Nazis. Meanwhile, his eldest son, Alfried—who had been a sponsoring member of the Nazi Schutzstaffel (SS) since 1931—was improving the anti-aircraft, anti-tank, anti-personnel 88-mm gun, which would first be used in the Spanish Civil War and which would, a decade later, become the most famous artillery piece of World War II.

The Lex Krupp

Alfried Krupp. Shortly after the outbreak of the war it became evident that Gustav was drifting into senility. Alfried assumed his duties, and in 1943 Hitler issued an unprecedented decree, the Lex Krupp, which, abolishing in this one case the laws of inheritance, preserved the firm as a family property. Alfried now assumed the name of Krupp and became the sole owner of his mother's vast holdings.

Even before 1939, the extent of the family's wealth had been staggering. Within the Reich, Krupp had wholly owned 87 industrial complexes, held a controlling interest in 110 firms, and possessed substantial investments in 142 other German corporations. Abroad, Krupp smokestacks had stained the sky over almost every continental country; the family owned over 50 percent of the stock in 41 foreign plants and large blocks of shares in another 25. There had been thousands of Krupp ore pits and coal mines, a chain of Krupp hotels, a group of Krupp banks, a Krupp cement works, and a score of private estates.

Now Alfried augmented this empire by seizing property in every country conquered by Germany. Already, in 1943, his salesmen were exporting finished machine products from his new Ukranian plants and selling them in Bulgaria, Turkey, and Romania. When Robert Rothschild refused to sign over his French holdings to Alfried, Rothschild was shipped to the Auschwitz concentration camp and gassed. It was incidents of this kind, together with his exploitation of slave labour, that put Alfried in the prisoners' dock at the Nuremberg war-crimes trials after the war.

Indictment of Alfried

At first the victorious allies, under the impression that Gustav had been in charge throughout the war, had indicted him. In fact, it was Alfried who had been the head of the family and the firm during the years when the inmates of 138 concentration camps worked for Krupp; Alfried who had had a fuze factory built inside Auschwitz to take full advantage of prison labour; and Alfried who had had Jewish prisoners at Auschwitz build a howitzer factory in Silesia. One Krupp camp was reserved for Jewish girls, who lived and worked under almost unendurable conditions before being shipped to the Buchenwald Camp to be killed.

The Nuremberg tribunal sentenced Alfried to 12 years in prison and ordered "forfeiture of all your property both real and personal." Seven months after the outbreak of the Korean War (1950–53), however, John J. McCloy, U.S. high commissioner in American-occupied Germany, granted Alfried amnesty and restored all his holdings. As a lawyer McCloy had been troubled by the confiscation, which, as he pointed out, had the effect of disinheriting Alfried's heir or heirs. Operating with tremendous skill and zeal, Alfried quickly restored the family firm to its former supremacy. By the early 1960s he was worth over a thousand million dollars.

End of the dynasty. Then the family suffered two blows from which it would not recover. The firm had extended long-term credit to firms in Eastern Europe, and Alfried's short-term notes balancing that credit became due in the German recession of 1966–67. The Krupp empire could survive only if he relinquished sole control and opened the firm to outside investment—selling stock on the Dusseldorf Exchange. At the same time Alfried's only son, Arndt, named after the family's founder, decided that he did not wish to take over the family business. In exchange for renouncing his succession rights, rights based upon the Lex Krupp of 1943, Arndt received $500,000 a year for the rest of his life. On the morning of July 31, 1967, Alfried was found dead in Essen, and in the following January the firm became a corporation. The dynasty that had ruled for almost four centuries had come to an end.

BIBLIOGRAPHY. WILLIAM MANCHESTER, ***The Arms of Krupp, 1587–1968*** (1968), an illustrated, annotated, and indexed chronology of significant events during the history of the dynasty, and the organization of the family firm prior to its dissolution, with appendixes on Krupp geneology; WILHELM BERDROW (ed.), ***Alfred Krupps Briefe 1826–1887, im Auftrage der Familie und der Firma Krupp*** (1928; Eng. trans., ***The Letters of Alfred Krupp, 1826–1887,*** 1930), an especially valuable selection from Krupp's correspondence; GERT VON KLASS, ***Die Drei Ringe: Lebensgesckichte eines Industrieunternehmens***, 2nd ed. (1953; Eng. trans., ***Krupps: The Story of an Industrial Empire,*** 1954), a presentation of Krupp history as seen from the family's point of view; BERNARD MENNE, ***Krupp: Deutschlands Kanonenkonige*** (1937; Eng. trans., ***Krupp, or the Lords of Essen,*** 1937), a documented, critical account of the dynasty's history; NORBERT MUHLEN, ***The Incredible Krupps*** (1959), a briskly written, somewhat superficial history of the Krupps; ***Trials of War Criminals Before Nuernberg Military Tribunals Under Control Council Law No. 10***, vol. 9, case 10 (1951), the indictment, prosecution, and defense of Alfried Krupp at Nuremberg, including the tribunal's judgment and excerpts from testimony; GORDON YOUNG, ***The Fall and Rise of Alfried Krupp*** (1960), a lively account of Alfried's comeback after his release from prison in 1951.

(W. Man.)

Kublai Khan

Kublai Khan, a grandson and the greatest successor of the Mongol conqueror Genghis Khan, was a brilliant general and a statesman of great imagination who conquered China and became the first emperor of its Yüan, or Mongol, dynasty. He was thus at one and the same time the overlord of all the Mongol dominions—which included areas as diverse as that of the Golden Horde in south Russia, the Īl-Khānate of Persia, and the steppe heartlands where Mongol princes were still living the traditional nomadic life—and the ruler of his own realm of China. To govern China, with its long and individual political and cultural history demanded statecraft of a special order. Kublai's career is interesting above all because of the way in which he interpreted—and finally failed to reconcile—his dual roles. As it turned out, he became a Chinese emperor of traditional type. China absorbed his interests and energies to the exclusion of the Mongol homeland, and for years he was actually engaged in civil war with rival Mongol princes of the steppes. Under him, China, and of course the privileged Mongols, enjoyed a brilliant spell of prosperity; but his politics, pursued with less skill by his successors, isolated the Mongols in China from their environment. With the collapse of the dynasty, the Mongols withdrew to the steppes and never again played any role of more than local importance.

By courtesy of the National Palace Museum, Taipei, Taiwan, Republic of China

Kublai Khan, portrait by an unknown artist. In the National Palace Museum, Taipei, Taiwan.

Kublai's major achievement was to reconcile the Chinese people to rule by a foreign and culturally inferior race. It was not until some 30 years after his death that any serious uprisings against Mongol rule were to occur in China. Yet even a man of his energy, will power, and political insight, ruling with the advantage of absolutism unfettered by the old bureaucratic apparatus of China, could not resolve the contradictions inherent in his situation.

Historical background. The Mongols were a parvenu nomadic power. Before the time of Genghis Khan they had been no more than a group of semibarbaric tribes, more or less unknown to history. They lacked all but the most primitive cultural traditions, and, except for some organized hunting and the management of their herds, they had little experience of economic activity. Until a few years before Kublai's birth, they had been illiterate. They had only the most elementary ideas of statecraft. Concepts such as taxation of urban societies were brought to their attention by their foreign advisers, upon whom they relied heavily. This political incompetence contributed much to the rapid collapse of their empire. With a few outstanding exceptions, such as Kublai himself (whom the Mongols always called Setsen Khan, the Wise Khan), the rulers of the Mongols seem to have looked upon power as a personal, at most a family, possession, to be exploited for immediate gain. Hence, except in areas where, like China, there was a firm native political tradition, they never succeeded in organizing a durable state. In China, too, everything depended ultimately upon the will power and ability of the ruler.

The Mongols had come to power in China, as elsewhere, by sheer force of arms; and with this prestige to back him, relying on his dominant personality, and building on the foundations of the brilliant civilization developed by the preceding Sung dynasty, Kublai for a while could maintain the illusion that Mongol supremacy was firmly based. Indeed, his reign must have appeared to be a period of solid expansion and lasting achievement to his contemporaries, including Marco Polo, the Venetian traveller who became Kublai's agent and whose book is the chief Renaissance source of information on the East. Yet Kublai Khan was faced at the outset of his reign by an insoluble dilemma, which was given vivid expression in a memorial presented to him by one of his Chinese advisers: "I have heard that one can conquer the empire on horseback, but one cannot govern it on horseback." In other words, to administer China the inexperienced Mongols would have to adopt Chinese methods, even live according to a Chinese pattern; and to the extent that they did so, they would be bound to become more and more assimilated and perhaps lose their identity altogether. If on the other hand they worked through Chinese and other agents they would become alienated from the mass of the population, which would reject them. In either case, the Mongols, culturally and numerically inferior and used to a different pattern of life, could not continue for long to rule China as a distinct and privileged caste; and only the brilliance of Kublai's personal achievement obscured this truth.

Dilemma of foreign rule in China

Rise to power. Kublai Khan was born in 1215, the fourth son of Tolui, the youngest of Genghis' four sons by his favourite wife. He began to play an important part in the extension and consolidation of the Mongol empire only in 1251, when he was in his middle 30s. His brother, the emperor Mongke, resolved to complete the conquest of Sung China, which had been planned by Genghis' third son, Ögödei, and also to subdue Persia—a task allotted to Kublai's brother Hülegü. Kublai was invested with full civil and military responsibility for the affairs of China. He appears never to have learned to read or write Chinese, but already he had recognized the superiority of Chinese thought and had gathered around himself a group of trustworthy Chinese advisers, Confucians. His attitude toward government was formed under the influence of these learned Chinese, who convinced him of the necessary interdependence of ruler and ruled and reinforced his innate tendency toward humanity and magnanimity. At home, in the fief allotted to him in the Wei River Valley (in modern Shensi Province), he established a competent administration and a supply base. In the field, he stressed to his generals the precepts of his mentors—the importance and effectiveness of clemency toward the conquered. This was a great advance in civilized behaviour compared to the methods of Genghis Khan and those of Kublai's contemporaries in Central Asia, where the massacre of the population was still the expected sequel to the capture of a city.

Kublai took Sung China in the flank, subjugating the Tai kingdom of Nanchao in present-day Yunnan before handing over command to his general Uriyangqadai. In 1257 Mongke assumed personal charge of the war, but he died in 1259. When Kublai, who with another army was besieging a city, heard that his brother, Arigboge, who had been left in charge of the homeland because he was the youngest, was planning to have himself elected khan, he patched up a truce with Sung. In April 1260 he arrived at his residence of K'ai-p'ing, or Shang-tu (the Xanadu of Samuel Taylor Coleridge's famous poem), in southeastern Mongolia. Here his associates held a *kuriltai*, or "great assembly"; and on May 5 Kublai was unanimously elected khan in succession to Möngke. Ten days later he announced his succession in a proclamation drawn up in Classical Chinese. Because primogeniture was not a recognized principle at the time, Arigboge, with some very powerful supporters, held a *kuriltai* at Karakorum and had himself declared khan, ignoring Kublai's action. In spite of Marco Polo's insistence that Kublai was the lineal and legitimate descendant of Genghis Khan and the rightful sovereign, there have always been doubts about this legitimacy. A legend recorded in Mongol chronicles to the effect that the dying Genghis designated the child Kublai as a future khan seems to have been contrived so as to provide retrospective justification of an act of usurpation.

Family feuds

In 1264 Kublai defeated Arigboge in battle and forced him to submit. He died two years later. But the family feud, of which this was one manifestation, continued throughout Kublai's reign. Against him were ranged those who resented the abandonment of the old ways of the steppe and the adoption of an alien, China-centred culture. The split was all the deeper because the leader of the opposition was Kaidu, who, as a grandson of Bgodei, who had been designated personally by Genghis as his successor, represented the cause of legitimacy. The throne had passed from the line of Ögödei to that of his brother Tolui in 1250 as a result of a coup d'etat. Kaidu never relaxed his hostility toward Kublai and remained master of Mongolia proper and Turkistan until his death in 1301.

The war with Kaidu showed how decisively Kublai had identified himself with the Chinese world and turned against the world of the nomads. Genghis had been strong and ruthless enough to compel the Mongols, always inclined to family feuds, to serve his cause; but Kublai, powerful though he was, could no longer control the steppe aristocracy effectively.

Unification of China. Kublai's achievement was to reestablish the unity of China, which had been divided since the end of the T'ang dynasty. This achievement was that much greater because he was a barbarian, nomadic conqueror. Even in Chinese official historiography the Mongol Kublai is treated with respect. As early as 1260 he instituted a reign period, in the Chinese manner, to date his reign; and in 1271, eight years before the disintegration of the Sung, he proclaimed his own dynasty under the title of Ta Yuan, or Great Origin. He never resided at Karakorum, Ögödei's short-lived capital in north Mongolia, but set up his own capital at what is now Peking, a city known in his time as Ta-tu, the Great Capital.

The final conquest of Sung China took several years. Kublai might well have been content to rule the North and to leave the Sung dynasty nominally in control of South China, but the detention and ill treatment of envoys he had sent convinced him that the declining régime in the south must be dealt with decisively. Military operations opened once again in 1267. The Sung emperor was apparently badly served by his last ministers, who are said to have kept him misinformed of the true situation, whereas many Sung commanders went over voluntarily to the Mongols. In 1276 Kublai's general Bayan captured the child emperor of the day; but loyalists in the south delayed the inevitable end until 1279.

Aims to make China centre of the world

With all China in Mongol hands, the Mongol conquests in the south and east had reached their effective limit; but Kublai, seeking to restore China's prestige, engaged in a series of costly and troublesome wars that brought little return. At various times tribute was demanded of the peripheral kingdoms: from Burma, from Annam and Champa in Indochina, from Java, and from Japan. The Mongol armies suffered some disastrous defeats in these campaigns. In particular, invasion fleets sent to Japan in

1274 and 1281 were virtually annihilated, though their loss was due as much to storms as to Japanese resistance. Kublai was never entirely discouraged by the indifferent results of these colonial wars nor by their expense, and they were brought to an end only under his successor. Marco Polo suggests that Kublai wished to annex Japan simply because he was excited by reports of its great wealth. It seems, however, that his colonial wars were fought mainly with a political objective—to establish China once more as the centre of the world.

"Four nationalities policy"

Social and administrative policy. By themselves the Mongols were incapable of ruling China, and, though at the lower levels they made use of Chinese civil servants, posts of importance were allotted to foreigners. Of these Marco Polo is a familiar example. Kublai instituted a "nationalities policy" under which the population of China was divided into four categories. At the top were the Mongols, forming a privileged, military caste of a few hundred thousands, exempt from taxation, and living at the expense of the Chinese peasantry who worked the great estates allocated for their upkeep. The foreign auxiliaries of the Mongols, natives for the most part of Central Asia, formed the second group, the *se-mu jen,* or persons with special status. This class furnished the higher officialdom; and its members, with their worldwide contacts and their privileged status, also formed a new breed of merchants and speculators. Like the Mongols, they were exempt from taxation and enjoyed preferential use of the official postroads and services. The bulk of the population belonged to the third and fourth classes, the Han-jen, or northern Chinese, and the *man-tzu,* or southern barbarians, those people who lived in what had been Sung China. The expenses of state and the support of the privileged bore heavily on these two classes. Kublai's continuing wars meant a heavy and useless burden, as did his showy and extravagant building operations at Ta-tu. Peasants were brought in as labourers, to the neglect of their farms. Food supplies in the north were inadequate for the new labour force and the unproductive Mongols; and large quantities had to be brought by sea, and, when the sea routes proved insecure, along the Grand Canal. The repair and extension of this canal also demanded the diversion of much labour. Kublai, in common with other Mongol rulers, was much preoccupied with religion. His reign was a time of toleration for rival religions and of economic privilege for the favoured religions. Clerics and their communities were exempted from taxation, and Buddhist temples especially were granted generous donations of land and of peasants for their upkeep. The arrogance of the many Tibetan lamas who enjoyed a special status in China was particularly detested.

Such a discriminatory social policy was eventually bound to arouse strong resentment. Moreover, it was only on the surface that Kublai's China, with its intense commercial activity, was economically strong and wealthy. Trade was mainly carried on in the interests of a privileged, foreign merchant class, not those of the community at large. The common people of China were becoming progressively poorer. The old examination system, which admitted to the civil service only men with a proper knowledge of Confucian philosophy, had lapsed; and customary restraints upon absolutism and arbitrary rule, such as would have been imposed by the censorate (a body that scrutinized the conduct of officials) and a professional public service, were lacking. The Chinese literati were excluded from public office and responsibility. As a result, adventurers could attain high positions; and even an emperor of Kublai's unique ability remained for years on end in ignorance of, and unable to check, the depredations of his dishonest foreign financial advisers. The extravagant policies that Kublai had countenanced and the financial ineptitude of later Mongol emperors, provoked, in the 14th century, the economically motivated uprisings that brought the dynasty down.

Kublai is celebrated, mainly because of Marco Polo's account, for his use of paper money. Paper money had, however, been in use in China under the Sung; and Kublai's innovation was merely to make it the sole medium of exchange. Toward the end of the dynasty, an incapable financial administration stimulated inflation by the over-issue of paper money; but in Kublai's time the use of banknotes was essential. The supply of copper was too small to form a metal currency in a period of expanding trade, and in any case large quantities were diverted to the temples to be made into statues and other cult objects.

Assessment. Though celebrated above all as a Chinese emperor, Kublai also helped to form the political traditions of his own Mongol people. To him and to his adviser, the Tibetan grand lama 'Phags-pa, is attributed the development of the political theory known as the "dual principle"; that is, the parity of power and dignity of church and state in political affairs. This theory was turned to practical account on more than one occasion in the subsequent history of Mongolia and, for example, underlay the constitution of the theocratic monarchy proclaimed in 1911, when Mongolia recovered its independence from China.

Kublai's character is difficult to assess. The only personal account of him is by Marco Polo, and this is more of a panegyric than a sober appraisal. Marco presents Kublai as the ideal of a universal sovereign. Yet he does not overlook his human weaknesses, above all, an excessive indulgence in feasting and hunting, a complicated and expensive sexual life. a failure always to exercise proper supervision over his subordinates, and occasional outbursts of cruelty.

BIBLIOGRAPHY. There is no good scholarly biography of Kublai Khan, in spite of his significance as a world figure. The best accounts of his life may be found in standard histories of China and of the Mongols, especially in OTTO FRANKE, *Geschichte des chinesischen Reiches,* vol. 4 (1948). Franke's account, though critical, tends to favour his subject; a contrasting though equally well-founded, viewpoint is put forward in WOLFRAM EBERHARD, *Chinas Geschichte* (1948; Eng. trans., *A History of China,* 1950). The relevant chapters in vol. 2 of M. D'OHSSON, *Histoire des Mongols* (1834, reprinted 1940), are still unsurpassed in spite of the lapse of time. RENE GROUSSET, *L'empire des steppes: Attila, Gengis-Khan, Tamerlan* (1939; Eng. trans., *The Empire of the Steppes: A History of Central Asia,* 1970), gives adequate coverage. A concise, but very competent account of Kublai's reign may be found in J.J. SAUNDERS, *The History of the Mongol Conquests* (1971), which has an annotated bibliography. Marco Polo's book, in any easily available edition, is indispensable reading. LEONARDO OLSCHKI, *L'Asia di Marco Polo* (1957; rev. Eng. trans., 1960), provides the most readable commentary upon it, with rich bibliographical references.

(C.R.B.)

Kunlun Mountains

The Kunlun Mountains (K'un-lun Shan), which extend approximately 1,675 miles from the Pamirs in the Soviet Union on the west to the Sino-Tibetan ranges on the east, constitute the longest mountain system in Asia, uniting dozens of ranges that are among the highest on Earth. Located in the People's Republic of China, within the autonomous regions of Sinkiang Uighur and Tibet and the province of Tsinghai, the Kunlun Mountains form the northern wing of the geologically uplifted region known as High Asia—the highest such region in the world.

The position of the Kunlun Mountains, between Tibet and the northern plains of Central Asia, determines the sharply asymmetrical structure of the system. Although the average elevation of the watershed ridges in the southern ranges is about 21,325 feet, when viewed from the south the ranges rise only from about 3,300 to 4,900 feet above the Plateau of Tibet, which itself has an average height of between about 14,500 and 16,500 feet. As seen from the Tarim and Ala Shan plains to the north, however, which have an average altitude of only about 2,600 to 3,900 feet, the watershed ridges of the northern ranges (the average elevation of which is about 19,650 feet) create the impression of gigantic mountains that tower up to 14,765 feet over the surrounding plains.

General alignment of the mountains

The natural environment. ***Relief and topography.*** General alignment is nearly latitudinal (from east to west), but individual segments change direction significantly, following the outlines of the Tarim Basin, the Ala

Shan Desert, and the Tibetan massifs (mountainous masses). The most significant such deviations occur in the Khotan–Keriya (Ho-t'ien–Yii-t'ien) sector (about 81° E), which faces the Lop Nor plain (about 90" E). Here, the strike of the mountains shifts abruptly from northwest–southeast to southwest–northeast, before resuming its northwest–southeast alig [illegible]

Almost the full length of the Kunlun Mou [illegible] consists of parallel chains of–ranges, the tallest of which are those closest to Tibet, separated by vast depressions and narrow valleys. Because they are higher in relation to the surrounding plains, the northern slopes (the Tarim and Ala Shan) are steeper and more complexly dissected, while the southern slopes (those facing Tibet) are shorter, sometimes taking the form of ledges that are only weakly dissected. Whereas the heights of the major Kunlun peaks are several thousand feet shorter than the famous Himalayan peaks, the average height of the Kunlun and Himalayan ridges are almost equal. The elevations of the Kunlun mountain passes (from 18,700 to 20,340 feet), however, surpass those of the deeply eroded Himalayas.

The western Kunluns. The Kunlun system is not uniform in structure, being subdivided into two unequal sections—the smaller western and the principal eastern part. The western Kunluns (between the Pamirs and the alignment change in the Khotan–Keriya region) form three parallel chains of ranges crowded closely together. The Sha-li-k'o-erh, T'a-shih-k'u-erh-kan (Tashkurgantag), Agyl, and Sugettag ranges form the southern chain, which adjoins the Karakoram Range (*q.v.*); the inner and highest chain is formed by the Mu-ssu-t'a-ko-a-t'e, Tokhtakoram, and Karangutag ranges; the Tiznaf and Sandzhutag ranges make up the northern chain. Because the mountain chains of the western Kunluns are divided only by narrow intermontane depressions, the width of this part of the system usually does not exceed about 60 miles.

The highest groups of peaks in the western Kunluns are found on its flanks—in the Gissar sector, the 24,865-foot (7,579-metre) Kung-ko-erh and the 24,462-foot (7,456-metre) Mu-ssu-t'a-ko-a-t'e massifs; in the Khotan sector, the 23,008-foot (7,013-metre) Karangutag and the 23,891-foot (7,282-metre) Mu-ssu massifs. In the intervening So-ch'e–Yarkand sector, however, all the ranges are lower; even the main peaks rarely reach 19,650 feet. Where the alignment changes in the Khotan–Kariya region, the northern chain of the Kunluns is interrupted, and ranges of the high inner chain border the Tarim Basin, on the side of which they are abruptly bounded by a gigantic ledge.

The eastern Kunluns. The eastern Kunluns are characterized by a complex branching of mountain chains that pass around broad intermontane valleys. The unitary direction characteristic of the western Kunlun ranges is lost here; individual ridges are often situated at an angle to one another. Moreover, the width of the mountain system increases sharply, reaching about 375 miles in places.

The Russian, the A-erh-chin Shan-mo (Astin Tagh), and the group of Nan Shan ranges form the northern chain of the eastern Kunluns, which many consider an independent mountain system. Near the Lop Nor plain these ranges describe an arclike curve in which their alignment changes from southwest–northeast to northwest–southeast. The western branch of the arc (with northeastern alignment) is made up of two ranges with a width of from 19 to 25 miles. The Russian Range, whose crest rises to 21,738 feet (6,626 metres), is the only one of noteworthy height within this branch; A-erh-chin Shan-mo (Astin Tagh) is as much as 11,500 to 13,000 lower, especially in the segment contiguous with the Tsaidam Basin.

The Nan Shan ranges

The eastern branch of the arc, which is formed by the system of Nan Shan ranges, is separated from the principal mountain chains of the eastern Kunluns by the vast Tsaidam depression. The five to seven ranges that constitute the Nan Shan system, which has an overall width of up to 186 miles, include Ch'i-lien Shan-mo (Richthofen), T'o-lai Shan (Khrebet Tkholo-shan'), Zyussa, Wu-lan-ta-pan Shan (Humboldt), Ta-k'en-ta-fan Shan (Ritter), and southern Koko Nor. While the height of their crests varies from about 16,700 to 20,700 feet, the elevation of the longitudinal (north–south) intermontane valley floors between them varies from 10,500 to 11,800 feet. The Zyussa, which rises to altitudes of 20,820 feet (6,346 metres), is the highest of these ranges; but the most developed, as a mountain range, is the Ch'i-lien Shan-mo.

The principal chains of the eastern Kunluns are located between the Tsaidam Basin and the Tibetan uplands. The northernmost of these, the Ch'i-man–Burkhanbudda chain, is the lowest; the level of its series of summits is from about 17,400 to 18,000 feet. It rises massively above the Tsaidam Basin but stands out in only mild relief when seen from the Tibetan side. Forming an extension of the inner chain of the western Kunlun ranges is the Przhevalsky Range (Arkatag), the highest and longest in the eastern Kunluns and the principal structural pivot of this part of the mountain system. It contains the tallest peaks of the entire Kunlun Mountains, the 25,341-foot (7,724-metre) Wu-lu-k'o-mu-shih Ling and the 25,328-foot (7,720-metre) T'ieh-k'o-li-k'o Shan (also called the Chong-Karlyk-Tag [Great Snow Range] or Shapka Monomakha [Hat of Monomakh]). The crest of the range is snow covered and steep walled, rising vertically almost 3,300 feet; its mountain passes (K'a-la-mu-lun Shan-k'ou, Rekviem, and others) have elevations of close to 18,000 feet. The southern chain, formed by the K'u-k'u-shih-li (K'o-k'o-hsi-li) and Pa-yen-k'a-la Shan (Bayan' Kara-shan') ranges, is about 3,300 feet lower than the Arkatag.

The high plains separating the inner mountain chains of the eastern Kunluns from those to the north and south lie at the same elevation as the plateaus in the foothills of the Tibetan uplands and have a similar landscape. The most extensive of these is the Kul'tala Plain, which lies between the Przhevalsky Range and the Ch'i-man chain and which is up to 59 miles wide. Covering its surface next to the mountains are fields of broken stone; in areas of lower elevation there are salt marshes, as well as the A-ya-ko-k'u-mu Hu (Aiag Kum Kul') and A-tz'u-k'o Hu (Achchik Köl) lakes.

Geological history. The principal folded structures and granitic rocks of the Kunluns are of Hercynian age, a period that lasted from about 230,000,000 to 250,000,000 years ago, during which there was much mountain building in the Eastern Hemisphere. The inner depressions of the Kunluns, however, are relatively recent structures in their entirety, being formed by deposits that are no more than 26,000,000 years old; only the largest of them, the Tsaidam depression, contains a thick sedimentary cover of which Jurassic deposits (136,000,000 to 190,000,000 years of age) represent the oldest strata. The Kunluns also represent a region of very recent movements of the Earth's crust, and great seismic (earthquake) activity.

Rivers. The Kunluns form a part of that region in Central Asia in which there is only internal drainage, associated mainly with the Tarim, Ala Shan, Tsaidam, and Kul'talin basins. Only the most easterly spurs of the mountain system, where the sources of the Huang (Yellow) and Yangtze rivers are located, have drainage systems that empty into the ocean.

Major rivers

Two groups of rivers compose the river network of the Kunluns: the large streams that rise in the Karakoram Range and in northern Tibet, cutting through the entire chain of Kunlun ranges by way of gorges, and the small streams that drain the slopes of the peripheral ranges. The major rivers—the Gez, Yarkand, K'a-la-k'a-shih (Kara Kash), Yu-lung-k'a-shih (Yurang Kash), K'o-li-ya (Keriya), Ha-la-mu-lab (Kara Muran), and Ch'e-erh-ch'en (Cherchen)—form lengthy, zigzag valleys. In the broad and open longitudinal valleys between mountain ranges, the rivers flow quietly and calmly for long distances; in the ravines that bisect the ranges, however, short sections of the rivers flow through narrow gorges in violent torrents.

Although they receive some rainwaters, the Kunlun rivers are fed mainly by snows and glaciers. Therefore, the volume of flow varies with the seasons; 60 to 80 percent of it occurs in the summer months, when intensive thawing of snow and ice in the mountains is combined with maximum precipitation. In winter, the discharge of the rivers is extremely insignificant; in spring and autumn it is somewhat greater.

Glaciation. In spite of the great elevation, there is little glaciation in the Kunluns because of the extreme dryness of the climate; external snows persist only along the deep crevices of the highest peaks.

The main centres of glaciation, in which are found dozens of glaciers that are usually not more than six miles long, are the Kung-ko-erh, Mu-ssu-t'a-ko-a-t', Wu-lu-k'o-mu-shih, and Tyumenlik massifs, where elevations approximate about 23,000 feet. All the glaciers are notable for their unusual steepness, dropping 20 to 30 feet in every 100 feet of length.

Soils. Because the surface layer of the Kunluns does not receive moisture during a large part of the year, soil formation proceeds at a very slow rate. In general, the soils have little productive capacity, contain many coarse skeletal soil elements, and are almost devoid of humus. Brown soils appear only in the vicinity of the Pamirs and in the eastern Kunluns, where the moisture levels are better and vegetation richer.

Aridity and extremes of temperature

Climate. Located within the arid region of Central Asia, the Kunluns are almost totally isolated from the climatic influence of the Indian and Pacific Ocean monsoons. Instead, they are under the constant influence of the continental air mass, which causes great annual and daily temperature fluctuations. Desert yields to mountain steppes only near the Pamirs (King Ata Tagh) and the Tibetan mountains (Pa-yen-Fa-la Shan [Bayan' Kara-shan'l and the eastern Nan Shans), where the amount of annual precipitation increases to about 18 inches. In these areas, up to 80 percent of the precipitation falls in summer—an indication of its relationship with the monsoons.

In the high-altitude zone, with its extremely sharp daily fluctuations of temperature, weathering from heat and frost reaches great intensity, accounting for the presence of an enormous quantity of loose material. Wind erosion and accumulations of loose debris are very apparent on the Gobi slope of the peripheral ranges. Maximum aridity occurs in the middle segment of the mountain system, between 78° and 93" E; to the west and east, however, the climate is somewhat moderated. Characteristic of the Kunluns as well as of the entire arid region of Central Asia are the winds; the strongest of which occur in autumn, when they often reach gale intensity.

Scanty cloud cover and prolonged sunshine are also characteristic of the Kunluns. Naturally, the amount of warmth in the Kunluns varies with the altitude: in the lower tier of mountains, (those bordering the northern plains), the average temperature in June is 77° to 82" F (25" to 28° C) and not lower than 16" F (—9° C) in January; in the upper tier of mountains and on the border of Tibet, however, the average temperature in July is less than 50° F (10" C) and often falls to —31" F (—35° C) in winter. Summers in the lower belt are long and hot; in the upper belt, they are short and cold.

In the most arid part of the Kunluns, precipitation is less than two inches annually in the foothills and about four to five inches in the high altitudes. The climate of the lower tier is that of the hot, Gobi-type desert; of the upper belt, it is that of the cold, Tibetan-type desert. Associated with these, in turn, are accumulations of loess —fine, wind-blown material—that forms a cover on the dissected mountain relief and is found at altitudes up to about 13,000 feet.

Soil characteristics

Vegetation. Because of the primitiveness of the soil cover, the extreme deficiency of moisture, and, in the high altitudes, the insufficient warmth, the natural conditions of the Kunluns are not very favourable for plant growth. Plants have a stunted appearance and possess a number of distinctive physiological peculiarities. Most often they are perennial, dwarfish semi-shrubs with stiff leaves and deeply penetrating root systems. The number of species of plants is very limited, and the plant cover itself extremely thin.

In the most arid part of the Kunluns, desert-like conditions prevail at high altitudes. Only two types of landscape are to be seen—the hot desert of the lower part of the mountains and the cold desert of the Alpine region. In both zones, completely barren spaces predominate, alternating with small areas that support a thin plant cover.

In the western and eastern extremities of the mountain system, however, the higher moisture levels and richer vegetation are more conducive to the formation of more fertile soil. In these areas, the vertical zonality of the plant cover is more complex and includes zones that do not occur in those parts of the Kunluns that consist entirely of desert. In the middle-altitude region there is a zone of desert-like steppe (a grassy and almost treeless plain) in which there is a somewhat thicker plant cover as one proceeds upwards. Mountain steppes then emerge, as well as scattered forests, small at first but then larger. A well-developed forest zone, interspersed with mountain steppes and meadow sections, appears in the extreme west and the extreme east of the Kunluns.

Animal life. In the totally arid part of the Kunluns, animal life is meagre and has little variety; it becomes somewhat richer and more varied in the approaches to the Pamirs and in the eastern spurs of the system, where mountain steppe and forest vegetation appears. The predominant animal life in these areas consists of hoofed animals and rodents. Especially characteristic of hoofed animals are the mountain sheep (in the west the ***arkhar*** and in the east the ***kukuyaman*** type); the mountain goat; the wild ass; and the wild yak. The rodents are represented by mouse hare, field vole, and, on the meadowland slopes, marmot. Beasts of prey commonly found include wolf, fox, bear, and, in the west, snow leopard.

The human imprint. Most of the Kunluns is unpopulated; only the large river valleys, up to an altitude of about 10,000 feet, contain any inhabitants. Settlements become fewer deep in the mountains and nonexistent in the high-altitude zone and on the Tibetan slope. The Uigurs, the most numerous population group, are concentrated in large settlements in the foothills bordering the Tarim Basin; Tajiks live in the western and Mogols in the eastern mountains. Agriculture and small-scale, nomadic animal husbandry are the basic occupations of the mountain population. The main agricultural crops are wheat and barley; domestic animals that are bred include sheep, goats, and yaks.

Transport. Throughout their extent the Kunluns are almost impassable, because of their deep and narrow ravines, steep slopes, mountain passes of great height, and torrential streams. In addition, there is an absence of fuel facilities for vehicles and of forage for pack animals. Even the caravan trails are rare and difficult, especially in the highly eroded middle belt of the mountains.

Of the three automobile roads through the Kunluns, one proceeds along the southern edge of the Tarim Basin, partially using the ancient Silk Route that until the 16th century connected China with Central and western Asia. The other two roads, both of which lead into Tibet, cut across the Kunluns: the western road passes along the T'i-shih-yuëh-fu (Tiznaf), Yarkand (Raskem), and K'a-la-k'a-shih (Kara Kash) river valleys and the eastern road reaches Tibet by a route that extends from Ko-erh-mu (Golmo; in the Tsaidam Basin) to the upper reaches of the Yangtze River and then runs through the Ch'i-man Shan, the eastern spur of the Przhevalsky (Marco Polo Range) and the K'o-k'o-hsi-li (Kukushili).

Exploitation of mineral resources. Because the interior of the Kunluns has not been studied extensively, information about the mineral resources of this vast and geologically heterogeneous territory is extremely scant. Evidences of gold-bearing deposits have been found in various parts of the mountain system, and gold was extracted until the beginning of the 20th century, principally along the upper course of the Yu-lung-k'a-shih Ho (Yurung Kash). Nephrite (a type of jade) is widely distributed and has been extracted from rock deposits and river alluvium. There are well-known rock deposits of this mineral in the Karangutag Range, while alluvial deposits are found in the river valleys of the K'a-la-k'a-shih and Yu-lung-k'a-shih Ho (Yurung Kash). Small deposits of lead and zinc have been discovered at a number of points in the northern chain of the western Kunluns, while in the

K'a-la-k'a-shih Ho Valley, above Sai-t'un-la (Shakhidul-la), indications of tin-bearing deposits have been found. Deposits of iron and chromitic ores and evidence of copper mineralization have been located in the Nan Shan and A-erh-chin Shan-mo (Astin Tagh) ranges. Oil is recovered in the Tsaidam Basin, and there are extensive coal-bearing deposits in the intermontane depressions of the western and eastern Kunluns. Although coal is mined in many of these areas, it is on a small scale and solely for local needs.

The exploitation of the mineral resources of the Kunluns is hampered by the harsh natural environment, the great height of the mountains, difficulties of access, lack of water supplies, and scarcity of population. (V.M.S.)

Kurosawa Akira

The Japanese film director Kurosawa Akira has been counted among the world's most prominent directors since the 1950s. His pictures have extremely powerful images and sometimes imply a vigorous moral message. With the immense success of his motion picture *Rashomon* in 1950, Kurosawa attracted a widespread international audience to the Japanese film, which previously had enjoyed little attention outside East Asia despite the high quality and vast number of the industry's products. Although other Japanese film makers subsequently acquired substantial followings, Kurosawa's films continue to command the greatest interest in the West. They represent a unique combination of elements of Japanese art—in the subtlety of their feeling and philosophy, the brilliance of their visual composition, and their frequent use of samurai and other historic Japanese themes—with a distinctly Western feeling for action and drama and a frequent use of stories from Western sources, both literary classics and popular thrillers.

© Rene Burri—Magnum

Kurosawa, 1961.

Assistant cinema director and scenario writer

Kurosawa was born in Tokyo March 23, 1910. His father, who had once been an army officer, was a teacher who contributed to the development of gymnastic education in Japan. After leaving secondary school, Kurosawa attended an art school and began painting in the Western style. Although he was awarded important art prizes, he gave up his ambition to become a painter and in 1936 became an assistant director in PCL cinema studio. Until 1943 he worked there mainly as assistant to Yamamoto Kajirō, one of Japan's major directors of World War II films. During this period Kurosawa became known as an excellent scenario writer. Some of his best scenarios were never filmed but only published in journals; yet they were noticed by specialists for their freshness of representation and were awarded prizes.

In 1943 Kurosawa was promoted to director and made his first feature film, *Sanshiro Sugata,* from his own scenario; this story of Japanese judo masters of the 1880s scored a great popular success. In 1944 he made his second film, *The Most Beautiful,* describing an episode of girls at work in an arsenal. Immediately thereafter, he married the actress who had played the leading part in the picture, Yoko Yaguchi; there were two children, a son and a daughter. In August 1945, when Japan offered to surrender in World War II, he was shooting his picture, *They Who Step on the Tiger's Tail,* a parody of a well-known Kabuki drama. The occupation forces, however, prohibited the release of most films dealing with Japan's feudal past, and this outstanding comedy was not distributed till 1952.

Kurosawa's *No Regrets for Our Youth* (1946) portrays the history of Japanese militarism during the period from 1933 through the end of the war in terms of a person executed on suspicion of espionage during the war. Of the many postwar films criticizing Japanese militarism, this was the most successful, both artistically and commercially. It was *Drunken Angel* (1948), however, that made Kurosawa's name famous. This story of a consumptive gangster and a drunken doctor living in the postwar desolation of downtown Tokyo is a melodrama in which desperation and hope, as well as violent action and melancholic atmosphere, are intermingled. The gangster was performed by a new actor, Toshirb Mifune, who became a star through this film and who subsequently appeared in most of Kurosawa's films.

Success of *Rashomon*

Rashomon (1950) was shown at the Venice Film Festival in 1951 and was awarded the Grand Prix. This was the first time a Japanese film had won such high international acclaim, and Japanese films now attracted serious attention all over the world. An adaptation of a short story written in 1915 by Akutagawa Ryūnosuke, the film deals with a samurai, his wife, a bandit, and a woodcutter in the 10th century; a rape and a murder are recollected by the four persons in distinctly different ways. This presentation of the same event as seen by different persons stimulated the imagination of the audience and advanced the idea of cinema as a means of probing a metaphysical problem.

Ikiru (1952) is regarded by many critics as one of the finest works in the history of the cinema. It concerns a petty governmental official who learns he has only half a year until he will die from cancer. He searches for solace in the affection of his family but is betrayed, then seeks enjoyments but becomes disillusioned, and, in the end, is redeemed by using his position to work for the poor. In this film, which abounds in strong moral messages, Kurosawa depicted in an extremely realistic manner the collapse of the family system, as well as the hypocritical aspects of officials in the postwar Japanese society. The picture was an outstanding document of the life and the spiritual situation of Japanese people, who were beginning to recover from the desperation caused by defeat in the war.

Seven Samurai (1954) is the most entertaining of Kurosawa's films and also his greatest commercial success. It depicts a village of peasants and a few leaderless samurai who fight for the village against a band of marauding bandits; although it was inspired by his admiration of Hollywood Western films, it was executed in an entirely Japanese style.

Record of a Living Being (1955) was a deeply honest film portraying the terror of a Japanese dentist of the atomic tests conducted by the United States and the Soviet Union; its pessimistic conclusion, however, made it a commercial failure.

Adaptations of European classics

Kurosawa was also noted for his adaptations of European literary classics into films with Japanese settings. *The Idiot* (1951) is based upon the Dostoyevsky novel of the same title, *The Throne of Blood* (1957) was adapted from Shakespeare's *Macbeth,* and *The Lower Depths* (1957) was from Gorky's drama; each of these films is skillfully Japanized. *The Throne of Blood,* which reflects the style of the sets and acting of the Japanese Nb play and uses not a word of the original, has been called the best film of all the countless cinematized Shakespearean dramas.

Kurosawa's pictures contributed a strong sense of style to the artistic Japanese film, which had been pursuing a naturalistic trend. The violent action of his more commercial works also exerted a powerful influence.

In 1960 he set up Kurosawa Productions, of which he became president, and began to produce his own works. As producer, however, he was continually embarrassed by economic difficulties. Throughout the 1960s, Kurosawa made a number of entertainment films, mainly with samurai as leading characters; Yojirnbo (1961) is a representative work. Red Beard (1965) combines the element of entertainment with a sentimental humanism. In the 1960s, however, Japanese cinema fell into an extreme depression, and Kurosawa's plans, in most cases, were found by film companies to be too expensive. As a result, Kurosawa attempted to work with Hollywood producers, but each of the projects ended in failure. At the Kyoto studio in 1968, for 20th Century Fox, he started shooting Tora, Tora, Tora!, a war spectacle film dealing with the air attack on Pearl Harbor. The work progressed slowly, however, and the producer, fearing an excess in estimated cost, dismissed Kurosawa and replaced him with another director. After a six-year interval, Kurosawa at last managed to present another of his films, Dodesuka*den* (1970); his first work in colour, a fantastic comedy of the poor people living in slums, it recaptured much of the poignancy of his best works.

MAJOR WORKS

Sanshiro *Sugata* (1943); ***Yoidore Tenshi*** (1948; Drunken Angel); ***Rashomon*** (1950); Hakuchi (1951; The Idiot); Ikiru (1952; Living); Shichinin no Samurai (1954; Seven Samurai); *Ikimono* no Kiroky (1955; Record of a Living Being); *Kumonosu-jo* (1957; The Throne of Blood); Donzoko (1957; The Lower Depths); Kakushi *Toride* no *San-akunin* (1958; The Hidden Forfress); Warzti Yafsu Hodo *Yoku* Nemuru (1960; The Bad *Sleep* Well); *Yojimbo* (1961); Tsubaki *Sanjuro* (1962; *Sanjuro*); Tengoku to *Jigoku* (1963; High and Low); Akahige (1965; Red Beard); Dodesukaden (1970).

BIBLIOGRAPHY. TADAO SATO, Kurosawa Akira no sekai (1969), the only book written in Japanese investigating carefully all the works of Kurosawa from the scenarios in his assistant days before he became director to his 1965 work, Red Beard; Complete Works of Akira Kurosawa, 5 vol. (1971), an incomplete compilation of the works of Kurosawa into books with Japanese and English scenarios and stills; DONALD RICHIE, The Films of Akira Kurosawa (1965), an excellent study (in English) written by an acquaintance of Kurosawa who is quite familiar with the language and people of Japan; SACHA EZRATTY, *Kurosawa* (1964), a study (in French) that attempts to place Kurosawa's works in the tradition of Japanese culture.

(T.S.)

Kuwait

Kuwait is an independent Arabian state, lying at the upper corner of the Persian Gulf and having an area of 6,880 square miles (17,818 square kilometres). It is bounded on the west and north by Iraq, on the east by the Persian Gulf, and on the south by Saudi Arabia. Its population in 1972 was estimated at 815,000.

Kuwait is largely a desert, except for al-Jahrah Oasis and a few fertile patches in the southeastern and coastal areas. The largest offshore islands are the uninhabited islands of Būbiyān and al-Warbah. The island of Faylakah, near the entrance of Kuwait Bay, has been populated since prehistoric times. The other islands, Umm an-Naml, Mischan, 'Awhah, al-Kubr, Qārū, and Umm al-Marādim, are small specks of uninviting sand. On some of them, police outposts are stationed; others are occasionally visited by fishermen.

The capital city of Kuwait, a true desert metropolis of 80,000 persons (1970), is located on the southern shore of Kuwait Bay and derives its name from the diminutive of kat, meaning "fort." It is the most important city, even though the suburb of Hawalli is now larger in population, with 107,000 persons (1970). Most of the population of the country is concentrated in the cities. Overall, Kuwait is one of the world's most highly urbanized states.

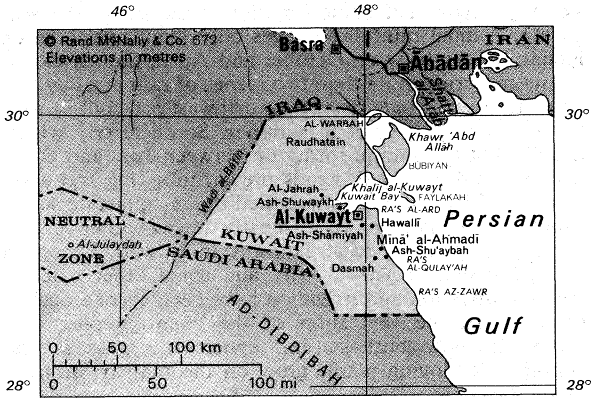
Oil dominates the economy of the country. Kuwait is the fourth largest oil producer in the world after the United States, the Soviet Union, and Venezuela and the second after Venezuela in the quantity of exported oil. It also ranks first among the producing countries in oil reserves.

Kuwait, Area and Population

	area		population	
	sq mi	sq km	1965 census	1970 census*
Governorates (*muḥāfaẓāt*)				
al-Ahmadi	1,983	5,138	59,000	100,000
al-Kuwayt	4,335	11,230	207,000	217,000
Ḥawallī	212	550	201,000	416,000
Total Kuwait	6,880†	17,818†	467,000	733,000

*Preliminary figures. †Total area includes Bubiyan Island (333 sq mi, 863 sq km) and al-Warbah Island (14 sq mi, 37 sq km).
Source: Official government figures.

History. The origin of the city of Kuwait—and of the State of Kuwait—is usually placed at about the beginning of the 18th century, when a number of families of the 'Anizah tribe in the interior of the Arabian Peninsula abandoned nomadism and migrated to the shores of the Persian Gulf. The foundation of the autonomous sheikhdom of Kuwait is dated from 1756, when the settlers decided to appoint a sheikh from the aṣ-Ṣabaḥ family.



KUWAIT

British protectorate

Kuwait first came to the attention of the great powers toward the end of the 19th century, at the time that Germany sought to extend the Berlin-Baghdad Railway to the port of Kuwait. Great Britain, eager to frustrate Germany's designs, and Kuwait's ruling sheikh, desirous of protecting himself against the Ottoman Empire's nominal suzerainty, concluded an agreement in 1899 whereby Britain assumed control of Kuwait's foreign affairs. Following the outbreak of war with the Ottomans in 1914, Britain established a protectorate over Kuwait.

Relations with Najd (later Saudi Arabia) were settled by the Treaty of al-'Uqayr (1922), which involved the creation of the compromise Neutral Zone. The northern frontier with Iraq was agreed upon in 1923.

On June 19, 1961, the British government announced its recognition of the full independence of the sheikhdom, and six days later the Iraqi prime minister claimed Kuwait as an integral part of Iraq, his argument being that Kuwait had been a part of the Ottoman Empire and that ethnically, geographically, and socially Kuwait and Iraq were one country that had been arbitrarily divided by Britain. Threatened with invasion, the ruler of Kuwait appealed to Britain for military aid. In early July, British troops landed in Kuwait, and on July 20 the Arab League admitted the sheikhdom to its membership, thus recognizing its independence and refusing to admit the Iraqi claim. Almost two years later, on May 14, 1963, Kuwait was admitted to the United Nations. Relations with Iraq meanwhile improved, and in October 1963 Iraq recognized Kuwait's independence. Today Kuwait is ruled by a sheikh who currently is both head of state and prime minister; under him is a cabinet of 14 ministers (ten appointed and four elected). Under the constitution

Government of independent Kuwait

framed in 1962 the suffrage was extended to all natural-born, literate Kuwaiti males over 21. The ruler promulgates laws and decrees, and the assembly of 50 members also has some legislative power. The state is divided into three governorates: Kuwait, al-Ahmadi, and Hawalli.

The landscape. *Topography and climate.* The topography of Kuwait and the Neutral Zone is generally flat or gently undulating, broken only by occasional low hills and shallow depressions. The elevations range from sea level in the east to nearly 1,000 feet in the southern corner of the country. The Jāl az-Zawr escarpment, one of the main topographic features, borders the northwestern shore of Kuwait Bay and rises to a maximum height of 475 feet above sea level. Elsewhere in coastal areas large patches of salty marshland have developed. Throughout the northern, western, and central sections of Kuwait there are desert basins called playas, which fill with water after winter rains, forming important watering places for the camel herds of the Bedouins.

The climate is semitropical, tempered somewhat in the coastal regions by the warm waters of the gulf. If there is enough rainfall, the desert turns green from the middle of March to the end of April. But during the dry season, between April and September, the heat is severe—the temperature ordinarily reaching 125" F (52" C) during the day and on occasion going as high as 165" F (74" C). The winter is more agreeable (even frost can occasionally occur in the interior, though never on the seacoast). Annual rainfall averages only from one to seven inches, chiefly between October and April, though cloudbursts can bring more than two inches of lain in a day.

The frequent winds from the northwest are cool in winter and spring and hot in summer. Southeasterly winds, usually hot and damp, spring up between July and October; hot and dry south winds prevail in spring and early summer. Fierce dust storms called the *tauz,* lasting sometimes for several days, occur mostly during winter.

True soils scarcely exist naturally in Kuwait. Except in the green belt of Kuwait city and some other patches, where cultivation and irrigation have been carried out for some years, the vegetation of the country consists of scrub and low bushes (and ephemeral grass in the spring). Salt-loving plants grow on the marshy stretches along the coast.

The old city. The old city of Kuwait thus was and is located on one of the harshest spots on the earth's surface. And it is the exploitation of oil that has given rise to an extraordinary metropolis that could never have grown so rapidly under normal conditions.

Nature of the old city

The old city of Kuwait---oriented toward the sea and bounded landward from 1918 to 1954 by a mud wall, with gates that led out only into the endless desert—was compact, only five square miles in area, its typical dwelling a courtyard house. Approaching the city by sea or desert route one could see in the distance a typically Arabian desert city. Although large parts of Kuwait have today changed radically, there are still quarters in old Kuwait that have not been disturbed, giving a true picture of urban conditions before the discovery of oil. The old city, referred to as ad-Dirah in Arabic, was like a huge apartment dwelling, its bazaars a huge department store, its alleys and streets an interesting network of highways. Its courtyards were centres for family socialization, industry, and relaxation. The culture was one of a seafaring, desert-daring, pearl-diving patriarchal society guided by the main driving force of the Arab—his Islāmic religion. Some Kuwaits can still remember the days when drinking water was imported by barge from Iraq.

Modern Kuwait. Today Kuwaitis drink distilled seawater and live and work in air-conditioned surroundings. In and around the city are schools, hospitals, and tall office buildings. There is planning for still further expansion of public and civic facilities. Engineers, doctors, contractors, and merchants from every corner of the earth can be seen busily at work in Kuwait. It is a booming city, a mushrooming city. Behind all this dynamism is oil.

Up to 1951 Kuwait city consisted of little more than the area enclosed within the semicircular city wall built in 1918. With the urban explosion of the years after 1951, the wall was demolished and new suburbs were formally laid out. Within these suburban blocks, building plots of 8,100 and 11,000 square feet were allocated to Kuwaiti citizens who had been displaced from the old city then in process of radical redevelopment. This left the growing non-Kuwaiti population the choice between living within the old city or living in one of the two more freely developing suburbs of Hawalli and as-Siilimiyah.

Urbanization and suburbanization

The results of this sorting process can be witnessed today, two distribution patterns being clearly visible. Kuwaitis are scattered at a relatively low density (16,000–18,000 persons per square mile) throughout the urban area, with minor concentrations in the new suburbs of Dasmah, ash-Shiimiyah, and al-'Idayliyah. On the other hand, non-Kuwaitis, excluded from the restricted suburbs, live at higher densities of between 21,000 and 31,000 persons per square mile in the old city and the suburbs of Hawalli and as-Siilmiyah—mostly in flats and apartments. Strong demographic and geographic contrasts have been introduced by the large-scale influx of migrants in recent years. In Kuwait today there is a veritable cosmopolitan mixture, each ethnic group adding colour and contribution to the fast-changing urban prospects and aspects of the country. A 1961 census of foreigners living in Kuwait showed 27,000 Iraqis, 18,000 Iranians, 31,000 Jordanians, and large numbers of Saudi Arabians, Indians, Pakistanis, and Lebanese. There were approximately 3,000 Britishers and a colony of 400 Americans. The major local firms have lndian or Pakistani clerks and Saudi Arabian, Iraqi, or Iranian workers. The 1970 census indicates a total of about 390,000 foreigners living in Kuwait. Although presumably many of the unskilled foreign construction workers will leave the country as major projects are completed, there are no signs yet of a letup in Kuwait construction programs, and thus there have been few departures of foreign workers. Both working conditions and wages are superior in Kuwait as compared with surrounding countries. Many of the foreign personnel and almost all of the Anglo-Americans are connected with either the oil companies or the government industrial development projects. It is a fair assumption that the population will continue to expand.

Development of oil production

The national economy. In 1934 the Kuwait Oil Company, the ownership of which is divided equally between the British Petroleum Company and the Gulf Oil Corporation (of the United States), obtained a concession covering the whole territory except the Neutral Zone. Oil was struck in 1938, but World War II deferred development until 1946. Thereafter, progress was spectacular. In 1953 the American Independent Oil Company and the Getty Oil Company, which jointly held concessions for the Neutral Zone, struck oil in commercial quantities; and in 1955 oil was discovered in northern Kuwait. European and Japanese companies operate offshore oil concessions. By 1971 total Kuwaiti production was over 1,000,000,000 barrels annually.

More than 90 percent of the income of the government of Kuwait is derived from royalties on the production of oil. Oil production accounts for roughly 57 percent of the country's gross national product—the remainder consisting largely of services dependent upon the flow of income from oil. Employment generated directly by oil production and export, however, accounts for 3.1 percent of the labour force. The government is now planning to spend about 18.5 percent on production sectors (oil and natural gas, industry, agriculture and animal husbandry, and fishing); almost 32 percent on the physical infrastructure (water and irrigation, power, transport, and communications); about 19.5 percent on housing and public buildings; about 22 percent on social, administrative, and commercial services (education, research, information, commerce and tourism, health, religious services, and social welfare); and the balance of about 8 percent on the build-up of commodity stocks.

The low cost of oil production in Kuwait stems from certain unique advantages. First, there are some 50 highly productive wells, whose output can be varied at short notice, thus eliminating the need for large numbers of storage tanks. Most of the storage tanks themselves are placed on a ridge set back a few miles from the seacoast at a height of some 300 feet; this enables loading operations to be carried out by gravity and not through the use of pumps; apart from the obvious savings, this promotes higher loading rates. There are also extensive refineries and bunkers for tankers.

Natural gas and other industries

Massive volumes of natural gas are produced in association with crude oil. Although this natural resource has great potential as a source of foreign exchange, its principal uses so far have been in re-injection in oil fields to maintain pressure, in the generation of electricity (as for water distillation), and in the production (as raw material) of petrochemicals and fertilizers. Other new industries in Kuwait include the production of cement, dry-cell batteries, electric cables, plastic tubes, woollen blankets, confectioneries, paints, liquid gas, and lime bricks. Most of these industries are of small size.

Under construction is ash-Shu'aybah Industrial Area, which is already provided with such facilities as roads, a harbour, gas, power, and water. There is also the National Fisheries seafood packing plant, ammonia and urea plants, a petrochemical complex, and a cement factory. Under consideration by ash-Shu'aybah Industrial Development Board are projects for the construction of a workers' town, a dry dock for the repair of ships, and a plant for liquefaction of natural gas.

Except for oil, there is scarcely any usable natural resource in Kuwait, at least none that is known. There are no other minerals, and even the large expanses of sand, being very solid, are not economically exploitable. The soil is very poor, and fresh water is almost nonexistent. Generally, Kuwait awaits further detailed geologic and stratigraphic studies to determine whether or not anything more fruitful can be extracted.

The government encourages the development of industry by granting land at nominal rent, by subsidizing key inputs like the production of electricity and water, by offering exemptions from customs duty, tariffs, and import controls, by providing loans at low interest rates, by providing capital in exchange for owners' shares, by training labour, and by conducting technical studies designed to devise ways and means of increasing productivity. This all-out government policy to accelerate industrialization has begun to show results. But despite whatever strides in industrialization Kuwait may eventually achieve, it seems unlikely that manufacturing will ever account for more than a small proportion of the total domestic output. The difficulties of a small market and a lack of raw materials and manpower obviously limit the scope for industrialization.

The possibilities of agricultural development, as already intimated, are severely limited. Only about 3 percent of the land is at all arable; and owing to the scarcity of water, deficiencies of soil, and lack of manpower trained in agricultural skills, barely 1 percent of the land area is under actual cultivation. Agriculture in Kuwait has remained a marginal activity. Its contribution to the output of the economy is insignificant at about 0.2 percent.

Fish are plentiful in the Persian Gulf, and in pre-oil days, fishing in Kuwait was a leading and flourishing industry. It is assuming more importance again and is one of the sectors planned to play a significant role in the country's drive for diversification. At present there are four major Kuwaiti fishing companies primarily engaged in exporting shrimp to Europe and the United States.

Water and power

For fresh water, in earlier days, people depended on a few artesian wells and on rainwater collected from the roofs of houses or from cisterns at ground level. Dhows manned by Kuwaiti seamen also brought fresh water from Shatt al-Arab near Basra, Iraq. With the rapid growth of population, however, the government of Kuwait built a seawater distillation plant at ash-Shuwaykh. Its progressive expansion now makes it the largest distillation plant in the world, producing up to 10,000,000 gallons daily.

The expansion of electric facilities has also been remarkable. Production is now concentrated in two large steam-power stations, one at ash-Shuwaykh and the other at ash-Shu'aybah, both fired with natural gas.

Aspects of Kuwaiti society. Unlike most Arab societies, Kuwaiti society was traditionally a compact, united, and virtually classless society. The whole population, to all intents and purposes, constituted one big family. This aspect could be attributed to many factors: (1) the small size of the original city of Kuwait, which made segregated living impossible; (2) the small population, which restricted one's circle of friends and companions; (3) the absence of non-Arab influence in Kuwait, which rendered the class concept incomprehensible; (4) the absence of agriculture and thus of hierarchies based on land ownership; (5) the influence of tribal relations and tribal pride, which tended to deny the importance of the unequal distribution of money and property. For these and other less defined reasons, a definite class system, with all its traditions and conventions, has never developed in Kuwait.

Social welfare services

Today there has arisen a comprehensive scheme of social welfare covering most of the misfortunes that might befall Kuwaiti citizens. For the needy there is financial assistance; for the handicapped there are loans to start a profitable business; for the invalid and deformed there are treatment and training; for the adult illiterates there is education; and for those of limited income there is comfortable housing.

A huge program called the Limited Income Housing Scheme to provide adequate houses, fully equipped with modern facilities, for limited income families has been one of the great cares of the Ministry of Social Affairs and Labour ever since it was entrusted with the task in 1962. About 2,000 new houses every year are contemplated to meet the increasing demand by eligible families.

There are also educational centres, where adults who grew up before the emergence of Kuwait's educational boom are not only taught to read and write but are also prepared for useful careers. General education in Kuwait is entirely free and includes not only tuition but also school meals, books, uniforms, transportation, and medical attention. The educational services provided by the state are arranged in five stages: kindergarten, elementary school, intermediate school, secondary school, and university. Kuwait university was founded in 1962.

The overwhelming majority of the population of Kuwait is Muslim, although there is a large group of Hindus as well as some Christian Arabs and Europeans. Full freedom of religious belief is guaranteed by the constitution.

The native and official language is Arabic. English is the second basic language taught in public schools. Hindi, Urdu, and Persian languages are also prevalent.

The Ministry of Guidance and Information is the principal organ of publicity, enlightenment, and information. It runs the government press and the radio and television broadcasting stations. In 1958 it began to publish *al-'Arab?,* the national and international Arabic monthly, now the largest selling magazine of any sort in the Arab world, with an average monthly circulation of about 125,000. The same ministry has also undertaken the great task of reviving some of the outstanding literary works in Arabic, in the "Arab Heritage" series.

BIBLIOGRAPHY. For the history and geography of the country, see H.R.P. DICKSON, *Kuwait and Her Neighbours* (1956); LAURENCE LOCKHART, "Outline of the History of Kuwait," *Royal Central Asian Journal,* vol. 34 (1947); JACQUES BERQUE, *Les Arabes d'hier à demain* (1960; Eng. trans., *The Arabs: Their History and Future,* 1964); and VIOLET DICKSON, *Forty Years in Kuwait* (1971). On the economy, see RALPH HEWINS, *A Golden Dream: The Miracle of Kuwait* (1963); INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT, *The Economic Development of Kuwait* (1965); and RAGAEI EL MALLAKH, *Economic Development and Regional Cooperation: Kuwait* (1968). Specialized works, also of interest, are ZAHRA DICKSON, "Kuwait and Its People," *Islamic Review,* 37:25–27 (1949); B.L. BURTT and P. LEWIS, "On the Flora of Kuwiet," *Kew Bull.,* 3:273–308 (1949); and D.I. MILTON, "The Geology of the Arabian Peninsula: Kuwait," *Prof. Pap. U.S. Geol. Surv. 560–F* (1967). For statistical data, see the KUWAIT PLANNING BOARD, CENTRAL STATISTICAL OFFICE, *Statistical Abstract* (annual); and for a detailed listing of additional sources, SORAYA M. KABEEL (comp.), *Selected Bibliography on Kuwait and the Arabian Gulf* (1969).

(D.A.Sa.)

Kwangsi

The Kwangsi Chuang (Guang-xi Zhuang-zu in Pin-yin romanization) Autonomous Region is located in southern China. It is bounded by the Chinese provinces of Yunnan on the west, Kweichow on the north, Hunan on the northeast, and Kwangtung on the southeast, and by North Vietnam and the Gulf of Tonkin on the southwest. It covers an area of 92,700 square miles (240,100 square kilometres), and in the early 1970s had an estimated population of 24,000,000. Nan-ning, the capital, is about 75 miles southwest of the region's geographic centre.

The name Kwangsi dates back to the Sung dynasty (960–1279), when the region was known as Kwang-nan Hsi-lu, or wide south, western route (western half of all territory south of the Nan Ling [Nan Mountains]). The Yuan dynasty (1279–1368) contracted the name to Kwangsi when it created a province out of the western half. In 1958 the People's Republic of China transformed the province into the Kwangsi Chuang Autonomous Region—a step designed to help foster the cultural autonomy of the Chuang people, who constitute the largest minority living in the region.

Under the People's Republic, the region has made considerable progress in its economic development. Dams, canals, and reservoirs have been built to help irrigate dry lands; and hydroelectric stations have been constructed and mineral resources exploited to stimulate modern industry. Rural industries have been developed in an effort to diversify village economy. Kwangsi has become self-sufficient in rice and, in fact, exports surplus rice to Kwangtung. This economic growth has given the region new strength and unity. For a related physical feature see HSI CHIANG (RIVER).

History. *Early history.* Kwangsi was known as the land of Pai Yiieh (the Hundred Yiieh—referring to the aborigines of South China) during the late Chou dynasty from *c.* 453–221 BC. A subgroup of the Tai people, known as the Chuang, inhabited the region and had an economy based on wet (grown under irrigation) rice. Eastern Kwangsi was conquered by the Chinese, or Han, people in 214 BC under the Ch'in dynasty, and the Ling Ch'ü (canal) was dug to link the Hsiang Chiang (Hsiang River) and the Kuei Chiang to form a north–south waterway.

The invasion of the Chinese

An independent state known as Nan Yiieh (South Yiieh), was created by Gen. Chao T'o, with Chuang support, at the end of the Ch'in dynasty and existed until it was annexed in 112–111 BC by the Han dynasty (206 BC–AD 220). The Chinese rulers reduced the power of the Chuang people by consolidating their own control in the areas surrounding the cities of Kuei-lin, Wuchow, and Yü-lin.

In AD 42 an uprising in Tonkin was quelled by an army under Gen. Ma Yüan, who not only sought victory on the battlefield but also showed concern for the well-being of the people. He reorganized Kwangsi's local government, improved public works, dug canals, and reclaimed land to increase production. Temples erected to his memory can still be seen in many places.

From the end of the Han to the beginning of the T'ang dynasty (618–907), the influx of Yao tribes from Kiangsi and Hunan added to racial tensions in Kwangsi. Unlike the Chuang, the Yao resisted Chinese culture. The hill country of Kuei-p'ing, Chin-hsiu, and Hsiu-jen in central eastern Kwangsi (the Ta-yao-shan Yao Autonomous Hsien) where they settled became a centre of chronic unrest. In subsequent dynasties there were further migrations of the Yao from Hunan and Kweichow provinces.

Under the T'ang dynasty, Kwangsi became a part of the Ling-nan Tao (large province). The noted scholar Liu Tsung-yüan was prefectural administrator at Liu-chou. Irked by Chinese expansion, however, the Chuang people moved to support the Tai kingdom of Nanchao in Yunnan. Kwangsi was then divided into an area of Chuang ascendancy west of a line from Kuei-lin to Nan-ning and an area of Chinese ascendancy east of the line. After the fall of the T'ang, an independent Chinese state of Southern Han was created, but it was liquidated by the Sung dynasty in 971.

The Sung dynasty, which lasted until 1279, governed Kwangsi by the alternate use of force and appeasement—a policy that neither satisfied the aspirations of the Chuang nor ended the savage warfare waged by the Yao against the Chinese. In 1052 a Chuang leader, Nung Chih-kao, led a revolt and set up an independent kingdom in the southwest. The revolt was crushed in 1053, but the region continued to seethe with discontent. The Yüan dynasty (1279–1368) imposed direct rule and made Kwangsi a province, but relations between the government and the people did not improve. To further complicate race relations, still another aboriginal people—the Miao—migrated from Kweichow, and additional Chuang also came from Kiangsi and Hunan.

Independent Chuang kingdom

Confronted with a complex situation, the Ming dynasty (1368–1644) actively promoted military colonization in an effort to undermine the tribal way of life. It governed the minority peoples through the hereditary *t'u-ssu* (tribal leaders serving as the agents of Chinese government). This led to some of the bloodiest battles in Kwangsi history—notably, the war with the Yao tribesmen at Giant Rattan Gorge, near Kuei-p'ing, in 1465.

The Chi'ng (Manchu) dynasty (1644–1911) placed the minorities under direct imperial rule in 1726. This, however, did not bring peace. Following a Yao uprising in 1831, the great Taiping Rebellion broke out in 1850—again near Kuei-p'ing and under minority leadership—lasting for over a decade.

Meantime, the execution of a French missionary in western Kwangsi led to an Anglo-French War against China that was concluded by the humiliating treaties of Tientsin in 1858. Then, following the Sino-French War of 1884 to 1885, French supremacy in Vietnam exposed Kwangsi to foreign encroachment. Lung-chou was opened to foreign trade in 1889, Wuchow in 1897, and Nan-ning in 1907; while in 1898 France obtained a sphere of influence which included Kwangsi.

Foreign encroachment

The *revolution.* Together with neighbouring Kwangtung, Kwangsi in the early years of the 20th century became the base of the nationalist revolution led by Sun Yat-sen (*q.v.*). Between 1906 and 1916, the provincial leaders of Kwangsi supported the establishment of a republic, and during the following decade played an active role in the reorganization of the Kuomintang (the Chinese Nationalist Party). Following the rise of Chiang Kai-shek (*q.v.*) to power in 1927, the Kwangsi leaders (notably Li Tsung-jen and Li Chi-sh'en) formed the Kwangsi Clique in opposition to Chiang. The group did much to modernize Kwangsi, but maintained a defiant posture against the central government. Although Chiang crushed their revolt in 1929. he was unable to end the semi-independent status of the region. The Chuang, on their part, formed a string of revolutionary soviets (elected Communist organizational units) between 1927 and 1931 that gave rise to new Communist leaders.

During World War II, Kwangsi was a major target of Japanese attack. The Japanese invaded southern Kwangsi in 1939 and occupied Nan-ning and Lung-chou. In this period, Kuei-lin became the principal base for the Chinese and Allied air forces, as well as the home of the patriotic press, the National Salvation Daily News. In 1944 the Japanese made a determined drive into Kwangsi; although they briefly took Kuei-lin, Liu-chou, and Wuchow, they were unable to maintain their position. Chinese forces subsequently recaptured the major cities. In the civil war that followed World War II, the Chinese Communist forces took Kuei-lin in November 1949 and declared Kwangsi a province of the Chinese People's Republic.

The landscape. *Relief.* Kwangsi forms a tableland that descends in elevation from the north and northwest to the south and southeast. Elevations between 3,000 and 6,000 feet above sea level are reached at the edge of the Yunnan-Kweichow plateau in the northwest, the Chiu-wan Shan (Chiu-wan Mountains) and the Feng-huang Ling (Feng-huang Mountains) in the north, and the Mao-erh Ling in the northeast. The greater part of the region is composed of hilly country lying at a height of between 1,500 to 3,000 feet. In the west, the Tu-yang

Shan rises to 6,500 feet. In the southeast, lowlands are situated at a height of between 300 and 1,500 feet.

The predominance of limestone gives many parts of Kwangsi a spectacular type of landscape, known to geographers as "karst," in which pinnacles and spires, caves and caverns, sinkholes, and subterranean streams abound. Picturesque rocky hills, spires of grotesque proportions, strangely shaped caves, and all manner of stalactites and stalagmites are to be seen in many different parts of this region.

The Hsi Chiang Basin

Drainage and soils. The Ch'in Chiang (Ch'in River) and the Nan-liu (or Lien) Chiang flow into the Gulf of Tonkin. The headwaters of the Hsiang Chiang flow into Hunan Province. The remainder of the region's numerous rivers—including the Hung-shui Ho, the Lui Chiang, Ch'ien Chiang, the You Chiang and Tso Chiang, the Yii Chiang, the Hsiin Chiang, the Meng Chiang, the Jung Chiang, and the Kuei Chiang—follow the general southeastward slant of the terrain. They rise from a profusion of sources and flow into one another in a succession of convergences until they merge into one giant river, the Hsi Chiang (*q.v.*). This mighty river rises in Yunnan Province and cuts across the entire width of Kwangsi before emptying into the South China Sea near Canton in Kwangtung Province.

The hilly areas are composed of red soil, while the lowlands are characterized by alluvial soil brought down by the many rivers.

Climate. Throughout the region, temperatures are warm enough to assure agricultural production throughout the year. The summer, covering a seven-month period from April to October, is marked by enervating heat and high humidity. Winters are mild and snow is rare. July temperatures vary between 80" F (27" C) and 90" F (32" C), while January temperatures range between 40" F (4° C) and 60" F (16" C). The north experiences slightly cooler temperatures than does the south.

Because of the influence of the rain-bearing monsoon wind, which blows from the south and southwest from late April to the end of September, precipitation is abundant; the drier areas are in the northwest, and the wetter areas in the south and east. The average annual rainfall varies from 35 inches in the drier areas to 68 in the wetter zones. Most of the precipitation occurs in the period between May and August. In the extreme south, rain bursts caused by typhoons occur between November and February.

Vegetation and animal life. Stands of fir, red pine, cedar, camphor, and rosewood are found in the north and west; oranges grow in profusion in the south; while the cassia tree, anise, and betel nuts flourish in many parts of the region. In central and south Kwangsi, many denuded hillsides have been taken over by tall coarse grasses, which are used for fuel or as pasturage for young water buffalo. Prominent types of wild life include the bison, boar, bear, gibbon (a kind of ape), hedgehog, and cockatoo.

Population. Demography. The population numbered 19,561,000 in 1953 and reached an estimated 24,000,000 in 1970. The average population density in 1953 was 230 per square mile but was unevenly distributed. Approximately two-thirds inhabit the eastern third of the region, while only one-third occupies the remainder of the territory to the west.

The Chuang, Yao, Miao, and T'ung peoples

Ethnic groups. The population in the early 1970s included over 14,500,000 Han or Chinese, 8,500,000 Chuang, 660,000 Yao, 250,000 Miao, and 200,000 T'ung. The Chuang were found largely in the western two-thirds of the region, while the Chinese were concentrated in the eastern third. Two distinct Chinese linguistic influences are noted—a southwest version of Mandarin is spoken in the Kuei-lin district in the northeast as well as in the north, while Cantonese is spoken throughout the remainder of the region. The Yao, Miao, and T'ung settlements are widely scattered.

The Chuang are a Tai people, who have inhabited Kwangsi since classical antiquity. Living on the plains and in the river valleys of the hilly west, they cultivate wet rice in paddy fields and practice an economy that easily merges with that of the Chinese. They are often referred to as "water dwellers" because their settlements are close to water and their dwellings are constructed on piles or stilts. For two millennia, the Chuang have coexisted with the Chinese; together they constitute the two largest racial groups in Kwangsi. The Chuang have absorbed Chinese culture, speaking both their own dialects and Cantonese. Under the People's Republic, a written Chuang alphabet has been created.

The origins of the T'ung are not clear, but they are generally considered to be a branch of the Chuang. They, too, have absorbed Chinese culture despite the fact that they live in the high mountains close to the Kweichow border to the north. The Miao and the Yao, however, have long resisted the absorption of Chinese culture. They belong to a separate linguistic branch of the vast Sino-Tibetan language family. Their dialects are marked by tribal variations and contain many words borrowed from Chinese and Tai. Neither the Miao nor the Yao dialects were written until alphabets based on adaptations of the Latin script were introduced in the late 1960s.

Upland dwellers who suffer from a scarcity of arable land, the Miao and the Yao practice subsistence agriculture. Characteristically, the Miao–Yao settlements are removed from transportation routes and are walled for defense. Besides farming and lumbering, which form the basis of their economy, the Yao make charcoal and bamboo basketry.

The region's cities

Urbanization. Although urbanization is proceeding slowly throughout Kwangsi, the region has many important cities. The capital city of Nan-ning, situated on the upper Yii Chiang and served by the Hunan–Kwangsi railroad, is the major city and industrial centre of the southwest. Kuei-lin, on the upper Kuei Chiang in the northeast, has a long history as the region's political and cultural centre; it is now a leading educational and commercial centre. In the north, Liu-chou is a hub of water and rail transport as well as the trading centre for the region's lumber and forest products and a burgeoning industrial area. Wuchow, in the southeast, is the gateway to trade along the Hsi Chiang and is also important industrially. Pei-hai (or Pakhoi), on the Gulf of Tonkin, has an excellent harbour; some of its commercial prosperity is attributable to the pearl trade of the nearby town of Ho-p'u. Located near the southern terminus of the Hunan–Kwangsi railroad, P'ing-hsiang guards the Vietnamese frontier and is also a major centre of regional and international trade.

Smaller important cities include Hsing-an, the site of the Ling Ch'ii (canal) connecting the Kuei Chiang and the Hsiang Chiang; Kuei-p'ing (or Hsün-chou), at the confluence of the Ch'ien Chiang and Yü Chiang; Yü-lin, on the Lit'ang–Chan-chiang railroad; Kuei-hsien, situated where the Lit'ang–Chan-chiang railroad crosses the Yii Chiang; and Li-t'ang, a booming town at the junction of the Hunan–Kwangsi and Lit'ang–Chan-chiang railroads. The city of Pai-se, formerly a transshipment point for goods moving to and from Yunnan, is a thriving industrial town.

Administration. Government structure. The region's administration is organized in a series of hierarchical levels. The top is the provincial level, directly under the central government in Peking. Below the top level are the six municipalities (shih) of Nan-ning, Liu-chou, Kuei-lin, Wu-chou, P'ing-hsiang, and Pei-hai, and the eight special districts (*chuan-ch'ü*) of Nan-ning, Pai-se, Liu-chou, Kuei-lin, Ch'in-chou, Wu-chou, Yü-lin, and Ho-ch'ih. Below these are 72 counties (hsien) and eight autonomous counties (tzu-chih-hsien). The lowest administrative units are the communes and towns.

People's congresses and councils

People's congresses are the organs of government authority at all levels. The deputies of the communes and towns are chosen by direct vote, but all other deputies are elected indirectly by the people's congresses to the next higher level. The executive functions of government are performed by people's councils, elected at each level by their respective people's congresses. The chairman of the Region's People's Council (or the Revolutionary Committee since the reorganization of the Council during the Cultural Revolution) is the governor of the region. The Communist Party, with its own congresses and committees at every level, exercises a controlling power over the government. The party supervises the nomination of candidates, and no more than one list of candidates is voted upon. Many positions are held by party members; and party cadres, or trained personnel, are attached to all levels of government.

Minority policy. The People's Republic attaches great importance to the treatment of national minorities and promotes the local autonomy of the Chuang, the Yao, the Miao, and other ethnic groups in an announced effort to assure their equality in an atmosphere of interracial cooperation. At the same time, while minorities enjoy cultural autonomy, minorities are required to participate in the nationwide Socialist transformation, especially at the lower levels of government.

Social conditions. ***Health services.*** Since the 1950s, Kwangsi has made impressive progress in public health and medicine. Such widespread diseases as malaria, smallpox, measles, and schistosomiasis (a parasitic infestation of the bladder or intestines) have been brought under control. The addition of iodine to water has ended the once-frequent occurrences of goitre, and the liver-fluke disease has been overcome by filling in old canals that were sources of infection and digging new ones. There is also a mass program to combat leprosy.

Clinics providing free medical care are available in the communes, but hospitals are found only in the major cities. The policy of "walking on two legs "—*i.e.*, using indigenous drugs and traditionally trained doctors to compensate for the shortage of Western medicine—has been promoted. Acupuncture—the practice of puncturing the body with gold or silver needles—is used extensively for the treatment of tumors, wounds, rheumatism, and arthritis.

Education. The educational system is directed toward raising the cultural level of the people and training technical personnel. Children from three to seven years of age attend nursery schools and kindergartens. Those from the ages of seven to 13 attend primary schools, which are run by production brigades in rural areas and by street committees in the cities. Upon entering the middle schools, all students are required to work and study at the same time; the curriculum stresses proletarian ideology, physical training, and science. The junior and senior middle schools cover three years each. Not all graduates from the junior middle schools enter the senior middle schools; many enter specialized technical schools or teacher training schools. In the 1960s there was a notable expansion in the number of agricultural middle schools and of industrial schools providing vocational-technical education.

Admission to higher education is based on entrance examinations. Since the Cultural Revolution of the late 1960s, students from peasant, worker, or soldier families have been accorded priority in college admissions. The work–study principle is also applied to higher education. Institutions of higher education include the Kwangsi Normal College at Kuei-lin, as well as the Kwangsi Agricultural Institute and the Kwangsi Medical College, both at Nan-ning. The Kwangsi Provincial Museum and the Provincial Library of Kwangsi are located in Kuei-lin.

National minorities education

A special educational feature is the program for the education of national minorities. Minority languages are used for instruction in primary and middle schools, written scripts are developed for spoken minority languages wherever needed, minority teachers are trained, and government subsidies are provided for minority students. The Institute for Minorities in Nan-ning trains both intellectuals and technical specialists of minority descent to work among the minority peoples below the county level.

Social services. A basic minimum of social welfare is available. Welfare funds of the communes guarantee care for the sick, disabled, and aged, and provide relief in times of drought or flood. For industrial workers, there are accident prevention and insurance programs that provide for hospital treatment, sick leave, disability compensation, maternity leave, old-age benefits, and death benefits. The government has been improving housing, expanding recreational facilities, and providing public-health centres.

The economy. ***Agriculture.*** Little more than 12 percent of the region is under cultivation. Agriculture is concentrated in the river valleys and on the limestone plains. The hillsides are terraced wherever feasible. Since the 1950s, the government has been seeking to bring new land under cultivation and to increase the yield of areas already cultivated.

Major food crops include rice, corn or maize, wheat, and sweet potatoes. The leading commercial crop is sugarcane; other important commercial crops include peanuts, sesame, ramie or China grass, tobacco, tea, cotton, and indigo. Kwangsi is also a rich producer of fruits—notably pomelos (a kind of grapefruit), tangerines, mandarin oranges, lemons, litchis (the fruit of a plant of the soapberry family), pears, papayas, bananas, and water chestnuts.

The raising of livestock in Kwangsi is ancillary to farming. Water buffalo are used as draft animals in the paddy fields. Pigs, chickens, and ducks are raised on farms and goats are raised in the hills. In many areas, silkworms are raised on mulberry leaves.

Fishing. Fishing is extensive. Both inshore and deep-sea fishing are carried on in the Gulf of Tonkin. Catches include croaker (a fish that makes a croaking noise), herring, squid, prawns, eels, perch, mackerel, sharks, and sturgeon. The catching of fish fry in the region's many streams is characteristic of the freshwater fishing industry. The fishing season begins in the spring when warm weather raises the water level of the rivers and the fish fry are sufficiently grown. The fry of ***wan yu*** (corvina) hatch first, that of pilchard next, and ***lien yu*** (bream) and others last. The fish-fry industry is heavily concentrated on the Hsün Chiang between Kuei-p'ing and Wuchow. Fish culture and the production of silkworms are complementary; the waste cocoons of silkworms are fed to the fish, and mud from fishponds is used as fertilizer for mulberry bushes.

Forestry. Kwangsi is an important producer of timber and forest products. In the north, large quantities of pine, fir, cedar, and giant bamboo are exploited and marketed at Liu-chou. Red and black sandalwood are also produced in the Yü Chiang Valley. Firewood, mushrooms, *mu-erh* (wood fungus), and bamboo shoots are also sold. More important, however, are sandarac (a resin used in making varnish and incense), star anise (Chinese anise), cassia bark (Chinese cinnamon), nutgall (a swelling on oak trees that produces tannin), and camphor. Tung oil, tea oil, and fennel oil are also produced. Some of these products are vital to traditional Chinese medicine, as are cardamom husks, cassia twigs, plantain seed, the seed of the wax tree, castor-oil seed, mugwort (a perennial herb) powder, dried lizard, mangosteen (a dark-brown fruit), and quinine.

Use of forest products in traditional Chinese medicines

Mining and quarrying. Kwangsi has sufficient coal and iron deposits to support moderate industrial development. Coal is mined in the Hsing-an-Ch'uan-chou region north of Kuei-lin and in the Lai-pin–Li-t'ang region south of Liu-chou. Iron is mined in the Ho-hsien-Chung-shan-Chao-p'ing area to the east, near the Kwangtung-Hunan border, as well as in Kuei-p'ing and Pei-liu. Exploited mineral resources include tin, tungsten, manganese, and antimony. Moderate amounts of bismuth, zinc, and lead are also produced.

Power. Major thermoelectric power stations are located in Nan-ning and Liu-chou. Since the mid-1960s, great emphasis has been placed on the construction of numerous small hydroelectric stations. A 1970 report states, for example, that a project at Tung-shan Commune (in Lu-ch'uan County in Yii-lin Special District) comprises 49 small hydroelectric power stations, each with a capacity of 600 kilowatts.

Industry. Light industries produce textiles, paper, flour, silk, leather, matches, chemicals, and pharmaceuticals as well as sandarac gum, sugar, dyestuffs, and oils and fats. Pine resin is a particularly notable export of

Wuchow. Heavy industries include the iron works and steelworks at Liu-chou and Lu-chai, machinery production at Nan-ning and Wuchow, and the cement works at Liu-chou. Among the numerous handicraft products, special mention may be made of the native cotton cloth of Yii-lin; the ramie cloth of Kuei-lin and Liu-t'an; the bamboo paper and rice paper of Chao-p'ing, Kuei-lin, and Tu-an; and the bamboo combs of Kuei-lin. Pin-yang produces such articles as ceramics, fans, felt caps, copper ware, combs, brushes, and straw bonnets.

Transportation and communications. ***Water transport.*** The elaborate system of waterways provides transportation almost throughout the region. Going north from Wuchow, the Kuei Chiang. is navigable for junks to its source as well as to the Hsiang Chiang via the Ling Ch'ii (canal). Motor launches sail up the Kuei Chiang to Ma-chiang in low water and to Kuei-lin in flood season. Westward from Wuchow, steam boats, motorboats, and junks navigate the hundred miles of the Hsiin Chiang to Kuei-p'ing at all times of the year. Beyond Kuei-p'ing, motor launches and junks navigate the Ch'ien Chiang to Shih-lung, and the Liu Chiang to Liu-chou and the Jung Chiang to Jung-an. The upper reaches of the Jung and Lung Chiang and of the Hung-shui Ho are navigable for junks only.

Conditions are far more favourable on the Yii Chiang, which is navigable by small steamers, motor launches, and junks from Kuei-p'ing to Nan-ning. At high water, the river is navigable to Pai-se on the You Chiang and to Lung-chou on the Tso Chiang. Farther inland, junks ply the headwaters of the You Chiang to Po-i in Yunnan and those of the Tso Chiang to Shui-k'ou-kuan on the North Vietnamese border. The south is served by the Ch'in Chiang and the Nan-liu Chiang (or Lien Chiang). There are proposals to construct a canal connecting the headwaters of the Nan-liu Chiang with those of the Hsiin Chiang, thus linking the south to Wuchow.

Railways. The Hunan–Kwangsi railroad from Heng-yang to Yu-i-kuan (Mu-nan-kuan) runs diagonally across the region from the northeast to the southwest. It forms a vital continental artery that connects with the Canton–Hankow railroad at Heng-yang and with the North Vietnamese railroad at Dong Dang. The Kwangsi–Kweichow railroad links Liu-chou with Kuei-yang, Kweichow Province, and is a great impetus to the development of northern Kwangsi. Another line leads from Li-t'ang on the Hunan–Kwangsi railroad to Chan-chiang, a seaport in southwest Kwangtung Province. It facilitates the trade of southeastern Kwangsi as well as the exploitation of the mineral resources of Hainan Island. Plans for future development include a major line from Nan-ning westward to I-liang, Yunnan Province; another major line from Pai-se northward to Kuei-yang; and a shorter line in the northeast to link Kuei-lin with Lien-hsien, Kwangtung Province.

Roads. The highway system has undergone continuous expansion since the 1950s. First-class highways are paved, and second-class roads have solid rock beds but no surface paving. In the early years of the People's Republic, highways in eastern Kwangsi radiated from Kuei-lin, Liu-chou, Nan-ning, and Yii-lin. The expanded network forms a central rectangle, with Nan-tan, Liu-chou, Nan-ning, and Pai-se at its four corners.

The growing highway system

Running almost due north and south, a trunk road connects Tu-yiin in Kweichow Province, Nan-tan, and Nan-ning with the coast of the Gulf of Tonkin. From Liu-chou, roads lead east to Shao-kuan, Kwangtung Province, north to Kuei-lin and Heng-yang and Yuan-ling in Hunan Province, and south to Nan-ning. From Nan-ning, roads lead south to P'ing-hsiang and Yu-i-kuan, southeast to Pei-hai, and west to Pai-se. From Pai-se roads fan out northeast to Nan-tan; northwest to An-lung, Kweichow Province; west to Wen-shan, Yunnan Province; and south to Cao Bang, North Vietnam. Serving the southeast, one major highway parallels the Lit'ang–Chan-chiang railroad; the coastal highway runs from P'ing-hsiang in the west to the Kwangtung border in the east.

The media. Kwangsi's leading newspapers include the ***Kwangsi Jih-pao*** ("Kwangsi Daily"), the ***Nan-ning Jih-pao*** ("Nan-ning Daily"), and the ***Kuei-lin Jih-pao*** ("Kuei-lin Daily"). The state-controlled New China News Agency is the source of all news. Radio broadcasting, however, is the most powerful means of mass communication, as receiver sets and loud speakers are installed throughout the region. A notable trend in the early 1970s was the use of mobile projection teams to show films and slides in the villages.

Cultural traditions and trends. The cultural traditions of the region are such that the primacy of Chinese culture is widely recognized. Because the minorities in Kwangsi possess neither a unified organization nor support by fraternal groups, their assimilation by the Chinese is far more advanced than in the other autonomous regions. The underlying causes of what appears to be Kwangsi's racial tensions are economic and geographic factors that have exerted a powerful influence on cultural trends.

Like the Chinese, the Chuang and T'ung are agriculturalists. Those who live in the fertile valleys and plains are scarcely distinguishable from the Chinese. They intermarry with the Chinese, speak Chinese in addition to their own dialects, and are at ease with their T'ai background, Before the revolution, many upper-class Chuang were landlords who exploited Yao or Miao tenant farmers. The Chuang and T'ung people in the mountainous regions, however, rarely intermarry, follow their own social customs, and have a definite sense of distinction from the Chinese. The Yao and the Miao, who live in the mountains and mingle even less with the Chinese, have a pronounced feeling of racial differentiation. Until the land reform of the early 1950s, the vast majority of the Miao and Yao were impoverished tenants of Chinese or Chuang landlords.

Under the present regime, a silent but powerful revolution is taking place in the lives of the peoples of Kwangsi. The vigorous development of highways and railroads has broken down transportation barriers and sectional antagonism. Communist authorities have taken steps to move many Yao and Miao to flatlands, to teach them to clear new land, and to allow them to form communes on land confiscated from Chinese or Chuang landlords. In many of the underdeveloped areas of the interior, dams have been built, canals dug, rivers led over arid lands, and reservoirs constructed to expand irrigation. Moreover, workshops for small industries are being set up in all rural communes. It is hoped that the economic measures will erase the racial tensions that have plagued the region for centuries.

BIBLIOGRAPHY. WANG CHUN-HENG, ***A Simple Geography of China*** (1958); and THEODORE SHABAD, ***China's Changing Map*** (1956), contain good introductory chapters on Kwangsi. More up-to-date information may be obtained from ***Nagel's Encyclopedia-Guide to China*** (1968). Maps for general reference may be found in the two CIA publications: ***Communist China Map Folio*** (1967) and ***Communist China: Administrative Atlas*** (1969). Pertinent information on certain periods of the history of Kwangsi may be gleaned from HAROLD WIENS, ***China's March Toward the Tropics*** (1954). For population, economy, and society, see REWI ALLEY, ***China's Hinterland in the Leap Forward*** (1961); and YUAN-LI WU, ***The Spatial Economy of Communist China: A Study on Industrial Location and Transportation*** (1967). Recent developments in the government's policy toward the national minorities are succinctly outlined in C.T. HU, ***The Education of National Minorities in Communist China*** (1970). Current developments can best be traced by following the issues of ***China Reconstructs, Peking Review, Union Research Service Reports,*** and ***Far Eastern Economic Review.***

(P.-c.K.)

Kwangtung

Kwangtung (Guang-dong in Pin-yin romanization), the southernmost of the 21 provinces of the People's Republic of China, also constitutes the region through which South China's trade is primarily channelled. The 12th largest Chinese province in area, it is also the province with the longest coastline—including Hainan, it is about 3,000 miles in length. The province has an area of 81,700 square miles (211,500 square kilometres), which amounts to more than 2 percent of the total area of the People's Republic. The province includes Hainan, which, with an area of 13,124 square miles, is the largest Chinese island except for Taiwan. It is bounded by the Kwangsi Chuang Autonomous Region to the west, and by the provinces of Hunan and Kiangsi to the north, and Fukien to the northeast, and by the South China Sea to the south. At its western extremity on the coast the province borders North Vietnam. Two foreign holdings remain on the coast of Kwangtung—the British crown colony of Hong Kong and the Portuguese territory of Macau. The population of Kwangtung was estimated to be 42,800,000 in 1970, representing a population density of about 525 persons per square mile. The capital is Canton.

Historically Kwangtung and Kwangsi, its neighbouring province to the west, often were jointly governed. Kwangtung was first administered as a separate entity in AD 997 during the Sung dynasty; it was from this time that the term Kwangtung (Chinese ***kuang tung,*** "eastern expanses") began to be used.

Kwangtung has its own characteristics, which manifest themselves in physiography, climate, language, folklore, and products. Its topography separates it somewhat from the rest of China, and this factor, together with its long coastline, its contact with other countries through its overseas emigrants, and its early exposure to Western influence through the port of Canton, has resulted historically in the emergence of a degree of self-sufficiency with which a tendency to separatism has also been associated. Canton dominates the economic, cultural, and political life of the province to an unusual extent. (See CANTON; for associated political units, see HONG KONG; MACAU.)

HISTORY

Physically separated from the early centres of Chinese civilization in North China, Kwangtung was originally occupied by non-Han ethnic groups. It was first incorporated into the Chinese Empire in 222 BC, when Ch'in Shih Huang-ti, first emperor of the Ch'in dynasty, conquered the area along the Hsi Chiang and Pei Chiang valleys down to the Canton Delta. In 111 BC Chinese domination was extended to the whole of what is now Kwangtung, including Hainan, by Emperor Wu of the Han dynasty. The conquest, however, was not followed by successful colonization, and Kwangtung remained part of the empire only politically.

During the five centuries of the Sui, T'ang, and Northern Sung dynasties, from AD 581 to 1126, the military and agricultural colonization of Kwangtung gradually took place. This, combined with increasing overseas trade through Canton, led to an increase of migration into Kwangtung, and to the rise of Canton as a metropolis with a population of hundreds of thousands. At the end of the period, however, Kwangtung was still occupied predominantly by its original ethnic population. The region was viewed as a semicivilized frontier, and disgraced officials often were exiled there as a punishment.

Occupation by the Han Chinese

The southward thrust of the Han Chinese was greatly intensified from 1126, when the Tatars captured the Sung capital. Another major population movement followed a century and a half later, in 1279, as China fell to the Mongols. These migrations marked the beginning of effective Han Chinese occupation and the rapid cultural development of Kwangtung. Especially after the 16th century, with the introduction of sweet potatoes and peanuts into China, the growth of population was so fast that, by the late 17th century, Kwangtung had already become an area from which emigration took place. Migrants from Kwangtung moved first to Kwangsi, Szechwan, and Taiwan and then in the mid-19th century began to pour into Southeast Asia and North America.

As a result of its early contact with modernizing forces, Kwangtung has, since the mid-19th century, produced a number of prominent political and military, as well as intellectual, leaders. Many of the leaders of political movements during this period, such as Hung Hsiu-ch'uan, leader of the Taiping Rebellion (1850–64); K'ang Yu-wei and Liang Ch'i-ch'ao of the Reform Movement (1898), and Sun Yat-sen, who led the Republican Revolution of 1911, had associations with Kwangtung.

PHYSICAL GEOGRAPHY

Physiography. The surface configuration in Kwangtung is diverse, being composed primarily of rounded hills, cut by streams and rivers, and scattered and ribbonlike alluvial valleys. Together with the Kwangsi region, Kwangtung is clearly separated from the Yangtze Basin by the Nan Ling, the southernmost of the three major Chinese mountain ranges running from east to west, which was formed in Late Mesozoic times, between about 70,000,000 and 130,000,000 years ago.

The greater part of eastern Kwangtung consists of the southerly extension of the Southern Uplands, which stretch down from Fukien and Chekiang provinces. Formed at the same time as the Nan Ling, a series of longitudinal valleys running from northeast to southwest extends as far as the vicinity of Canton.

Relief and drainage. Smooth, low hills with slopes of up to 10° cover about 70 percent of the province. Most peaks range in elevation from 1,500 to 2,500 feet, with a few reaching 5,500 feet. Level land of any size is primarily found in the alluvial deltas, formed where rivers empty into the South China Sea.

The Canton Delta

Of great extent and importance, the Canton Delta, measuring about 3,500 square miles, is marked by hilly outliers and by a labyrinth of canalized channels and distributaries totalling some 1,500 miles in length. The delta marks the convergence of the three major rivers of the Chu Chiang system—the Hsi (West), Pei (North), and Tung (East) rivers. The Chu Chiang, or Pearl River, is the name given to the lower course of the Hsi beyond the confluence. Entirely rain-fed, these rivers, which are subject to extreme seasonal fluctuations, collect so much water that, anomalously, the Chu Chiang system discharges annually six and a half times as much water as the Huang Ho (Yellow River), although its basin area of about 173,000 square miles is only about half as large.

Altogether, Kwangtung has 1,343 large and small rivers, with drainage basins covering a total area of about 400,000 square miles. The Han is the most important river outside the Chu Chiang system. Other important rivers and lowlands are located in the southwest and on Hainan.

Hainan

Hainan, separated from the Luichow Peninsula by a shallow channel 15 miles wide, possesses rivers that discharge at least twice as much water as do the rivers in mainland Kwangtung. The discharge rates, moreover, show even greater fluctuations. Alluvial plains on Hainan are largely confined to the narrow coastal margins. A quarter of the island is covered by mountains, the highest of which is the Wuchih Shan, which reaches 6,165 feet.

Soils. In general, soils are poor, as high temperatures and plentiful rainfall result in podzolization (bleaching) and leaching. Almost all of western Kwangtung, together with Hainan, is covered with mature red soils, whereas the rest of the province is covered with a mixture of old and young red soils that have usually been subjected to a high degree of podzolization. In the main, young red soils cover sloping hillsides up to an elevation of over 3,000 feet. In the wettest and hottest parts of Kwangtung, lateritic (heavily leached, iron-bearing) soils are common; like the red soils, they do not resist erosion and require substantial fertilization in their cultivation. The yellow soils are found in the wettest and coolest parts of Kwangtung, occurring in small pockets of flatland with imperfect drainage.

Of more limited distribution, but of evident economic significance, are the alluviums deposited relatively recently in the river valleys and deltas. As a result of the cultivation of rice, the alluviums have developed special morphological characteristics, the most striking of which is the formation of iron hardpans (hard impervious layers composed chiefly of clay) in the zone of the fluctuating water table.

Climate **and** vegetation. Since much of Kwangtung lies south of the Tropic of Cancer, it is the only Chinese province with tropical and subtropical climates. The average July temperature in the Hsi Valley, which is from

82" to 86° F (28° C to 30° C), is little different from temperatures in the lower Yangtze and on the Huang Ho, but the average January temperature is considerably higher, ranging from 55° to 60° F (13° to 16° C). Except on higher elevations, frost is rare, so that almost the whole province lies within the area where two crops of rice can be grown. True winter does not occur in the province, but the hot summer varies in length from ten months in Hainan to six months in northern Kwangtung.

Rainfall. The rainfall regime shows a pronounced summer maximum, with the rainy season lasting from mid-April, when Kwangtung starts to be dominated by moisture-laden tropical air masses from the Equator and the Indian Ocean, until mid-October. Over half of the total precipitation falls between June and August. The months between July and September form the main typhoon season, which ordinarily is accompanied by heavy rains and widespread destruction. From December to February, only 8 percent of the precipitation occurs. Kwangtung's annual rainfall is approximately 60 to 80 inches, decreasing with distance from the coast to the northwest but increasing with altitude and exposure to the prevailing summer monsoon winds.

The typhoon season

Natural vegetation. Abundant moisture and moderate to high temperatures support luxuriant plant growth. Broad-leaved evergreen forests, intermixed with coniferous and deciduous trees, originally covered much of the land, while a more tropical type of vegetation predominates on the south coast and on Hainan. With the exception of the more remote mountainous areas and the interior of Hainan, much of this natural vegetation cover has been stripped by fire and by the use of trees and shrubs for fuel. This circumstance, together with millennia of uninterrupted cultivation, has resulted in much of the natural vegetation now taking the form of secondary forests of hardwoods and horsetail pine. On the more severely eroded hills, coarse grasses and ferns have taken hold. Bamboo groves, varying greatly in height and extent, are widespread, particularly in humid river valleys. The most productive and least disturbed forests cover the mountainous areas. Certain trees, notably camphor, have been revered and protected for centuries, and are found around grave plots and cultivated fields. Since 1949, as a result of massive afforestation programs, the once bare hills of Kwangtung have been reported as resembling "a mist of green."

Plant and animal life. The combination of a propitious climate and a variegated physiography favours considerable diversity in plant and animal life. In the highlands, where coniferous and deciduous species thrive together, the broad-leaved evergreen forests are characterized by tropical oaks, tan oaks (oaks that yield tanin), and chestnut oaks (or chinquapins). The more significant coniferous species of economic value include horsetail pine, Chinese fir, and Chinese hemlock. Some of the species of cypress and pine are little known outside China.

Truly tropical monsoon rain forests are common in the south, particularly in Hainan. In tropical Kwangtung, a number of industrial crops have been introduced and successfully raised. These include rubber, sisal, palm oil, hemp, coffee, and black pepper.

Kwangtung's more traditional agricultural products, however, are rice, sweet potatoes, barley, sugarcane, peanuts, tea, mulberry, tobacco, and no less than 300 types of fruits, among the more representative of which are oranges, tangerines, pomelos (a grapefruit), lychee, longan (a pulpy fruit related to lychee), pineapples, and bananas.

Among the mammals found in Kwangtung are many tropical bats, including fruit-eating bats and several leaf-nosed, insect-eating species. Squirrels, mice, and rats of many species are abundant. Insectivores are generally more diverse than in other regions of China, and carnivores are exemplified by civet cats and small-clawed otters.

Types of birds vary according to their habitat. In the tropical forest, wildfowl, peacocks, and silver pheasants are common. Reptiles are more restricted in distribution. Blind snakes are found on Hainan, while mainland Kwangtung has a number of pit vipers, including the huge and deadly Chinese vipers and bamboo vipers, as well as nonpoisonous pythons which are up to 20 feet long.

Insects of every description—crickets, butterflies, dragonflies, grasshoppers, cicadas, and beetles—are found in profusion. Amphibians include ground burrowers and many types of frogs and toads, some of which rival the American bullfrog in size.

Tigers, rhinoceroses, panthers, wolves, bears, and foxes used to roam the hills of Kwangtung, but their numbers have been decimated by forest fires and persistent deforestation. In the tropical monsoon forest, however, a great number of animals, many of which live in the trees, still remain.

POPULATION

Ethnic composition and distribution. Kwangtung is largely populated by the Han Chinese, the other ethnic minorities totalling less than 2 percent of its population. Numbering around 400,000, the Li are the largest ethnic minority in Kwangtung. They live almost entirely on Hainan, particularly in the mountainous southern portion, designated administratively as the Hainan Li Miao Autonomous District.

Next in importance, the Yao total 41,000. They are concentrated principally near Kwangtung's northwestern border in areas designated as the Lien-shan Chuang Yao Autonomous County (*hsien*), the Lien-nan Yao Autonomous County, and the Ju-yüan Yao Autonomous County. There are 16,000 Miao living with the Li in the autonomous district in Hainan. Some ethnographers, however, maintain that "Miao" is a misnomer on the grounds that, according to their language, customs, and traditions, the Miao are actually the Yao. A heavily Sinicized group, the Chuang live in the Lien-shan Chuang Yao Autonomous County, and there are also about 120,000 Chung in Kwangtung, now redefined as the Ch'in-chou Chuang Autonomous County in Kwangsi.

Other minority groups include the She, who live in the northeast around Ch'ao-an, and the Hui (Chinese Muslims), who are scattered in Hainan. The Ching, roughly 4,500 in number, were transferred to Kwangsi in 1965, when the multinational Tung-hsing Autonomous County in extreme southwestern Kwangtung changed its provincial jurisdiction.

The so-called Tan, or Tanka, the Boat People, are not officially designated as a national minority. Whereas some scholars believe they are descendants of aborigines, others regard them as simply a people who live on boats and speak Cantonese. They generally live along the rivers in the Chu Chiang Basin as well as along the coast.

Linguistic patterns. The relative ethnic homogeneity prevailing in Kwangtung stands in contrast to its linguistic heterogeneity. The Han Chinese speak a wide variety of dialects forming distinct areal patterns. Within the Cantonese-speaking region, however, there exists a great diversity of other dialects.

By far the most important dialect is Cantonese, spoken by about 27,000,000 people in central and western Kwangtung. There is considerable variety among the Cantonese speakers, but the form spoken in Canton is generally regarded as the standard. Hakka is another important dialect; it predominates in the north and northeast of the province; offshoots of Hakka are common in central Kwangtung and northern Hainan. A third major dialect, Fukienese (Southern Min dialect), is spoken mostly along an eastern coastal area centred on Swatow, but Fukienese speakers also are found in northern Hainan. Sometimes Hainanese is treated as a distinct dialect, commonly spoken in the coastal plains of Hainan.

In addition to these Han Chinese dialects, there are the languages and dialects of the ethnic minorities. Notable among these are the dialects of the Li people on Hainan. It should be noted that, although linguistic unity has been the policy orientation of the Chinese Communists, considerable attention has been paid to the popularization of minority languages. New scripts have been created for a number of these languages. They not only are taught in minority-area schools but also are used in conjunction with Chinese in official communications in minority communities.

Patterns of rural and urban settlement. The settlement pattern in Kwangtung is based on a traditional settlement hierarchy ranging from nuclear agglomerated villages, through small and large towns, to county (hsien) capitals and the provincial capital. This basic structure appears to have been little altered since 1949.

Urbanization

Most of the people of the province live in more than 3,000 villages, which remain the basic functional units in the countryside. Population distribution bears almost a one-to-one correlation with agricultural productivity. The greatest numbers of villages are in the fertile river deltas and along the waterways. To an even greater extent, towns and cities are located in the deltas and coastal areas and along major communication lines. The most highly urbanized area within the province is the Canton Delta, where some 60 percent of the population live in urban areas. Kwangtung is a relatively highly urbanized province for China, with about 18 percent of its population being classified as urban.

The urban hierarchy is headed by Canton, with a population of about 3,100,000 (1970). It is far greater in size than the second largest city, Swatow (280,000 in 1959). Other municipalities of over 100,000 inhabitants are Sh'ao-kuan (232,000 in 1965), Chiang-men (180,000 in 1962), Chan-chiang (170,000 in 1959), Fo-shan (120,000 in 1958), Hai-k'ou (110,000 in 1959), and Ch'ao-an (101,000 in 1958). In addition, there are nine cities with populations between 30,000 and 60,000; 46 at the 10,000–30,000 level; and 278 townships of nearly 10,000 persons each.

ADMINISTRATION AND SOCIAL CONDITIONS

Government. The administrative system in Kwangtung has undergone many changes since 1949. "Autonomous" administrative units were established in the early 1950s for areas with large ethnic minority populations. The status of Canton was changed in 1954 from a centrally administered municipality (t'eh-pieh-shih) to a municipality under the jurisdiction of the provincial government. Another important change has been the development of an intermediate layer of administration between the province and the county (hsien), the special districts (chuan-ch'ii), assisting the province in supervising the county.

In addition to the municipality of Canton, Kwangtung is subdivided into seven special districts, one administrative district (hsing-cheng-ch'ii), and one autonomous district (*tzu-chih-chou*). The headquarters of the provincial government is located in Canton. Altogether, there are 94 counties, three autonomous counties, and nine county-level municipalities.

In each special district, there are from seven to 15 counties including the autonomous counties. Every special district also has under its jurisdiction a municipality where its headquarters is situated, except for one having none and two having two municipalities each. The Hainan administrative district includes one municipality and nine counties, while the municipality of Canton administers two counties.

The autonomous district and autonomous county as administrative units are equivalent to the special district and the county, respectively. Though labelled autonomous, they are in fact wholly integrated into the regular administrative hierarchy, possessing no important powers that other administrative units lack. They differ from the other units in the concessions extended to local minority customs and languages. Moreover, special efforts are made to absorb minority cadres into the local administration.

Before the communes were introduced in 1958, the basic administrative units below the county level were the administrative villages (hsiang) or market towns (*chen*) in counties and the districts (ch'ii) in municipalities.

The communes

Rural administration was fully reorganized when communization was implemented in 1958. Especially in Kwangtung and other coastal areas of South China, the communes took over the apparatus and territory of the administrative districts and have become the basic organ of state power. Administrative villages were converted into production brigades under the communes, and the natural villages within administrative villages became the production teams of the brigades. There are 1,500 communes in Kwangtung, integrating the state administration and the economic administration at the subcounty level. A recent development has been greater decentralization of economic management to the production brigades and teams.

Social conditions. Education. Education, health, and other social conditions in Kwangtung have generally been improved since 1949. There are now many more kindergartens and nurseries for preschool education. The number of primary schools in a commune varies from about 15 to more than 50, and, although primary education is not compulsory, over 80 percent of children of primary school age attend these schools. Many communes also have one general middle school, and some have one or two vocational schools. The establishment of many vocational schools, especially the agricultural middle schools, on a part-work, part-study basis, has markedly widened the opportunities for middle school education. There are more than 30 universities and colleges and roughly 20 technical schools.

Special attention has been given to the education of the ethnic minorities. New schools, including one national minority college, have been established in minority communities where none existed before. There are also schools, including one university in Canton, specifically built for returned overseas Chinese students and students from families of overseas Chinese living in Kwangtung. Repeated campaigns have succeeded in reducing illiteracy to well below 50 percent in many areas.

Hospitals, clinics, and medical services

Health. In general, a commune possesses one hospital, up to 15 clinics, and many health stations, including maternity centres. There may be 80 or 90 sickbeds in one hospital, and some hospitals are equipped with X-ray facilities and operating theatres. Better equipped and better staffed hospitals are maintained at the county and provincial levels. All schools and factories, except for the smallest ones, have their own clinics.

Medical education has been greatly expanded. There are eight medical colleges and schools, with one college devoted to Chinese medicine (acupuncture and herbal medicine). Many short-term medical-training classes are organized for health workers who are assigned to rural areas.

Medical services are available to large sections of the population. The cost of a worker's medical treatment is borne by his factory's welfare fund, to which he subscribes. Cooperative medical service resembling the factory's welfare fund is also widely provided in rural communes.

Although epidemics of influenza and liver inflammation struck in 1959 and 1960, the development of medical services, coupled with the general improvement in sanitation and health education, has succeeded in eliminating many previously common diseases such as malaria, schistosomiasis (a waterborne parasitic disease), and filariasis (a parasitic disease infecting the blood and lymph glands).

Welfare. Welfare programs have been organized for children and workers. Nurseries and child-care centres are usually free of charge for children of working mothers. Workers, in addition to the medical funds already mentioned, are provided with maternity benefits, life insurance, and retirement pensions. Their dependents are entitled to partial health benefits.

Economy. Agriculture. The economic foundation of Kwangtung is primarily agriculture. Inasmuch as only about 15 percent of its land is under cultivation, agriculture is of necessity extremely intensive. The limited extent of sown land available is, however, partly offset by repeated uses of it. Two crops of rice a year can be grown on most cultivated land, and in Hainan three crops are not unusual. Thus, although average yields per harvest are below the national average, annual yields resulting from two harvests exceed the national average and are bettered by only four other provinces.

Rice is the leading crop, occupying about 76 percent of the total cultivated area and accounting for over 80 percent of Kwangtung's total food production.

Although food-grain crops (including rice) occupy almost 90 percent of the total cultivated area, the industrial and fruit crops grown on the remaining land are of national importance. Kwangtung annually produces more than 6,720,000 tons of sugarcane—about half of China's total output. Fruit production reaches 6,000,000 piculs (828,000,000 pounds), a significant proportion of which is consumed in north or central China.

Good rice yields depend upon careful water control. Progress in irrigation and flood control has made such water control possible for about 84 percent of the cultivated area. Farming has become increasingly mechanized, and increasing reliance has been placed on the use of chemical fertilizers.

Rural reorganization

Many of these changes have been brought about by rural reorganization since 1949. In a land traditionally plagued by a system of tenant farming that was more widespread than in most other provinces, large-scale agrarian reform was not introduced in Kwangtung until 1952. Rural cooperatives began to be formed in 1954, and by 1958, 98.5 percent of all rural households were organized into people's communes. Each commune has an average population of 30,000, consisting on the average of 7,000 to 8,000 households. Commune production teams are commonly composed of 40 to 50 households each. Small plots of private land are retained for raising vegetables and livestock, particularly hogs.

Since the river valleys and the deltas have been intensively tilled for untold generations, it is in the remote areas that land-reclamation schemes are pursued. In the uplands, for example, efforts have been concentrated on afforestation, fruit growing, and livestock raising.

Reclamation of wasteland in peripheral areas has been largely undertaken by the state farms, which engage in collective farming under the direct control of public authorities.

Fishing. Kwangtung, with its long coastline, produces about one-quarter of China's fish. Fish production accounts for as much as one-third of the income of some communes. Fish farming yields over 110,000 tons of pond fish every year. More than 400 species of saltwater fish, including yellow croaker, white herring, mackerel, golden thread, and pomfret are caught from 57 fishing ports.

Industry. Because of the paucity of its iron deposits, Kwangtung received only scant attention when China's heavy industry was developed during the first five-year plan (1953-57). Light industry, however, has always been of significance in the province. Apart from handicrafts, light industry--especially food processing and the manufacture of textiles—accounts for about three-quarters of industrial production. Almost all of the major light industries are located in the Canton Delta. Sugar refining is centred in Canton, Tung-kuan, Shun-te, Chiang-men, and Swatow, while silk filature (the reeling of silk from cocoons) and weaving are well developed in Canton, Fo-shan, Shun-te, and Fu-shan. The largest and most widespread industry is rice milling, which takes place in nearly every county and municipality.

Existing heavy industries include metal processing, the manufacture of machinery, shipbuilding and ship repairing, the production of hydroelectricity, and mining. A large proportion of these industries is concentrated in Canton, where 85 percent of Kwangtung's machinery-manufacturing and metal-processing industries are located. The province has 32 fertilizer-producing plants.

Despite a shortage of iron in mainland Kwangtung, extensive deposits of high-quality hematite (the principal iron ore) with up to 63 percent iron content, exist in Hainan at Ch'ang-chiang and T'ien-tu. Kwangtung also has nonferrous resources. Coal reserves are estimated at about 4,000,000,000 tons. The province also has about 60 percent of China's manganese deposits, located on the Luichow Peninsula and nearby. Tungsten, which is associated with bismuth, molybdenum, and tin deposits, is mined near the Kiangsi border at Shih-hsing, Lo-ch'ang, and Weng-yüan. Oil-shale deposits have been discovered in quantity on the Luichow Peninsula. Oil refineries have been established in Mao-ming.

Water transport

Transportation and communications. Economically and culturally, the different regions of Kwangtung are linked together by the waterways associated with the Chu Chiang ssytem, which comprises 80 percent of the province's 9,000 miles of navigable inland waterways, of which 2,440 miles are used by motor vessels. In addition, a number of coastwise and international shipping routes are variously linked to more than 100 large and small ports. The leading ports, including Whampoa (Canton's seaport), Chan-chiang, Swatow, and Hai-k'ou, are of national significance. Water transport accounts for approximately 70 percent of Kwangtung's total traffic tonnage. The waterways are maintained by a continual dredging, widening, and clearing of channels.

Connections with other provinces depend principally on land transportation. Kwangtung has developed the best highway network in China, running primarily along river valleys. There are 10,690 miles of highway and 466 miles of railroad. Interprovincial links, both for highways and railroads usually run north–south. The vital Canton–Hankow railroad was double-tracked in the early 1960s. The low priority placed on east–west transport is indicated by the absence of a railroad running parallel to the Hsi Chiang and by the delay in completing the Canton–Chan-chiang line, which was finally opened in 1963.

Kwangtung provides a crucial link in China's domestic and international civil-aviation routes. Air services connect the province to London, Paris, Hanoi, Dacca, Djakarta, Pyongyang, and Dar es Salaam. To cope with the increasing traffic, Canton's White Cloud (Pai-yun) airport was enlarged and modernized in 1964.

Telecommunications, which previously were restricted to the cities, now extend not only to the county capitals but to all communes and to 95 percent of the production brigades as well. Mail from Canton can reach half the province within the same day, while the slowest deliveries do not take more than four days.

Cultural life. Kwangtung has long been noted for the distinctive cultural traits of its people, as evidenced by the variety of dialects spoken. Kwangtung is famous for its two types of local opera: the Yueh Opera and the Ch'ao Opera, which are popular among the Cantonese and Fukienese communities, respectively. Kwangtung also has some characteristic puppet plays. The hand puppets of Canton are distinguished by their size—they are between three and four feet high—and by the beautiful carving of their wooden heads. Many places in Kwangtung have distinctive forms of folk art; the woodcuts of Ch'ao-an and the stone engravings of Shun-te may be mentioned as examples.

Cantonese food is widely recognized as among the best in China. Living in a coastal province, the people are particularly fond of seafood. Especially in winter, the "big-headed fish" (tench) is often served raw in a fish salad—a departure from habitual Chinese culinary practice. Some other food habits, such as the eating of newborn rats, live monkey's brain, and fried snake, are regarded as highly revolting by most Chinese in other provinces. The chewing of betel nut wrapped in ***Celosia*** (cockscomb) leaves has been popularized by Chinese who have returned from Southeast Asia.

Although following essentially the same annual festivals as people in other provinces, many ceremonies in Kwangtung are associated with characteristic celebrations. During the Chinese New Year the performance of lion dances is especially popular. The Dragon Boat Festival in early summer, during which dragon-boat races are held on rivers and lakes, provides another example.

Religions

Ancestor worship, folk religions, and all the institutional religions of Taoism, Buddhism, Christianity, and Islām coexist in the province, as in most places in China. Among these, ancestor worship has the most pervasive influence. Although some folk religions are national in outlook, others are of a more local character, such as the worship of the goddess of fishing and navigation, T'ien-hou Sheng-mu. With the possible exception of Muslims and Christians, people in Kwangtung are polytheistic, visiting temples or priests of different faiths as occasions demand.

Related to folk religions and Taoism are the various folktales and superstitions among the people. There is, for example, a folk belief among the peasants that the work of repairing river dikes should be restricted to men, since the participation of women in the work might cause the dikes to collapse.

Kwangtung is a province where lineage—an important social institution in China—has been particularly emphasized. This is often reflected in the settlement pattern of lineage groups. The inhabitants of many villages belong exclusively to one or two lineages. In such villages, community and lineage organizations are virtually identical. Conflicts between lineages that once were common often took the form of community strife, with bitter fighting sometimes lasting for long periods of time.

Many of these cultural patterns had begun to change long before the founding of the new regime. Since 1949, however, systematic efforts have been made to change them in accordance with governmental ideology and policy. Ancestor worship and folk religions have been repudiated as antisocial and superstitious. The institutional religions, with the possible exception of Islām, have also been ridiculed, criticized, and finally reorganized. The political and economic reorganization of rural areas that culminated in the establishment of communes has deprived the lineage organizations of much of their structural basis and social functions. On the other hand, many aspects of traditional culture, especially the folk arts and the theatre, have been revived and extolled. Provincial songs and popular plays, all in local dialects, are now flourishing as never before.

BIBLIOGRAPHY. Few published sources deal specifically with Kwangtung. One which describes the immediate postwar situation is KWANGTUNG RECONSTRUCTION RESEARCH COMMITTEE, ***Post-War Kwangtung: Reconstruction and Prospects*** (1947). Focussing on the early developments after 1949 is a useful work by LUANG RHIN-CHOY, ***Economic Geography of Kwangtung*** (Eng. trans. 1958).

Books containing sections on Kwangtung include GEORGE B. CRESSEY, ***China's Geographic Foundations*** (1934), and ***Land of the 500 Million: A Geography of China*** (1955); T.R. TREGEAR, ***A Geography of China*** (1965), and ***An Economic Geography of China*** (1970); NORTON S. GINSBURG (ed.), ***The Pattern of Asia*** (1958); THEODORE SHABAD, ***China's Changing Map*** (1956); TUAN YI-FU, ***China*** (1970); and KEITH BUCHANAN, ***The Transformation of the Chinese Earth*** (1970).

(Y.–m.Y./C.-T.C.)

Kweichow

Located in the southwestern part of the People's Republic of China, Kweichow (Gui-jou in Pin-yin romanization) Province is bounded on the north by Szechwan Province, on the west by Yunnan Province, on the east by Hunan Province, and on the south by the Kwangsi Chuang Autonomous Region. The province measures over 350 miles from east to west and about 320 miles from north to south. It has an area of 67,200 square miles (174,000 square kilometres), accounting for some 1.8 percent of China's total land area. The smallest province in the southwest, it had, at the beginning of the 1970s, a reported population of about 20,000,000, or an average density of 298 persons per square mile. The provincial capital, centrally located at Kuei-yang, had an estimated population of more than 530,000 in 1958.

Kweichow shares the frontier character of other southwestern plateau lands. It is characterized by rough topography, poor communication and consequent isolation, and the presence of many ethnic minority groups.

History. Although the area has been known to the Chinese since time immemorial, Kweichow came under large-scale Chinese influence only in the modem era, particularly during the Ming dynasty (1368–1644), when it was made a province. The colonization policy of the Ming and Ch'ing dynasties encouraged a large number of Chinese immigrants from Hunan, Kiangsi, and Szechwan to move into the eastern, northern, and central parts of Kweichow.

Kweichow's history of rebellion

During the Ch'ing dynasty (1644–1911), when the government decided to replace local chiefs by officials appointed by the central government, struggles broke out between the minorities, especially the Miao ethnic group, and the Chinese. Rebellions and suppressions were so common that there was a saying, "a riot every 30 years and a major rebellion every 60 years." In 1726 at the Battle of Lei-kung Shan (Lei-kung Mountain), more than 10,000 Miao people were beheaded and over 400,000 starved to death. The Pan-chiang Riot of 1797 was said to have been started by the Puyi people, and thousands of them were either burned to death or beheaded. The most important popular revolt against the central government was one led by Chang Hsiu-mei, a Miao, in 1854. He and his followers united with the Taiping revolutionaries, and the joint army with a centralized command that was organized soon controlled eastern and southern Kweichow and won numerous victories under the Miao leaders Yen Ta-wu and Pa Ta-tu. When the Miao were eventually defeated in 1871, however, countless numbers of them were massacred. The most recent revolt, known as the Ch'ien Tung (eastern Kweichow) Incident, occurred between 1941 and 1944 as a result of exploitation and suppression by the war-lord Wu T'ing-chang. Bitter struggles between the Miao and Wu's armies went on until 1944.

The landscape. ***Relief, drainage, and soils.*** The Kweichow Province is part of an old eroded plateau, variously known as Kweichow Plateau or Yunkwei Plateau, which connects with plateau areas in Yunnan Province. Situated between the mighty Plateau of Tibet and the hilly regions of Hunan and Kwangsi, the plateau forms part of a continuously ascending profile of the southwest, its altitude increasing from about 2,300 feet in eastern Kweichow to about 6,600 feet in the west. The Szechwan Basin to the north and the Kwangsi Basin to the south are both the results of faulting. The entire terrain of Kweichow thus slopes at a steep angle from the centre toward the north, east, and south. In areas adjacent to Szechwan and Hunan in the north and east, the elevation is about 2,300 feet, while the province's southern slopes descend 1,600–2,000 feet into Kwangsi. Accordingly, rivers in the province flow in three directions, north, east, and south.

Karst features of the landscape

The plateau, which is mostly of limestone and basalt, has undergone complicated and extensive folding, faulting, and stream erosion and consequently has abrupt relief, an example of which is the famous Huang-kuo-shu Waterfall in the southwest. Incised valleys, steep gorges, and cliffs are very common. In the limestone areas, the landscape is karstic (characterized by precipitous slopes, abrupt, protuberant mountains, caverns, and subterranean streams). Only the anticline (upfold of stratified rock) and syncline (downfold of stratified rock) of the plateau in central Kweichow are broad and relatively flat.

Most of the rivers in Kweichow are the upper streams of various large rivers, such as the Yangtze River (Ch'ang Chiang) and the Hsi Chiang. The abrupt change of gradient, the great fluctuation in the flow volumes, and the numerous rapids and reefs make them unsuitable for navigation, though they may have enormous hydroelectric-power potential.

Because of the high humidity, a yellow soil with a yellowish-brown subsoil originated from sandstone, shale, and clay, constitutes the largest area in the province. In the limestone area in the south, there are broad areas of red soil. In the west, the red soils are originated from basalt and sandstone and developed under a relatively drier climate.

Climate. Kweichow enjoys a mild climate with warm summers and mild winters. Kuei-yang has a mean July temperature of about 76" F (24" C), lower than that of any other cities to the east on the same latitude. This is due to its high altitude and the cloudiness of the summer months. In winter, cold air from Siberia cannot easily reach Kweichow because of the barrier effect of the

Tsinling Shan (Ch'in Ling or Tsin Ling in Chinese) to the north of the Szechwan Basin. In spite of its high altitude, Kweichow thus has few snowy days and even fewer freezing days. The mean January temperature at Kuei-yang is about 40° F (4° C).

Rainfall is fairly uniform and plentiful, with an annual average of 31–51 inches, decreasing toward the north and west. The southern and eastern parts of Kweichow are open to the influence of the moist maritime air mass in summer. For the same reason, there is a summer maximum in rainfall, averaging 45 percent of the annual total. About 25 percent falls in spring and 23 percent in autumn. Typically, the province has high relative humidity, lengthy cloudy and rainy days, and little sunshine. The capital, Kuei-yang, has over 260 rainy and cloudy days and the province is said to be without three consecutive rainless days. Most of the precipitation results from frontal activity, though some is a result of convection or condensation.

Forest resources

Vegetation. Partly because of the steep gradient, and partly because of the exposure of limestone, wasteland accounts for nearly 50 percent of the total area. Yet part of the province's natural wealth lies in its forests, The plateau surface is mostly dry and barren, but the peripheral valleys have rich and valuable woodlands. About 9 percent of the land area is under natural forest. There are four main forested areas: the drainage areas of the Ch'ing-shui Chiang in the east, the Jung Chiang in the southeast, the Nan-p'an Chiang and Pei-p'an Chiang in the southwest, and the Ta-lou Shan area in the north.

The forests of the northern valleys, still among the most important in China, consist chiefly of conifers and other trees, such as the tung tree, lacquer tree, camellia, birch, maple, pine, and fir. Forests in the southeast produce camphor, banyan tree, and other broad-leaved varieties. Trees of the southern subtropical valleys typically include willow, cedar, bamboo, and various species of pine and fir. Oak, Yunnan pine, *Hua-shan* pine, and camphor are grown in the west near Yunnan. Cedar, cypress, poplar, and palm trees are also found in the province.

Fauna. In addition to domesticated animals, such as buffalo, horses, donkeys, asses, and pigs, the province's fauna includes leopards, otters, foxes, badgers, tigers, sea otters, and squirrels. In most of the larger rivers such as the Ch'ing-shui Chiang, Tu Chiang, and Wu Chiang, carp and savoury fish are abundant.

Main ethnic groups

Population. Demography. In 1970 the province had a total population of some 20,000,000, of which about 75 percent were Han Chinese. Non-Han Chinese tribesmen account for about 25 percent of the total. An ethnic frontier, Kweichow has a large number of minority peoples. At least 30 different groups have been identified. Of these the most important are the Miao (1,800,000), the Puyi (1,600,000), the Shui (Shui-chia; 200,000), the T'ung (200,000), and the Yi (100,000).

All the minority groups intermingle with Han people. Only at the low hsiang, or village level, can one find any exclusive racial grouping. Generally, very few minority people live in northern Kweichow, particularly in areas north of the Wu Chiang. The Miao are mainly found in southeastern Kweichow, especially in the drainage area of Ch'ing-shui Chiang and in the Miao Ling (mountain range). The Chuang people live near the Kwangsi Chuang Autonomous Region. Most of the Puyi live in south central and southwestern Kweichow in the P'an Chiang drainage area, including the suburbs of Kuei-yang. The T'ung are found mainly in the southeastern areas adjacent to Hunan Province and the Kwangsi Chuang Autonomous Region. The Shui concentrate in southern Kweichow, around San-tu and Li-po, while the Yi, who once were rulers of this frontier region, are scattered in western Kweichow. The Muslim Hui in Kweichow migrated there from Yunnan in the late Ch'ing dynasty after the defeat of a local rebellion. They are found chiefly in towns and cities along the main lines of communication in western and southern Kweichow, especially in Wei-ning.

About 88 percent of the total population is rural, and agriculture is the chief occupation. Rice cultivators dominate the peripheral valleys of the plateau. On the plateau itself, the Miao practice a more primitive form of agriculture, growing subsistence upland crops including wheat, maize, and potatoes. Most of the Puyi live on level lands in the valleys and cultivate rice. While the T'ung are experienced lowland rice cultivators, they are also skillful in forestry and in growing upland crops. The Shui, living together in large families and tribes, are also rice cultivators. Besides growing upland crops, the Yi undertake animal husbandry.

The density of the rural population in general decreases from north to south. It averages 500–800 persons per square mile in the valley lowlands and 250–500 persons per square mile on the lower river courses in northern Kweichow. Further south, the density decreases to an average of only 125 to 250 persons even in the basin areas, and to less than 125 on the southern mountains.

There are few cities in Kweichow. Kuei-yang is the largest (1958 population: 530,000), and Tsun-i (with some 200,000 people in 1958) is a poor second. Only two other cities had populations of more than 50,000 each, namely, Tu-yiin (1959: 80,000) and An-shun (over 50,000), while another six cities had between 20,000 and 50,000 people each (Pi-chieh, Ch'ien-hsi, Ch'ih-shui, T'ung-jen, Kuei-ting, and Tu-shan). Most of these cities are the sites of provincial, special district (chuan-ch'ii) people's councils, and are the economic and, often, communication centres for the various regions of the province.

Linguistic patterns. Chinese is the common language of the Han Chinese and the Muslims in Kweichow, Mandarin being spoken almost exclusively by the former group. Among other minority peoples, only the Miao, the Chuang, the Yi, and the Puyi have their own languages; the newly-created Yi and Chuang scripts are increasingly used by the Yi and the Puyi respectively. The spoken languages of the Puyi, Shui, T'ung, and Chuang are common to the Sai-Chuang group; those of the Miao and the Yao belong to the Miao-Yao group, and that of the Yi to the Tibeto-Burman group.

Administration and social conditions. Government structure. Administratively, Kweichow is divided into one subprovincial municipality (shih), Kuei-yang; five special districts (chuan-ch'ii), An-shun, Hsing-i, Pi-chieh, Tsun-i, and T'ung-jen; and two autonomous districts (tzu-chih-chou), the Southeast Kweichow Miao-T'ung Autonomous District and the South Kweichow Puyi-Miao Autonomous District.

Below the subprovincial level, there are three hsien- (municipal people's congress) level cities (shih) of Tsun-i, Tu-yiin, and An-shun; 69 counties (hsien); and nine autonomous counties (tzu-chih-hsien).

By September 1968, when the turmoil of the Cultural Revolution had subsided, all of the special districts, autonomous districts, subprovincial and hsien-level cities had established Revolutionary Committees. In May 1971, the re-formation of a provincial party committee in Kweichow was announced. The Kweichow Provincial Party Committee was the 21st established in China's 29 administrative regions up to that date.

Health. Since 1949 great strides have been made in the field of public health. Between 1952 and 1955, the province established 118 district health stations, 35 mother-and-child-care centres, and 250 maternity homes. Health-work teams were established, and nearly 3,000 health-work cadres were trained and organized.

Medical education

In 1953 the Hua-ch'i Workers Convalescent Hospital was completed with a capacity of 100 beds. A year later, it was modified to become a preventive medical institution. In addition, the Kuei-yang Medical College turns out doctors and medical workers. Its affiliated hospital not only provides teaching and practical facilities to the college but also takes care of 50,000 outpatients and some 3,000 inmates annually.

Education. In 1954 there were over 879,000 primary school students, 25,800 secondary students, and 2,032 college or university students, representing an 84 to 350 percent increase over the pre-1949 peaks at the three different educational levels.

There are nine institutions of higher learning in

Kweichow. In addition to Kweichow University and the Kuei-yang Medical College, there are the Kweichow Agricultural Institute (which, founded in 1939, is perhaps the oldest of the nine), the Kuei-yang Normal College, the Kuei-yang Nationalities Institute (for training members of ethnic minority groups), Kweichow Chinese Medical College, Kweichow College of Forestry, Kweichow College of Engineering, and the Kuei-yang Part-Study Part-Work Normal College.

Under the Socialist leisure-time education program, winter and other study classes provided some form of education to a total of more than 1,600,000 peasants between 1950 and 1954.

Economy. *Mining.* Kweichow has rich mineral resources. Its metallic minerals include mercury, manganese, zinc, lead, antimony, aluminum, copper, iron, and gold. Its nonmetallic minerals include coal, petroleum, oil shale, phosphate, gypsum, arsenic, limestone, and fluorite.

Extractive industries are consequently very important in Kweichow. Mercury, of which distribution is widespread and reserves are great, is mined at T'ung-jen; manganese at Tsun-i and T'ung-tzu; zinc and lead in many areas (but processed at Ho-chang); antimony at Ch'ing-lung and Tu-shan; coal at Kuei-yang, Shui-ch'eng, Tu-yün, and T'ung-tzu; and iron ore and phosphate at Kuei-yang, Weng-an, and Tsun-i.

Manufacturing. With a sufficient local supply of raw material and fuel and an expanding market in southwestern China, an iron-and-steel industry has been developed in Kuei-yang, Tsun-i, Shui-ch'eng, and, more recently, in Wei-ning and Tu-yun. Machinery manufacturing has been established at Kuei-yang, Pi-chieh, Tsun-i, Cheng-an, and other towns. Mining machinery, agricultural implements, and irrigation equipment are among the main products. Kuei-yang produces steel-rolling machines, steel-smelting converters, and other smelting equipment.

Based on the local supply of phosphate and other raw materials, such as by-products of the iron-and-steel industry, the chemical industry in Kuei-yang and Tsun-i produces chemical fertilizers, soda acid, and other chemicals. Both cities also have a textile industry. Silk textile production is concentrated in Tsun-i; cotton textiles are also produced. Again based on abundant local supply of raw materials, paper mills are found in Kuei-yang, Ch'ih-shui, K'ai-li, and Tu-yiin.

Kuei-yang also has cement, cigarette, glass, battery, and food-processing industries, while Tsun-i produces matches for southwestern China. Tu-yiin produces leather and leather goods, and Hsing-i specializes in sugar refining.

Main food crops

Agriculture. Twelve percent of Kweichow's land area is under cultivation. Over 75 percent of the cultivated area is under grain crops, particularly the rice and corn (maize) which together constitute 90 percent of the total grain output. The rest is under wheat, barley, sweet potato, potato, oat, and broad bean, etc. Due to increases in output, from 6 to 10 percent of the grain production is available for export to other provinces.

Approximately 2,000 square miles and increasingly more of the cultivated area are under industrial crops, of which the most important is rapeseed, followed by cured tobacco, peanuts, cotton, sugarcane, and sesame. Kweichow is the second largest producer of rapeseed in China, and its cured tobacco is well known. In spite of increasing production, Kweichow is never self-sufficient in cotton. Its small production of cotton and other fibres, such as ramie and raw silk, is mainly for local consumption.

Timber and other forestry products are plentiful. Among all the provinces of China, Kweichow ranks first in the production of raw lacquer and third in tung oil. Other important forestry products include camelia oil, cypress oil, gall-nut extract, and lichens.

Kweichow is also known for its production of *mao-tai* liquor—made from wheat and kaoliang—which has won a number of international prizes. Its annual production of about 300 tons is split between domestic consumption and foreign export.

Transportation and communications. *Rivers.* River transportation is of little importance in Kweichow due to the presence of reefs and rapids. The section of Ch'ih-shui Ho between Yu-tang Ho and Erh-lang-t'an is navigable for junks of two to eight tons and the section below Erh-lang-t'an for craft of 50 tons and shallow-water steam tugs. Wu Chiang, the largest river in Kweichow, is another prosperous waterway. The section between Ssu-nan and Kung-t'an, Szechwan Province, is a waterway of 132 miles, navigable for shallow-water steam tugs. The section between Ssu-nan and Shui-k'ou is navigable for junks of 12–40 tons. In addition, the Wu Shui, Chin Chiang, Sung-tao Ho, Ch'ing-shui Chiang, and Tu-Chiang are partially navigable.

Roads. In terms of highway transportation, the province is relatively well-developed. Since 1958 all counties have been connected by roads, the majority of which are all-weather roads. The main interprovincial and intraprovincial arteries are: Kweichow to Kwangsi, Kweichow to Yunnan, Szechwan to Kweichow, Hunan to Kweichow, Ts'e-heng to San-tu, Ch'ing-chen to Pi-chieh, Tsun-i to T'ung-jen, and Kuei-yang to Lo-tien. Prior to the completion of the Kweichow to Kwangsi railway, the Kweichow-Kwangsi highway was the principal freight and passenger route to Kwangtung and Kwangsi and to Central and East China.

Railroads. The Kweichow-Kwangsi railway, opened to traffic in 1959, was built on an old roadbed. Since its completion, it has been responsible for about 40–50 percent of the province's grain shipment. A few years later, railway-transportation service was greatly extended by the completion of the Kweichow–Yunnan railway (about 1963), connecting Kuei-yang with Kunming, and the Kweichow–Szechwan railway (1964 or later), connecting Kuei-yang and Tsun-i with Chungking and Ch'eng-tu. When the Kweichow–Hunan railway is completed, Kweichow will have direct rail links with all of its four neighbouring provinces and will be completely linked up with the national railway network.

Folk dramas and dances

Cultural life. The minority peoples in Kweichow are among the most artistic and musical in China. The Han Chinese also have a long and mixed cultural background. Various types of folk dramas with varying degrees of elaboration, some of which are combined with folk dances, are popular among different nationalities in different areas. Some of the Han Chinese folk dramas, *hua teng* ("Flower Lantern") in northern Kweichow and *ti-hsi* ("Floor Plays") in southern Kweichow, are also popular among the minority groups. Buffalo fighting is part of the festival activity over the New Year, especially among the Miao, Yao, and Chuang peoples. The Miao often sing of their revolutionary history and heroes, and both the Miao and the T'ung folk songs are well known. Embroidery and paper cutting are both important forms of folk art among all minority peoples. The Puyi and Ch'ilao are particularly known for their wax dyeing, the Miao and Puyi for their intricate, coloured cross-stitch work, and the Miao for their heavy silver ornaments.

Since 1949 the Propaganda Bureau of the Provincial Party Committee, the cultural and educational departments of the provincial government, and the Provincial Cultural Workers Union have been responsible for most of the cultural affairs in Kweichow. The Union organizes training classes for artists and publishes literary and art journals. Newspapers and periodicals, such as *Hsin Ch'ien Jih Pao* ("New Kweichow Daily"), *Kuei-chou nung-min* ("Kweichow Peasants"), *Kuei-chou Ching-nien* ("Kweichow Youths") and *Kuei-chou Chiao-yu* ("Kweichow Education"), are published by the Kweichow People's Press.

Since July 1953 a music work team has been established by the provincial cultural department and has maintained contact with some 135 music workers. Kweichow has 50 cultural centres, 36 cultural stations, two public libraries (one at Kuei-yang and one at Tsun-i), and a museum with a collection of over 22,000 articles.

BIBLIOGRAPHY. KWEICHOW PEOPLE'S PRESS, *Tsu kuo ti Kuei-chou* (1955), is an important sourcebook on Kweichow (in Chinese), covering its physiography, nationalities, mineral resources, industries, agriculture, transportation, water conservancy, commerce, and art. The latter part of the book is

devoted to its cities, legends, historical incidents and heroes, and its scenic and historical spots. SUN CHING-CHIH (ed.), *Economic Geography of Southwest China: Szechwan, Kweichow, Yunnan* (1962; orig. pub. in Chinese, 1960), provides an adequate account of the province's economic geography.

(C.K.L.)

Kyōto

For more than 1,000 years—from 794 to 1868—Kybto was the capital of Japan and the place of residence of the Imperial family; and it remains today the principal centre of Japanese culture. It is also the heart of Buddhism in Japan and is a leading manufacturing source of fine textiles and other traditional Japanese products. It is located on the island of Honshu some 29 miles northeast of the great industrial city of Osaka and about the same distance from Nara, which, like Kybto, is another ancient centre of Japanese culture. The capital of the Kyoto Urban Prefecture (Kyoto-fu), in Kinki district, is the city of Kybto, which had a population of about 1,420,000 in 1970.

Plan of the ancient city

History. The history of the city begins in 794, when the Emperor Kammu transferred the capital from Nara to Nagaoka (south of Kyoto) and then to Kybto, where he constructed Heian-kyb ("the capital of peace and tranquillity"; later known simply as Kybto, "the capital"). The plan of the city, like that of its predecessor, was made in imitation of the capital of the T'ang dynasty (618-907) in Ch'ang-an (modern Sian) China. Following the principles of geomancy, or divination by means of landforms and lines of direction, the emperor's palace and court was located in the northern centre of the town, and the central street ran from the north to the south. The city east of the central street was called Sakyb, meaning the left side of the capital, and the west of the same street was called Ukyb, or right side. The plan provided streets and districts for each craft. The central area of the present city of Kybto nearly follows the original street plan, although the city has since developed far to the east and to the north of the old city.

During the Heian period (794–1185), Kybto was the centre of politics and aristocratic culture. At the beginning of the 9th century, its population may have numbered 500,000—which would have made the city larger than any European city of its day. In the 12th century Japan entered into its medieval era, which was characterized by the growth of the samurai class. The first samurai clan to seize the political leadership of Japan was the Taira (Heike), who placed their government in Kybto. But, when Minamoto Yoritomo, chief of the Genji clan, bested the Taira in 1185, he established the shogunate at Kamakura, near Tokyo, after which Kybto experienced a political decline for a time.

After another civil war, at the close of the 14th century, the Ashikaga clan rose to power and placed the government again in Kyoto. During the Ashikaga period the aristocratic culture of the Heian era blended with the culture of Zen Buddhism, which had developed under the samurai and gave birth to a new cultural period called Muromachi (after the site of the Ashikaga government in Kybto). Many of Japan's magnificent architectural landmarks, such as Kinkaku-ji (Golden Pavilion; destroyed by fire in 1950, rebuilt 1955) and Ginkaku-ji (Silver Pavilion), and its famous landscape gardens, paintings, Nb plays (the classical drama of Japan), the arts of the tea ceremony and flower arrangement, and its pottery industry were created here at this time.

The reign of the Ashikaga government ended after the Ōnin War (1467–77), whereupon Japan entered upon a century of civil war fought between powerful daimyos (feudal lords). The warlord Oda Nobunaga and his successor, General Toyotomi Hideyoshi, however, completed the unification of the country, after which Hideyoshi reconstructed Kybto, which had been devastated during these wars. It was at this time—in 1594—that the Momoyama Castle in the southern part of the city was built. Mainly because of Toyotomi's excellent and aristocratic taste, a brilliant civilization called the Momoyama period flourished in his reign (1582–98).

Tokugawa Ieyasu (1542–1616), founder of the third and last military dynasty (shogunate), made Edo (present-day Tokyo) the site of his government in the beginning of the 17th century, but the Kyoto Imperial Palace still remained in Kybto. Throughout the Tokugawa regime, which lasted until 1868, Kyōto remained one of the three main cities of Japan together with Edo and Osaka.

With the beginning of the modern period in Japanese history—*i.e.*, with the Meiji Restoration in 1868—the capital of the country and the residence of the Imperial families was moved to Edo, which was re-named Tokyo.

The contemporary city. City site. Kybto is located in the northern part of the great Kybto (Yamashiro) fault basin and is surrounded on three sides--north, east, and west—by low mountains. The city covers an area of about 236 square miles (611 square kilometres). Its northern edge lies about 290 feet and the south about 41 feet above sea level. The hilly area to the east of the city was called Higashiyama and includes Mt. Hiei (2,782 feet), bordering Shiga Prefecture. To the north extends the Tanba plateau, called Kitayama; it includes Mt. Mikuniga (3,146 feet). To the west is the Nishiyama region, which includes Mt. Atago (3,031 feet).

Many streams flow down from the mountains, including the Kamo, the Katsura, and other rivers, which run through the city to join the Yodo-gawa to the south.

The climate

The climate is typical of inland Japan; it is hot in summer and cold in winter. The yearly mean temperature of Kybto is about 59° F (15" C); the highest monthly mean (80° F, or about 27" C) is in August and the lowest (38° F, or about 3° C) in January. Average yearly rainfall is about 62 inches, with much of the rain arriving in summer (June, July, and August).

The streets of Kybto cross at right angles in a checkerboard pattern, in accordance with the original Heiankyo plan. The centre of the present city, however, has moved northeastward since the time of the original city.

In the centre of the town, Shijb street runs east to west, and Kawaramachi and Karasuma streets run from north to south. The intersection of Shijb and Karasuma streets is the centre of the business district, while the area around the intersection of Shijb and Kawaramachi streets is the shopping and amusement centre. Most of the old city downtown consists of small stores, workshops, and residences, all standing side-by-side. There are many Buddhist temples and Shintb shrines, whose old wood roofs jut through the city's many trees. There are also several parks and small public squares.

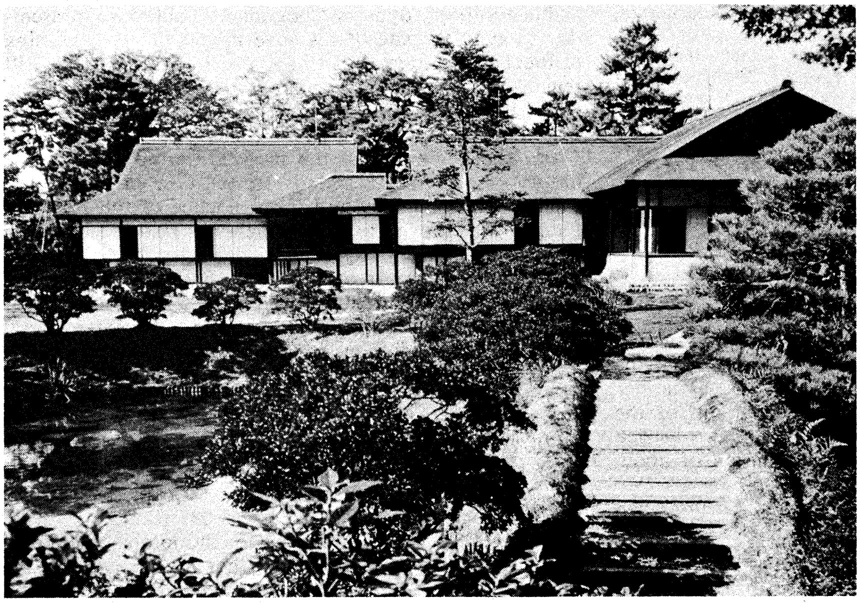
Although the city is divided into nine wards (ku), in popular usage the central part of Kybto is known as the Rakuchu district, the eastern part as Rakuto, the north as Rakuhoku, the west the Rakusei, and the south as the Rakunan district. Suburban communities include Kameoka, a town in a small basin to the west, while to the south the city of Uji lies on the Uji-gawa (a headstream of the Yodo-gawa).

Early transport routes

Transportation. Kybto has always occupied a key position in Japan's transportation system. In the Tokugawa period (1603–1867), it was the terminal of the Tōkaidb Road, which ran westward from Edo, and it was also linked with Osaka by water transport on the Yodo-gawa.

The New Tbkaidb Railway Line runs parallel to the old Tōkaidō Line from Tokyo to Osaka via Kybto. Between Kybto and Osaka there are two tramcar lines. The journey between the two cities takes 40 to 50 minutes, and many who live in Kybto commute to Osaka daily. Kybto is linked to Osaka and Nagoya by the Meishin Expressway, which continues from Nagoya to Tokyo as the Tbmei Expressway. Many national roads also enter the city, including routes running between Osaka, Nagoya, and Tokyo and the main road to Nara. Within the city itself streetcars and buses provide the most serviceable means of transportation.

Demography. In 1970 Kybto had a total population of about 1,420,000 and was the fifth largest city in Japan. The annual rate of population increase was about 1 percent, and the population density was about 6,000 per square mile. The highest density occurs in the central part of the city, where three wards—Kamigyb, Nakagyō, and Shimogyb—each have a density of more



Katsura Imperial Villa, Kyōto.
Camera Tokyo

than 44,000 per square mile. In all other wards, the density is under 18,000.

Economic life. There are about 16,000 factories in Kybto, but 60 percent of them are very small, having under four employees. This reflects the fact that Kybto industry is characterized by the manufacture of traditional products and that its heavy and chemical industries rank relatively low in the scale of importance.

The textile industry accounts for 53 percent of the total number of factories and provides employment for 34 percent of all workers. Its products include silk fabric and embroidery, for which Kybto is famous. They are manufactured in the Nishijin district, the silk-weaving industry of which dates back to the time of the founding of the capital.

The food and drink industry ranks second in importance, principally because of the brewing of sake (rice beer), mainly in Fushimi ward.

The central part of the city is crowded with many small factories or workshops that produce such typical Japanese goods as fans, dolls, Buddhist altar fittings, and lacquered wares. The heart of Kyōto's famed porcelain industry is the Kiyomizu district.

The central business district

In 1971 the city had about 34,000 stores, more than 80 percent of which were retail outlets. The ratio of wholesale stores to retail is lower than that of other cities in Japan, indicating that Kybto is a consuming city. The more expensive stores are concentrated along Shijb street, the business centre, where the main banks and head offices of many companies are located. Wholesale stores, dealing in fine textiles, silk fabrics, and embroidered products, constitute the national centre for the industry.

Machine and chemical industries are developing in the southern part of the city, in the Rakunan industrial area.

Government. Kybto has two main governmental offices; these are the city hall and the prefectural office, both in the central part of the city. The prefectural governor and the mayor are elected by popular vote.

Education. Kybto is also one of Japan's traditional centres for education and for training in arts and sciences. There are 22 national and private universities and colleges in the city, some of which, such as Kyōto University (formerly Kybto Imperial University), founded in 1897, are among the leading educational institutions in Japan. Dbshisha University, founded in cooperation with the American Board of Foreign Missions, has been a centre of Christian higher education since 1875. In addition, as befits the seat of Japanese Buddhism, there are several Buddhist universities.

Culture. As the ancient capital and the former residence of the emperors, Kybto contains many historical and cultural treasures, including paintings, carvings, fine examples of calligraphy and the industrial arts, architecture, and exquisite Japanese-style gardens. Further, there are a number of major cultural institutions, including Kybto Public Hall, Kybto Municipal Museum of Art, Okazaki Zoological Garden, Kyōto Municipal Commercial Museum, and the Kyōto Prefectural Central Library. The Kyōto National Museum, near the Kyōto railway station, is another landmark; it was erected in 1897 by the Imperial family as the Kyōto Imperial Museum. Kyōto also has two important Nō theatres and one Kabuki theatre, as well as many smaller theatres and cinemas and a sports stadium.

Many of the city's houses are built in the Japanese style, characterized by wooden pillars, tiled roofs, mud walls, and sliding inner doors made of wood and paper. And many of the houses also have small Japanese gardens. These characteristics are even more apparent in the examples of historical architecture. Among the most famous examples are Higashi-Hongan-ji, Nishi-Hongan-ji, Saihō-ji, Tenryū-ji, Ninna-ji, and Daitoku-ji (all of which are Buddhist temples); and Heian, Yasaka, and Kitano Tenjin (Shintb shrines).

The old Kybto Imperial Palace, the present buildings of which date from 1855 but which closely resemble the original that they replaced, is one of the most representative examples of Japanese architecture. Nijō Castle, originally built in 1603 by the first Tokugawa shogun, is the most famous castle of that era. The castle itself, which contains many cultural treasures, is one of the 38 buildings in the city registered as national monuments; in 1939 it was given to the-municipality and has since been open to the public.

The facts that the city is the repository for these treasures and that annual festivals and other seasonal events are held there make Kybto the cultural centre of Japan. The city's physical setting also provides another attraction, drawing many tourists to the city. In 1969 alone, some 20,000,000 tourists, including several hundred thousand visitors from abroad, visited Kybto.

BIBLIOGRAPHY. On the history of Kyōto, see WENDELL COLE, ***Kyoto in the Momoyama Period*** (1967); and R.A.B. PONSONBY-FANE, ***Kyoto: The Old Capital of Japan, 794–1869*** (1956). Descriptive works include ***Kyoto,*** compiled by the city of Kyōto, 2nd ed. (1970); YOSHIKAZU IZUMOJI, *Kyoto* (1963); and TADASHI ISHIKAWA, ***Palaces of Kyoto*** (1968) and ***Imperial Villas of Kyoto*** (1970). The KYOTO TOURIST BOARD, ***Photographic Kyoto*** (1951); and MARTIN HURLIMANN, ***Kyoto*** (1962), are mainly pictorial works. On social life and customs, see RUTH L. GAINES, ***City-Royal: A Memory of Kyōto*** (1953).

(T.O.)

Labour Economics

Labour economics is the study of the labour force as an element in the process of production. The labour force comprises all those who work for gain, whether as employees, employers, or as self-employed, and it includes the unemployed who are seeking work. Labour economics involves the study of the factors affecting the efficiency of these workers, their deployment between different industries and occupations, and the determination of their pay. The article deals with the labour force of contemporary industrialized economies.

Obviously, the economist cannot study the capabilities, jobs, and earnings of men and women without taking account of human nature, social structures, cultures, and the activities of government. These forces, indeed, often play a more conspicuous part in the field of labour than do the market forces with which the economist is ostensibly and chiefly concerned. The difference, for instance, between the rates of pay of a craftsman and the labourer who works alongside him may be governed by custom or determined by an arbitrator or maintained by the bargaining power of a trade union. These determinants may seem to override the forces of supply and demand. but a distinction must be made between the agency by which a given result is brought about and the factors that circumscribe the agency's freedom of action and to that extent determine the outcome. Thus an arbitrator may be very conscious that his award will be effective only if it lies within certain limits. Even when the initial action may be regarded as arbitrary, reactions to it bring market forces into play. The comparative study of wage movements in different periods and countries, moreover, shows similarities and regularities that are more marked than the variety of their settings would ever lead one to expect and that can only be regarded as the product of persistent forces working within an equilibrating system. It is this that justifies speaking of a labour market. Though there is much in labour that can be understood only with the aid of the psychologist, the sociologist, or the political scientist, the forces they study do not supersede the market forces that are the special province of the economist; rather, they provide the setting within which the market forces take effect.

QUANTITY AND QUALITY OF THE LABOUR FORCE

The size of a country's labour force, within a given total population, depends on two factors: the proportion of the total population that is of working age, and the proportion of these who work for gain.

The limits of working age are usually taken to be established by the minimum school-leaving age and the prevailing pensionable age. Allowance must then be made for those persons who continue to work for gain after attaining pensionable age. Typically, some two-thirds of the population of an industrial country lies within these limits.

Activity **rates.** The employed labour force may be analyzed in terms of particular activity rates. An activity rate is the proportion of the whole number in a given age and sex group—for example, females aged 30–34—who work for gain. Among males, activity rates in the earlier years of working age are as a rule low, because so many remain in education and training. Between the ages of 25 and 50, male activity rates approach 100 per cent, but from 50 onward they fall as men begin to retire. The pattern of female activity rates is very different, and has changed greatly in the second half of the 20th century. Formerly, female rates ran higher than male in the earlier years because fewer girls enjoyed extended education, but from the age of 20 onward they fell sharply as women married and withdrew to domestic duties. Women so occupied remain by far the largest contingent of persons of working age but not in the labour force. Since World War II, however, it has been less usual for women to leave paid employment immediately on marriage. A fall in the age of marriage, moreover, together with a smaller size of family, has enabled many women to return to paid work in their 30s, and female activity rates have come to show a second peak between the mid-30s and the mid-40s, after which they decline more steeply than male rates. From these various activity rates there emerges an overall proportion of the gainfully occupied among all of working age that is typically in the region of two-thirds.

Quality of labour. The quality of the labour force depends on education and training, physique, and health. There is evidence that physique has been greatly improved by increases in the standard of living in the 20th century. Because of the reduction in family size, this rise has been even more marked for children than for adults, and the effects have been seen in the greater height and weight attained by children at a given age. The beneficial effects of stronger physique on health have been enhanced by the advance of medical knowledge and the increased availability of medical services. Better health has raised productivity by a reduction in absenteeism and by a prolongation of the working life during which the economy reaps the benefit of the education and training the worker has received.

Education and training

Education and training can be regarded as a kind of investment, and the rate of return it yields can be estimated. The amount of the investment is the value of the student's use of resources—buildings, equipment, and instructors—together with the output that the economy would have enjoyed from his work had he been gainfully occupied rather than studying. The yield, in turn, is calculated by assuming that the average subsequent earnings of those who completed a given course of education, compared with the average earnings of those who stopped just short of it, provide a valuation of the increase in productivity that the course confers. From this difference in earnings there must be deducted the contributions to the sinking fund required to replace the amount of the investment by the end of the student's working life. The net yield so calculated can then be expressed as a rate of return on the investment. Estimates suggest that this rate of return is not less than that generally obtained from investment in physical capital. They also indicate that a great part of the productive resources of the economy consists in the education and training embodied in its labour force.

Though estimates of this kind are subject to some objections in principle, they do serve a useful purpose in stressing the potential of education and technical training in raising productivity and the risk of investing too little in them relative to other forms of investment. There is no less a risk of underinvestment in training in industry. The great obstacle there is that the employer is not assured of retaining the services of workers in whose training he has invested. In traditional apprenticeship, the employer's cost of training was offset by the apprentice's binding himself to stay with the employer at a low wage for a number of years. But apprenticeships persist today only in limited fields. Governments have increasingly exerted themselves to improve the training of non-apprenticed entrants, as well as that of older workers in need of retraining.

DEPLOYMENT OF THE LABOUR FORCE

Shifting patterns of employment

The contribution of education and training to economic development is apparent in the changes that have taken place in the deployment of labour in the developing economies. When the deployment of the labour force is followed over a period of time, certain patterns appear. One of these arises from changes in methods of production. In farming, improvements in technique and equipment have made possible an increasing output from a declining labour force. In industry, the extension of research and development, the increased complexity of products and equipment, and new methods of collecting, storing, and processing information, along with other developments of management procedures, have all acted to increase the numbers of administrative, clerical, and technical workers relative to manual workers. A second course of change has affected occupations linked with particular industries, when those industries have contracted or expanded as compared with others. Coal mining and cotton textiles are examples of contraction. The service industries, on the other hand, have expanded: a greater proportion of household expenditure is devoted to services; education has extended; governments have provided more social services. A third course of change has its origins in relation to supply. Domestic services, for instance, have contracted because improved education and the opening up of other occupations to women has enabled many to take up work that they prefer. The aggregate effect of these changes has been to decrease the relative number of manual workers in the labour force and increase that of the administrative, clerical, and technical. One general tendency is that as standards of living rise the service industries absorb a greater proportion of the labour force, because the extension of the demand for their output is not generally offset, as in manufacturing, by a progressive reduction in the amount of labour required to produce a given output.

The many far-reaching changes that have come about in the relative numbers in different occupations and industries have called for corresponding changes in the training and allocation of young entrants to employment and for the movement of workers already in employment to other kinds of work and, often, other places. Though part of this adaptation has been unplanned and undirected, a number of governments have undertaken to foster the process of adaptation by a labour-market policy. One means of applying this policy is the provision of information to job seekers as to vacancies immediately available, and to workers at large as to the prospects and requirements of particular occupations. Labour-market policy also tries to guide entrants toward those occupations for which an expansion of demand is expected. One way of doing this is by promoting the training and retraining of selected persons for selected occupations. The function of retraining may be extended, as in Sweden, to offer all workers opportunities to qualify themselves for better-paid jobs throughout their working lives.

FIXING RATES OF PAY

Customary rates

Wages may be fixed by collective bargaining between unions and management or by individual bargaining between worker and employer or simply by custom. When the status of wage earner became distinguished from other forms of labour, it was marked by the existence of an individual agreement about the rate of pay between each wage earner and his employer. The law still recognizes the individual contract of service even where the rate of pay has been fixed collectively. In earlier days there was often not even individual bargaining, because customary rates of pay prevailed that might be unchanged for many years at a time. In southern England, for instance, the prevailing rate for building craftsmen remained at sixpence a day for 120 years after 1412; for most of the 500 years after 1412, the building craftsman's rate was half again as great as the labourer's, or nearly so. After industrialization had set in, custom continued in some measure to regulate rates of pay and to protect workers who entered into individual agreements. But its sway was much less extensive: from time to time rates changed. Though there was at first no reference to the cost of living, when price increases were general and sustained, there must have been informal understandings among the wage earners of a locality that each in making his own agreement would hold out for a higher rate. At times of increased demand for labour, moreover, the employer would have to offer a rate sufficient to attract and retain the wage-earners against the competition of other employers. The necessity of holding needed labour is today the governing factor for employers who have workers with whom they do not negotiate either collectively or individually—generally clerical and administrative workers. Individual bargaining also persists in those higher appointments in which the worker has access to a number of potential employers, and the employer will have to pay more if he is to secure a worker of higher quality.

But frequently, where the safeguards of both custom and of competition for workers have been missing, workers have felt the need to combine in order to bargain collectively. The force of custom declined as industrialization created new jobs and moved workers into new localities. Business fluctuations brought unemployment so that instead of employers competing for labour, workers were often competing for jobs. Thus industrialization has been universally associated with the rise of trade unions.

Forces making for uniformity

Trade unions and bargaining areas. A main purpose of the trade union was to maintain a minimum rate of pay for its members, a purpose that led unions to extend or delimit both their membership and the number of employers with whom they bargained. The starting point was typically the club of craftsmen in a certain locality, concerned to ensure that none of its members worked for less than the rate it recognized from time to time as a minimum and to raise that rate when opportunity offered. By bringing all who worked in the same craft and district into membership, the club could reduce the risk of their bidding against each other; and if it could also limit the number entering the craft—by controlling the number of apprentices—it would be more likely to be able to raise the rates. But since it was still likely to be subject to the competition of members of the same craft coming in from other places, and some of its own members might move in search of work; it had an interest in extending its coverage over all members of the craft throughout the labour market. If the labour market was not coextensive with the product market, however, the union might still find itself exposed to the competition of workers at a distance if these worked at lower rates and so enabled their products to be sold at lower prices. Thus there was reason to extend the coverage of the union up to the boundary of the market for the product, though it was not practicable to organize workers in other countries. But the union would see no advantage in bringing men of other occupations into membership: on the contrary, it was felt that one could expect employers to concede a rise more readily if it would have to be paid to only a restricted membership. What has been said here of the craft union applies to all unions insofar as their aim is to maintain and raise the pay of members of a given occupation: the pursuit of that aim will lead them to embrace all the members of the occupation throughout the market for their product and to establish a basic rate throughout this bargaining area.

The reactions of employers both reinforce and modify this tendency. The ability of any one employer to pay a given rate will depend largely on what rates are being paid by other employers who compete with him in the product market. When competition is close and labour costs are a substantial proportion of total costs, all employers selling in a given product market have a strong inducement to negotiate only through an employers' association that embraces them all. Most employers' associations are in fact industry-wide, though some are limited to particular regions or sectors of an industry. Employers also know that what is conceded to employees in one occupation will commonly be demanded by those in others, unless they are divided by such a gulf as used to separate the manual from the clerical workers. Employers therefore commonly prefer to reach an agreement with all their workers in common, and may make this a condition of negotiation. They thereby put pressure on occupational unions either to extend, amalgamate, and divide up until they form industrial unions each embracing all the manual workers in a given industry, as the Swedish unions have done, or to enter into confederations that provide all the unions having members in a given industry with a common front for the purpose of bargaining—the course followed by British unions.

Many semiskilled and unskilled workers are unable to seek bargaining advantage by restricting the membership of their unions to one defined occupation: they have to seek it rather through the accumulation of funds and the force of numbers—for them, "unity is strength." Some unions have therefore adopted the principle of industrial unionism from the outset, in accordance with the tendency noted above toward establishing industry-wide bargaining areas. Others, the general unions, have set out to recruit workers from every occupation and industry; but for bargaining purposes they have commonly had to act on behalf of their members in each industry separately. In any clash between the forces delimiting the bargaining area and those delimiting the trade union, the former generally prove the stronger.

Effects of collective bargaining. Collective bargaining developed with the growth of trade unionism, especially from 1890 onward. It impinged upon labour markets in which the trend of money wages was upward: in years of good business, money wages generally rose, and though in the years of falling or low activity they were often cut, the cuts were generally smaller than the preceding rises had been.

Levelling of pay rates. A first effect of the extension of collective bargaining was to reduce pay differences, which had been large, between the wages a given grade of labour received at any one time in different regions and in different firms in the same region, and even between one man and another under the same employer. The unions at first had to accept the prevailing regional differences, but their pressure to bring up the lower-paid regions has reinforced the effect of improved communications and information in reducing these differences greatly, especially since World War II. Assurance of "the rate for the job" raised the wages of particular groups or individuals who lacked access to alternative employers, either spatially or because of their lack of information and mobility. In general, the extension of collective bargaining brought about greater uniformity in the rates of pay received by workers of a given grade, and it did so by raising the lower rates.

"The rate for the job"

Collective bargaining has also affected the forms in which improvements in pay are realized. It has borne particularly on those parts of the terms and conditions of employment that of their nature require to be regulated collectively. Chief among these are the hours of work. The extension of fringe benefits, such as insurance and pensions paid for by the employer, has also reflected trade union pressure.

Other effects. Studies of differences between the movements of wages in unionized and non-unionized sectors of employment, especially in the United States, have brought out three other effects of the extension of collective bargaining. One is an impact and once-for-all effect: the introduction of collective bargaining has raised the wages of the workers concerned, relative to the general level prevailing around them, by some 10–15 percent. A second effect has been in the timing of changes: when wage rises were the order of the day, unionized workers achieved them earlier than non-unionized; and when the market was moving the other way, cuts of unionized workers were put off longer. When the cost of living has risen rapidly, as in wartime, the unionists' ability to secure compensatory rises in money wages more promptly promoted the extension of unionism, especially among white-collar workers who had previously stood aloof from it. The third effect has been in the ability not only to defer wage cuts in depression but also to reduce their amount. In the United States, for example, the differential between wages in the unionized and non-unionized sectors was at its highest in the 1932 depression trough. A major effect on the general level of pay in terms of purchasing power and on its share in the product of industry seems to have stemmed from the resistance to pay cuts in the world economic depression of 1921: though pay was cut severely, often after protracted struggles, it could not be brought down as far as product prices had fallen, and in more than one country the distribution of the product of industry between pay and profit seems to have been permanently shifted.

Impact of the union on pay rates

In these ways, by raising the pay of particular workers and by modifying fluctuations in the workers' favour, over a period of time collective bargaining has made the total of pay higher than it would have been otherwise in the same conditions of the market. But the effect has been limited. Before World War II, the movements of the general level of pay continued to depend mainly on market conditions, and the points at which the effects of collective bargaining can be distinguished clearly are fewer than might be expected. Collective bargaining provided the arena in which market forces took their effect, rather than a shelter from or alternative to them. Full employment after World War II, however, greatly modified the bearing of those forces on the collective bargain by engendering the expectation among employers that if they agree to rises in pay that exceed the rise in productivity and so raise unit costs, they will still be able to preserve profit margins by raising the prices of their products, and do this without loss of business, provided only that the initial rise in pay is not greater than what is generally being conceded at the time by other employers. In these circumstances of cost inflation, the rate of rise of money wages depends in great part on the policies of the trade union.

A second limitation is that even where collective bargaining has affected the movement of money wages, it has had only transient effect on the division of the national income between pay and profits. Whatever the course from time to time of rates of pay in money, the pay per man in real terms (*i.e.*, in terms of purchasing power) has risen with remarkable regularity in much the same proportion as output per man, save for the one major exception of the displacement in favour of pay in the early 1920s. This fact is well-established. The reasons for it will be touched on in the last section of this article. Here an explanation may be offered of its consistency with what will seem, especially to the trade unionist, to be equally well established facts—that in a particular case a trade union is able to obtain a rise in pay that would not have come about at that time had the union not entered a claim and backed it with the threat of a strike; that when the rise begins to be paid, it comes out of current profits; and that at that time also the real wage (*i.e.*, the wage in terms of what it will buy) is raised in the same proportion as the wage in money. It is necessary to look beyond these immediate effects. As time goes on, firms will take the opportunity to restore profit margins either by maintaining their selling prices while productivity rises or by raising those prices. The statistical record makes it clear that firms have generally been able to do this, except during phases of downward pressure by the market upon prices, and it is in these phases that the successful prosecution of pay claims is least likely. When market conditions favour such prosecution, moreover, many unions are likely to be moving at much the same time, and insofar as higher rates of pay are followed by price rises, these rises will be general. What holds of one pay rise in isolation does not hold of a number occurring together: the rise in real pay initially conferred by any one rise in money pay will be reduced as the cost of living rises.

Restoration of profit margins

Theory of bargaining. Limitations on the scope of bargaining are also suggested by theory. Collective bargaining can be seen as the reduction of two risks to which the worker is exposed if he bargains individually. There is first the risk that he will find himself merely one of a number of applicants for a single vacancy and that competition between them will force the pay down. Even if he is the sole applicant for the vacancy, there remains the second risk that the job will be offered to him only on terms that he is not willing to accept; in the event of failure to agree, going without the job will inflict more hardship on him than not filling the vacancy will on the firm. Bargaining through a trade union removes the first risk by ensuring that whichever applicant the firm engages it must pay not less than the union rate: in this sense the union exercises monopoly power. Membership in a trade union reduces the second risk by increasing the workers' relative power to change proffered terms by withholding consent: in this sense the union confers bargaining power.

Constraints. The scope of the monopoly power that the union exercises by maintaining the rate for the job may be seen by supposing that this rate is simply announced by the union, which leaves firms to hire as many or as few men as they choose at that rate. In deciding how high it can set the rate, the union must have regard for the consequences for employment. Firms may be able to alter the design of the product and the method of production so as to use less labour. To the extent that they cannot economize in the use of labour and that the pay of this labour enters into the total cost of production, a higher cost arises that firms may be obliged to pass on to their customers through higher product prices; the customers are then likely to buy less from them, especially if there is international competition in the markets for the product, and again employment will suffer. Thus a union that dictates its own terms is still subject to the constraint of the demand curve for the labour concerned. Equally, if the employers dictate the rate of pay, they could not set it so low as to make it impossible to attract and retain the required labour force: they would be subject to the constraint of the supply curve of the labour concerned.

Work stoppages. When neither side dictates the terms and an agreement must be negotiated, failure to agree results in a stoppage that causes losses to both parties. Attempts have been made to develop the pluses and minuses of these losses into a theory of bargaining. If, for example, it is assumed simply that the continuance of a stoppage progressively increases the wish of the parties to end it, and so causes firms to raise and the union to lower the rate at which each is prepared to settle, then the stoppage will end on the day when the two rates have been brought into equality. Further, if the parties agree in their forecasts of how the wish to settle will be affected by the continuance of the strike, they will find it in their interests to reach agreement on what would be the terms of the ultimate settlement without resorting to coercion by stoppage. A more elaborate theory has been developed in which each party is seen as weighing the cost to itself of a stoppage of given length, the benefit to it of a given concession by the other party, and its estimate of the effect of a given extension of the stoppage on the willingness of the other party to make a concession.

When stoppages are unlikely

In practice much more is involved—internal political pressures, for instance, personal prestige, or the tactic of involving the government and public opinion. Many of the costs of a stoppage, moreover, are hard to express in terms of money. But the above three variables must always figure prominently in the parties' consideration. A stoppage is unlikely when on a consideration of these variables it appears that there will be no net gain; this situation will exist when bargaining power is evenly balanced or when negotiation has already brought the parties' positions close together. One party is likely to see a clear advantage in a stoppage only when market forces are working in its favour, and these will have told already in the course of negotiation. In particular the cost of a stoppage will be high to employers when they are busiest, whereas in a recession a stoppage may be a positive benefit to them. Insofar as bargaining power is thus conferred by market forces, it injects no distinct factor into the determination of rates of pay. But bargaining power may also be conferred by determination, loyalty, and leadership on either side. It has also been conferred on trade unions by the expectation, engendered among employers by the experience of sustained full employment, that rises in pay can be covered by higher prices so as to maintain profit margins without loss of business.

Methods of rate setting. Industrialization has greatly increased the variety of the jobs performed, even among workers who work alongside one another. Rates of pay have been adjusted to particular jobs and to the performance of particular workers by a combination of procedures initiated by management and of bargaining pressures exerted by the workers.

Job evaluation. One such procedure is job evaluation, the aim of which is to measure the aggregate requirements of different jobs on a common scale and to use these measurements to adjust the relative rates of pay for particular jobs, always within the range of prevailing rates. The requirements considered include the amount of previous experience and of training that constitute the threshold for entry to the job, the demands the job makes on mental and bodily faculties, the responsibilities it imposes, and any unpleasant working conditions or risks of accident associated with it. Of the various ways of measuring these requirements, the simplest and most common procedure is to begin by allotting a maximum number of points to each requirement, and then assigning each job a number of points within that maximum according as the intensity of the requirement appears in a given job in comparison with others. The totals of the points awarded in this way to the various jobs are then compared with the jobs' rates of pay, and a formula for turning points into money is derived and applied to the total of points for each job in order to obtain the appropriate rate of pay for it. There may thus be as many rates as there are jobs, but it is often found more convenient to draw up a wage structure of, say, seven or eight grades and to put into each grade, at a common rate of pay, all the jobs whose evaluated requirements fall within a certain range.

Graded structure of wages

Merit rating and the "straight piece rate." The adjustment of rates of pay to the performance of individual workers may be made by merit rating, that is, by the payment of higher rates to those whose personal conduct and capabilities are adjudged to be superior to the general run. The most widely applied and far-reaching method, however, adjusts the worker's earnings to the output he achieves week by week. Such payment by results can be applied only where output is measurable, and this may require applying it to all the members of a working group in common and not individually. Where output cannot be measured in physical units at all, indicators of performance such as economy of materials or the running hours achieved by machines for whose maintenance the worker is responsible, may be used as the basis of bonus payments. The simplest connection between output and earnings is provided by the "straight piece rate," under which the worker receives the same payment for each piece or unit he produces; but the connection may also be provided by a formula under which earnings vary less, or in some instances more, than in proportion to output. The common aim of methods of payment by results is to give the worker an incentive to reach and maintain higher levels of output, and experience shows that workers moved from hourly rates to payment by results are likely to raise their output by from 15 to 30 percent.

The piece rate is commonly arrived at after time study, in which the time actually taken to achieve a unit of output is measured by stopwatch, and the observed time is adjusted according to the observer's judgment of whether the worker studied was working faster or more slowly than the average worker might be expected to when habituated to the job. If the adjusted time in a given case was three minutes, the worker on a straight piece rate would be credited with three minutes' pay, at his hourly rate, for each unit produced, even if he produced it in less than three minutes. Obviously there is room for argument here; rates are often the subject of dispute, and many are in practice negotiated rather than fixed by the rate setter alone. A rate that is appropriate when first installed is apt to become less so as minor improvements in materials and methods cumulate. Some firms that looked to methods of payments by results to lower their costs have thus found them increasingly costly as well as vexatious as time goes on, and some have moved back to hourly rates with reliance on supervision or the payment of a high hourly rate subject to the maintenance of a certain output in order to maintain productivity.

Single-plant contracts and pay adjustments. The rates paid in a given plant or firm may be the subject of collective bargaining when the bargaining area is not wider than the firm; negotiations over a wider area can deal with them only when they are relatively few or are capable of being classified in a limited number of grades. One of the advantages of the plant contract prevalent in United States manufacturing is that it enables the wage structure of the plant to be specified and negotiated in detail. United States trade unionism lends itself to this procedure in that in manufacturing the unions are predominantly industrial in scope, so that most of the manual workers in a plant are members of the same union; and the local, or branch, of the union coincides with the plant. In the absence of such arrangements, bargaining over particular rates is likely to go on nonetheless, but in a piecemeal and unregulated way. A leading part in the negotiation of particular rates will be played by the shop steward. When a number of unions have members in the same firm, the committee of shop stewards provides a common front, and its convener may be heavily engaged as a negotiator; such a committee acts in considerable independence of the parent unions.

The wage drift phenomenon

A number of possible ways have been indicated in which the effective rates of pay obtaining at the place of work may be raised during the intervals between rises negotiated by recognized procedures: employers may voluntarily extend payments over and above the negotiated rates; the standards applied in fixing rates under payment by results may become looser; any single rate once fixed may become looser as time goes on; and pressures exerted by workers individually or in small groups may enforce claims for higher rates for particular jobs. The outcome of these processes is known as wage drift. It has been fostered by conditions of full employment, which have given the individual worker the indispensability that enables him to press his claim under threat of withdrawing his labour, and has made management more anxious to maintain and raise output than to keep costs down. In a number of countries wage drift has accounted for a sixth or more of the total increase in the payment made for the hour of work. Wage drift is precluded where rates are specified in detail in a collective agreement such as the plant contract and will not be varied until that agreement is renegotiated.

Public regulation of rates of pay. Governments have intervened in three ways to enforce minimum rates for workers who lacked both the protection of trade unions and competition between employers for their services and whose wages in consequence were regarded as needlessly low. One way has been to provide by law that "recognized terms and conditions of employment," such as those reached by collective bargaining for workers of a particular description, shall be applied to all others engaged in the same kind of work. A second way, followed by the United Kingdom since the Trade Boards Act of 1909, and by a number of state legislatures in the United States, has been to set up boards of representatives of the workers concerned and their employers, together with independent members, charged with determining rates of pay and hours of work that will be legally binding as minima on all employers within the scope of the board. The board discusses and negotiates wage claims in much the same way as in collective bargaining, albeit if the parties cannot reach agreement, the independent members have a deciding vote. A third way, followed notably by the United States in its Fair Labor Standards Act since 1938, has been to specify by statute the actual minimum wage applicable to wide categories of employment—the amount set being such that only a relatively small number of workers, namely the lowest paid, are immediately affected.

When such third-way measures were first proposed, critics argued that they would only result in the workers they were intended to protect losing their jobs. In some cases this has happened, as when the United States minimum wage was applied to the needleworkers of Puerto Rico. More often, however, the workers concerned were receiving lower pay than a competitive market would have afforded them—that is, if they had had more access to alternative employers. The first two forms of intervention were calculated to raise the pay of particular groups of unorganized workers only to the extent that it would have been raised by the extension of collective bargaining to cover them. It is a disadvantage of the third method that the pay of all workers below the statutory minimum is raised irrespective of the effect it may have on their continued employment.

Practice in Australia and New Zealand

Arbitration. Another way of regulating rates of pay is a by-product of arbitration systems set up originally as a means of avoiding strikes and lock-outs. In Australia and New Zealand it has become the practice, accepted by both employers and trade unions, to have the main proportions of the wage structure and the movements of the general level of wages determined by the awards of arbitrators to whom these issues are submitted in the form of disputes. In setting rates for particular occupations or industries relatively to others, arbitrators must in practice have regard to what is acceptable to the parties; for even where arbitration is compulsory, its awards would cease to be observed if either party had cause to believe that the terms of the awards were persistently less favourable than it could obtain by its own bargaining power. In regulating the movement of the general level of pay, the arbitrators have more discretion; but the government, and the employers in so far as they meet international competition at home and abroad, will make them aware of the effects of the awards on the level of domestic costs and prices and on the balance of payments.

National incomes policy. Under full employment the rise in effective rates of pay, whether through negotiation or through wage drift, has generally been inflationary in that it has exceeded the rise of productivity. The consequent rise in costs and prices has been disturbing domestically and has been particularly embarrassing to governments that face difficulties in balancing their external payments. Governments in general have been unwilling to check the rise of inflation by applying fiscal and monetary restraints to the degree that unemployment would be substantially raised. In the belief that at least part of the rise is due not to excess purchasing power but to the pushing up of costs and prices, governments have appealed to those who make decisions affecting labour costs and product prices to moderate the rise in pay and profits. Some governments have formulated norms that would in theory keep the general level of prices constant, and would keep the general level of pay rising only at the rate of the expected rise in productivity, allowing, of course, for specific exceptions. Agencies have been set up to apply these principles, but usually only by way of investigation, assessment, and advice. Governments have preferred to rely on the acceptance of the policy in principle by employers and trade unions, and on their efforts to secure its observance by their affiliates. Even where statutory powers of control exist, they have usually been kept in reserve. When the experience of sustained full employment has made employers generally expect that they can raise pay by an inflationary amount without harm to their own businesses, it is hard to persuade workers to accept smaller rises in pay than employers are willing to give. Whatever moderating effect incomes policy may have had upon cost inflation, it has not stopped it. The policy remains in the stage of improvisation and experiment.

THE STRUCTURE OF PAY

Factors in pay, variations

Systematic differences are found in the average earnings obtained in different regions, industries, and occupations. The average earnings prevailing in different regions of a country show a considerable range between the highest and the lowest, even when the same procedures for fixing rates apply everywhere. Much of the dispersion is due to differences in the localization of industry: if the relatively high-paying iron and steel industry, for instance, is concentrated in a particular region, then the average pay of the region will be raised to that extent. But regional differences also exist because work of the same kind frequently commands different rates of pay in different regions.

Such differences may correspond to regional differences in the quality of workers, so that there is no real difference in what is being paid per unit of work. But there may be real differences as well. Such differences are sometimes necessary to maintain the balance of payments between regions; they may also be in some measure a legacy of history and are likely to be reduced as communications improve and labour becomes more mobile. As noted above, this process of reduction has been expedited by trade-union pressures.

Average earnings also vary from industry to industry, and the considerable range that appears can again be attributed largely to differences in the composition of the labour force: such an industry as printing that by the nature of its processes employs a high proportion of skilled workers will on that account alone show higher average earnings than, say, the textile industry, which employs a higher proportion of the semiskilled. The similarity between the structure of earnings by industry in different countries—with printing, iron and steel, and engineering near the top, and textiles and food processing low down—is thus attributable to common processes requiring similar compositions of the labour force (assuming, of course, that the structure of rates of pay by occupation in the given countries is also similar).

Occupational pay theories. This occupational structure, therefore, presents the main object of economic analysis. International comparisons show that the ranking order of the rates of pay prevailing in different occupations is similar in different countries, but that the range, whether between professional and manual occupations or, within the manual, between the skilled and unskilled, is much wider in economies at an early stage of development and diminishes in the course of development. These are the principal observations to be accounted for by any theory of the differences in the rates of pay that different kinds of work command. Several such theories have been propounded.

Relation between esteem and pay

Status. One theory stresses the link between occupations and their status in the community, some having higher status than others. The community believes, according to this theory, that pay should correspond to status; and the rate of pay for each occupation is assigned to it by common consent, reflecting the place it occupies in the hierarchy of esteem. The implications are that the community's discretion is not as a rule subject to other factors such as the market forces of supply and demand and that, if people came to make less distinction of status between occupations, then the rates of pay for different jobs could be more nearly equal. Doubtless, many people do think in the way the theory supposes, feeling it anomalous, for example, when an occupation that is commonly accorded a higher status than another ceases to command a higher rate of pay; and arguments appealing to this habit of thought are sometimes used in support of claims. But the correspondence between status and pay is ambiguous; it is not clear to what extent pay is made to fit the status and to what extent status follows from pay, if not today's pay then yesterday's. Moreover, the occupational pay structures of different countries show more similarity than do their social values. Nor does the theory explain why the pay structure has generally become compressed in the course of development, and the ranking order sometimes inverted, as when some clerical occupations drop below some manual ones: changes that can otherwise be explained as the effect of extended education in increasing the relative supply of more qualified labour.

But if the theory is not acceptable as an explanation of the pay structure as a whole, it does call attention to a factor that appears to affect parts of that structure. One of these parts is that of the higher administrative posts. It is generally accepted that any such post must carry a higher salary than any post below it in the chain of command; and when this chain is long, as it is in a big corporation, the salaries set for the posts in it, and the high level reached at the top, are to be accounted for by this principle. The same theory also suggests a cause of prevailing differences between men's and women's rates of pay. Most women's work is different in kind from men's, irrespective of the fact that it is done by women; and where men and women both do work of the same description, some disabilities attaching to women as employees, in particular the likelihood that they will not stay in the job as long as men, may make them worth less to the employer. But there are some jobs in which these considerations do not apply and in which there is no difference in the productivity of men and women adequate to account for the actual difference between men's and women's rates. The difference seems attributable rather to customary attitudes and valuations: in particular, the assumption that women's productivity is lower in all jobs and also the belief that pay should be proportioned to need (women workers generally needing less than the man who has a family to support). If such factors as these account for differences in pay between men and women where there is no corresponding difference in the work they do or the efficiency with which they do it, one may speculate that the same factors account for some part of the pay differential where the two kinds of job are distinct.

Control of the labour supply

Power. A second theory lays its stress on power: the ways in which organized groups can protect and advance the pay of their members. Any group that restricts entry into its occupation can keep its labour relatively scarce and thereby support the rate of pay that that kind of labour commands. The discussion above of the economic effects of the trade union indicated the circumstances in which a trade union would be able to raise the relative pay of its members by the exercise of monopoly and bargaining power. The general increase in the pay of the less skilled relative to that of the skilled manual worker has been attributed in part at least to the increased unionization of the unskilled. Evidently, policies of organized groups will account for some part at least of the position of particular occupations in the pay structure, but they cannot alone account for the main proportions of that structure: it cannot be supposed that the hierarchy of pay corresponds with and is due to a hierarchy of the power of organized groups.

Value. A third theory treats the differences in pay for different jobs as corresponding to differences in their content or requirements. The simplest form of this theory was embodied in the labour theory of value, whether in the system of Adam Smith or of Karl Marx, by the assumption that different kinds of labour can be reduced to different quantities of "homogeneous labour time," and that rates of pay are then simply proportional to those quantities. Job evaluation, discussed above, purports to condense the varied requirements of each job to a single figure in a common scale in order that the ranking order of the rates of pay of the jobs may be brought into conformity with that of those figures. But the assumption that if two articles are priced in the same currency they must contain quantities of a common substance is gratuitous. The impossibility of establishing the existence of such a substance and measuring the amount of it in any article drives both the labour theory of value and job evaluation into the circular argument of inferring the job content from the rate of pay and then explaining the rate of pay by the job content. That job evaluation is nonetheless useful in practice may be accounted for by its ability to arrive at an estimate of the extent to which the requirements of each job tend to raise the rate of pay it commands in the present state of the market.

Costs and benefits to workers

The supply price of labour. The foregoing directs attention to the supply price of labour to the job—the rate that must be paid if employers are to be able to attract and retain the quantity of labour that they wish to employ at that rate. Entry into an occupation generally imposes certain monetary costs; there may also be subjective costs; for example, in the effort of concentration required by preparation for examinations. The exercise of any occupation may be attended by disadvantages that require monetary compensation or may provide satisfactions and amenities that make workers willing to accept lower pay. For each occupation the various costs and benefits can be set off against the pay, and entrants will choose the one in which the prospective balance of advantage seems greatest. If more workers are to be attracted to and retained in a given occupation with unchanged conditions on the side of supply, the rate of pay in that occupation must be raised relative to others. An extension of supply will work to the opposite effect: for instance, if there is more public provision for secondary and tertiary education, and if rising standards of living enable more families to hear the costs of training, then a given number of workers will come to be available in a given occupation at a lower relative rate of pay. Here is to be found the reason for the occupational pay structure extending over a smaller range in developed than in poor countries and for the reduction in the margins for skill and the relative rate of pay for clerical work in the developed economies during the present century.

A number of considerations thus indicate that the rate of pay in any occupation tends to equality with the long-run supply price of labour to that occupation. In the absence of an extension of supply, a fall in the relative rate of pay of an occupation will bring a check to recruitment, followed by some withdrawal to other jobs of those already in the occupation. A rise in the relative rate of pay needs longer to take effect where proficiency takes long to acquire. Some types of proficiency may be limited by nature, and the rise in the rate of pay that follows on an extension of demand for them constitutes an economic rent; *i.e.*, a payment that is not required to maintain supply. In general, however, given time, the number of proficient workers available to follow a given occupation will be increased by a rise in the relative rate of pay it offers.

It is, then, to differences in the long-run supply price of labour to different occupations that the differences between the rates of pay they offer appear to be mainly attributable; though considerations of status and the power of organized groups set their mark upon the occupational pay structure at particular points.

Marginal productivity and **the** demand for labour. The above statement is incomplete, because the supply price of labour in a given occupation generally varies with the number of workers supplied, and what that number will be depends in turn on the demand for labour in that occupation. The rate of pay must not only be such as will maintain the supply of a given number of workers but such also as will lead employers to employ that number. The theory of marginal productivity analyzes the demand for labour in a given employment in the same way that it derives the demand for any factor of production from the demand for the product and from technical conditions of production that determine how much output will be increased by using one more unit of the factor in the presence of given inputs of the other factors; that is, what the marginal product of the factor in the given setting will be. If the employer seeks to maximize his profit, he will presumably adjust the input of each factor so as to bring the value of its marginal product into equality with its price (see DISTRIBUTION, THEORY OF).

Reality versus theory

The theory of marginal productivity has been challenged on the ground that it is unrealistic. The marginal product can be measured, it is pointed out, only for direct labour; that is, for work in which physical output varies directly with labour input; but here the number of workers required is determined simply by the size of current output. Yet though this holds at any one time, a rise in the relative pay of a given kind of labour puts managers under pressure to change equipment, processes, or the design of the product as opportunity offers, so as to use less of that labour relatively to other factors of production. Insofar as this cannot be done, moreover, the rise in pay will result in a higher relative price for the product, which will tend to reduce sales and hence the number of workers.

With indirect labour, on the other hand, the marginal product is not measurable. It is the function of management, however, to assess the contribution made by such workers, so as to decide whether employing one more at a given rate of pay would yield a balance of advantage to the business, and managers are efficient in proportion as they make such assessments rightly. In general, insofar as managers are concerned to produce a given output at the lowest possible cost, they will avoid paying for a unit of labour if the same payment used to hire more of another factor would result in a greater increment of output; and insofar as they are concerned to increase profit, they will not fail to engage a unit of labour whose contribution to proceeds would exceed its pay. While, therefore, the adjustment supposed by the marginal productivity theory of the demand for labour is seldom capable of being effected exactly, there is no reason to doubt that market forces do operate to check deviations from that adjustment and retain them within a zone of tolerance.

MOVEMENT OF THE GENERAL LEVEL OF PAY

A wage is a price, and the rise of the general level of wages or rates of pay in the course of time has, to some extent, been part of the long-term rise in the general level of prices; that is, of the cumulative depreciation of the purchasing power of money, largely attributable to increases in its quantity. In another way, however, the movements of rates of pay have been an independent cause of the rising trend of prices. At times those rates rose in common with prices under the pull of monetary demand (in times of inflation, during war, or in the rising phase of the trade cycle), but when the demand fell off they were resistant to cuts; and though they were cut somewhat, they commonly remained at a higher level than when the preceding rise began. A graph of product prices shows big falls as well as big rises, and sometimes a falling trend persisting for many years together; but a graph of money wage rates is more like a flight of steps. This characteristic of wage movements puts a floor under prices and provides a higher starting point for the next upward movement, so that the fluctuations of monetary demand impose a rising trend on prices. In addition, the analysis of cost inflation under full employment, noted above, has shown that when employers generally expect demand to be sustained, rises in pay may occur in the absence of excess demand and so initiate rises in prices; and it is possible that the same process may have played some part in the rising phase of the trade cycles of earlier years.

The level of real earnings

The rise of real earnings may be traced by comparing the movements of earnings in money with those of an index number of the prices of the articles on which pay is typically expended. Such comparisons indicate that between 1860 and 1960 the real earnings of manual workers rose fourfold in France, Germany, and the United Kingdom; more than fivefold in the United States; and more than sevenfold in Sweden. In considering the standard of living attendant on these movements, it is necessary also to take account of the prevailing reduction in the size of the family, the complex effects of urbanization on the amenities of life, the effects of changed techniques and deployment between occupations on the strains and satisfactions experienced in work, and the reduction of hours of work. The last element has been extensive: it appears that down to World War II the wage earners of the five countries mentioned, save the United States, gave up from a third to a half of the potential increase in annual purchasing power in favour of a shorter working week and longer vacations.

To the extent that real earnings are measured simply by the quantity of consumables that money earnings will buy, their rise has depended on three factors: productivity, or the output per worker in terms of his own product; the share of this product that accrues to the worker; and the rate of exchange between the worker's own product and the goods and services he buys. In the industrialized countries, the last factor has presented itself largely in the form of the terms of trade between manufactured products and primary products, especially foodstuffs: real earnings have risen faster or slower according as a representative consignment of manufacturers may be exchanged, at the prices of the day, for a greater or smaller quantity of foodstuffs and raw materials. There have also been variations from time to time in the second factor: the share of the product accruing to the worker. The effect of the last two factors, however, has been small in comparison with that of the first, the rise of productivity. The salient finding from the statistical record of the last hundred years is that real earnings per worker have risen very nearly in the same proportion as output per worker.

How this can have come about is a question for the theory of distribution (*q.v.*). Here the comparative stability of the share of the product accruing to the worker may be noted. The statistical record shows this stability in its relation with the behaviour of two other factors: the rate of return on capital, and the amount of capital employed per unit of output. The stability of any two of the three factors implies that of the third.

BIBLIOGRAPHY. Some textbooks that apply economic analysis to labour in the particular circumstances of labour markets are: G.F. BLOOM and H.R. NORTHRUP, *The Economics of Labor Relations,* 6th ed. (1969); A.M. CARTTER and F.R. MARSHALL, *Labor Economics: Wages, Employment and Trade Unionism* (1966); L.C. HUNTER and D.J. ROBERTSON, *Economics of Wages and Labour* (1969); L.G. REYNOLDS, *Labor Economics and Labor Relations,* 5th ed. (1970); and E.H. PHELPS BROWN, *The Economics of Labor* (1962).

The relation of the labour force to the population of a number of countries is surveyed in UNITED NATIONS, *Sex and Age*

Patterns of Participation in Economic Activities, Population Studies no. 33, rept. 1 (1962); and in C.D. LONG, *The Labor Force Under Changing Income and Employment* (1958). Education is treated as a form of investment in G.S. BECKER, *Human Capital* (1964).

The deployment of the labour force is traced historically in S. LEBERGOTT, *Manpower in Economic Growth: The American Record Since 1800* (1964). The mechanism of recent changes is discussed in two OECD reports: P. de WOLFF *et al., Wages and Labour Mobility* (July 1965); and L.C. HUNTER and G.L. REID, *Urban Worker Mobility* (1968).

For trade union structure and bargaining areas, see A.R. WEBER (ed.), *The Structure of Collective Bargaining* (1962). Numerous studies of the effects of collective bargaining are digested in: L.G. REYNOLDS and C.H. TAFT, *The Evolution of Wage Structure* (1956); and in the papers by C. KERR and L.G. REYNOLDS, in *The Theory of Wage Determination,* ed. by J.T. DUNLOP (1957). Theories of collective bargaining are surveyed in A. CODDINGTON, *Theories of the Bargaining Process* (1968); see also R.E. WALTON and R.B. MCKERSIE, A *Behavioral Theory of Labor Negotiations* (1965).

Forms of minimum wage regulation are surveyed in S.B.L. NIGAM, *State Regulation of Minimum Wages* (1955). The experience of a number of countries with incomes policy is reported by D.C. SMITH, *Incomes Policies: Some Foreign Experiences and their Relevance for Canada* (1966); and by H.A. TURNER and H. ZOETEWEIJ, *Prices, Wages, and Incomes Policies in Industrialized Market Economies* (1966). Regional, industrial, and occupational structures are treated in L.G. REYNOLDS and C.H. TAFT, *The Evolution of Wage Structure* (1956); and E.M. HUGH-JONES (ed.), *Wage-Structure in Theory and Practice* (1966). Theories of differences in the rates of pay for different jobs are developed in B.F. WOOTTON, *The Social Foundations of Wage Policy: A Study of Contemporary British Wage and Salary Structure* (1955); and H.F. LYDALL, *The Structure of Earnings* (1968). The movement of the general level of pay, in money and real terms, is recorded and analyzed in France, Germany, Sweden, the United Kingdom, and the U.S. since 1860 in E.H. PHELPS BROWN and M.H. BROWNE, *A Century of Pay* (1968).

(E.H.P.B.)

Labour Law

Labour law is the term used to describe the varied body of law applied to such matters as employment, remuneration, conditions of work, trade unions, and labour management relations. In its most comprehensive sense it also includes old age and disability insurance. Unlike the laws of contract, tort, or property. the elements of labour law are less homogeneous than the rules governing a particular legal relationship. In addition to the individual contractual relationships growing out of the traditional employment situation, labour law deals with the statutory requirements and collective contractual relationships that are increasingly important in mass-production societies, the legal relationships between organized economic interests and the state, and the various rights and obligations related to some types of social services.

Labour law has now won recognition as a distinctive branch of the law within the academic legal community, but the extent to which it is recognized as a separate branch of legal practice varies widely depending partly on the extent to which there is a labour code or other distinctive body of labour legislation in the country concerned, partly on the extent to which there are separate labour tribunals, partly on the extent to which an influential group of the legal profession practice specifically as labour lawyers.

In the early phases of development the scope of labour law is often limited to the most developed and important industries, to undertakings above a certain size, and to wage earners; as a general rule, these limitations are gradually eliminated and the scope of the law extended to include handicrafts, rural industries and agriculture, small undertakings, the self-employed, and office workers. Thus a body of law originally intended for the protection of the working class is gradually transformed into broader welfare legislation.

Statute
versus
contract

The general tendency in the modern development of labour law has been the strengthening of statutory requirements and collective contractual relations at the expense of individual employment relationships. How important these latter remain depends, of course, on the degree of personal freedom in the given society as well as the autonomy of both employer and worker allowed by the actual operation of the economy. In such matters as hours of work, health and safety conditions, or labour-management relations, the statutory or collective elements may define most of the substance of the rights and obligations of the individual worker, while with respect to such things as the duration of his appointment, his level and extent of responsibility, or his place in the scale of remuneration, these elements may provide what is essentially a framework for individual agreement.

The scope of labour law also grows beyond the employment relationship as important parts of it are increasingly made applicable to the self-employed. This has been done partly, as in the case of laws regulating business hours, to prevent the undercutting by the self-employed of fair labour standards for employed persons and partly, as in the case of the extension of old age and disability insurance to the self-employed and the protection of the tenancy rights of small farmers and sharecroppers, to improve the living and working conditions of disadvantaged groups who need such protection as much as employed persons.

HISTORICAL DEVELOPMENT

The origins of labour law can be traced back to the remote past and the most varied parts of the world. While European writers often attach importance to the guilds and apprenticeship systems of the medieval world, some Asian scholars have identified labour standards as far back as the Laws of Hammurabi and rules for labour-management relations in the Laws of Manu; Latin American authors point to the Laws of the Indies of the conquistadors. None of these can be regarded as more than anticipations with no influence on subsequent developments. Labour law as it is known is essentially the child of successive industrial revolutions from the 18th century onward. It became necessary when customary restraints and the intimacy of employment relationships in small communities ceased to provide adequate protection against the abuses incidental to new forms of mining and manufacture on a rapidly increasing scale at precisely the time when the 18th-century Enlightenment, the French Revolution, and the political forces that they set in motion were creating the elements of the modern social conscience. It developed rather slowly, chiefly in the more industrialized countries of western Europe, during the 19th century and has attained its present importance, relative maturity, and worldwide acceptance only during the present century.

Impact of
the
Industrial
Revolution

The first landmark of modern labour law was the British Health and Morals of Apprentices Act of 1802, sponsored by the elder Sir Robert Peel. Similar legislation for the protection of the young was adopted in Ziirich in 1815 and in France in 1841. By 1848, the first legal limitation of the working hours of adults was adopted by the Landsgemeinde (citizens' assembly) of the Swiss canton of Glarus. Sickness insurance and workmen's compensation were pioneered by Germany in 1883 and 1884, and compulsory arbitration in industrial disputes was introduced in New Zealand in the 1890s. The progress of labour legislation outside western Europe, Australia, and New Zealand was slow until after the First World War. The more industrialized states of the United States began to enact such legislation toward the end of the 19th century, but the bulk of the present labour legislation of the United States was not adopted until after the 1929 depression. There was virtually no labour legislation in Russia prior to the October Revolution of 1917. In India the hours of work of children between 7 and 12 were limited to nine per day in 1881 and the hours of adult males in textile mills to ten per day in 1911, but the first major advance was the amendment of the Factory Act in 1922 to give effect to conventions adopted at the first session of the International Labour Conference at Washington in 1919. In Japan rudimentary regulation? on work in mines were introduced in 1890, but a proposed factory act was controversial for 30 years before it was adopted in 1911, and the decisive step was the revision of this act in 1923 to give effect to the Washington Convention. Labour legislation in Latin America began in Argentina in the early years of the century and received a powerful impetus from the Mexican Revolution, which ended in 1917, but, as in the North, the trend became general only with the impact of the world depression. In Africa the progress of labour legislation became significant only from the 1940s onward.

The legal recognition of the right of association for trade union purposes has a distinctive history. There is no other aspect of labour law in which successive phases of progress and regression have been more decisively influenced by political changes and considerations. The legal prohibition of such association was repealed in the United Kingdom in 1824 and in France in 1884; there have been many subsequent changes in the law and may well be further changes, but these have related to matters of detail rather than to fundamental principles. In the United States freedom of association for trade union purposes remained precarious and subject to the unpredictable scope of the labour injunction, by means of which the courts helped restrain trade-union activity until the 1930s. In many other countries the record of progress and regression with respect to freedom of association falls into clearly distinguished periods separated by decisive political changes. This has certainly been the case with Germany, Italy, Spain, Japan, and much of eastern Europe; there have been many illustrations of it, and there may well be more, in the developing world.

Labour codes and ministries of labour were not introduced until the 20th century. The first labour code (which, like many of its successors, was a consolidation rather than a codification) was projected in France in 1901 and promulgated in stages from 1910 to 1927. Among the more advanced formulations affecting the general condition of labour were the Mexican Constitution of 1917 and the Weimar Constitution of Germany of 1919, both of which gave constitutional status to certain general principles of social policy regarding economic rights. Provisions of this kind have become increasingly common and are now widespread in all parts of the world.

Departments or ministries of labour responsible for the effective administration of labour legislation and for promoting its future development were established in Canada in 1900, in France in 1906, in the United States in 1913, in the United Kingdom in 1916, and in Germany in 1918. They became general in Europe and were established in India and Japan during the following years and became common in Latin America in the '30s. A labour office was established in Egypt in 1930, but only in the '40s and '50s did similar arrangements begin to take root elsewhere in Asia and Africa. Under differing political circumstances there continue, of course, to be wide variations in the authority and effectiveness of such administrative machinery and in the extent of its influence in matters of general economic policy and labour conditions.

THE ELEMENTS OF LABOUR LAW

The basic subject matter of labour law can be considered under nine broad heads: employment; individual employment relationships; wages and remuneration; conditions of work; health, safety, and welfare; social security; trade unions and labour management relations; the administration of labour law; and special provisions for particular occupational or other groups.

The worker in new economic thought

Employment. Employment considered as a basic concept and category of labour law is a relatively recent development. Prior to the Great Depression and World War II the emphasis was upon the prevention or reduction of excessive unemployment rather than upon long-term employment policy as part of a comprehensive scheme to promote economic stability and growth. The new approach, arising from changes in political outlook and contemporary economic thought, has increasingly found expression in legal provisions that establish the creation of employment opportunities as a general objective of policy. To this end, legislation has established the necessary legal framework for the forecasting of manpower needs and availability and the provision of employment services including recruitment, vocational training, and apprenticeship. Freedom from forced labour, equality of treatment in employment and occupation, and even unemployment compensation may, in a broad sense. be regarded as part of the same general subject.

Individual employment relationships. This includes the making, modification, and termination of such relationships and the resulting obligations for the parties. It may also involve certain aspects of promotion, transfer, and dismissal procedures and compensation. The law on the matter was at one time often described as the law of master and servant. It implied a contractual relation in which one party agreed to be under the control of the other in the sense that he was bound to obey orders not only as to the work that he would execute but also as to the details of the work and the manner of its execution. As the law developed, the implied terms and statutory incidents attached to this relationship concerning such matters as termination of employment, dismissal procedures and compensation, minimum wages, conditions of work, and social security rights began to limit freedom of contract. The individual employment relationship continues, however, to be the subject matter of labour law to which general legal principles, as opposed to statutes and collective agreements, still have the greatest application.

Wages and remuneration. The substantive law here covers such elements as forms of wages and methods of payment, the protection of wages against unlawful deductions and other abuses, minimum wage arrangements, the determination of wages, fringe benefits in relation to wages, and, in highly sophisticated economies, incomes policies designed to protect the real value of wages against the erosion of inflation. The concept of wage regulation as a restraint upon extreme social evils has gradually been superseded by wage policies as deliberate instruments of positive management designed to promote economic stability and growth.

Wage protection and abuse

Legal requirements concerning the forms of wages and methods of wage payment deal with such matters as the proper notification of wage conditions, the payment of wages in legal tender or by check, the limitation and proper valuation of payments in kind, the freedom of the worker to dispose of his wages, regularity in wage payments, the treatment of wages as a privileged debt, and restrictions upon the attachment or assignment of wages.

The protection of wages against deductions and related abuses covers such topics as placement fees, payments at company stores or for tools or other supplies, and fines for alleged breaches of discipline. Minimum-wage regulation takes varied forms; it may, following the pattern originally set by the British Trades Boards Acts from 1909 onward, provide for wages councils or similar bodies to fix wages in trades that have no arrangements for collective agreements and where wages are exceptionally low; it may consist, as in Australia and New Zealand, essentially of arbitration arrangements; or it may, as in the United States under the Fair Labor Standards acts, provide a comprehensive statutory rate or criteria for determining such a rate. Statutory provisions and collective agreements for determining wages may embrace such varied matters as skill differentials, the elimination of race and sex differentials, payment according to results and the relationship of wages to productivity, and wage guarantees for agreed periods of time. Fringe benefits, such as bonuses payable in varying contingencies, are typically a matter for collective agreements. Incomes policies remain the subject of much controversy. Their general purpose, sometimes embodied in legislation and sometimes expressed in collective agreements or statements of government policy, is to restrain inflationary pressures resulting from wage increases unrelated to increased productivity and to do this in a manner that promotes a fairer distribution of income.

Conditions of work. The conditions of work involve hours, rest periods, and vacations; the prohibition of child labour and regulation of the employment of young persons; and some special provisions concerning the employment of women. This part of the law originated in legislation for the protection of children, young persons, and women against the worse evils of the Industrial Revolution. It originally dealt particularly with such matters as admission to employment, night work, and excessive hours, but the elements of its content and their relative importance have been wholly transformed during the present century.

As economic and educational progress and changed social habits have limited recourse to child labour in the advanced countries, and to an increasing extent in the modernized sectors of developing economies, the special concern of labour law with regard to the young has shifted from the prohibition of child labour to such areas as vocational guidance and training, career planning and advancement, and medical protection.

As employment opportunities for women have become more varied and responsible, there has been a similar shift of emphasis from protective legislation, which has now come to be regarded as discriminatory since it tends to limit such opportunities, to legal guarantees of equal pay and equal employment, coupled with adequate maternity protection and the provision of facilities to enable women with family responsibilities to continue to be employed.

Another, but no less important, change has taken place in the gradual slackening of the so-called morality of work. Whereas previously any statutory limitation of the hours of work of adult males was regarded as being highly questionable, except in mines where it had been introduced on safety grounds, in a society of much increased leisure it has now become a general practice to fix hours of work by statute or collective agreement. In many countries the eight-hour day has been superseded by the forty-hour week as the statutory maximum for a wide range of occupations, and collective agreements providing for substantially shorter working hours are not uncommon. The details of hours regulation, whether by statute or collective agreement, include such matters as exceptions allowed for managerial and supervisory personnel and adjustments necessary for continuous shift working. In addition, such regulations cover the extensions permitted for preparatory, complementary, and intermittent work; the special rules for *force majeure* (work of absolute necessity), accident, maintenance, and repair work; and the limitation, authorization, and remuneration of overtime.

The principle of resting one day of the week, sanctioned as it is by religious practice in many places, was widely incorporated in legislation at an early date; the lengthening of this weekly rest through the creation of the five-day week has been strongly influenced by statutory requirements and collective agreements.

Holidays

Legislation granting annual holidays with pay and comprehensive collective agreements providing for such holidays are almost entirely a development of the last half-century but are increasingly common; moreover, there is a marked tendency for the minimum annual holiday to be increased and for collective agreements to provide holidays longer than the statutory minimum. Complex questions may arise concerning the qualifying period of service required for entitlement, breaks in the continuity of service, the calculation of average or normal remuneration for the purpose of the holidays, the extent to which holidays may be divided, and the liability for holidays where there has been a change of employer. The question of educational leave is receiving increasing attention.

Health, safety, and welfare. This is a category that includes such general matters as occupational health and accident prevention regulations and services; special regulations for particularly hazardous occupations such as mining, building and dock work; and provisions concerning such health and safety risks as poisons, dangerous machinery, dust, noise, vibration, and radiation. The law on the subject began with elementary safety rules for mines and other undertakings with particularly high accident rates and with the precautions taken against a number of dramatic poisonings and infections brought on by such agents as phosphorus, lead, and anthrax germs. The efforts of organized safety movements and the progress of occupational medicine have produced comprehensive occupational health and accident-prevention services and regulations no longer limited to a few specially acute risks but covering the full range of dangers arising from modern industrial processes. A major recent development has been the increased concern with health and safety regulations for agriculture and forestry made necessary by mechanization and the widespread use of chemicals. Welfare benefits, in the sense of communal services and special schemes and facilities for disadvantaged groups of society, go well beyond the scope of labour law; but there is increasing provision by law or collective agreement for welfare facilities related to employment, including feeding, rest, recreation, and transport facilities.

Social security. In the present stage of development of labour law, social security ranges from relatively straightforward employers' liability for occupational accidents to the most comprehensive schemes of payment, including income security in the form of sickness, unemployment, old age, employment injury, maternity, family, invalidity, and survivors benefits and medical care. As with other aspects of labour law, a progression from the particular to the general has been characteristic of the development of social security legislation. By the time of World War I, workmen's compensation schemes were general in industrialized and industrializing countries, but they still tended to define restrictively the cases regarded as occupational accidents and they were highly restrictive in their provisions for occupational diseases. Pension insurance was part of Otto von Bismarck's legacy to Germany, but elsewhere there was little more to be found than pension funds for the privileged and in some countries noncontributory pensions for the aged. Great Britain, as befitted its early entry into industrialization, had been the pioneer in health and unemployment insurance. But social insurance remained a pragmatic experiment limited to a few countries advanced in both economic development and social policies. Such coverage as there was in these countries was limited to specific risks for certain categories of protected persons. Its object was to protect the worker against the sort of hazards of life for which preindustrial societies provide by some form of community or family responsibility, but the approach was piecemeal and was limited to the most manageable cases of acute hardship. Eventually, the impact of the world economic depression of the '30s and World War II in the industrial nations and the increasingly apparent inadequacy of earlier forms of community responsibility in developing countries transformed the position. The concept of social security, first given statutory expression in the United States in 1935 and in New Zealand in 1938, superseded that of social insurance, and the Beveridge Report of 1943 developed it even further to provide a basic income for all in need of such protection, along with comprehensive medical care. The concept has continued to broaden since that time and has found increasing acceptance, with varying degrees of practical application, in countries in the most varied stages of economic development. Acute problems, particularly cost and efficiency of administrative organization and medical care, remain almost everywhere and sometimes become highly controversial, but far-reaching progress has been made in many countries in developing higher standards of medical care available as a legal right and in converting the guarantee of a basic income as a protection against want into provision for effective income maintenance in the event of the loss of employment or breadwinner. The idea has not yet reached the end of its development. The current trend is to broaden it to the point at which it includes all the varied hazards of life in industrialized societies, including accidents of any kind, with the idea of facilitating economic growth by reducing the human cost of structural change. Some attempt is being made to adapt these same ideas to the needs and practical possibilities of developing societies and rural communities. The pattern varies widely in different countries, partly as a reflection of different relationships between social security and private life, retirement and health insurance, partly because of differences in economic and social conditions.

The right to security

Trade unions and labour-management relations. A number of complex legal relationships fall under this heading, including the legal status, rights, and obligations of trade unions and employers' organizations, collective bargaining and collective agreements, the representation of employees at plant and enterprise level (including joint consultation and, where it exists, co-determination and other forms of worker's participation in management), work rules and labour discipline, the prevention and settlement of labour disputes in general, and strikes and lockouts in particular. As might be expected, there are wide variations both in the extent to which such matters as the representative character and capacity of trade unions, their legal status, the obligation to recognize and bargain with them, the enforceability of collective agreements, the scope of activities permitted to trade unions, and their obligations in contract and tort are subject to legal rules and in the content of such rules. In the United States, for instance, there is a considerable body of law on these subjects, the most important enactments being the National Labor Relations Act of 1935 (the Wagner Act) and the Labor Management Relations Act of 1947 (the Taft-Hartley Act); in the United Kingdom the law has hitherto remained marginal to most of these trade union issues, though changes may be in prospect. In general, the trend is to expand the role of law in labour-management relations to reduce the increasing disruption caused by industrial conflict in a more complex society. How to reconcile freedom of association with the stability and growth of the economy remains the most challenging and difficult problem of labour law.

The administration of labour law. This is an area involving the organization and functioning of administrative authorities concerned with labour problems, including labour inspection services and other organs of enforcement. Administration of the law also encompasses the organization, jurisdiction, composition, and procedure of labour courts and other bodies for the settlement of grievances arising from existing contracts or collective agreements.

Labour law and the economic goal

The principal problem in many countries is to relate the process of labour administration and its special intimacy with labour and management to overall economic and social planning in a manner that gives proper weight to social considerations in economic policy. This problem falls mostly outside the scope of labour law, but its solution does depend in part on the extent to which labour law provides for and secures effective standards of administration.

Special categories of workers. Labour law includes a wide range of special provisions for particular occupational or other groups. These sometimes appear as special parts of a general code or as special legislation; sometimes they take the form of provisions that limit the applicability, or vary the mode of application, of specific legislative provisions with regard to particular groups. The broad sectors of the economy for which such special provisions are particularly common and important include mines, transport (and in particular maritime transport), commercial occupations, and agriculture. Cutting across these broad sectors of economic activity are the traditional legal distinctions made in some countries between workers and salaried employees and certain newer distinctions, such as that between employees who earn annual salaries and have rights of tenure and persons with no such rights engaged and remunerated on a monthly or weekly basis, or that between white collar and manual workers—the latter often resting on a factual rather than a legal basis.

LABOUR LAW IN DIFFERENT LEGAL SYSTEMS

Among the distinctive elements of labour law that reflect the political, socio-economic, and legal differences from country to country are variations in the relative importance of statutory regulation and collective agreements, the prevalence of national or industrial collective agreements as opposed to company or plant agreements, the exceptional importance in certain countries of arbitral awards, and the extent to which the labour law of certain countries has been affected by the constitutional structure, especially with regard to judicial review of constitutionality of legislation and judicial interpretation of constitutional powers, limitations, and guarantees. How significant these distinctive elements will remain is uncertain.

In the United Kingdom and Scandinavia, for instance, the tradition has been to allow a maximum of initiative and freedom to employers' and workers' organizations in the regulation of their mutual relations and the determination of conditions of work.

In the United Kingdom, however, the reluctance to legislate is becoming less marked; there is now legislation concerning industrial training and discrimination in employment, until recently essentially matters for collective agreement, and legislation concerning collective bargaining and safeguards against unfair dismissal is under consideration. It is anticipated that membership in or closer association with the European Economic Community will accentuate this tendency. In virtually all of the developing countries the absence of an established tradition of responsible collective bargaining and the importance of the part played by the state in economic development have placed a premium on legislative action.

Variations in collective agreements

The coverage and scope, term of validity and legal effect of collective agreements varies widely. In Sweden there has been a practice of national negotiations covering the whole of industry; in the United Kingdom agreements generally cover an industry or occupation in the country as a whole or a particular industrial area; in the United States the unit of negotiation is generally the company or plant. The contrast may be less significant in practice than in principle, since an important company or plant agreement tends to set an industry-wide pattern of negotiation; nevertheless, the difference is important. The future trend may well be toward mutually complementary and interlocking national or industrial and company or plant agreements, but uniformity in this matter among countries with widely varying traditions and problems seems unlikely, and there remain many uncertainties. Still more important is the distinction between a collective agreement that is accepted as binding for an agreed term and one that has no such agreed term. There may well be a far-reaching tendency to prefer agreements for agreed terms during the coming years, but this also is by no means a certainty.

In Australia and New Zealand conciliation and arbitration tribunals determine by awards matters normally dealt with in other countries by legislation or collective agreement, such as wages, hours, and conditions of work. The example has had some influence on systems of arbitration courts established in developing countries, notably in Singapore and East Africa, but with this exception there is no tendency for it to be widely imitated elsewhere except as a device for avoiding deadlocks in negotiation, especially in essential public services.

Constitutional impediments

The United States and Canada are examples of the development of labour law being affected by questions of constitutionality, which, in this instance, not only influenced its ultimate form but also retarded its development. In the United States the constitutionality of workmen's compensation laws was a much debated issue until it was favourably settled by the Supreme Court in 1917, and child-labour and minimum-wage regulation was retarded for a generation by judicial decisions holding them to be outside federal competence and, in some cases, inconsistent with the constitutional guarantee against deprivation by the state of life, liberty, and property without due process of law (the guarantee conceived of here as applying to the factory owner). The first attempt of Pres. Franklin Roosevelt's administration to regulate hours and wages by codes of fair competition during the Great Depression was also held to be unconstitutional as an improper delegation of legislative power by Congress to the executive branch of the government. But thereafter the temper of judicial review changed and the validity of federal legislation guaranteeing free collective bargaining, regulating wages and hours, and establishing social security was upheld. In Canada, a pioneer in establishing a labour department, restrictive judicial interpretations of the powers of the federal government had a somewhat similar effect, and only after World War II did federal-provincial cooperation afford a basis for achieving greater uniformity and more rapid progress. In both cases the impress of constitutional law upon labour law remains but is perhaps less significant than a generation ago.

UNIFYING TENDENCIES

It is no less true in labour law than in other branches of the law that the possible solutions for identical or similar problems are not infinite in number. Indeed, the range of possible solutions often consists of variations of detail on a limited number of options. It is therefore natural that many common elements, often expressed in identical or almost identical texts, should recur in the law of different countries. These elements derive partly from the legislation of other nations but increasingly from the influence on the law of the international standards evolved by the International Labour Organisation.

From the beginnings of modern labour legislation in the early 19th century, the law of certain countries has been extensively used as a model. A full picture of this pattern would require a detailed account of the interplay of external influences in the law of many countries, but, by way of illustration, it may be pointed out that United Kingdom factory legislation was widely copied at an early date and German social-insurance legislation provided a prototype dating back to the time of Bismarck's reforms. United Kingdom legislation has continued to serve as a model in the contemporary period for the basic legislation of many states that were formerly British dependencies and remains in force subject to modifications made since independence. Much of the French Labour Code became applicable through the 1952 Labour Code for Overseas Territories to the states that were formerly French dependencies and remains the basis of their labour law. The United States legislation of the period from the 1930s onward has been taken as a model in Japan, the Philippines, Liberia, and other countries. The Mexican Labour Law of 1917, varied by elements derived primarily from European models, had considerable influence on the early development of labour law in a number of Latin American countries. Through quite another process, the labour law of the U.S.S.R. has reshaped without replacing the earlier elements in the labour law of the other Communist states. But on the whole these national influences of particular countries and legal systems are declining.

The ILO: international standards

During the last half-century the standards evolved by the International Labour Organisation (ILO) have gradually become the leading external influence upon the labour law of many countries. They have had a far-reaching impact in virtually all of the advanced countries except the U.S. and the U.S.S.R., where external influences have been secondary. In much of the developing world they were of great importance even before independence, since a large part of the legislation sponsored there by the metropolitan powers was based essentially on ILO standards; in more recent years they have become a main source of new legislation in these countries. The ILO, created in 1919 as an autonomous partner of the League of Nations, and since 1946 a specialized agency associated with the United Nations, adopts international standards in the form of conventions and recommendations that each member state has an obligation to submit for consideration to the national legislative authority. Conventions when ratified become binding obligations of the member states ratifying them; recommendations are designed as guides for legislation, collective agreements, administrative measures, etc. Elaborate follow-up arrangements, including examination of regular reports and commissions of enquiry into complaints, are provided for to ensure that the obligations assumed are fulfilled. These standards, which already cover, in varying degrees of detail and at varied stages of development, virtually all the more important branches of labour law, are constantly amplified and revised at the annual sessions of the International Labour Conference. By 1970, 130 conventions and 134 recommendations had been adopted. The conventions dealing with such fundamental human rights as freedom from forced labour, freedom of association, and freedom from discrimination have been ratified by from over 70 to over 100 countries. The conventions and recommendations are supplemented by regional and technical standards, approved in the form of resolutions but often given effect by law, regulation, or collective agreement. In the field of health and safety model codes, codes of practice, manuals, and guides have become particularly important; in social security matters the technical resolutions adopted by the International Social Security Association, which was created on the initiative of ILO, play a comparable role.

Unification of labour law is one of the professed purposes of a number of regional economic organizations in different parts of the world, but only in the Council of Europe, the European Communities, and the Organization of American States have tangible measures been taken, and these have been largely based on ILO standards and framed with its technical cooperation and assistance. The more important instruments adopted are the European Social Charter, the European Social Security Code, the Social Security Regulations of the European Economic Community, and the Central American Convention on Social Security for Migrant Workers. The Conferences of American and African Labour Ministers, sponsored by the Organization of American States and the Organization of African Unity respectively, and the Conference of Asian Labour Ministers, which has developed without any comparable sponsorship, discuss matters of general policy and the coordination of action in the ILO rather than the formulation of specific standards. An Arab Labour Organization was created in 1970.

FUTURE OF LABOUR LAW

Labour law differs from the older branches of the law in that its history has been in some cases so much influenced by the ebb and flow of political change, its development so rapid, and its expansion on a world scale so recent, that it is exceptionally difficult to predict its future. But the general trend is clear.

Nowhere is labour law becoming less important. While some types of protective legislation, notably special provision for the protection of women workers, are losing their importance, the general tendency continues to be toward more comprehensive legislation embracing a wider range of subjects and often dealing with matters previously left to collective agreement, individual contract, or the discretion of the employer.

Irrespective of ideology or economic or social structure, the transition everywhere has been from a class law protecting the weakest segment of society to a community law designed to serve the common interest. This is a development that can be seen not only in the elimination of limitations and exceptions to the law but in the increasing emphasis being given to matters of general interest, including full employment, equitable distribution of wealth, and community responsibility for the incidence of misfortune in individual lives.

The place of labour law in legal evolution

Labour law must also be said to serve the social interest in promoting civilized labour-management relations. This evolution of labour law is an important contribution to the evolution of the law as a whole, from a law for the propertied and trading classes with a special chapter for the working class to a common law for the entire community.

The importance of a body of law that has a dynamic and progressive impact rather than a restrictive influence is now widely understood, and the need for legal flexibility to facilitate economic development and change is increasingly appreciated. In addition, the value of delegated powers and procedures of consultation with interested groups and organizations to achieve such flexibility is more generally recognized. Social objectives remain the test of the validity of economic policy, and labour law plays a major part in defining these objectives and ensuring that economic policy respects them in the interest of the whole community.

BIBLIOGRAPHY. The more important labour laws from all countries are reproduced in the *International Labour Office Legislative Series* (1920–), continuing the *Bulletin of the International Labour Office* (1906–19). The international standards are collected in ILO, *The International Labour Code,* 2 vol. (1951), supplemented for the more recent period by *Conventions and Recommendations, 1919–1966.* For the impact of ILO standards see C.W. JENKS, *International Protection of Trade Union Freedom* (1957), *Human Rights and International Labour Standards* (1960), and *Social Justice in the Law of Nations* (1970). For the wide range of comparative international studies available at any particular time see the current *Catalogue of Publications* of the Inter-

national Labour Office. The *International Encyclopaedia of Comparative Law* (in prep.), sponsored by the International Association of Legal Science, is to contain a volume on labour law. The following are some of the more important works on this subject.

Introductions: K.W. WEDDERBURN, *The Worker and the Law,* new ed. (1966); A. ROUAST and P. DURAND, *Précis de droit du travail,* 2nd ed. (1961); F. SANTORO-PASSARELLI, *Nozioni di diritto del lavoro,* 19th ed. (1967); M. ALONSO OLEA, *Introducción al derecho del trabajo* (1963).

Treatises and reports: (United Kingdom): F. TILLYARD, *The Worker and the State,* 3rd ed. (1948); A. FLANDERS and H.A. CLECG (eds.), *The System of Industrial Relations in Great Britain* (1964); *Report of the Royal Commission on Trade Unions and Employers' Associations* (1965–68). (United States): C.O. GREGORY, *Labor and the Law,* 2nd rev. ed. (1958); J.S. WILLIAMS (ed.), *Labor Relations and the Law,* 3rd ed. (1965). (France): G.H. CAMERLYNCK (ed.), *Traitd de droit du travail,* 6 vol. (1966); P. DURAND *et al., Traité de droit du travail,* 3 vol. (1947–56). (Germany): A. HUECK and H.C. NIPPERDEY, *Lehrbuch des Arbeitsrechts,* 7th rev. ed. (1963); P. WALINE, *Cinquante ans de rapports entre patrons et ouvriers en Allemagne, 1918–1968* (1968). (Italy): U. BORSI and F. PERGOLESI (eds.), *Trattato di diritto del lavoro,* 3rd ed., 5 vol. (1958–60). (Spain): G. BAYON CHACON and E. PEREZ BOTIJA, *Manuel de derecho del trabajo,* 7th ed., 2 vol. (1967–68). (Sweden): F.F. SCHMIDT, *The Law of Labour Relations in Sweden* (1962). (U.S.S.R. and Eastern Europe): V. GSOVSKI and K. GRZYBOWSKI (eds.), *Government, Law and Courts in the Soviet Union and Eastern Europe,* pt. 6, *Worker and Factory,* vol. 2, pp. 1411–1601 (1959); E.C. BROWN, *Soviet Trade Unions and Labor Relations* (1966). (Australia): J.E. ISAAC and G.W. FORD, *Australian Labour Economics* (1967). (Japan): K. ISHIKAWA, "The Regulation of the Employer-Employee Relationship: Japanese Labour Relations Law," in A. TAYLOR VON MEHREN (ed.), *Law in Japan: The Legal Order in a Changing Society* (1963); I.F. AYUSAWA, *A History of Labor in Moderrz Japan* (1966). (India): G.M. KOTHARI, *Labour Law and Practice in India* (1964). (Iran): AHMAD ALI BAHRAMY, *La Législation internationale du travail et son influence sur le droit iranien* (1963). (Argentina): M.L. DEVEALI, *Tratado de derecho del trabajo, 5* vol. (1966). (Brazil): A.F. CESARINO, *Direito social brasiliero,* 5th ed. (1963); A.L. SUSSEKIND, D. MARANHAO, and J. DE SEGADAS VIANNA, *Instituicões de direito do trabalho,* 4th ed. (1966). (Mexico): M. de la CUEVA, *Derecho mexicano del trabajo,* 6th ed. (1961). (Venezuela): R. CALDERA, *Derecho del trabajo,* 2nd ed. (1960). (Africa): ILO, *African Labour Sztrvey* (1958), *Labour Survey of North Africa* (1960).

(C.W.J.)

Lacquerwork

Lacquerwork is a general term for certain metallic and wood objects to which coloured and frequently opaque varnishes called lacquer are applied. The term lacquer is derived from lac, which is the basis of some lacquers. This article deals with artistic lacquer ware; for a discussion of lacquers used as protective coatings, see PAINTS, VARNISHES, AND ALLIED PRODUCTS.

The lacquer of East Asia, China, Japan, and Korea must not be confused with other substances to which the term is generally applied; for instance, the lac of Burma, which is the gummy deposit of an insect, *Coccus lacca,* and the various solutions of gums or resin in turpentine of which European imitations of Eastern lacquer have been and are concocted.

TECHNIQUES

Obtaining and preparing lac. Lacquer, as used in China and Japan, is a natural product, the sap of a tree, *Rhus vernicifera;* subject to the removal of impurities and excess water, it can be used in its natural state, though it was frequently adulterated. The tree, which is indigenous to China and has certainly been cultivated in Japan at least since the 6th century AD, is tapped at about the age of ten years, when lateral incisions are made in the bark and the running sap is collected during the months of June to September. Branches of a diameter of one inch (about three centimetres) or more are also tapped, the bark having first been removed. Smaller branches are cut off and soaked in water for ten days, and the sap is collected, producing a lacquer *(seskime)* of particular quality, used for special purposes. These processes kill the tree, but the wood, when of sufficient size, is of some use for carpentry. From the roots five or six shoots spring up, which become available for the production of lacquer after about six years, and the operation can be thus continued for a considerable length of time before the growth is exhausted. The Chinese and Japanese methods are practically identical in this respect, but the cultivation of the tree does not seem to have been as systematic in China as in Japan.

Collecting
the sap

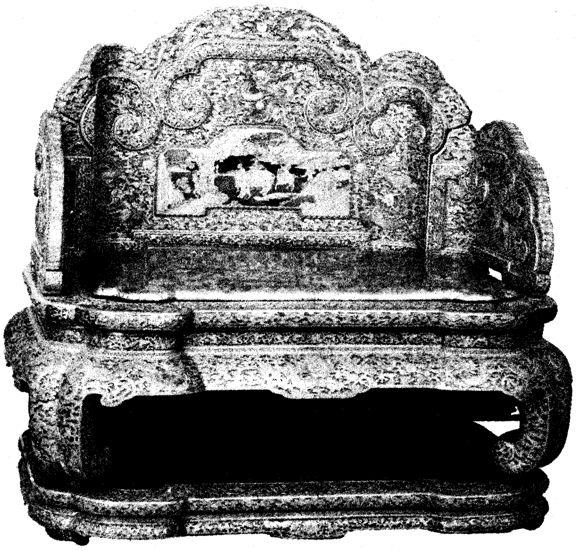
The sap is white or grayish in colour and about the consistency of molasses. On exposure to the air it turns yellow-brown and then black. It is strained through hempen cloth to remove physical impurities, after being pounded and stirred in shallow wooden tubs to give it uniform liquidity. It is then slightly heated over a slow fire or in hot sunshine, and stirred again to evaporate excess moisture, and stored in airtight vessels.

The basis of lacquer ware, both in Japan and in China, is almost always wood, although it was also occasionally applied to porcelain, brass, and white metal alloys. In some instances, objects were carved out of solid lacquer. The wood used, generally a sort of pine having a soft and even grain, was worked to an astonishing thinness. The processes that follow are the result of extraordinary qualities of lacquer itself, which, on exposure to air, takes on an extreme but not brittle hardness and is capable of receiving a brilliant polish of such a nature as to rival even the surface of highly glazed porcelain. Moreover, it has the peculiar characteristic of attaining its maximum hardness in the presence of moisture. To secure this result, the Japanese place the object in a damp box or chamber after each application of lacquer to the basic material (wood, etc.). The Chinese are said (in an account of the industry dating from 1621–28) to use a cave in the ground for this purpose and to place the objects therein at night in order to take advantage of the cool night air. It may, indeed, be said that lacquer dries in a moist atmosphere.

Hardening
with
moisture

Application. The joiner's work having been completed and all knots or projections having been most carefully smoothed away, cracks and joints are sealed with a mixture of rice paste and *seshime* lacquer, until an absolutely even surface is obtained. It is then given a thin coat of *seshime* lacquer to fill up the pores of the wood and to provide a basis for succeeding operations, which may number as many as **20** or **30** or even more, of which any one of the following may be taken as typical. On the basis, as above described, is laid a coat of lacquer composition which is allowed to harden and is then ground smooth with whetstone. Next comes a further coat of finer composition, in which is mixed some burnt clay, which is again ground and laid aside to harden for at least 12 hours. On this is fixed a coat of hempen cloth (or, rarely in Japan but more often in China, paper) by means of an adhesive paste of wheat or rice flour and lacquer, which needs **24** hours at least to dry. The cloth is smoothed with a knife and then receives several successive coats of lacquer composition, each demanding the delay necessary for hardening. On this is laid very hard lacquer, requiring a much longer drying interval, afterward being ground to a fine surface. Succeeding coats of lacquer of varying quality are now laid on, dried, and polished.

This preliminary work, requiring for artistic lacquer at least 18 days, produces the surface on which the artist in lacquer begins his task of decoration. **A** large number of processes have been at his command, especially in Japan, but the design was first generally made on paper with lacquer and transferred to the object while still wet or drawn on it direct with a thin paste of white lead or colour. In carrying it out the artist used gold or silver dust applied through a quill, a bamboo tube, or, for equal distribution, a sieve. Larger fragments of the precious metals were applied separately by hand, with the aid of a small, pointed tool. In one typical instance approximately 500 squares of thin gold foil were thus inserted within one square inch (six square centimetres). These decorative processes each entailed prolonged hardening periods and meticulous polishing. Relief was obtained by model-



Imperial Chinese throne of the Ch'ien-ling emperor (1736-96), red lacquer carved in dragons and floral scrolls, Ch'ing dynasty. In the Victoria and Albert Museum, London. 1.19 m X 1.26 m X 91 cm.

By courtesy of the Victoria and Albert Museum, London; photograph. A.C. Cooper Ltd.

ling with a putty consisting of a mixture of laquer with fine charcoal, white lead, lampblack, etc., camphor being added to make it work easily. Lacquer was sometimes engraved, both in China and Japan.

Chinese carved lacquer. The carved lacquer of China (tiao-ch'i), which was imitated but never equalled in Japan (as the Chinese have never reached the perfection of the Japanese gold lacquer ware), needs particular notice. In this the lacquer was built up in the method described above, but to a considerable thickness; when several colours were used, successive layers of each colour of uniform thickness were arranged in the order in which they were to predominate. When the whole mass was complete and homogeneous, it was cut back from the surface to expose each colour as required by the design. When the lacquer was cold and hard, the carving was done with a V-shaped tool kept very sharp. The cutting was done with amazing precision—no correction of faults was possible, for each layer had to be exactly and accurately reached and the final result precisely foreseen and allowed for from the beginning of the work. The red lacquer (t'i-hung), so well known and justly appreciated, was coloured with cinnabar (red mercuric sulfide). Other colours which were employed include a deep and a lighter olive-green, buff, brown, black, and purple (aubergine).

Use of successive layers of different coloured lacquer

Japanese processes. In Japanese lacquer, the following are the chief processes used: nashiji (pear skin), small flakes of gold or silver sunk to various depths in the lacquer; *fundame*, fine gold or silver powder worked to a flat, dull surface; *hirame*, small, irregularly shaped pieces of sheet gold or silver placed on the surface; togidashi, the design built up to the surface in gold, silver, and colours with many coats of lacquer and then polished down to show them; *taka-makie*, decoration in bold relief; hiramaki-e, decoration in low relief: *rō-iro*, polished black; chinkin-bori, engraved lacquer; kirikane, square dice of sheet gold or silver, inserted separately on the surface; and *raden*, inlaid shell and metal. From the earliest recorded times shell was used in the adornment of lacquer in China as well as in Japan, being inlaid on the surface in patterns as well as in small squares like *kiri*kane and dust. For this purpose various shells were used, mother-of-pearl (for larger work), nautilus, pear shell, sea-ear (Haliotes, Japanese Awabi) and Turbo *cornutus* (Japanese Sazae). For a very charming form, called laque *burgauté*, the iridescent blue and green shell of the sea-ear was delicately engraved with gold and silver as early as the Ming period (1368–1644) and also in Japan. Chinese lacquer was also inlaid with jade, malachite, coral, soapstone, ivory, porcelain, and other substances.

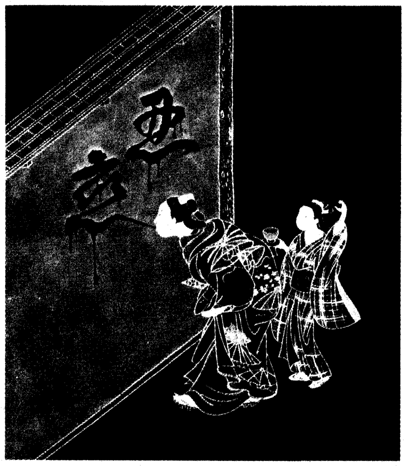
Laque burgautk

HISTORICAL DEVELOPMENTS

China. The use of lacquer in China goes back traditionally to legendary times. A late Ming manuscript, the Hsui *shih lu*, states that it was first employed for writing on bamboo slips, then for utensils for food, made of black lacquer, and subsequently for vessels for ceremonial use, of black with red interiors. During the Chou dynasty (1122–249 BC) it served for the decoration of carriages, harnesses, bows and arrows, etc., and was the subject of official regulations. At this time, gold and colours are said to have come into use. About the 2nd century BC buildings were decorated with lacquer, as were musical instruments. Under the Han dynasty (206 BC–AD 220) further development took place. Pot covers of paper, covered with lacquer and found near Port Arthur, are attributed to this period.

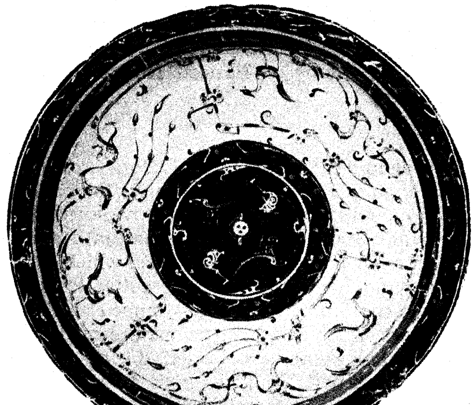
Of the lacquer of the T'ang dynasty (618–907) more reliable information is available, for the collections still preserved in the Hōryū-ji in Japan, founded AD 607, and those collected by the Japanese emperor Shōmu (724–748), deposited after his death in the Imperial treasury (Shdsd-in) at Nara, contain many objects of Chinese origin; in particular, musical instruments with inlay of cutout figures of gold and silver inserted on the surface, covered with lacquer, which was then rubbed down until the metal ornaments were again visible.

By courtesy of the Victoria and Albert Museum, London; photograph, A.C. Cooper Ltd.



Japanese box with "ardent lover" theme in *togidashi* on a *rō-iro* background, signed Katsukawa Shunshō, early 19th century. A young woman takes black tooth stain from a bowl held by her attendant and squirts from her lips the characters for "perseverance in love." In the Victoria and Albert Museum, London. 21.6 × 25 × 5 cm.

By courtesy of the Seattle Art Museum, Washington



Wood bowl decorated in red and black lacquer with stylized birds and animals, from Ch'ang Sha, China, late Chou dynasty, 3rd century BC. In the Seattle Art Museum, Washington. Diameter 25 cm.

Under the Sung dynasty (960–1279) the industry further developed, and the use of gold and silver lacquer in the utensils made for the palace is particularly recorded. The chief seat of manufacture was Chia-hsing, between Hangchow and Su-chou, the latter city being also an important centre of the industry. A lacquer box of the early Sung period, probably once of rhinoceros-horn colour, black and red, with gold dust and silver wire, is one of the very few known examples of the period. Toward the close of the period (c. 1220) it is stated that lacquer wares were exported from Ch'uan-chou, Fukien to Java, India, Persia, Japan, Mecca, and other places. Chinese writers record the existence of carved red lacquer during the time of the Yüan dynasty (1279–1368) as well as of pierced ware and that inlaid with shell.

Ming lacquerwork

Of the state of the industry under the Ming dynasty (1368–1644) there are contemporary Chinese descriptions; for instance, in the *Ko* ku yao lun, published in 1388, the Ch'ing pi-ts'ang, published in 1595, and the Hsiu *shih lu*, which has been handed down in manuscript. This last work was written by a celebrated lacquerer, Huang Ch'eng, and bears a preface by Yang Ming, another lacquerer, dated 1625. The work itself was probably written towards the end of the 16th century. From these works one can ascertain the excellence of the carved lacquer made during the reigns of the Yung-lo (1403–24) and Hsiian-te (1426–35) emperors. Examples of carved lacquer that can be attributed to both these reigns are extant. They are bold in design and free from the superabundance of small detail that characterized later productions; the colour also is generally deeper and richer than that of the 18th-century pieces. In the 16th century there were special factories for carved lacquer at Ta-Li in Yünnan, which also produced spurious imitations. Lacquer with designs painted in gold outlines were made, early in the Ming dynasty, at Nanking and afterward at Peking, and lacquer inlaid with mother-of-pearl was made at Chi-chou in Kiangsi. In the reign of Hsiian-te, lacquer decorated in sprinkled gold was introduced from Japan and excellent copies were made by Chinese lacquerers. Toward the end of the Ming dynasty there was a decline in lacquer manufacture as a result of the troubles accompanying the fall of the last Ming emperor.

Importation into Europe

The first and perhaps the greatest of the Manchu emperors, K'ang-hsi (1654–1722), revived the lacquerwork industry in 1680, when he established a series of 27 workshops for artistic handicrafts in the precincts of the palace at Peking. Carved lacquer was, however, also made at Canton, Su-chou, and Foo-chow; and the Jesuit Louis le Comte, who arrived in China in 1687, gave a good account of the flourishing state of the industry at that time. In this connection it is worth noting that the period of K'ang-hsi was that which saw the first considerable importation of lacquer ware (and other objects of industrial art) into Europe. This led to the development of imitation lacquer applied to furniture and other objects, which were conspicuous features of the chinoiserie craze of the late 17th and 18th centuries. A screen, c. 1700 in the collections of Earl Spencer and R. Freemer Smith, Esq., for example, was made by command of K'ang-hsi for presentation to the Holy Roman emperor Leopold I, whose badge, the double-headed eagle, is incorporated in the design. Carved lacquer of this period, though far from negligible, hardly attains to the rich colour, breadth, and simplicity of that of the Ming period.

In technique the K'ang-hsi ware shows an advance and is generally free from the small cracks too often found to have developed in the Ming products. The perfection of this quality, apart from other considerations, is found in the lacquer ware of Emperor Ch'ien-lung (1736–96), a devoted admirer of this art, who employed it on a large scale for the furniture and fittings of his palaces, as well as for ceremonial and commemorative gifts. The workmanship of objects made under his auspices is brilliant in the extreme, but the colour is hard as compared to earlier work, and the design tends to a somewhat stereotyped formalism.

Still, the 18th century can hardly be called a period of decadence in the decorative arts of China: the superb execution of its productions, a characteristic that commands admiration, redeems it from adverse criticism. The downward course began in the 19th century, with loss of originality and a falling off, due to adulteration, in the quality of the material. What was left of the Imperial factories was burnt in 1869, and, though carved red lacquer was made after that date, the industry had already ceased to have artistic importance.

Japan. Although the earliest reference to the manufacture of lacquer accepted by all Japanese authorities is a code of law (known as Taihō code) dated 701, there can be no doubt that the manufacture was brought to Japan from China via Korea at the time of the introduction of Buddhism in the middle of the 6th century. At the same time, according to tradition, the Chinese lacquer tree was introduced. The earliest piece of lacquer known today that is accepted as having been made in Japan is the Tamamushi Shrine in the Hōryū-ji, which is attributed to the 7th century. This piece shows strong Korean influence. Many fine pieces of the late 7th and the 8th centuries, inlaid in gold, silver, or mother-of-pearl, of Chinese origin, have been preserved in the Shōsō-in, as already mentioned in the section above on Chinese lacquer. But there is one piece in the Shōsō-in, a sword-scabbard of black lacquer decorated in gold, formerly belonging to the emperor Shōmu (724–748), that is undoubtedly Japanese. This is listed in an old catalog dated 756. There are also two arrows in the Tokyo National Museum that belong to the same period. These can be regarded as the real beginnings of a Japanese style in lacquer.

The emperor Kammu (781–806) removed the capital from Nara to a new city, Heian-kyō—the modern Kydto; and an increased luxury in the style of living brought about further developments in the art, especially in the use of gold lacquer, largely because of the spread of Buddhistic influence. This period, however, saw the beginnings of a Japanese national style as distinct from the Chinese methods and manner, imported by Buddhist missionaries. Lacquer was used at this time in the decoration of important buildings, and inlay of shell also became popular. The organization of the industry was extended, and, as early as 905, sumptuary edicts began to be issued regulating the dimensions and quantities of material to be used in the domestic utensils—chiefly of black or red polished lacquer—which now began to come into general use. From this time, it is no exaggeration to say that, to a considerable extent, lacquer filled the place occupied in China by ceramic wares. A remarkable development of this period that must not be overlooked was the production of statuary of considerable merit, made with lacquer composition (kanshitsu), a process derived from China but carried to a high standard in Japan for a brief period, until it was superseded by wood sculpture. Some few authentic examples remain of the fine lacquer of the Heian period, notably a case for Buddhist scriptures in the Ninna-ji at Kyōto, made at the beginning of the 10th century, which bears an inscription dated 919. The case is in black lacquer, sprinkled with gold dust and with a pattern of flowers, clouds, birds, and Buddhist winged genii in gold and silver togidashi.

Production of statuary

During the Kamakura period (1192–1333), in spite of the disturbance caused by the famous struggle between the Minamoto and Taira clans and the establishment of the feudal shogunate at Kamakura, which gives its name to the period, the art of making fine lacquer continued to progress under the patronage of the Fujiwara family, who maintained the Imperial court at Kydto with ever increasing luxury. Marked features of this time are improved methods of inlay of precious metals and shell and, especially, an attractive form of design in which beautifully written poems are interwoven with the pattern (*ashide*). The process called Kamakura-bori, carved wood thickly lacquered with red or black, also dates from this period and continued to flourish for another two centuries or so. During this epoch occurred the beginnings of the characteristic Japanese treatment of landscape and flower subjects in design, generally in flat gold lacquer with nashiji and pewter inlay.

"Priest Ganjin," hollow dry lacquer (*kanshitsu*) statue, Late Nara period (724–794 AD). In the Tōshōdai-ji, Nara, Japan. Height 80 cm.
By courtesy of the Toshodai-ji, Nara, Japan

Muromachi lacquerwork

The Muromachi period (1338–1573) saw a further technical and artistic development, largely under the patronage of the shogun Ashikaga Yoshimasa (reigned 1443–73). He gave great impetus to the tea and incense ceremonies, the latter of which brought about a whole series of new applications of the art because of the exquisitely wrought small utensils required by that ritual. The ostentatious simplicity of the Zen sect of Buddhists was displayed in the use of black lacquer of the first quality with little or no ornament. Excellent work in shell inlay was also a characteristic of the time. The gold lacquer of the Muromachi craftsmen gained so great a reputation in China that artisans from that country went to Japan to learn the methods by which it was produced, though they seem to have had little success in introducing it into their own country. Among the leading Japanese craftsmen of the period may be mentioned Kbami Dbchb, Taiami, Seiami, and Igarashi Shinsai, but attribution of specific works to them is largely a matter of conjecture.

The civil wars which continuously infested Japan during the later Middle Ages checked the growth of the industry for a while, but the short Azuchi-Momoyama period (1574–1600) that followed saw at least the work of one of the greatest of Japanese artists in lacquer, Hon'ami Kbetsu. He was the founder of a striking and original style of ornament, essentially national in character. His designs were bold and simple in detail, generally executed in high relief with masses of shell or metal inlay. The great shogun Toyotomi Hideyoshi (died 1598), who secured the peace of the country with a strong hand, was an enthusiastic patron of the arts, and under his patronage a real revival took place. When he died, his widow erected the Kbdai-ji at Kybto, in which distinctive lacquer decoration called tata maki-e (Koda-ji maki-e) was used. This temple still contains examples of this ware that were presented by her.

The Edo school

In 1603 began the rule of the Tokugawa shogunate, which continued without a break until the restoration of the Imperial family to actual power in 1867. The first of the line, Ieyasu, established at Edo (the modern Tokyo) the great school of lacquer artists that is responsible for almost the whole of the artistic ware known outside Japan. Technical processes were still further developed with additions such as engraved lacquer (chinkinbori) derived from China, carved red and black lacquer from the same source, and the so-called somada ware of shell inlay of black, different in character from the Chinese *laque burgauté* already mentioned above.

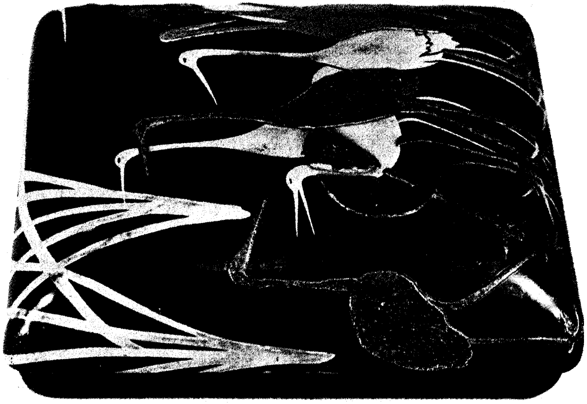
This period also saw the introduction of the now well-known inrd, or portable medicine case, worn on the girdle and an indispensable addition to the national costume so long as the latter was uncontaminated by Western influence. An inrd consisted, as a rule, of from two to five compartments, beautifully fitted into each other and held together by silken cords running along each side, secured by a bead (ojime) and kept in place on the sash by a kind of toggle (netsuke), sometimes of lacquer but more often of cunningly carved wood, ivory, bone, or other material. On this class of work was lavished some of the finest artistry of the Japanese craftsmen, and the convenient size and intrinsic charm of these dainty utensils (originally, perhaps, made for seals) have caused them to be much favoured by collectors.

Chinese influence on Tokugawa lacquerware

The earlier years of the Tokugawa period saw a considerable Chinese influence in the design of lacquer, especially in inrd; but the work of the greatest Japanese lacquer artists, Ogata Kbrin (*c.* 1658–1716), followed and extended in the late 17th century the style originated by his master, Kbetsu. Ritsuō and Hanzan in the 18th century maintained this tradition, and a considerable revival of the style took place in the early years of the 19th century, when memorial volumes of the designs of the great master were published. To the latter period belong not a few objects which have been accepted as the original work of Kbrin himself. The more formal school of lacquerers included Kbami Chbgen (1572–1607) and Komo Kitb-ye, who was appointed court lacquer artist to the shogun Iemitsu in 1636 and died in 1674. One of the most important lacquerers of the Kbami family was Kbami Nagashige, whose masterpiece, consisting of three connected cabinets with numerous writing cases, paper cases, and toilet cases, a mirror stand, and other accessories, was completed in 1639. It was made for the dowry of Tokugawa, Iemitsu's eldest daughter, on her marriage to Mitsumoto, prince of Bitchū, whose coat of arms is affixed to all the pieces. The lacquer is now in the Tokugawa Art Museum, Nagoya. As did other craftsmen in Japan, lacquer artists followed the practice of transmitting their names to sons or selected pupils. Thus, there were ten generations of the family of Yamamoto Shunshō, who died in 1682, aged 63. The Kajikawa family continued the tradition of its founder well into the 19th century, and the same must be said of Shiomi Masanari in the 18th century, whose work was notable for the quality of the rubbed-down gold and colour lacquer called togidashi.

The Genroku period (1688–1703) saw, perhaps, the ultimate perfection of style and technique; but the work of the later 18th and, to some extent, of the early 19th centuries has many exquisite qualities. The later periods were characterized by more elaborate detail, but adulteration of the gold with bronze and other metallic powders was often prevalent. A fiery brown tint of the nashiji is a certain mark of quite late date. Nevertheless, there is plenty of good work of the 19th century, and to this period belongs the last of the great artists of the industry,

By courtesy of the Seattle Art Museum, Washington



Tokugawa period writing box of black lacquered wood decorated in gold paint, inlaid lead, and pewter. Attributed to Ogata Kōrin (*c.* 1658–1716). In the Seattle Art Museum, Washington. 23 X 22 X 8 cm.

Lacquered cabinet by Kōami Nagashige (17th century). In the Tokugawa Art Museum, Nagoya. 101 X 77.5 X 40 cm.

By courtesy of the Tokugawa Art Museum, Nagoya

Shibata Zeshin, whose work bears comparison even with some of the greatest of his predecessors, both in technique and in design.

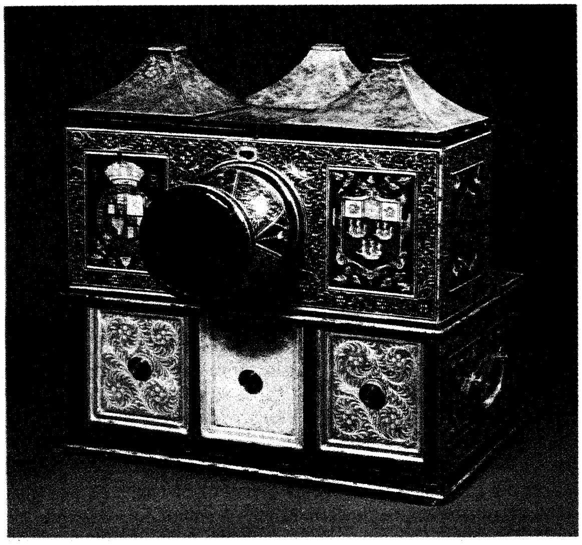
Modern industrial conditions, however, have practically killed this ancient and beautiful art. It would not have survived so long had not the country been closed to alien influences for two and a half centuries. (E.F.S.)

Europe. Although East Asian objects of art were taken to Europe in considerable quantity during the 16th century, it was not until after 1600 that a real trade with China grew up, fostered by the East India companies of the Netherlands, England, and France. Porcelain and lacquerwork then became so fashionable that European craftsmen undertook to make their own. The secret of porcelain manufacture was discovered shortly after 1700, and by that time the imitation of lacquer had become established in various parts of Europe. Jesuit missionaries had cooperated with scientists and master craftsmen to create formulas, and a small body of writing had appeared, which was to be augmented in the 18th century —the golden age of European lacquer.

Origins of the European lacquer-work industry

Among the earliest surviving examples of this art is the ballot box of the Saddlers Company (1619; Saddlers Company, London). Information on the lacquer process seems first to have been published by the Italian Jesuit Martin Martinius (*Novus Atlas Sinensis*, 1655). John Stalker and George Parker's *Treatise of Japanning and Varnishing* (London, 1688) was the first text with pattern illustrations. The English term japanning was inspired by the superiority of Japanese lacquer, which Stalker found ". . . in fineness of Black, and neatness of draught . . . more beautiful, more rich, or Majestick" than the lacquer of other places, which came to be known as "Indian" or "Bantam" work.

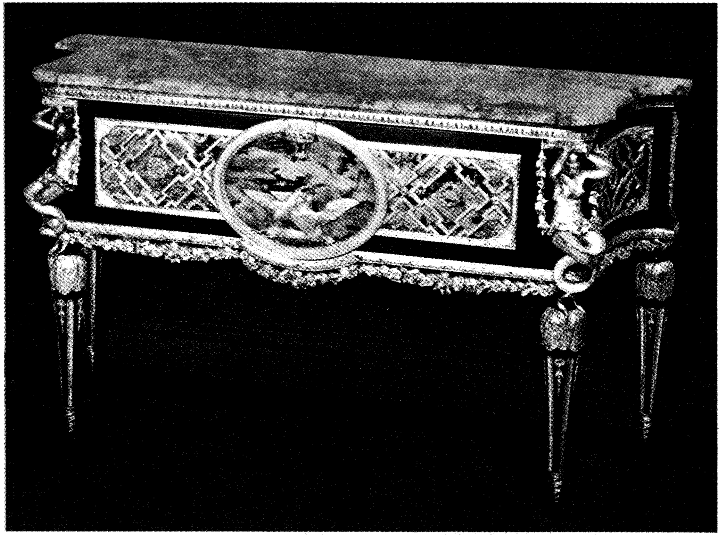
By courtesy of the Saddlers Company, London



Ballot box of the Saddlers Company, 1619. In the collection of the Saddlers Company, London. 45 X 44 X 33 cm.

In France, on the other hand, *ouvrage à la **Chine*** was the term for the imitation of lacquer practiced at the Gobelins factory in Paris from 1672. By the end of the century Berlin had become another centre of experimentation, from which a Fleming, Jacques Dagly, brought secrets that were to lead to the 18th-century innovations of the Martin brothers: Guillaume, Simon, Étienne, Julien, and Robert. They created the lustrous *vernis Martin,* which was praised by Voltaire. The Martins decorated rooms at Versailles, and Robert's son Jean Alexandre worked for Frederick the Great II at Potsdam. French lacquer was further improved through new information provided by the *Mémoire sur le vernis de la Chine*, which the French missionary Pierre d'Incarville wrote in 1760 and which appeared as an appendix to *L'Art du peintre, doreur, vernisseur* of the Sieur Jean-Felix Watin (1772), the most precise account of lacquerwork that appeared in the 18th century. In this book Watin examined the recipes of his predecessors and recommended the best formulas for lacquering objects to be used indoors, such as furniture, and outdoors, such as carriages. Although nothing could equal the excellence of Oriental resins, he found sandarac from Western juniper trees the best substitute. This, together with various gums dissolved in alcohol and turpentine and mixed with bitumens, produced the different varnishes Watin relied upon. His book gives detailed instruction for preparing the wood, covering it with cloth, varnishing this with from eight to 20 coats, polishing the surface, drawing and painting the designs, and making relief decorations. The best results of this process (*e.g.*, an 18th-century commode attributed to René Dubois in the Wallace Collection, London) were never as hard and brilliant as real Oriental lacquer, but they provided an admirable substitute; on occasions it is not easy to distinguish them, especially when East Asian designs were imitated.

Reproduced by permission of the Trustees of the Wallace Collection, London



Lacquered commode attributed to Rene Dubois (18th century). In the Wallace Collection, London. 1.57 m X 91 cm.

Introduction of European motifs

The chief interest of European lacquer of the 18th century lies, however, in the fact that it was not a purely imitative art, as it had been at its beginning. Until about 1720, European craftsmen sought to reproduce exactly the figures, the architectural settings, and the stylized vegetable forms of the imported lacquers. Then, in keeping with the playful spirit of the rococo, they modified these designs by introducing European figures, exotic animals such as monkeys, draperies and arabesques, and cartouches and ribbon compositions. Along with this transformation, in place of the conventional black and gold of the Oriental products, came a wide choice of background colours, including scarlet, yellow, white, blue, and green, sometimes flecked with gold. Particularly in Venice, where craftsmen followed the rules of a treatise by Filippo Buonanni (1720), a great originality was achieved by the informal spacing of bouquets of flowers around gracefully posed figures set against delicate hues of yellow and bluish green.

The uses of lacquer in Europe reflect the changing tastes of the 17th and 18th centuries. In the late 17th century it was principally employed for decorating the cases of cabinets set upon carved baroque bases or for ornamenting leather wall coverings. In the 18th century, bookcase desks, tall clocks, and tea tables became the most fashionable articles of lacquered furniture in England and Germany, while in France and Italy the chest of drawers and corner cabinet were preferred. Whole sets of furniture with Orientalizing lacquer decoration were made in England by Giles Grendey and other cabinetmakers of the Chippendale era (1754–68). Rooms with lacquered walls have survived at the palaces of Nymphenburg outside Munich, Bamberg, and other German cities. Small lacquered boxes called ***Boite de Spa*** became a specialty of that Belgian town and the nearby centres of Liège and Aachen, where a member of the Dagly family was active.

By courtesy of the Victoria and Albert Museum, London

Lacquered armoire belonging to David Garrick; by Giles Grendey (1693–1780). In the Victoria and Albert Museum, London. 1.4 m × 1.68 m × 62 cm.

In the early 19th century the solemn grandeur of the international Classical style with its insistence on plain wood or gilt surfaces almost eliminated the taste for lacquered furnishings. It survived in Victorian times in the vogue for painted tin or toleware (as in Pontypool, England) and the decoration of small tables and chairs of papier-maché inlaid with mother-of-pearl. The making of this furniture, dominated by the London firm of Jennens and Bettridge, rapidly declined after reaching the height of its popularity about the time of the Great Exhibition held in London in 1851. In the period 1925–30, two European sculptors, the Belgian Marcel Wolfers and the Franco-Swiss Jean Dunand, successfully approximated the true Oriental lacquer techniques in a few art objects of clay, wood, and bronze. By this time, however, the making of lacquer surfaces had become a part of the chemical industry, which today produces effects of depth and hard translucence surpassing the finest products of the European lacquer masters of the past. (R.C.Sm.)

BIBLIOGRAPHY

European lacquer: HANS HUTH, ***Lacquer of the West: The History of a Craft and an Industry, 1550–1950*** (1971), a concise, well-written statement, the most complete to date, based on a life-time study, with many excellent illustrations.

Chinese lacquer: FILIPPO BONANNI, ***Trattato sopra la vernice comunemente detta cinese*** (1720); PERE L. D'INCARVILLE, "Mémoire sur le Vernis de la Chine," in the *Mémoires* of the Académie Royale des Sciences (1760); STEPHEN W. BUSHELL, ***Chinese Art,*** vol. 1 (1904); EDWARD F. STRANGE, ***Catalogue of Chinese Lacquer in the Victoria and Albert Museum*** (1925) and ***Chinese Lacquer*** (1926); R.S. JENYNS, "Chinese Lacquer," ***Transactions of the Oriental Ceramic Society,*** vol. 17 (1939–40); FRITZ LOW-BEER, "Chinese Lacquer of the Early 15th Century" and "Chinese Lacquer of the Middle and Late Ming Period," ***Bulletin of the Museum of Far Eastern Antiquities,*** no. 22 and 24 (1950–52); KURT HERBERTS, ***Das Buch der ostasiatischen Lackkunst*** (1959; Eng. trans., ***Oriental Lacquer,*** 1962); WERNER SPEISER, ***Lackkunst in Ostasien*** (1965); HARRY M. GARNER, "Diaper Backgrounds on Chinese Carved Lacquer," ***Ars Orientalis,*** 6:165–190 (1966), and "A Group of Chinese Lacquers with Basketry Panels," ***Archives of Asian Art,*** 20:6–24 (1966-67); LEE YU-KUAN, ***Oriental Lacquer Art*** (1972).

Japanese lacquer: JOHANN J. REIN, ***Japan, nach Reisen und Studien,*** vol. 2 (1886; Eng. trans., ***The Industries of Japan,*** vol. 2, 1889, reprinted 1969); MICHAEL TOMKINSON, ***A Japanese Collection,*** 2 vol. (1898); ***L'Histoire de l'art du Japon,*** Paris Exhibition (1900); FRANK BRINKLEY, ***Japanese Temples and Their Treasures,*** 3 vol. (1910); ***Official Catalogue of the Japan-British Exhibition*** (1910); EDWARD F. STRANGE, "The Incense Ceremony and Its Utensils," Japan Society ***Transactions,*** 21:28–38 (1923–24), and ***Catalogue of Japanese Lacquer and Inrō in the Victoria and Albert Museum*** (1924); MARTHA BOYER, ***Japanese Export Lacquers from the Seventeenth Century in the National Museum of Denmark*** (1959); BEATRIX VON RAGUE, ***Geschichte der japanischen Lackkunst*** (1967).

(R.C.Sm./E.F.S.)

Lacrosse

Lacrosse is the oldest organized sport in North America and, inasmuch as baseball and U.S. and Canadian football were developed respectively from pre-existing rounders and rugby prototypes, is the only modern team sport, except for basketball, that originated on that continent. The game resembles the Irish national game of hurling and the hockey games (and thus also the football games, especially soccer) in the flow of play, in which two sides or teams each attempt to propel an object (ball) into a goal defined by uprights and a crossbar framing a loose net and to prevent the other side from doing so.

The crosse

The distinctive feature of the game is the crosse, the implement used by the players to carry, catch, and pass the ball. The crosse is a staff of wood, usually hickory, the top being sharply bent to form a hook from the end of which a thong is drawn and fastened to the shaft about two or three feet (0.6 or 0.9 metres) from the end of the handle, forming an oval triangle that is woven with a loose network of leather, nylon, or gut to form the pocket with which the ball is handled.

The following discussion traces the history and development of the game and describes the playing field and equipment and the principles of play and points out the distinctive features of the women's game.

History. Lacrosse was played by The Six Nations of the Iroquois in upper New York State and lower Ontario long before Columbus landed in the New World. The sport was then played in a much rougher form than it is today. Among some tribes as many as a thousand players took part on each side, goals were miles apart, and a game could last as long as three days (thus somewhat reminiscent of the Romans' *harpastum* and the *mêlées* or mellays indigenous to ancient Briton, battle-like pastimes held to be ancestral to modern football). The strategy of each player was to disable as many opponents as possible with the stick he carried and afterward concentrate on scoring a goal. The Cherokees called their version of the game "little brother of war." Because of the endurance required and the injuries that had to be borne with fortitude it was considered excellent training for combat. Among many tribes the game was as much a mystic ceremony as a sport and was preceded by complex rituals and a solemn dance. Indians on government reservations in the United States and Canada still field strong teams.

Origin of the name

To the first French settlers in Canada who saw the game, called "baggataway" by the Indians, the shape of the implement used to catch, carry, and throw the ball suggested a bishop's crozier *(la crosse)*, giving the sport its name; the stick itself becoming known as a crosse.

Europeans in Canada started playing the game in about 1840, and the first lacrosse organization, the Olympic Club, was founded in Montreal in 1842. In playing Indian teams, white players lost so frequently they were allowed to field extra men. Members of the Montreal Lacrosse Club, founded in 1856, modified the rules somewhat, and four years later George Beers of Montreal, called "the father of lacrosse," made further changes that included replacing the Indian ball of deerskin stuffed with hair by a hard rubber ball, limiting the number of players on a team to 12, and improving the stick for easier catching and throwing of the ball. The 12 players were designated at that time as goal, point, cover point, first defense, second defense, third defense, centre, third attack, second attack, first attack, out home, and in home. The year 1867 was an important one for the sport: Parliament declared it Canada's national game, the National Lacrosse Association was formed, and the game was introduced to England. Capt. W.B. Johnson of Montreal toured with a team of Caughnawaga Indians, appearing before Queen Victoria at Windsor Castle, who found the game "very pretty to watch." The English took to the sport and the game achieved popularity in various parts of the country—notably in Lancashire, Cheshire, Yorkshire, Manchester, Bristol, and London. The English Lacrosse Union was founded in 1892. English teams have exchanged visits with teams from the United States and Canada from time to time, and combined Oxford–Cambridge teams have frequently exchanged visits with college or all-star teams from the United States. The game is also popular in Ireland, Australia, New Zealand, and the Republic of South Africa. In both Canada and the United States it has grown with extreme rapidity since the mid-1950s, being played at the high-school and elementary-school levels, as well as by colleges and club teams. Lacrosse was included in the Olympic Games in 1904 and 1908 with teams representing Canada, the U.S., and Great Britain. Teams from those countries also demonstrated the sport in exhibitions at the Games in 1928, 1932, and 1948, but lacrosse did not attract international interest and is no longer recognized as an Olympic sport.

Arrival of the game in England

Although preoccupied with baseball in its embryo form in the late 1800s, the United States was not long in developing an interest in the game. A team of Indians introduced it to Troy, New York, in about 1868, and a few years later teams were started in that city, New York City, and Brooklyn. In the 1880s Eastern colleges including New York University, Princeton, Yale, and Harvard took up the sport. In 1884 a successful tour of Europe was made by a team of American collegians and in 1906 the U.S. Intercollegiate Lacrosse League was formed. The game received its greatest impetus, however, when it was introduced to Baltimore, Maryland, by some track men from that city who were intrigued when they saw the game being played by some Canadians on Long Island. Returning, they actively promoted the sport, aiming at all age levels, with the result that Baltimore is the main U.S. centre of lacrosse. Strong teams are invariably fielded by Johns Hopkins University and St. John's of that city and by the nearby University of Maryland and the U.S. Naval Academy at Annapolis. Two other colleges with fine records are the U.S. Military Academy at West Point, New York, and the University of Virginia. The Mt. Washington Club of Baltimore and the Long Island Athletic Club in Wantagh, Long Island, are perennial winners of annual national club championships sponsored by the U.S. Lacrosse Club Association. The Intercollegiate Lacrosse League was reorganized in 1926 to become the U.S. Intercollegiate Lacrosse Association, which had about 120 member colleges. In 1970 the National Collegiate Athletic Association (NCAA), with more than 500 member colleges, undertook sponsorship of intercollegiate lacrosse competition and championships, reflecting the growth of the sport in the country outside its traditional Eastern Seaboard stronghold. The college team considered the best in the country is awarded the Wingate Trophy, Johns Hopkins being its most frequent recipient.

The field and equipment. The basic elements and dimensions of the field are shown on the accompanying diagram. The field is 110 yards (about 100 metres) long and 60 yd (55 m) wide. The goals are 80 yd (73 m) apart, the goal posts being 6 feet (1.83 m) high and the same distance apart, surmounted by a crossbar. The posts are fitted with a pyramid-shaped netting fastened to the ground behind the goal to stop passage of the ball after a successful shot. The goal-area lines, wing lines, and centre line serve to act as restraints on certain players during play; *i.e.*, their range of movement is limited. Passing over a line into a prohibited area results in a penalty.

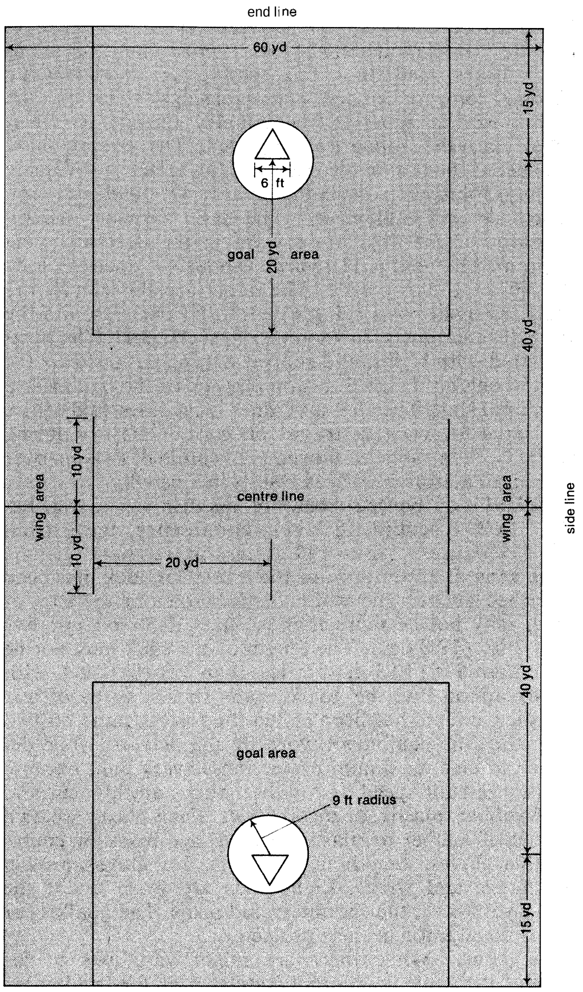
The ball is of sponge rubber, not less than 7% to 8 inches (19.7 to 20.3 centimetres) in circumference, and is from 5 to 5% ounces (142 to 149 grams) in weight.

The general appearance of the crosse, or stick, has been described above. The width of the crosse at the top, or head, may not be more than 12 in. (30.5 cm) nor less than 7 in. (17.8 cm). The length of the stick may not be more than 6 ft (1.83 m) nor less than 3 ft (0.91 m), with the exception that the goalkeeper's stick may be of any length. Lacrosse has been called the fastest game on two feet: thus the equipment worn by the players, while designed to give maximum protection, is very light because of the need for speed and agility. Shoes are low cut and have rubber, plastic, or metal cleats. Each player wears a helmet of leather or plastic, with a face mask or guard. Leather gloves, something like ice-hockey gloves, protect the hands and wrists. Light pads are worn under the jersey to protect the shoulders and arms. The goalkeeper wears, in addition, a chest protector.

Use of lightweight equipment to enhance speed of play

"Cradling" the ball

The game. While there are minor variations in the rules of the game in different countries, all are fundamentally similar. The object of the game is to send the ball through the opponents' goal as many times as possible and to prevent one's opponents from scoring. A goal counts one point. Under most rules for men, there are ten players per team. They are designated as the goalkeeper or goaltender, three defensemen, three midfielders (one of whom is the centre), and three attackmen. During play each team must have at least four players in its defensive half of the field and no fewer than three in its offensive half of the field. The purpose of this rule is to prevent excessive crowding around a goal when it is under attack. Conventionally, the goalkeeper and the three defensemen stay in the defensive half, while the three attackmen stay in the offensive half. The midfielders are permitted to roam the field, reinforcing the attack or defense as needed. There are two officials, a referee and a judge. The game is divided into four periods of 15 minutes each, with intervals of one minute between the first and second quarters and between the third and fourth quarters and a ten-minute rest at half time. If the score is tied at the end of regulation time, play is resumed after an intermission of five minutes for two four-minute periods, with a one-minute rest in between. Free substitution is allowed. A player may run with the ball, pass it in any direction, and catch it, but—with the exception of the goalkeeper—he may not touch it with his hand. A player may kick the ball or bat it, but not into the opponents' goal. A unique feature of the game is "cradling," in which the player rapidly rotates the stick in half-turns while holding it nearly upright as he runs. The centrifugal force developed keeps the ball in the pocket of the crosse and also puts it in position for accurate throwing. Defensive players are allowed to poke the ballcarrier in the body with their sticks or slap at his stick to dislodge the ball. Blocking the ballcarrier—*i.e.*, hitting him with the shoulder in an attempt to throw him off-balance or knock him down—is legal. For minor infractions of the rules the penalty is either suspension



Playing field for men's lacrosse

from the game for 30 seconds or an exchange of the ball. In the case of personal fouls of a more serious nature—for example, an illegal block (as below the knees from the rear)—the offender is suspended from the game for one, two, or three minutes, and his team plays one man short for that period of time. Other serious fouls in the game of lacrosse include tripping, slashing, and unnecessary roughness.

Play is started at the beginning of each quarter and after the scoring of a goal with a face-off at midfield, as in hockey. The two centres face each other, the heads of their sticks touching the ground. The referee places the ball between the two crosses, and at his signal each player tries to gain control of the ball. He may keep it himself or bat it to a teammate. The player with the ball tries to advance it toward the opposing goal by running or passing to a teammate in the open. The defenders try to force him into making a poor pass, intercept the ball when it is thrown, knock the ball from his stick, or occasionally knock it loose with a block. The game is lively since the players are in constant movement: dodging, hurling, or flipping the ball to a teammate; scooping up the ball while running at full speed; or making quick, deceptive shots at the goal. A unique ceremony at the beginning of the game consists of the teams lining up in the centre of the field opposite each other. Each player introduces himself to his particular opponent, shakes hands, and wishes him luck.

Women's lacrosse. This version of the game, popular in the British Isles, Australia, Canada, New Zealand, South Africa, Switzerland, and the United States, was first played in Scottish and English private schools in the early 1900s. It was introduced to schools and colleges in the eastern United States by English women educators. Frequent exchange visits have been made by American and English women's teams, the latter showing themselves to be the best in the world. British-Irish touring teams have continued to maintain excellent records against American teams.

The game is one of speed and skill in passing, no body contact or rough play with the stick being permitted. With the exception of the goalkeeper, who wears a chest pad and leg guards, the players do not wear protective equipment. There are 12 players on each side. The goals are 90 to 110 yards (82 to 100 metres) apart, and there are no sidelines or end lines, the goal creases and centre circle being the only ground markings. A game consists of 25-minute halves, with a 10-minute intermission. There is no overtime in the case of a tie. The rules established by the All-England Ladies' Lacrosse Association are universally accepted by other nations. In the United States, the Philadelphia, Baltimore, and Boston areas traditionally have had strong teams, while the main English centres of women's lacrosse are Hertfordshire, Kent, Norfolk, Essex, and Suffolk.

BIBLIOGRAPHY. W.K. MORRILL, *Lacrosse*, rev. ed. (1966), instruction for the more advanced player, describing techniques and team play, with an extensive bibliography; MARGARET BOYD (ed.), *Lacrosse*, rev. ed. (1969), a study of women's lacrosse, including fundamentals, positional play, tactics, equipment, and rules; MIKE HANNA, *Lacrosse for Men and Women* (1980), information on every facet of the game, with differences between men's and women's rules noted; BOB SCOTT, *Lacrosse: Technique and Tradition* (1976), descriptions of various methods and styles of play; JIM HINKSON, *Box Lacrosse* (1974), a text on this version of the game; A.M. WEYAND and M.R. ROBERTS, *The Lacrosse Story* (1965), a splendid, detailed coverage of the history of the game and its more noted players and innovators, supplemented by Roberts' *The Lacrosse Story, 1965–1976* (1977); *Lacrosse Guide* (annual), a handbook covering U.S. teams, along with the latest rules; AMERICAN ALLIANCE FOR HEALTH, PHYSICAL EDUCATION, RECREATION, AND DANCE, NATIONAL ASSOCIATION FOR GIRLS AND WOMEN IN SPORT, *NAGWS Guide: Lacrosse* (biennial), rules on women's lacrosse in the United States; REX LARDNER, "The Assault and Battery of Lacrosse," *Holiday*, 45:66, 108–109 (1969), an article on the game's history and techniques.

(R.La.)

Lactation, Human

Lactation, the secreting and yielding of milk, normally occurs in human beings and other mammals after the birth of an offspring. The secreting organs are the mammary glands, contained within the breasts. The structure and diseases of these glands, and of the breasts, are discussed in the article MAMMARY GLANDS, HUMAN. Breast development begins at puberty, though the rudiments of the mammary glands are present much earlier and can manufacture a secretion in the first few days after birth. The mammary glands of boys and girls are similar before puberty, and most of the increase in size evident in girls upon sexual maturation is due to the accumulation of fat and connective tissue within which the secretory elements, called alveoli, and the duct system of the mammary glands are embedded. The ovarian hormones, estrogen and progesterone, are necessary for mammary gland growth, and in their absence, as in girls with malfunction of the pituitary gland or whose ovaries have been lost before puberty, breast development is minimal. Rhythmic changes in the microscopic structure of the breasts are detectable during the menstrual cycle, and many women describe a feeling of heaviness, fullness, and tenderness in the breasts just before menstruation (see MENSTRUATION). These changes increase during pregnancy, when each mammary gland may increase in weight by a pound or more, largely because of the development and elaboration of the milk-secreting apparatus.

Development of the breasts and the mammary glands

Enlargement of the breasts during pregnancy. The growth of the breasts during pregnancy is controlled by the hormones prolactin, from the hypophysis, or pituitary gland, and estrogen and progesterone, from the ovaries and the placenta (the temporary organ whose prime purpose is to provide nourishment for the fetus and to eliminate its wastes). The placenta also produces a substance that acts like prolactin. Estrogen promotes development of the ducts of the mammary glands, whereas alveolar growth is most greatly influenced by progesterone; both hormones are secreted in increased amount during pregnancy. Milk formation follows from the action of prolactin. Experiments on rats have shown that additional hormones aid in maintaining and augmenting mammary growth and lactation. These include thyroid hormone, adrenal hormone, insulin, and growth hormone. Because of the dual supply of estrogen and progesterone, from the ovaries and the placenta, the ovaries can be removed between the first and second months of pregnancy without causing loss of the fetus or arresting breast growth. Sufficient prolactin-like hormone for mammary development can also be produced by the placenta, as was shown in cases in which the removal of the pituitary gland during pregnancy did not stop mammary growth and development in monkeys or in a woman, and milk secretion occurred after delivery.

Initiation of lactation. A hormonal change that initiates lactation occurs during pregnancy. (It is no longer possible to believe, as did the ancient Greeks, that milk is produced in the uterus and carried to the mammary glands by special vessels, so that the uterine milk that has nourished the fetus can continue to be utilized after birth.) Although mammary growth occurs during pregnancy under the influence of ovarian and placental hormones and some milk is formed, copious milk secretion sets in only after delivery. Since lactation ensues after a premature birth, it would appear that milk production is held back during pregnancy. The mechanism by which this inhibitory effect is brought about, or by which lactation is initiated at delivery, has long been the subject of an argument that revolves around the opposing actions of estrogen, progesterone, and prolactin, as studied in laboratory animals, goats, and cattle. During pregnancy the combination of estrogen and progesterone circulating in the blood appears to inhibit milk secretion by blocking the release of prolactin from the pituitary gland and by making the mammary gland cells unresponsive to this pituitary hormone. The blockade is removed at the end of pregnancy by the expulsion of the placenta and the loss of its supply of hormones, as well as by the decline in hormone production by the ovaries, while sufficient estrogen remains in circulation to promote the secretion of prolactin by the pituitary gland and so favour lactation. It is also possible that a relative increase in the blood level of adrenal hormones favours the production of milk.

Milk-producing complex of hormones

Continuation of lactation. For lactation to continue, necessary patterns of hormone secretion must be maintained; and disturbances of the equilibrium by the experimental removal of the pituitary gland, in animals, or by comparable diseased conditions in human beings, quickly arrest milk production. Several pituitary hormones seem to be involved in the formation of milk, so that it is customary to speak of a lactogenic ("milk-producing") complex of hormones. To some degree, the role of the pituitary hormones adrenocorticotropin, thyrotropin, and growth hormone in supporting lactation in women is inferred from the results of studies done on animals and from clinical observations that are in agreement with the results of animal studies. Prolactin, growth hormone, and adrenal hormone seem of greatest value in restoring lactation after removal of the pituitary, although the precise response varies from species to species. The greatest difficulty with respect to the observations on women has been experienced in separating the actions of prolactin and growth hormone, for highly purified samples of growth hormone prepared from human pituitary glands still possess prolactin-like activity and can greatly enhance milk production.

The release of prolactin and of other hormones from the pituitary is controlled by the brain through its hypothalamic area. In the case of prolactin, however, the control consists in holding back output; whereas brain activity serves to stimulate the secretion of the other pituitary hormones. Thus, surgical separation of the connections between the pituitary gland and the brain, or depression of the functioning of the hypothalamus by drug treatment, favours the release of prolactin and the secretion of milk while it reduces the secretion of the other pituitary hormones. Tranquillizing agents, such as chlorpromazine and reserpine, can cause breast enlargement and lactation, probably by increasing the output of prolactin. In the condition known as pseudocyesis, or false pregnancy, the affected person firmly believes that she is pregnant and shows many of the signs of pregnancy such as breast enlargement, milk secretion, and abdominal swelling, although the uterus is empty. This effect of a psychological disturbance is an excellent example of the influence of the mind on the secretion of hormones.

Suckling and the release of oxytocin

The stimulus of nursing or suckling supports continued lactation in two ways: it promotes the secretion of prolactin (and possibly other pituitary hormones of value in milk formation), and it triggers the release of yet another hormone from the pituitary gland—oxytocin, which causes contraction of special muscle cells around the alveoli in the breast and ensures the expulsion of milk. It is in this way that a baby's sucking at one breast may cause an increase in milk flow from both, so that milk may drip from the unsuckled nipple. About 30 seconds elapse between the beginning of active suckling and the initiation of milk flow.

The nerve supply to the mammary glands is not of great significance in lactation, for milk production is normal after the experimental severing of nerves to the normal mammary glands in animals or in an udder transplanted to the neck of a goat. Milk ejection, or "the draught," in women is readily conditioned and can be precipitated by the preparations for nursing. Conversely, embarrassment or fright can inhibit milk ejection by interfering with the release of oxytocin; alcohol, also, is known to block milk ejection in women, again by an action on the brain. Beyond its action on the mammary glands, oxytocin affects uterine muscle, so that suckling can cause contractions of the uterus and may sometimes result in cramp. Since oxytocin release occurs during sexual intercourse, milk ejection in lactating women has been observed on such occasions. Disturbance of oxytocin secretion, or of the milk-ejection reflex, stops lactation just as readily as a lack of the hormones necessary for milk production, for the milk in the breast is then not extractable by the infant. Many instances of nursing failure are due to a lack of milk ejection in stressful circumstances; fortunately, treatment with oxytocin, coupled with the reassurance gained from a successful nursing, is ordinarily successful in overcoming the difficulty.

Suckling can initiate lactation in nonpregnant women. This has been seen most often in women of child-bearing age but also has been observed in older persons. A baby who had lost his mother was suckled by his 60-year-old grandmother, who had borne her last child 18 years before. The grandmother produced milk after a few days and continued to nurse the baby until he was a year old and could walk. Rarely, lactation has been reported to set in after operations on the chest; in such instances it is attributed to injury or irritation of the nerves in this region. Such observations argue against the possibility that lactation continues simply as a consequence of emptying the breasts.

The production of milk during lactation seems to decline with age, regardless of the number of previous pregnancies. Some correlation exists between the size of the breasts and milk secretion, but women with small breasts secrete milk normally, so that output could probably be raised by increasing the frequency of nursing. Drinking water beyond the demands of thirst impairs lactation and has been advocated as a means of suppressing this process.

Varying customs with respect to weaning

Weaning and the cessation of lactation. There is no typical age at which human infants are weaned, for this varies from country to country and among the social classes of a nation. In the United States in 1946 only 38 percent of babies were totally breast-fed at the time of discharge from the hospital. By 1958 the figure had fallen to 21 percent, but by the early 1970s had risen to about 23 percent. Breast-feeding rates vary with region and culture and may be higher in Arizona, Colorado, and Georgia than in Connecticut, Maine, and Massachusetts. In India women in the higher socioeconomic groups tend to use artificial feeding, while the reverse relationship holds in Britain and the United States. Most commonly, weaning is a gradual process, with a gradual increase in the proportion of solid food supplied to the infant together with breast milk. Pediatricians in general have concluded that, on the basis of present knowledge, no nutritional superiority or psychological benefits result from the introduction of solid foods into the infant diet earlier than 2% to 3½ months and that normal full-term infants can be expected to thrive for the first three months of life on a diet consisting exclusively of milk, either normal human milk or properly modified milk from other sources.

With the reduced demand of the baby, lactation slowly declines and stops. Estrogen treatment is often used to suppress lactation, and the high doses used may accomplish this; but there is often a rebound effect at the end of treatment. Lactation may be slightly depressed when oral contraceptives are being taken in high dosage.

Nursing as a contraceptive technique

Apart from providing nourishment for the infant, nursing may be continued as a means of contraception. In a recent study in which the subjects were women of European stock, menstruation was found to return in a high proportion of the women who had lactated for three months or more, while nearly all women who had discontinued lactation some two months after delivery resumed menstruation within six weeks. Different results emerged from studies on Punjabi women, for whom the mean duration of lactation was 21 months and the median length of time from the date of giving birth to the return of menstruation was 11 months. The lesser effect of lactation in delaying menstruation in women of European stock may have been due to the provision of supplementary food in addition to the breast milk so that the suckling stimulus was reduced in intensity.

Composition and properties of milk. Milk can be regarded as an emulsion of fat globules in a colloidal solution of protein together with other substances in true solution. Two constituents of milk, the protein casein and milk sugar, or lactose, are not found elsewhere in the body. The general assumption that breast milk provides the ideal food for human babies, in that it contains all the elements required in a good diet and in the most appropriate proportions, though entirely reasonable, is extremely difficult to prove, for controlled experiments are hardly possible with human infants. It is also extremely difficult to disprove, for the faster growth that may occur with supplemented diets is not necessarily optimal growth. While supplying an excellent infant diet, breast feeding confers other benefits, for the milk is normally delivered from a sterile container, and the feelings of well-being and the emotional bond generated between mother and child are important in the psychological welfare of both.

Colostrum

The milk released from the breast when lactation starts differs in composition from the mature milk produced when lactation is well established. The early milk, or colostrum, is rich in essential amino acids, the protein building blocks essential for growth; it also contains the proteins that convey immunity to some infections from mother to young, although not in such quantity as among domestic animals. The human infant gains this type of immunity largely within the uterus by the transfer of these antibody proteins through the placenta; the young baby seldom falls victim to mumps, measles, diphtheria, or scarlet fever. For a short time after birth, proteins can be absorbed from the intestine without digestion, so that the acquisition of further immunity is facilitated. The growth of viruses and bacteria in the intestines is probably inhibited by immune factors in human milk. After childbirth the composition of milk gradually changes; within four or five days the colostrum has become transitional milk; mature milk is secreted some ten days after delivery.

Some Constituents of Human Colostrum, Transitional, and Mature Milk and of Cow's Milk
(average values per 100 millilitres whole milk)

	colostrum (1–5 days)	transitional (6–10 days)	mature (after 30 day?)	cow's milk
Energy, kcal*	58	74	71	69
Total Solids, g	12.8	13.6	12.4	12.7
Fat, g	2.9	3.6	3.8	3.7
Lactose, g	5.3	6.6	7.0	4.8
Protein, g	2.7	1.6	1.2	3.3
Ash, g	0.33	0.24	0.21	0.72
Minerals				
Calcium, mg	31	34	33	125
Magnesium, mg	4	4	4	12
Potassium, mg	74	64	55	138
Sodium, mg	48	29	15	58
Iron, mg	0.09	0.04	0.15	0.10
Casein, g	1.2	0.7	0.4	2.8

***Kilocalorie; sufficient energy to raise the temperature of one kilogram of water one degree Centigrade.**

Some variations between human colostrum, transitional milk, and mature milk and cow's milk are shown in the accompanying table. The greater amount of protein in unmodified cow's milk is largely responsible for its dense, hard curd, which the infant cannot digest; the difficulty can be avoided by heat treatment or dilution of the milk. Ordinarily, when cow's milk is fed to young infants, it is modified so as to match its composition as far as possible to breast milk.

BIBLIOGRAPHY. S.J. FOLLEY, *The Physiology and Biochemistry of Lactation* (1956), a compact and readable survey; S.K. KON and A.T. COWIE (eds.), *Milk: The Mammary Gland and its Secretion*, 2 vol. (1961), the most comprehensive work currently available, and an excellent starting point for further reading; G.W. HARRIS and B.T. DONOVAN (eds.), *The Pituitary Gland*, 3 vol. (1966), a treatise covering many aspects of the structure and function of the pituitary, with chapters on the control of milk secretion and milk ejection; R.H. WILLIAMS, *Textbook of Endocrinology*, 4th ed. (1968), an up-to-date summary of the control of mammary growth and lactation in women; B.T. DONOVAN, *Mammalian Neuroendocrinology* (1970), a short text that deals with many of the interactions between the brain and the endocrine system, including the control of lactation. The following articles are also recommended: N. and M. NEWTON, "Psychologic Aspects of Lactation," *New Eng. J. Med.*, 227:1179–1188 (1967); L.F. HILL, "Infant Feeding: Historical and Current," *Pediat. Clin. N. Amer.*, 14:255–268 (1967); M. GUNTHER, "Diet and Milk Secretion in Women," *Proc. Nutr. Soc.*, 27:77–82 (1968); and P.A. DAVIES, "Feeding the Newborn Baby," *ibid.*, 28:66–72 (1969).

(B.T.D.)

Ladoga, Lake

Lake Ladoga (Ladozhskoye Ozero in Russian; Ladozhskoje Ozero in the transliteration system of the Akademiya Nauk), in the northwest European part of the Union of Soviet Socialist Republics, is the largest lake in Europe. Located about 25 miles east of Leningrad in the Russian Soviet Federated Socialist Republic and in the Karelian Autonomous Soviet Socialist Republic, it is 6,826 square miles (17,678 square kilometres) in area --exclusive of islands—and 136 miles (219 kilometres) long, with an average width of 51 miles and an average depth of 167 feet. Its greatest depth is 754 feet.

The lake basin

Lake Ladoga's basin has a total area of about 100,000 square miles. The northern part of its basin and depression lies within the Baltic crystalline shield, and the southern part lies within the so-called Russian Platform of Precambrian rocks (more than 570,000,000 years old). The depression of the lake was produced by the action of glaciers. The northern shores are mostly high and craggy and are broken by deep, ice-covered, fjord-like inlets. There are numerous, mostly wooded islands here, with cliffs. The southern shores, which have many sandy or rocky beaches, are primarily low, slightly indented, and overgrown with willows and alders. In some places there are ancient coastal embankments overgrown with pines. The contour of the northern part of the lake bottom is complex, with deep depressions alternating with shallower areas. Depths of more than 300 feet predominate. The greatest depth, 754 feet, is located west of Valaam Island (Ostrov). In the southern section of the lake the bottom is more uniform, with depths decreasing from 300 to 30 feet and less; there are also many sandy and rocky slopes and sandbanks. In the deep areas of the northern and central sections of the lake, the bottom is muddy, whereas in the south it is mainly sandy and covered with detritus. In many areas there are boulders near the shore.

The lake also contains approximately 660 islands of more than 2½ acres in extent, occupying a total of 176 square miles. About 500 of the islands are located near the northwestern shore. In the central section of the lake, approximately 65 islands make up the Valaam and Western archipelagoes. The largest islands are Riyekkalan-Sari, Mantsinsari, Kilpola, Tulolansari, and Valaam.

The climate in the Lake Ladoga region is moderately cold, with a mean air temperature of 14° to 18" F (–10° to –8" C) in February and 61" to 64" F (16" to 18° C) in July. The absolute minimum temperature is –65" F (–54" C) and the absolute maximum 90" F (32" C). Mean annual precipitation is 24 inches.

There are approximately 50,000 lakes and 3,500 rivers of more than six miles length in the Lake Ladoga Basin. The largest tributaries are the Volkhov from the south, the Svir from the southeast, and the Vuoksa from the west. The discharge of river waters into the lake, mainly from the Volkhov, Svir, and Vuoksa, accounts for 85 percent of its incoming water balance (an average of 16.3 cubic miles per year), with 13 percent coming from precipitation and 2 percent from groundwater. Of the outgoing water balance, 92 percent flows into the Neva River (an average of 17.8 cubic miles per year), and the remaining 8 percent is accounted for by evaporation from the water surface. The lake is highest in June and July and lowest in December and January, with the average annual range of variation being 2.6 feet and the absolute maximum being about 9.8 feet. Taken together, the northern section has variations from about two to four inches, and, less frequently, up to a range of seven to 15 inches, with the southern section having level variations of up to 35 inches. Seiches, or changes in the water level, can be observed; in the north and centre of the lake, the seiches average ten to 11.5 feet and occasionally rise to 16 or 20 feet; in the south, they measure about eight feet. During the fall there are frequent storms.

Water temperatures

Thermal conditions differ from the deep central to the shallow coastal regions of the lake. The coastal regions and inlets usually freeze at the beginning of December, and the open central area freezes in January or February; the average ice thickness is 20–23 inches, with a maximum of 35–39 inches. The central part of the lake opens in late March or early April and the northern part not until the beginning of May. The temperature of the bottom layers varies from 35.5" to 36.5" F (2° to 2.5" C) in the winter and from 39" to 41° F (4° to 5° C) in the summer. In August the mean temperature of the surface is about 61° F (16" C), with a maximum of 77" F (25" C).

The water of Lake Ladoga is yellow brown in colour. Its transparency averages 15 feet in the centre, 6.5 to eight feet near the western shore, three to 6.5 feet near the eastern shore, and one to three feet in the estuarine areas. The greatest transparency east of Valaam Island is from 26 to 33 feet. Lake Ladoga's water is fresh, with an average mineralization of 56 milligrams (0.002 ounce) of calcium hydrocarbonate per litre. Its dissolved oxygen content is 14 to 15 milligrams per litre in winter and in summer from ten to 11 milligrams per litre in the surface layers and from 12 to 13 milligrams per litre in the depths. The lake abounds in fish; of commercial importance are salmon, trout, bull trout, whitefish, pike perch, golden shiner, bass, roach, pickerel, and frostfish.

Water transportation and fishing are the principal commercial uses of Lake Ladoga. The lake is part of the Volga–Baltic water route and of the Baltic–White Sea Waterway system, through which freight is carried, without the need for transshipment, to points within the U.S.S.R. and to Finland, the German Democratic Republic, and other countries. A bypass canal was cut along its southern shore to protect shipping against storms.

During World War II, when Leningrad was under siege from September 1941 to March 1943, Lake Ladoga was the lifeline connecting it with the rest of the country. Supplies and military equipment were brought to the city across the water and ice, and the sick and wounded were evacuated over the same route.

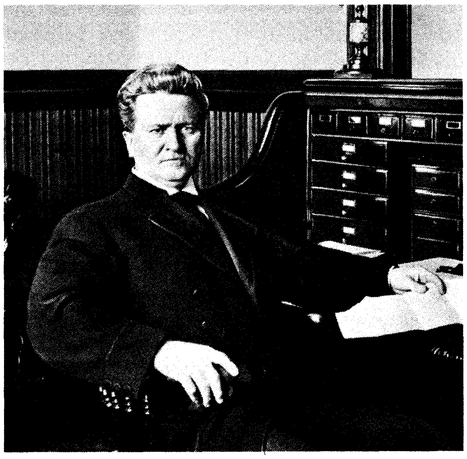
The cities of Priozyorsk, Petrokrepost, and Sortavala are located on its shores. (B.B.Bo.)

La Follette, Robert M.

A congressman, governor, and senator from Wisconsin, Robert Marion La Follette was a prominent leader of the United States Progressive movement. A major critic from the left of five piesidents, La Follette was most famous for his courage and eloquence as a speaker and writer. He championed progressive legislation in Wisconsin and Washington and was the first United States politician to rely heavily on the help of university professors to draft and administer reform laws.

La Follette's progressivism developed slowly. Born on a farm in Primrose, Wisconsin, on June 14, 1855, he did not advocate progressive measures until 1897. In the meantime, as a boy growing up in moderately prosperous rural areas, as a student at the University of Wisconsin (1875–79), as a county district attorney (1880–84) and congressman from southwestern Wisconsin, La Follette developed the personality and style that made him a popular leader. He combined an unusually outgoing personality, which made it natural for him to absorb the ideas and prejudices of his constituents, with an extraordinary flair for zealous oratory. As an eloquent spokesman for popular causes, La Follette exalted his constituents' wishes—even when those wishes ran counter to the desires of party leaders. His principal concerns in his three terms as congressman were economical government and protection for his district's farmers. He married his college sweetheart, Belle Case, on December 31, 1881, after his first year as district attorney.

By courtesy of the Library of Congress, Washington, D.C.



La Follette, 1906.

Defeated for re-election to Congress in a Democratic landslide of 1890, La Follette returned to Madison to practice law and develop the political organization that within ten years would elect him governor and allow him to dominate Wisconsin politics until his death. His reputation as an enemy of political bosses began in 1891 when he announced that the state Republican boss, Sen. Philetus Sawyer, had offered him a bribe. For the next six years La Follette built a competing Republican faction on the support of other party members (Scandinavians, dairy farmers, young men, disgruntled politicians) with grievances against the dominant "stalwart" faction. His oratorical talents, combined with his natural charm, organizational skill, and driving ambition to become governor made him the leader of his new group of Republicans.

Campaign for governor

In 1897 La Follette began to advocate programs that local-level progressives had popularized during the legislative session a few months earlier. Following their lead, he demanded tax reform, corporation regulation, and political democracy. In particular, he promoted steeper railroad taxes and a direct primary. Elected governor on this platform in 1900, he was re-elected in 1902 and 1904.

As governor he developed new political techniques, which he later took to the Senate. The first, which received national attention as the "Wisconsin Idea," was the use of professors from the University of Wisconsin—57 at one point—to draft bills and administer the state regulatory apparatus created by the new laws. The second innovation was his public reading of the "roll call" in districts in which legislators had opposed his reform proposals.

With these new methods he secured the passage of several progressive laws. Believing that the railroads were the principal subverters of the political process, he persuaded the legislature to tax them on the basis of their property (1903) and to regulate them by commission (1905). The legislature enacted the direct primary in 1903 and state civil service reform in 1905. His appointees to the Tax Commission, given new power by the legislature, equalized tax assessments. Wisconsin's leadership in these areas gave La Follette his reputation as a pioneering progressive.

United States senator

Resigning as governor in 1906, he was elected to the Senate at a time when that institution was widely believed to be a refuge for millionaires. La Follette acquired instant fame as a new type of senator, one who was not controlled by "the interests," and in his first three years there La Follette achieved the passage of laws aimed against the freight rates, labour policies, and financing practices of the railroads.

These laws reflected an emerging ideology that dominated La Follette's Senate activities thereafter. Politics, he believed, was a never-ending struggle between "the people," all men and women in their common roles as consumers and taxpayers, and the "selfish interests" for control of government; law-given privileges allowed "selfish interests" to dominate all facets of American life. He supported labour legislation because unions were battling the same enemies that menaced consumers, and because consumers benefitted directly from improvements in working conditions. He believed, for example, that his most famous achievement, the La Follette Seaman's Act of 1915, would increase the safety of passengers while it also improved working conditions for sailors. Beginning in 1908, with elaborate documentation during debate on the Aldrich–Vreeland Currency Act, La Follette argued that the nation's entire economy was dominated by fewer than 100 men who were, in turn, controlled by the J.P. Morgan and Standard Oil investment banking groups. Thereafter, he shifted his concern from the power of railroads to the power of their "owners," namely the large banks.

In 1909 La Follette founded *La Follette's Weekly,* later a monthly, and, much later called *The Progressive.* The highpoint of his national popularity came in 1909–11 when he emerged as the leader of newly elected and newly converted progressives in Congress. Directing Republican opposition to the tariff, conservation, and railroad policies of Pres. William Howard Taft, La Follette was widely boomed for the presidency in 1912. Most progressives backed La Follette because their first choice, Theodore Roosevelt, had refused to run; later, when Roosevelt entered the race early in 1912, they deserted La Follette. The bitterness of La Follette's attacks on Roosevelt cost him his reputation as a leader and left him an independent figure in the Senate. Although he had backed Woodrow Wilson in 1912 for the presidency, he was disgusted that the new president ignored the ideas of progressive Republicans and shaped most legislation in the Democratic caucus. While applauding the social justice laws, he believed that most of Wilson's regulatory acts—particularly the Federal Reserve Board—constituted government sponsorship of big business.

Anti-war position

Foreign affairs catapulted La Follette back into a leadership position in 1917, this time of the anti-war movement. Since 1910 he had argued that U.S. interventions in the problems of foreign governments were intended to protect the investments of U.S. corporations and to smash revolutions. Now he believed that the U.S. entered World War I in 1917 because U.S. businessmen needed protection for their investments and because Wilson had become isolated from public opinion. Confident that the majority opposed U.S. involvement, La Follette led the campaign for a popular referendum on war in 1916–17. He led the 1917 Senate filibuster against arming U.S. merchant ships and voted against the war declaration. Once war was declared, he opposed the draft, defended the civil liberties of the war's opponents, and insisted that wealthy individuals and corporations pay the costs of a war that mainly benefitted them. Pro-war groups demanded his expulsion from the Senate for treason, but a Senate investigating committee exonerated him. As a martyr to the war hysteria, La Follette once again became a popular hero to millions of Americans.

Believing that the war had given large corporations nearly complete control over the federal government, La Follette concentrated on exposing the most flagrant corruption of the postwar years. His most significant contribution was his major role in publicizing the oil scandals of the Warren Harding administration.

As disillusionment with the war deepened and as labour and farm groups despaired of the conservatism of Democrats and Republicans alike, La Follette was frequently mentioned as a presidential candidate for a third party. Declining the pleas of the Farmer–Labor convention that he run in 1920, La Follette accepted the Progressive Party's nomination in 1924. His 1924 candidacy was supported by several farm groups, by organized labour (particularly the railroad brotherhoods, La Follette's oldest friends in the labour movement), by many old Progressives, by the Socialist Party, and by the Scripps–Howard newspaper chain. In the end La Follette carried only the state of Wisconsin, although he placed second in 11 states, and polled about one-sixth of the national total. He died in office, in Washington D.C., on June 18, 1925.

BIBLIOGRAPHY. The most detailed biography is BELLE CASE LA FOLLETTE and FOLA LA FOLLETTE, *Robert M. La Follette, June 14, 1855–June 18, 1925,* 2 vol. (1953). His *Autobiography* (1913), a campaign document for 1912, reveals his ideology. DAVID P. THELEN, *The Early Life of Robert M. La Follette, 1855–1884* (1966); ROBERT S. MAXWELL, *La Follette and the Rise of the Progressives in Wisconsin* (1956), on the governorship; HERBERT F. MARGULIES, *The Decline of the Progressive Movement in Wisconsin, 1890–1920* (1968), on Wisconsin politics; and KENNETH C. MACKAY, *The Progressive Movement of 1924* (1947), are useful monographs.

(D.P.T.)

La Fontaine, Jean de

The 17th-century poet Jean de La Fontaine is remembered chiefly for his *Fables,* which rank among the greatest masterpieces in French literature. Too wise to suppose that moral truths can ever be simple, he wrote stories that offer no rudimentary illustration of a certain moral but a subtle commentary on it, sometimes amending it and hinting that only the naïve would take it at face value. Thus, what the *Fables* teach is trivial in comparison with what they suggest: a view of life that, although incomplete (for it takes little account of man's metaphysical anguish or his highest aspirations), is mature, profound, and wise. Enjoyed at many different levels, the *Fables* continue to form part of the culture of every Frenchman, from schoolchildren to such men of letters as André Gide, Paul Valéry, and Jean Giraudoux, who have given fresh luster to La Fontaine's reputation in the 20th century.

He was born in 1621, probably on July 8, at Château-Thierry in Champagne into a bourgeois family. There, in 1647, he married an heiress, Marie Héricart, but they separated in 1658. From 1652 to 1671 he held office as an inspector of forests and waterways, an office inherited from his father. It was in Paris, however, that he made his most important contacts and spent his most productive years as a writer. An outstanding feature of his existence was his ability to attract the goodwill of patrons prepared to relieve him of the responsibility of providing for his livelihood. In 1657 he became one of the protégés of Nicolas Fouquet, the wealthy superintendent of finance. From 1664 to 1672 he served as gentleman-in-waiting to the dowager duchess of Orléans in Luxembourg. For 20 years, from 1673, he was a member of the household of Mme de La Sablière, whose salon was a celebrated meeting place of scholars, philosophers, and writers. In 1683 he was elected to the Académie Française after some opposition by the King to his unconventional and irreligious character.

The Fables. La Fontaine did not invent the basic material of his ***Fables;*** he took it chiefly from the Aesopic tradition and, in the case of the "second collection," from the Oriental. He enriched immeasurably the simple stories that earlier fabulists had in general been content to tell perfunctorily, subordinating them to their narrowly didactic intention. He contrived delightful miniature comedies and dramas, excelling in the rapid characterization of his actors, sometimes by deft sketches of their appearance or indications of their gestures and always by the expressive discourse he invented for them. In settings usually rustic, he evoked the perennial charm of the countryside. Within the compass of about 240 poems, the range and the diversity of subject and of treatment are astonishing. Often he held up the mirror to the social hierarchy of his day. Intermittently he seems inspired to satire; but, sharp though his thrusts are, he had not enough of the true satirist's indignation to press them home. The ***Fables*** occasionally reflect contemporary political issues and intellectual preoccupations. Some of them, fables only in name, are really elegies, idylls, epistles, or poetic meditations. But his chief and most comprehensive theme remains that of the traditional fable: the fundamental, everyday moral experience of mankind throughout the ages, exhibited in a profusion of typical characters, emotions, attitudes, and situations.

By courtesy of the Bibiiotheque Publique et Universitatre, Geneva: photograph, Jean Arlaud

La Fontaine, oil painting by François De Troy (1645–1730); in the Bibliothèque Publique et Universitaire, Geneva.

Diverse themes of the ***Fables***

Countless critics have listed and classified the morals of La Fontaine's ***Fables*** and have correctly concluded that they amount simply to an epitome of more or less proverbial wisdom, generally prudential but tinged in the "second collection" with a more genial epicureanism. Simple countryfolk and heroes of Greek mythology and legend, as well as familiar animals of the fable, all play their parts in this comedy; and the poetic resonance of the ***Fables*** owes much to these actors who, belonging to no century and to every century, speak with timeless voices.

What disconcerts many non-French readers and critics is that in the ***Fables*** profundity is expressed lightly. La Fontaine's animal characters illustrate the point. They are serious representations of human types, so presented as to hint that human nature and animal nature have much in common. But they are also creatures of fantasy, bearing only a distant resemblance to the animals the naturalist observes; and they are amusing because the poet skillfully exploits the incongruities between the animal and the human elements they embody. Moreover—as in his ***Contes,*** but with far more delicate and lyrical modulations—the voice of La Fontaine himself can constantly be heard, always controlled and discreet, even when most charged with emotion. Its tones change swiftly, almost imperceptibly: they are in turn ironical, impertinent, brusque, laconic, eloquent, compassionate, melancholy, or reflective. But the predominant note is that of ***la gaietk,*** which, as he says in the preface to the "first collection," he deliberately sought to introduce into his ***Fables;*** and "Gaiety," he explains, is not that which provokes laughter but is "a certain charm . . . that can be given to any kind of subject, even the most serious." No one reads the ***Fables*** rightly who does not read them with a smile—not only of amusement but also of complicity with the poet in the understanding of the human comedy and in the enjoyment of his art.

Verse and metre

To the grace, ease, and delicate perfection of the best of the ***Fables,*** even close textual commentary cannot hope to do full justice. They represent the quintessence of a century of experiments in prosody and poetic diction in France. The great majority of the ***Fables*** are composed of lines of varying metre and, from the unpredictable interplay of their rhymes and of their changing rhythms, La Fontaine derived the most exquisite and diverse effects of tone and movement. His vocabulary harmonizes widely different elements: the archaic, the precious and the burlesque, the refined, the familiar and the rustic, the language of professions and trades and the language of philosophy and mythology. But for all this richness, economy and understatement are the chief characteristics of his style, and its full appreciation calls for keener sensitivity to the overtones of 17th-century French than most foreign readers can hope to possess.

Miscellaneous writings. La Fontaine's many miscellaneous writings include much occasional verse in a great variety of poetic forms and dramatic or pseudodramatic pieces such as his first published work, ***L'Eunuque*** (1654), and ***Climène*** (1671), as well as poems on subjects as different as ***Adonis*** (1658, revised 1669), ***La Captivitk de saint Malc*** (1673), and ***Le Quinquina*** (1682). All these are, at best, works of uneven quality. In relation to the perfection of the ***Fables,*** they are no more than poetic exercises or experiments. The exception is the leisurely narrative of ***Les Amours de Psiché et de Cupidon*** (1669), notable for the lucid elegance of its prose, its skillful blend of delicate feeling and witty banter, and some sly studies of feminine psychology.

The Contes. Like his miscellaneous works, La Fontaine's ***Contes*** ("Short Tales") considerably exceed the ***Fables*** in bulk. The first of them was published in 1664, the last posthumously. He borrowed them mostly from Italian sources, in particular Boccaccio, but he preserved none of the 14th-century poet's rich sense of reality. The essence of nearly all his ***Contes*** lies in their licentiousness, which is not presented with frank Rabelaisian verve but is transparently and flippantly disguised. Characters and situations are not meant to be taken seriously; they are meant to amuse and are too monotonous to amuse for long. The ***Contes*** are the work far less of a poet than of an ingenious stylist and versifier. The accent of La Fontaine the narrator enlivens the story with playfully capricious comments, explanations, and digressions.

Personality and reputation. Though he never secured the favour of Louis XIV, La Fontaine had many well-wishers close to the throne and among the nobility. He moved among churchmen, doctors, artists, musicians, and actors. But it was literary circles that he especially frequented. Legend has exaggerated the closeness of his ties with Molière, Nicholas Boileau, and Jean Racine; but he certainly numbered them among his friends and acquaintances, as well as La Rochefoucauld, Mme de Sévigné, Mme de La Fayette, and many less well-remembered writers.

Friendships with Molière, Boileau, and Racine

The true nature of the man remains enigmatic. He was intensely and naïvely selfish, unconventional in behaviour, and impatient of all constraint; yet he charmed countless friends—perhaps by a naturalness of manner and a sincerity in social relationships that were rare in his age—and made apparently only one enemy (a fellow academician, Antoine Furetière). He was a parasite without servility, a sycophant without baseness, a shrewd schemer who was also a blunderer, and a sinner whose errors were, as one close to him observed, "full of wisdom." He was accommodating, sometimes to the detriment of proper self-respect; but he was certainly not the lazy, absent-minded simpleton that superficial observers took him for. The quantity and the quality of his work show that this legendary description of him cannot be accurate: for at least 40 years La Fontaine, in spite of his apparent aimlessness, was an ambitious and diligent literary craftsman of subtle intelligence and meticulous conscientiousness.

He was an assiduous and discriminating reader whose works abound in judicious imitations of both the matter and the manner of his favourite authors. He was influenced by so many 16th- and 17-century French writers that it is almost invidious to mention only **François Rabelais**, **Clément Marot**, Franfois de Malherbe, Honor6 **d'Urfe**, and Vincent Voiture. The authors of classical antiquity that he knew best were Homer, Plato, Plutarch —these he almost certainly read in translation—Terence, Virgil, Horace, and Ovid. Boccaccio, **Machiavelli**, Ariosto, and Tasso were his favourites among the Italians. La Fontaine was no romantic; his work derives its substance and its savour less from his experience of life than from this rich and complex literary heritage, affectionately received and patiently exploited. He died in Paris on April 13, 1695.

MAJOR WORKS

Les Amours de Psiché et de Cupidon (1669; *The Loves of Cupid and Psyche.* 1744); *Contes et nouvelles en vers* (the first published 1664, the last posthumously; *Tales and Novels in Verse,* 2 vol., 1884); *Fables choisies, mises en vers,* in verse ("premier recueil," first six books, 1668; "second recueil," the next five, 1678–79; the twelfth book 1692–94); *La Fontaine's Fables,* verse trans. by Sir Edward Marsh, 1952; and *The Fables of La Fontaine,* 1954, abridged Eng. ed., *Selected Fables,* 1955, free trans. by M. Moore.

BIBLIOGRAPHY

Editions and translations: The standard edition of the complete works is still that of HENRI REGNIER, *Oeuvres,* 11 vol. (1883–92); the best edition of the *Fables* is by FERDINAND GOHIN, 2 vol. (1934); and of the *Contes* by PIERRE CLARAC, 2 vol. (1933); the best edition of the miscellaneous works (*i.e.*, all the works except the *Fables* and the *Contes)* is *Oeuvres diverses* by PIERRE CLARAC (1942 and 1958). *La Fontaine's Fables* were translated into English verse by SIR EDWARD MARSH (1931; Everyman's Library, 1952); also by the U.S. poet MARIANNE MOORE (1954; new ed., 1964), whose versions may, however, seem to some readers to be adaptations rather than translations of the original text.

Biography and .criticism: All of the studies listed below, with the exception of John C. Lapp's, are written in French. LOUIS ROCHE, *La Vie de Jean de La Fontaine* (1913); GUSTAVE MICHAUT, *La Forttaine,* 2 vol. (1913–14), mainly biographical; PIERRE CLARAC, *La Fontaine* (1947; rev. ed., 1959), a basic reference book; ODETTE DE MOURGUES, *O Muse, fuyante proie: essai sur la poésie de La Fontaine* (1962); RENEE KOHN, *Le Goût de La Fontaine* (1962); and J.-P. COLLINET, *Le Monde littéraire de La Fontaine* (1970), three general appraisals of La Fontaine's works; FERDINAND GOHIN, *L'Art de La Fontaine dans ses Fables* (1929); and JOHN C. LAPP, *The Esthetics of Negligence: La Fontaine's Contes* (1971), deal more specifically with the *Fables* and the *Contes.*

Lagomorpha

The terrestrial mammals of the order Lagomorpha include the relatively well known rabbits and hares and also the less frequently encountered pikas, or mouse-hares. Rabbits and hares characteristically have long ears, a short tail, and strong hindlimbs that provide a bounding locomotion. In contrast, pikas have shorter, rounded ears, no external tail, and less well developed hindlimbs associated with scampering locomotion.

GENERAL FEATURES

Wild lagomorphs are small to small-medium sized, ranging from the smallest pikas, about 150 millimetres (about 6 inches) in length and 100 grams (3% ounces) in weight, to the largest hares, 700 millimetres (about 28 inches) and 4.5 kilograms (10 pounds). Wild rabbits range between pikas and hares in size, while some varieties of domestic rabbit may reach up to seven kilograms in weight.

Economic importance of lagomorphs

Nearly worldwide in distribution, lagomorphs are absent only from most of Southeast Asian islands, Australia, New Zealand, Madagascar, southern South America, and Antarctica. Man has introduced rabbits and hares into areas outside their original ranges for purposes of sport and to provide a readily available food supply. The natural range of the Old World or European rabbit, *Oryctolagus cuniculus,* appears to have been southwestern Europe and North Africa, but in Roman times this species was often introduced to islands, where it flourished, and spread north and east, partly following man's agricultural activities, reaching Britain in Norman times. More recently *Oryctolagus* was introduced for sport and food into New Zealand and Australia, where the absence of native predators and other factors soon led to its becoming an agricultural pest and a threat to some of the native fauna. In North America several species of the cottontail rabbit, *Sylvilagus,* have shown an increase in abundance and a broadening of range in areas disturbed by man.

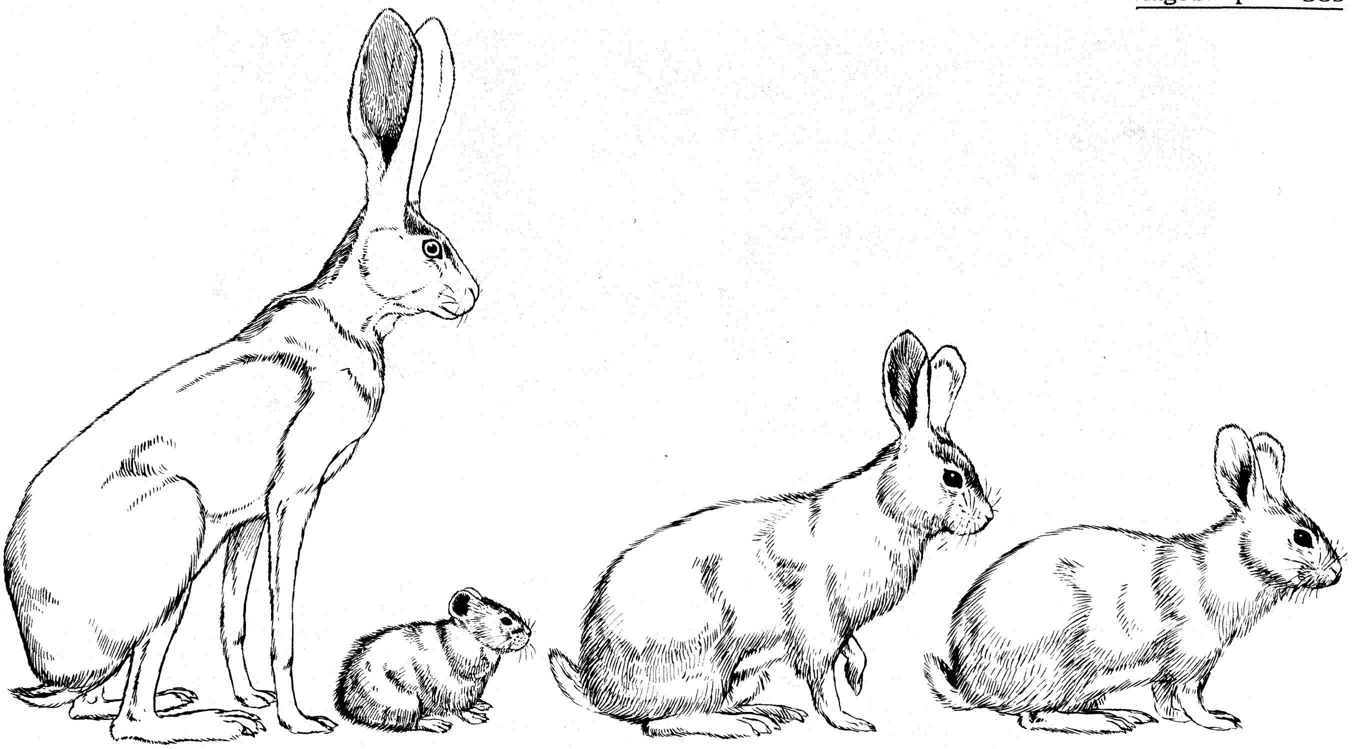
Wild lagomorphs are popular with hunters for sport as well as for food and fur. Domestic rabbits, all descendants of the Old World rabbit, O. *cuniculus,* are raised for meat and skins, the latter of which are used both as pelts and for making felt. In addition, these rabbits are important laboratory animals for biological and medical research. They make good and relatively undemanding pets. Their attractive appearance and quiet manners have made them favourite characters with children's storytellers. On the other hand, wild rabbits and hares locally may become pests, depleting vegetation available to domestic grazers or damaging young trees and orchards, especially in winter. The problem becomes especially acute in areas where man has removed the natural predators of lagomorphs. Attempts have been made to control Old World rabbit populations in western Europe and Australia by introducing the virus disease **myxomatosis**, which exists naturally in populations of certain South American rabbits of the genus *Sylvilagus,* but in some areas the rabbits have developed a certain amount of resistance to the disease. Use of this virus in controlling rabbit numbers has been fought by some humane organizations because of the distressing symptoms of the disease. Pikas, usually found in areas remote from human activity, have little economic importance for man.

NATURAL HISTORY

Rabbits and hares occupy a wide variety of habitats, including grassland, desert, forest, marsh, brushland, and tundra. They are found in mountains up to 4,900 metres above sea level. Rabbits frequent areas where cover is readily accessible and do not venture far from it. Some rabbits dig their own burrows; others use burrows dug by other mammals. Although most rabbits are solitary, some live in pairs. The Old World rabbit, *O. cuniculus*, usually is gregarious, forming breeding colonies called warrens, with extensive burrow systems. The Mexican volcano rabbit, *Romerolagus,* maintains runways through dense vegetation. Although rabbits are typically terrestrial, two North American species of *Sylvilagus,* the marsh rabbit and the swamp rabbit, take readily to water when necessary. Hares tend to live in more open areas than do rabbits. They are solitary and do not use burrows but rest and take shelter in shallow depressions made in soil or vegetation. Pikas occur in northern steppes, semideserts, some forests and scrub thickets of Asia, and in rocky terrain, especially on rock debris on slopes; they range in mountainous areas to 6,000 metres in the southeastern U.S.S.R., northern and central Asia, and western North America. Steppe-dwelling pikas construct and inhabit burrows in flat areas. Rock-dwelling pikas usually seek shelter among rocks and boulders. Some pikas are colonial, whereas others are solitary and may exhibit some degree of territorial behaviour.

Vocalizations

Rabbits and hares are customarily nonvocal, making cries or screams only when frightened or injured. One rabbit, the South African red "hare" *Pronolagus,* is known to utter a warning call. Pikas are highly vocal, having a whistle or bark and a chattering call, reflected in



antelope jackrabbit (*Lepus* alleni) pika (Ochotona princeps) European rabbit (*Oryctolagus cuniculus*) eastern cottontail rabbit (*Sylvilagus floridanus*)

Diversity among lagomorphs.

such common names as whistling hare and piping hare. The voice is used in giving alarm signals and in maintaining territorial boundaries. Mountain hikers or climbers frequently hear their calls and may see these attractive little creatures sunning themselves on rocks and going about their daily activities.

Rabbits and hares are active mainly from dusk to after dawn, whereas most pikas are active during the day. All lagomorphs are plant eaters, grasses and soft vegetation being their dietary staples, but they may also eat twigs and bark, especially in winter. Arctic hares are known to break through crusted snow with their forelimbs to feed on the buried vegetation. Rabbits and hares are not known to store food, but pikas store and cure piles of vegetation for use during the winter. Some pikas cut grasses, carry them to the pile, and allow the vegetation to dry before adding more, whereas other pikas actually turn the contents of a pile to aid in the curing. Pikas that live among rocks first pile the vegetation in the sun and open air before moving it under the shelter of rocks for the winter. Steppe-dwelling pikas pile their hay in the open. Other pikas may store hay not in piles in the open but in crevices in trees, stumps, and among rocks. Some of the Asian pikas unwittingly supply hay for domestic livestock during the winter.

Some lagomorphs regularly reingest fecal pellets (coprophagy). Two kinds of pellets are produced: dry and hard, and soft and moist. The latter kind, which appear to contain vitamins and metabolic products, are eaten, often directly from the anus. The nutritional effect of this practice has been compared to that of rumination among cows.

Rabbits and hares differ from one another in the condition of their young at birth. Young rabbits are born and cared for in a nest; they are naked, blind, and relatively helpless at birth. Hares are born in the open; they are furred, have open eyes, and can run shortly after birth. Confusion of the terms "rabbit" and "hare" in common parlance is shown by such popular names as "jack rabbit" for a hare and "Belgian hare" for a variety of domestic rabbit.

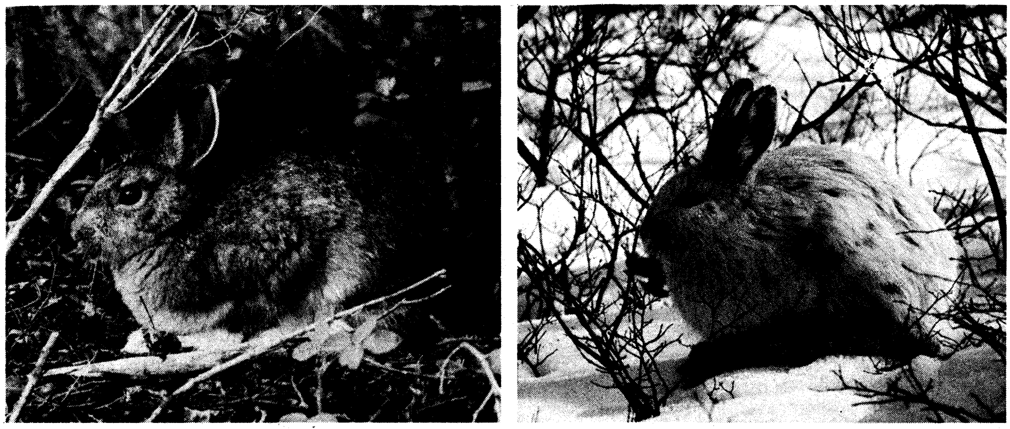
Reproductive potential

The reproductive abilities of rabbits are proverbial. Rabbits and hares usually produce several litters during each breeding season, two or three litters being common among hares and three to six among rabbits. The gestation period ranges from about 28 days in rabbits to 47 in hares, and litter size is usually between two and eight. A female cottontail rabbit from an early spring litter can produce offspring by the end of the summer at age ten weeks. In contrast to most mammals, female rabbits and hares are usually larger than males. Reproductive cycles of pikas are less well known than those of rabbits and hares. Their breeding season occurs in late spring and summer, with a gestation period of approximately 30 days. Litter size normally ranges from two to six, and two or three litters may be born each year. In some pikas, the newborn young are only lightly furred, have eyes and ears closed, but can move about after about eight days.

Lagomorphs, especially hares and rabbits, are important elements in many food chains. They are preyed upon by a variety of carnivores, both mammals and birds, who rely upon them as dietary staples. Wolves, foxes, bobcats, weasels, predatory hawks, and owls all take their toll. More northerly species of rabbits and hares frequently exhibit cyclic fluctuations in numbers, which are reflected in the cycles of the carnivores dependent upon them. The snowshoe hare–snowy owl cycle is one of the prominent ones (see PERIODICITY, BIOLOGICAL). Pikas are preyed upon by various carnivores, among which weasels and birds of prey are probably most important.

Externally most rabbits and hares are not dissimilar from one another, the main variation occurring in length of ear and strength of hindlimb. The coat is usually brownish or reddish brown above and lighter to white below. An exception to this coloration is found in the Arctic hare, in which the winter coat is pure white with only the eartips dark. In more northerly representatives of the Arctic hare, individuals retain this colour pattern all year, whereas more southerly forms of it may become grayish or buff-coloured. In other northern hares, the varying hare (Lepus *timidus*) of Eurasia and the snowshoe hare (L. *americanus*) of North America, the winter coat is white and the summer coat brown. Some varying hares in Ireland and snowshoe hares from the Pacific Coast do not become white in winter. The most unusual coloration in a modern rabbit occurs in the Sumatran rabbit ***Nesolagus***, which has brown stripes on its buff-gray pelage and a bright red rump and tail. Pikas are usually brownish to reddish above and somewhat lighter below.

Lagomorphs illustrate adaptation to particular habitats in an interesting gradational series of features for detection of, and escape from, enemies. At one end of the series



The summer fur (left) of the snowshoe hare (Lepus *americanus*) blends with the browns of woodland and prairie. A mottled effect, created by white hairs gradually replacing the brown, is similar to patches of new snow, while the white winter coat (right) is completely camouflaged in a snowy landscape.
National Audubon Society, (left) Charlie J. Ott, (right) Leonard Lee Rue

are the long-eared, wary, and alert hares, inhabitants of open country, which detect enemies at considerable distance and rely for escape on their strengthened hindlimbs and bounding locomotion that takes them up to 80 kilometres per hour. Rabbits, shorter eared and with weaker hindlimbs, do not react to such distant threats and are scamperers or bounders who do not venture far from cover. They are quick over short distances but lack the endurance of hares. With still shorter ears and relatively weaker hindlimbs, pikas live in areas in which a quick scamper takes them the short distance that usually separates them from secure cover.

FORM AND FUNCTION

Dentition

Lagomorphs are well adapted to a herbivorous diet. Like rodents, they have well-developed incisors that grow continuously from the roots, while they are worn down at the cutting edges. These teeth are extremely effective for severing plant stems and for gnawing on bark. The very mobile lips each consist of two lobes, which meet behind the chisel-like incisors when the mouth is closed. There are no canine teeth, a gap (diastema) showing where they might have been. Cheek teeth farther back in the jaw are also ever growing, wearing away as they grind abrasive vegetation. The upper tooth rows are more widely separated than the lower rows, and chewing is done with a transverse movement. The chewing muscles of the jaws, though strong, are less well developed than in rodents.

Vegetation passes through the long small intestine, which has a spiral valve (like a wood screw), providing a much greater surface area for digestion of food and absorption of nutriment. A large pouch, or cecum, located at the point of attachment of the large intestine contains bacteria that aid digestion and that produce the nutritional soft fecal pellets.

Skeletal adaptations for speed and agility are evident and are discussed as taxonomic characters in the *Annotated Classifications* section, below. The bones of the hindlimb are fused where they move against the anklebone (calcaneum), affording strength and leverage employed in bounding or scampering locomotion. Lagomorphs move about as if they walked on their toes (digitigrade locomotion). There are five digits on the forefoot and four or five on the hindfoot.

EVOLUTION, PALEONTOLOGY, AND CLASSIFICATION

The order Lagomorpha appears to have originated in northern Asia, probably by the end of the Paleocene (about 55,000,000 years ago), and has been relatively stable morphologically throughout the approximately 40,000,000 years since the end of the Eocene, when its fossil record first becomes well documented. Lagomorphs seem to have established early in their history the patterns of dental and skeletal development that were to change, for the most part, only by gradual advances in the basic pattern. Within the order several trends can be observed from more primitive to more advanced lagomorphs: development of ever-growing cheek teeth, simplification of the pattern of the occlusal surfaces of the upper cheek teeth, and shortening of the shaft of the lower incisor where it passes back into the jaw. When dealing with fossil forms, often known only from fragments of jaws and teeth, the most useful characters that differentiate taxa are ordinarily found in the structure of the anterior two upper, and anterior lower, premolars. Presence or absence of the last upper molar may also be important, as may the pattern of the upper cheek teeth. When preserved in fossils, structure of palate and of lower jaw also provide characters that assist in determining affinities.

Evolutionary trends

The family of rabbits and hares (Leporidae) entered North America by the end of the Eocene and underwent most of its Middle Tertiary development there. By the Pliocene (about 7,000,000 years ago) it had become reestablished in Asia and had moved also into Europe. The leporids now extend throughout those ranges and down to the tip of South Africa and as far south as northern Argentina in South America.

The pika family (Ochotonidae) spread from Asia to Europe, where they developed into a number of types in the Middle Tertiary. One of these lines persisted until the late Pleistocene or early modern times (about 1,000,000 years ago) on Corsica and Sardinia. Other ochotonid branches reached Africa and North America in the Middle Tertiary. The extant genus *Ochotona* appeared in Asia in the Pliocene and spread from there, reaching western Europe and eastern North America in the Pleistocene. The current range of the genus represents a considerable reduction from that during the Pleistocene.

Distinguishing taxonomic features. Within the order Lagomorpha living members of its two extant families are clearly separable on morphological grounds, the differences being in skull, jaw, and dental structures, as well as in postcranial morphology.

Morphological trends in the rabbits and hares include increase in arching of the skull, correlated with the development of bounding locomotion and of a relatively upright posture of the head; reduction in the ability to perform lateral movements, caused partly by fusion of bones and partly by modifications related to specialization for anteroposterior movements; strengthening of the hindlimbs and pelvic girdle; and elongation of the limbs distally. Pikas lack these skeletal modifications and are adapted rather for a scampering locomotion.

Annotated classification. In older classifications lagomorphs were usually placed in the order Rodentia. Both rodents and lagomorphs are gliriform mammals; that is, they have enlarged, ever-growing incisor teeth that function as effective tools for gnawing. Under this scheme, the lagomorphs were the suborder Duplicidentata, and the true rodents the suborder Simplicidentata. Early workers, however, recognized many features separating these suborders. For one, there are two pairs of upper incisors in lagomorphs (thus the name Duplicidentata) as opposed to one pair of upper incisors in rodents (Simplicidentata). It is now known that lagomorphs and rodents have long separate histories and that even their earliest representatives did not closely resemble one another. Accordingly, it is more accurate and currently generally accepted to classify the two groups as distinct orders, Lagomorpha and Rodentia. The following classification is the one in general use.

ORDER LAGOMORPHA

Small to medium-sized mammals with two pairs of upper incisor teeth, one pair large, ever-growing, and grooved, in front of a pair of small, peglike incisors, and one pair ever-growing lower incisors. In skull the jawbone, or maxilla, pierced or with open lacework laterally; palatal bone with large, deep-cut openings and relatively short bridge. Of muscles used in chewing, masseter muscles not greatly enlarged, temporalis muscle reduced, pterygoid muscles well developed. Twelve extant and approximately 35 known extinct genera are included in three families, one extinct (†).

†**Family** Eurymylidae

Late Paleocene and (?) Eocene. This extinct family of two genera known only from northern Asia may or may not be closely allied to the other two families of lagomorph. Several anatomical features, such as the presence of only two pairs of upper premolars, show that it was not directly ancestral to later lagomorphs, but it may have been near a group from which later lagomorphs originated.

Family Leporidae (rabbits and hares)

Late Eocene to Recent. Almost worldwide. Distinguished by elongated ears, hindlegs much longer and stronger than forelimbs, often greatly strengthened, and short external tail. Most anterior premolars unlike others in pattern, the following two premolars similar; usually three upper and three lower molars in each jaw, reduced to two upper molars in *Pentalagus;* shafts of upper cheek teeth extend upward into socket in bony capsule (orbit) surrounding eye. Skull with jawbone having open lacework of bone laterally; bony process present over orbit. The palate has one pair of large deep-cut holes; bony palate with longer maxillary and shorter palatine components. Tendency toward arching of skull in profile, related to increasingly upright posture of head. Clavicle rudimentary. Strong, well-developed lumbar region of back, and strong pelvic girdle for articulation with hindlimb and attachment of limb muscles. Five digits on forefoot and on hindfoot.

The subfamily Palaeolaginae is an extinct group that includes primitive Old and New World forms. Descendants of the Palaeolaginae were the extinct Archaeolaginae, including most North American Middle Tertiary forms as well as a group that reached Eurasia in the Pliocene. The most advanced leporids are in the subfamily Leporinae, which includes all living leporids (11 genera), including the American *Sylvilagus*, the Old World *Oryctolagus, Lepus* of the Northern Hemisphere, *Romerolagus* of central Mexico, *Pentalagus* of the Ryukyu Islands south of Japan, *Bunolagus* and *Pronolagus* of South Africa. Other leporids showing some relatively primitive characters are *Nesolagus* of Sumatra and *Brachylagus* of the western United States. The subfamilial divisions are based partly on the pattern of the most anterior upper and lower premolars, which seem to be valid indicators of leporid relationships. Size 25 to 70 centimetres.

Family Ochotonidae (pikas)

Oligocene to Recent. Asia and western North America. Distinguished by short, rounded ears, hindlegs slightly longer than forelimbs, no external tail. Three upper premolars exhibit three different patterns; two upper and three lower molars in each jaw in modern forms, ranges from three upper and lower molars in each jaw to two in each jaw in fossil forms. Shafts of upper teeth curve outward into arch of the zygomatic bone beneath eye socket. Skull with single large opening in each jawbone, other holes present in some fossil forms. Bony process over orbit absent in modern ochotonids but present in some fossils. On palate one or two pairs of deep-cut holes; bony palate with larger palatine component and smaller maxillary component. Skull typically more flattened than in leporids. Well-developed clavicle. Five digits on forefoot and four on hindfoot. One extant genus, *Ochotona,* and 14 or 15 extinct genera. Size 15 to 28.5 centimetres.

Critical appraisal. Perhaps the most important unanswered question in lagomorph classification is that of affinities to other mammals. Part of the difficulty is that the fossil record from the Early Tertiary of Asia is incomplete for large parts of the Paleocene and Eocene. The Eurymylidae, known from some Paleocene and possibly Eocene deposits in Asia, have some characters similar to those later lagomorphs but other characters showing that they were not the direct ancestors of later forms. Other fossil mammals now known from the Paleocene of Asia may represent forms related to eurymylids. The evidence is incomplete, but perhaps eurymylids and these possibly allied forms were in a generally ancestral position to leporids and ochotonids. The wider affinities of this possibly ancestral group to other mammals are not yet clear. No transitional forms are known between the eurymylids and later lagomorphs, which had attained their distinctive characters by the late Eocene. Earlier attempts to arrive at the solution to the problem of lagomorph affinities often emphasized similarities between various primitive herbivorous animals, such as the condylarths, and primitive lagomorphs. These similarities, however, are probably due to the presence, in both groups, of morphological characters common to primitive mammals in general.

Relationships with other mammals

A further point causing some difficulty in studies on fossil leporids and ochotonids is the assignment of various primitive forms to one family or the other. The problem is especially acute when the fossil record is based only on a few teeth or jaw fragments. Morphologically these primitive forms seem to be intermediate between the two families.

Some differences of opinion exist in regard to the subfamilial classification of leporids. Some students of these animals have taken the step of subdividing leporids into several families. Other problems in the classification of modern leporids are matters of specific assignment of various populations. A sound subfamilial arrangement for the ochotonids has not yet been established.

BIBLIOGRAPHY. J.N. LAYNE, "Lagomorphs," in *Recent Mammals of the World,* ed. by S. ANDERSON and J.K. JONES (1967), general review of characters, distribution, and habits of living lagomorphs; E.P. WALKER *et al., Mammals of the World,* 2nd ed. (1968), photographically illustrated articles on living genera of lagomorphs; R.M. LOCKLEY, *The Private Life of the Rabbit: An Account of the Life History and Social Behaviour of the Wild Rabbit* (1964), semipopular work dealing with the rabbit in Britain and the disease myxomatosis; H.E. BROADBOOKS, "Ecology and Distribution of the Pikas of Washington and Alaska," *Am. Midl. Nat.,* 73:299–335 (1965), discussion of habits and habitats of North American pikas; M.R. DAWSON, "Lagomorph History and the Stratigraphic Record," in *Essays in Paleontology and Stratigraphy,* ed. by C. TEICHERT and E.L. YOCHELSON (1967), review of fossil lagomorphs and their usefulness for stratigraphic studies; A.A. GUREEV, "Lagomorpha," in *Mammals,* vol. 3, no. 10, *Fauna S.S.S.R.,* Zoologicheskii Institut Akademii Nauk S.S.S.R., New series no. 87 (1964), comprehensive treatment of the morphology and relationships of fossil and living lagomorphs (in Russian); E.R. HALL, "A Synopsis of the North American Lagomorpha," *Univ. Kans. Publs. Mus. Nat. Hist.,* 5:121–202 (1951), on the taxonomy and distribution of North American forms; A.S. LOUKASHKIN, "On the Pikas of North Manchuria," *J. Mammal.* 21:402–405 (1940), on the behaviour and habitats of Asian pikas.

(M.R.D.)

Lagoons

Lagoons are areas of relatively shallow water situated in a coastal environment and having access to the sea but separated from the open marine conditions by a barrier. The barrier may be either a sandy or shingly wave-built feature, or it may be a coral reef. Thus, there are two main types of lagoons, namely, elongated or irregular stretches of water that lie between coastal barrier islands and the shoreline, and circular or irregular stretches of water surrounded by coral atoll reefs or protected by barrier coral reefs from direct wave action. Lagoons of the first type are characterized by quiet water conditions, fine grained sedimentation, and, frequently, brackish salt marshes. Water movements are related to discharge of river flow through the lagoon and to the regular influx and egress of tidal waters through the inlets that normally separate the barrier islands. Lagoons of the second type are best exemplified by the roughly circular quiet waters that are surrounded by coral atoll reefs. Lagoon depths are maintained at a moderate level by sedimentation, and

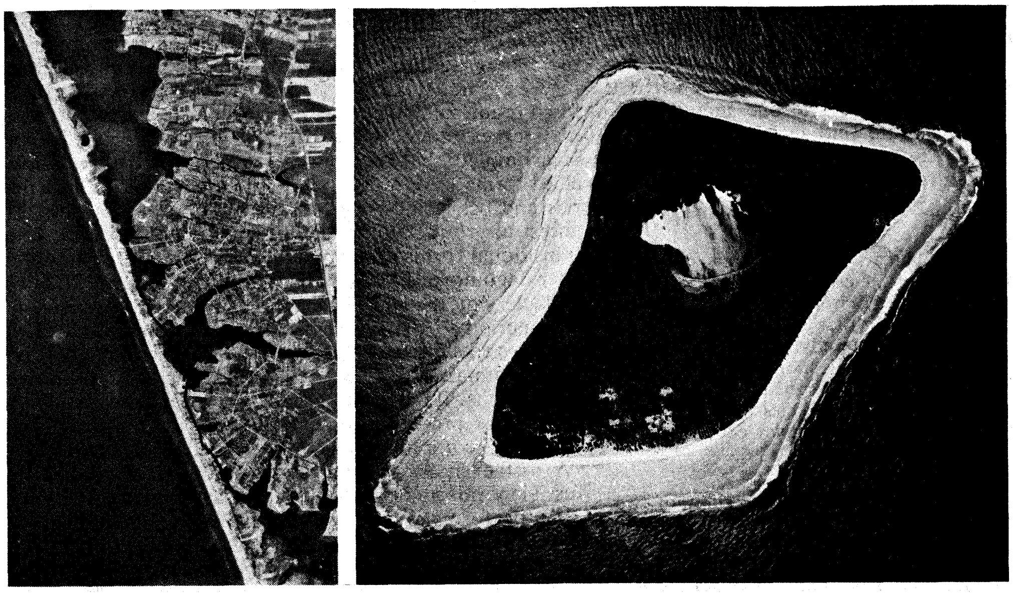


Figure 1: Two *main* types of lagoon.
(Left) Barrier island lagoon at Westhampton, Long Island, New York. (Right) Coral lagoon surrounding Lib Island, Kwajalein area, Marshall Islands.
By courtesy of (left) U.S. Geological Survey. (right) U.S. Navy.

this compensates for the subsidence that commonly attends reef formation (see CORAL ISLANDS, CORAL REEFS, AND ATOLLS). Because the reef is an organic structure, the lagoonal sediments contain much calcareous material. The sheltered waters support a distinctive flora and fauna. Lagoons also can occur in lakes, where transport of sediment from deltas or other sources creates barrier bars.

Coastal lagoons

Coastal lagoons are widely distributed throughout the world and have been estimated to constitute about 13 percent of the total world coastline. Lagoons are more common on coasts with moderate to low tidal ranges; for example, they occur widely on low coasts of the southern Baltic, southeast North Sea, Black Sea, Caspian Sea, and Mediterranean Sea, and of the southeastern United States, and the Gulf of Mexico. Lagoon coasts also occur along the southeastern coast of South America; the east coast of Madagascar; northeastern U.S.S.R.; Japan; and isolated parts of Africa, India, Australia, and New Zealand.

Lagoons are generally characteristic of coasts of low or moderate energy, occurring especially on the east coasts of continents where the swells are less violent and in high latitudes where offshore ice provides some protection. They also are associated with low coasts and rarely occur where high cliffs form the coast. They can form only where there is abundant sediment for construction of the protective barrier islands. Too much sediment from the mainland, however, can lead to delta formation rather than lagoons, although lagoons frequently occur along the outer delta margin and between delta distributaries (see also RIVER DELTAS).

Coral lagoons

Coral lagoons are restricted to tropical open seas that provide the conditions necessary for coral growth: the water temperature must equal or exceed 20" C (68" F), and it must be clear, moving, and not deeper than seven metres. Coral lagoons occur widely in the western Pacific, in parts of the Indian Ocean, and in isolated places in the Caribbean, mainly within 25° latitude of the Equator. Coral lagoons are of great importance to many island communities in the Pacific, particularly where they provide the only quiet water for use as harbours, although the passage through the reef into the lagoon is often perilous.

The clarity of the water of the coral lagoon contrasts with the considerable amount of fine sediment in barrier island lagoon water. The extensive growth of salt marsh in such lagoons often detracts from their use as harbours, but where the water is deep enough they provide sheltered anchorage and good conditions for small-boat sailing in quiet water (see also CORAL ISLANDS, CORAL REEFS, AND ATOLLS; SWAMPS, MARSHES, AND BOGS).

THE NATURE OF THE LAGOON ENVIRONMENT

Dimensions. Coral lagoon dimensions range from small atolls to those so wide that the coral reefs on the far side cannot be seen across the lagoon. Atoll widths range from about 2.5 kilometres to nearly 100 km, but the mean value is about 20 km. Depths rarely exceed 60 metres and many are shallower, usually less than 20 m deep. The lagoon of Mayotte Island in the Comoro Archipelago in the Indian Ocean attains a maximum depth of about 92 m but generally is shallower, and it is about 16 km in width at its widest point; this lagoon lies behind a barrier reef.

Barrier island lagoons are typically elongated, although irregular ones may occur where river estuaries are flooded behind the barriers. This pattern occurs on the east coast of the U.S., where lagoons extend intermittently for nearly 1,500 km along the coast. The Gippsland Lakes in Victoria exemplify a complex lagoon system formed behind the 90-mile beach (Figure 2). Elongated lagoons up to 64 km (40 miles) long lie behind the beach barrier, and larger lagoons, such as Lake Wellington, which is 50 km wide, lie behind the northern end of the barrier. Postglacial subsidence has flooded the lowland in this area. The lagoons are shallow; Lake Wellington is less than 3.5 m deep and much of King Lake is less than 6 m deep. Scour holes as deep as 16.5 m do occur, however. The elongated lagoon behind the barrier is only 1 to 1% m deep, typical of barrier island lagoons.

Current velocities

Water circulation. The degree of water circulation depends on the width of the tidal inlets, the tidal range, and on the amount of runoff from adjacent land areas. Maximum velocities are attained at the points where the water passes through the barriers. In the entrance to the Gippsland lagoons, tidal currents reach three knots, but river floods that escape to the sea can raise the velocity to seven knots. Water may be blown into the lagoon by strong winds, which raise its level and cause an outflowing current when the wind drops. Seiches can be created in this way. Small waves can be generated within lagoons when the wind blows along their maximum dimension. These may reach 1% m (4 ft) in height and 1½ to 9 m in length in the Gippsland lagoons. In coral atoll lagoons there is little or no runoff; and seawater moves in and out through the passes, where tidal currents reach their maxima. Velocities of 10 to 12 knots have been recorded in the Hao Channel of the Tuamotu Island.

Water temperature and salinity. In the Mayotte Lagoon the water has the same properties as the upper layers of the open sea; the salinity is close to 35 parts per 1,000 (‰), and the temperature varies between 27° and 24° C. This is typical of coral lagoons, but the temperature

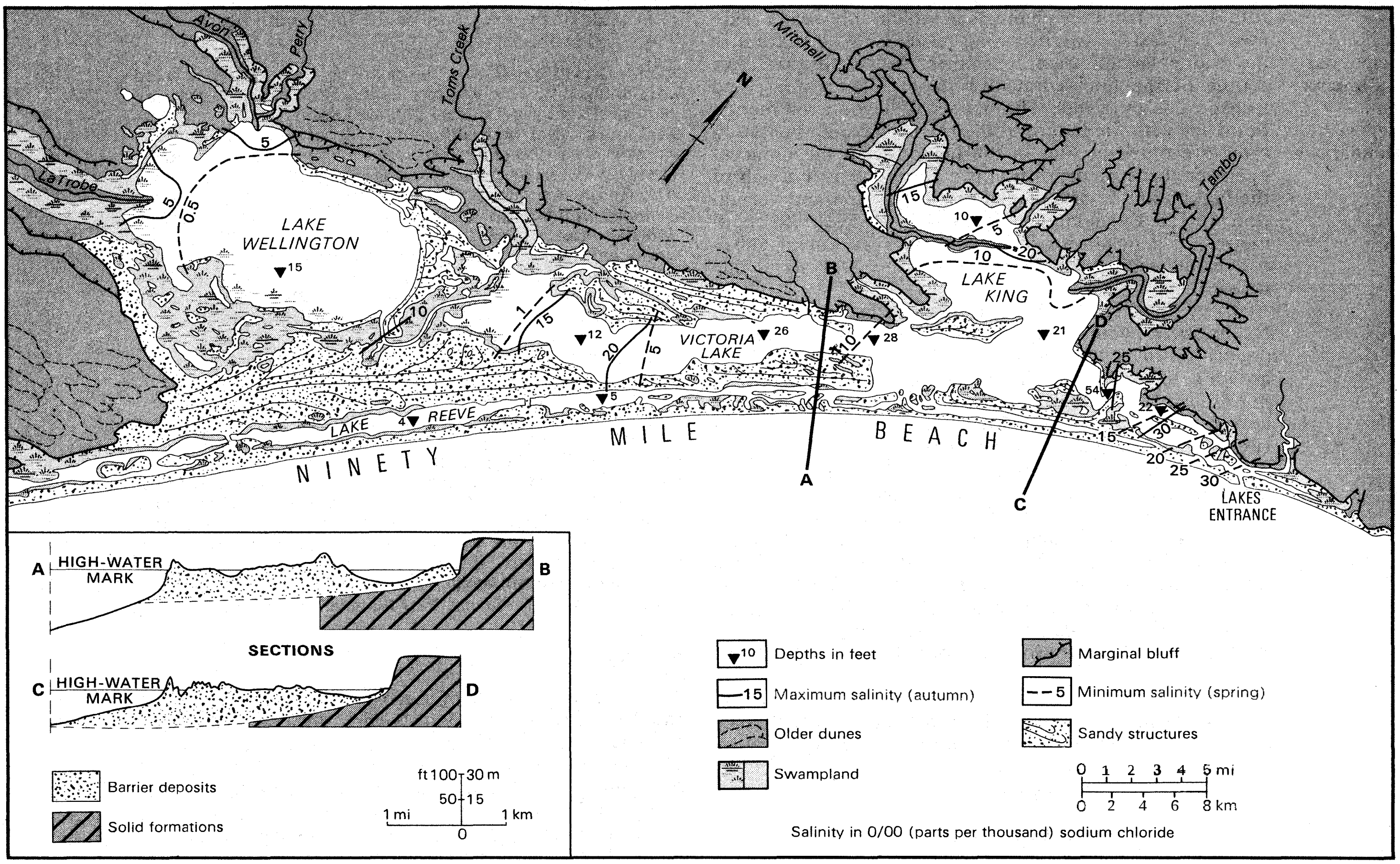


Figure 2: Sandy coastal barriers enclose the Gippsland Lakes, a lagoon region in Victoria, Australia.

and salinity of barrier island lagoons are more variable because of their wider climatic range. Because they are shallow, lagoon waters approximate the air temperature: colder than the open sea in winter, warmer in summer.

Salinities decrease as a function of the amount of runoff entering the lagoon in relation to the tidal influx. Lagoons may be considered brackish, marine, or hypersaline. Brackish lagoons receive much runoff, and salinity increases toward the tidal inlets. The Gippsland lagoons exemplify this type. The salinity at the inner end varies from 0.5‰ to 5‰ according to season, and central values vary between 5‰ and 20‰. Hypersaline lagoons occur where evaporation exceeds inflow. Lagune Madre in Texas and Sivash Lagoon in the Black Sea have salinities of 65 and 132‰, respectively. Salt deposits may form in these conditions. The denser saline water tends to lie beneath the fresher water where mixing is not severe.

Equilibrium bottom profiles. Lagoons behind coastal barriers normally are zones of fine sedimentation. Their bottom profiles, therefore, build up gradually with time. Typical depths of the Texas lagoons vary between 1¼ and 3½ m, and their floors are flat. Early theories that attempted to relate the form of the offshore and lagoon profile are no longer held; and because the lagoon profile changes with continued deposition, it cannot be used to establish the process of lagoon formation. The profile is usually gently undulating, but it may contain deeper channels, especially near the tidal inlets. Profiles across coral lagoons are either smooth and flat from calcareous sedimentation, or they contain knolls of growing or dead coral. There are 2,300 such knobs in the Eniwetok Lagoon in the Marshall Islands.

Flora and fauna. The flora depends on the climate of the area and the salinity of the lagoon waters. In tropical areas mangroves colonize the quiet lagoons, whereas in temperate latitudes salt marsh vegetation is common. The species depend on their salt tolerance and the bottom conditions. In less saline conditions *Phragmites,* or reed swamp, may be dominant but, where salinity approaches that of the open sea, halophytes are found. The plants accelerate the deposition of sediment on the lagoon floor, resulting in a decrease in area and water volume.

The fauna of the barrier-type lagoon are those that can live in soft mud or sandy sediment. Oysters and mollusks, such as *Scrobicularia,* can flourish in lagoon conditions. Organisms must be able to withstand a greater range of conditions than those of the open sea.

Corals and algae

In coral lagoons fauna are fundamental to the formation of the reefs. Coral polyps live in symbiosis with the unicellular algae *Zooxanthellae,* hence they require light and are restricted to the upper layers of water. Calcareous algae, *Lithothamnion,* normally grow in coral reefs. Corals and algae grow in the lagoons, often forming knolls. Sediment may prevent coral growth, although various species of *Acropora* may grow in this zone.

FACTORS INVOLVED IN LAGOON FORMATION

The essential feature that causes the lagoon to exist is the barrier that separates it from the sea. In the coral lagoon the formation of the reef depends upon the existence of suitable conditions for reef growth, which have already been mentioned briefly.

Barrier bars and sediment sources. The barrier island lagoons, on the other hand, depend not on organic processes but on waves, which act in a suitable environment on an adequate supply of bottom material, most commonly sand. Barrier islands are formed in those areas where the coastal slope is flatter than the equilibrium slope required by the long constructive swells, the waves that build up the foreshore in front of their breakpoint. They are, therefore, found on low coasts. They may occur in areas of subsidence, stability, or emergence, wherever sufficient sand exists.

The barrier islands that form the lagoons are made of sand, but the sediments of the lagoon are usually finer, as conditions are quieter. The lagoonal muds differ from shelf muds. Glauconite is rare in lagoon muds, but oyster reefs may occur as along the Texas coast. The muds are

Lagoonal sediments and depositional rates

found away from the channels in which only coarse sediment can rest, owing to strong currents. Flocculation in the saline lagoon water expedites clay deposition. The source of the fine sediment is from inland areas, and transport is by rivers. The details of lagoon sedimentation vary with the nature of the river load. Sedimentation rates are much greater in the lagoon than the adjacent open sea, because deposition is continuous over much of the lagoon and is often aided by flocculation and vegetation. In the Texas lagoons from 1875 to 1936, 20 centimetres of deposition occurred, despite 30 cm of subsidence; the sedimentation rate, therefore, was about one metre in 100 years.

Waves, tides, and surf. The barrier islands are formed by the waves, which build up their equilibrium profile by deposition on a gradient that is too flat. The level of the growing accumulation may be raised by the wind, forming dunes. Where the land behind the growing barrier is low, it will become flooded to form a lagoon if sea level rises slowly. Such a rise of sea level has occurred during the last 20,000 years. As long as the barrier island can maintain its level above the sea, as sea level rises, the lagoon will exist until it is filled with sediment. Not all lagoons and barrier island complexes have formed during rising sea-level conditions, but where there is evidence that no open sea foreshore sediments lie on the land side of the barrier this hypothesis seems likely. In some barriers, however, outbuilding of material by glacial outwash streams or rivers may provide a suitably low gradient and enough sediment to form a barrier, as along the south coast of Iceland. In other areas material carried alongshore to form a spit may develop into a bay-mouth barrier, enclosing a lagoon. Such features can be of sand or shingle. The Fleet behind Chesil Bank in southern England is an example of the latter type.

Waves within the lagoon may develop cuspate spits along the land side of the barrier and the inner edge of the lagoon. These features may eventually break the lagoon into almost circular or oval water bodies. Examples occur on the Chukchi Sea lagoons in the U.S.S.R. and elsewhere where vegetation does not form marshland.

Storms and catastrophic events. Storms or tsunami exert an effect on lagoons when they breach or overtop the barriers around the lagoon by raising the water level temporarily. Major changes in configuration can occur in a short time. Hurricanes, for instance, can cause serious effects on the coast of Texas, along which lagoons are common. Padre Island was lowered to below mean low tide at its southern end by one storm in 1919, and several washover channels were cut. The mainland shore also suffered erosion. Deposition may also occur; saline marls have been laid down on freshwater marsh, and small beach ridges may be built inside the lagoon where the high water level drives sand inland over mud. Coral reefs are more resistant to storms than are mobile sandy barriers.

Darwin's hypothesis

The effect of time. Lagoons of both types change with time. In both a relative rise of sea level with time is important in the development of the lagoon. In coral atolls there is evidence from deep boring that Charles Darwin's original subsidence hypothesis of atoll formation, via barrier reefs from fringing reefs around a subsiding volcanic peak, is substantially correct in many cases. As long as the coral can maintain its growth at a suitable level as its foundation subsides, the atoll will continue to enclose a lagoon, which is floored by coral or calcareous sediment, derived from the reef, and that maintains its depth by deposition or growth.

The postglacial rise of sea level also has influenced barrier island formation in many instances. When sea level rises too fast a barrier may be drowned, and its lagoon will cease to exist.

Experimental studies include work on the closing of tidal inlets into lagoons across barrier beaches. Such experimental work has shown the importance of storm tides in maintaining the openings through barriers, where longshore movement of material is great and runoff from land areas is not abundant. These experiments show that the deepest water occurs where the channel cuts through the barrier. Lobate shoals form areas of shallow water on either side of the deep part of the inlet, both on the open shore and in the lagoon.

BIBLIOGRAPHY. There are relatively few works devoted only to lagoons. The literature is widely spread in biological, geological, and geographical publications. The following classical works deal with different types of lagoons: CHARLES DARWIN, *The Structure and Distribution of Coral Reefs* (1842); W.M. DAVIS, "The Coral Reef Problem," *Spec. Publs. Am. Geogr. Soc.*, no. 9 (1928); E. DE BEAUMONT, *Leçons de geologie pratique,* pp. 223–252 (1845), and D.W. JOHNSON, *Shore Processes and Shoreline Development* (1919), two works on coastal barrier island lagoons. A recent study of coral lagoons and their ecology is D.R. STODDART, "Ecology and Morphology of Recent Coral Reefs," *Biol. Rev.*, 44:433–498 (1969), with an extensive bibliography. Useful references on barrier island lagoons include: K.O. EMERY, R.E. STEVENSON, and J.W. HEDGPETH, "Estuaries and Lagoons," in *Treatise on Marine Ecology and Paleoecology,* ch. 23, pp. 673–749 (1957); A. GUILCHER, *Coastal and Submarine Morphology* (1958; orig. pub. in French, 1954); F.P. SHEPARD, *Submarine Geology,* 2nd ed. (1963); and V.P. ZENKOVICH, *Processes of Coastal Development* (1967; orig. pub. in Russian, 1962). H.G. GIERLOFF-EMDEN, "Nehrungen und Lagunen," *Petermanns Geogr. Mitt.,* 105:81–92, 161–176 (1961), deals in detail with early theories of lagoon formation and includes a map of lagoon and barrier shorelines of the world. A more detailed study of one lagoon is E.C.F. BIRD, "Coastal Lagoons of Southeastern Australia," in J.N. JENNINGS and J.A. MABBUTT (eds.), *Landform Studies from Australia and New Guinea,* pp. 365–385 (1967).

(C.A.M.K.)

Lagos

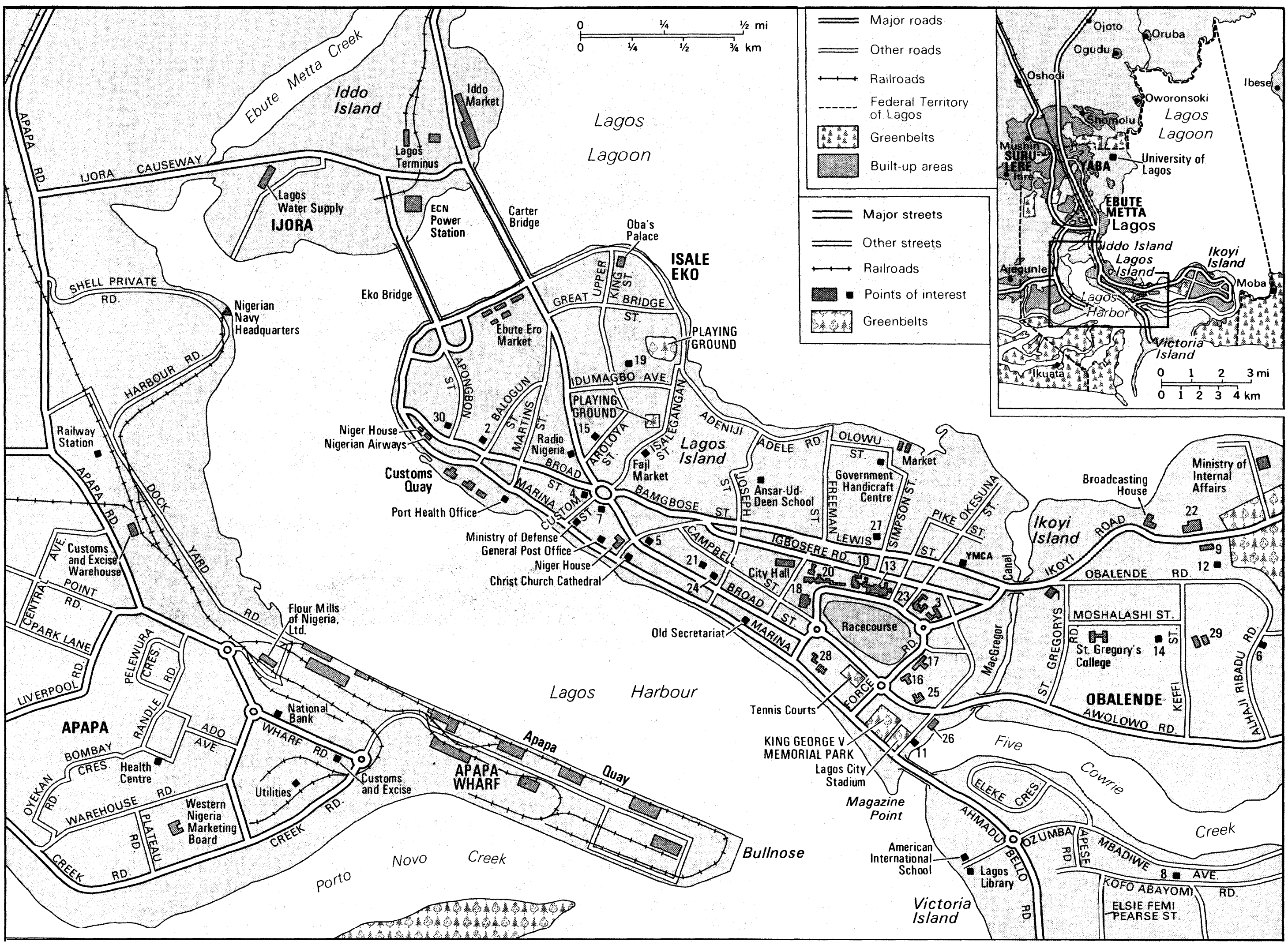
The Federal Territory of Lagos, the national capital of the West African country of Nigeria, located at the western end of its Atlantic coastline, covers 27 square miles (70 square kilometres); and, at the start of the 1970s had a population of 945,000 in the city proper and about 1,600,000 in the metropolitan area. The territory is also the administrative centre for Lagos State and is the country's industrial and commercial centre, as well as its most important port. Originally known to the conquering Bini people as Eko (War Camp), it was successively designated by the Portuguese as Lago de Curamo and Onin, before finally being named Lagos, after a harbour in southern Portugal.

The city itself is the second largest—after Kinshasa—in tropical Africa. It is afflicted with industrial pollution, slums, and traffic congestion. It is also a thriving centre of economic and social activity that has an aura of progress. Its cultural life is cosmopolitan; tourists are attracted to its sandy beaches, and foreign capital investment is encouraged. The basic character of the city, however, remains African.

The city's origins

History. Some 300 years ago, the first Yoruba fishermen and hunters settled on Lagos Island, where the lagoon, swamps, and forests helped to protect the settlement from the neighbouring kingdoms of Dahomey to the west, Benin to the east, and the Yoruba states to the north. The area was, however, under Benin's domination from the late 16th to the mid-19th century, during which time it was governed as a part of the kingdom's westernmost province. It was during this period that power over land in Lagos was exercised by the Idejo, or White Cap Chiefs, who wore white hats as a symbol of their office. When lesser functionaries were later given the privilege of wearing the caps, two classes of chiefs were created. Their conflicting claims to power are still part of the city's political life.

The Portuguese first landed on Lagos in 1472; trade developed slowly, however, until the Portuguese were granted a slaving monopoly a century later. The city's position on the main lagoon of the West African coast and its proximity to the Yoruba states that were then ravaged by war contributed to its emergence as a major slaving centre, and this in turn attracted British interest in Lagos when British policy was to favour suppression of the slave trade. In about 1841, Kosoko became ***oba*** (king) of Lagos. Ten years later his support of the slave trade led to a clash with the British, who deposed him in favour of his rival, Akitoye. The slave trade continued to grow under Docemo, successor to Akitoye, and in 1861 he was forced to cede Lagos to the United Kingdom.



1 Administrator General s Office
2 Bank of Lagos
3 Cabinet Office
4 Central Bank of Nigeria
5 Central Library
6 Federal Electoral Commission
7 Federal Ministry ot Finance
8 Federal Ministry of Industries
9 Federal Office of Statistics
10 Federal Supreme Court
11 Governor's Office
12 Health Office
13 High Court
14 Holy Child College
15 Holv Qur'an Mosque
16 House of Representatives
17 House of Senate
18 Independence Building
19 Jankara Market
20 King s College
21 Ministry of Health
22 Ministry of Labour
23 Ministry of Works
24 Natonal Library of Nigeria
25 National Museum
26 Prime Minister's Residence
27 Send Grouse Marker
28 State House
29 Veterinary Centre
30 West End Coliseum

Central Lagos and (inset) its metropolitan area.

Under British administration. Lagos continued to thrive as a trade and administrative centre. Much of the population increase of the period was due to the repatriation of emancipated Yoruba slaves from Brazil, the West Indies, and Sierra Leone. Originally governed as a crown colony, Lagos was part of the United Kingdom's West African Settlements from 1866 to 1874, when it became part of the Gold Coast Colony (then the coastal region of modern Ghana). In 1886 it again achieved separate status under a British governor, and in 1906 it was amalgamated with the Protectorate of Southern Nigeria. When Southern and Northern Nigeria were amalgamated as the Protectorate of Nigeria in 1914, Lagos was made the capital of the Colony and Protectorate of Nigeria. In 1954 most of the hinterland was incorporated into the region of Western Nigeria, while the city itself was designated as federal territory. In 1960 it became the capital of independent Nigeria. In 1967, control of its hinterland was returned to the city with the creation of Lagos State which has an area of 1,381 square miles.

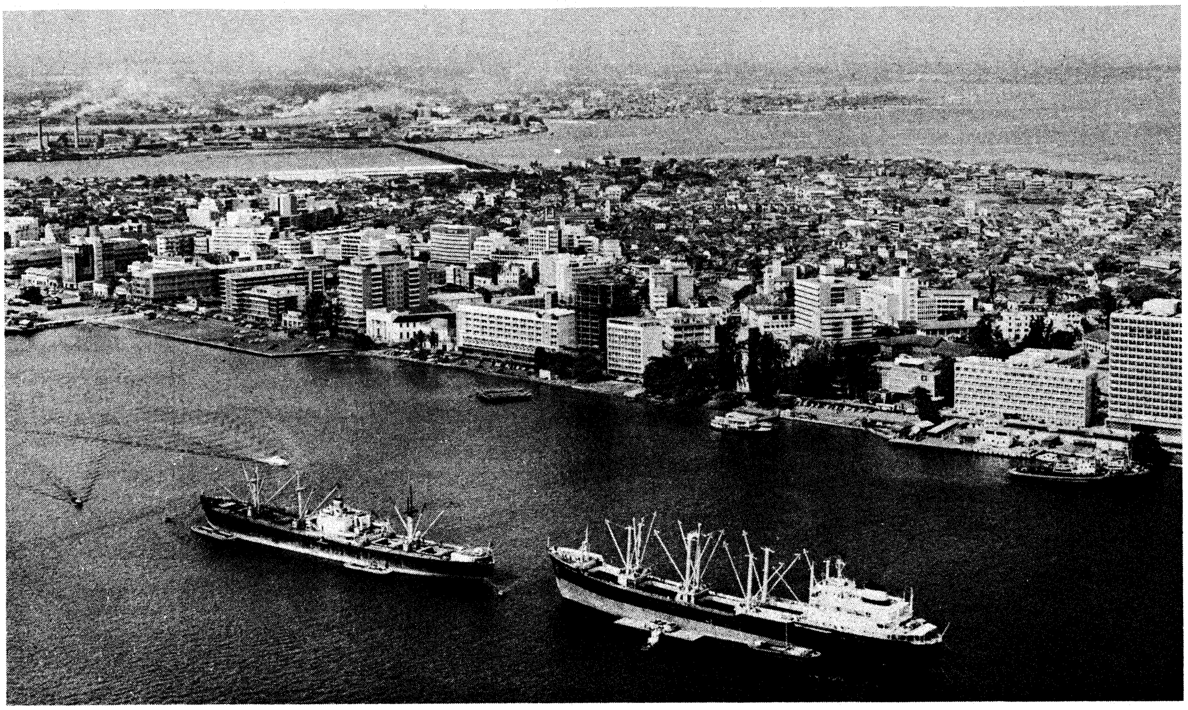
The contemporary city. ***The environment.*** The Lagos landscape is dominated by its system of islands. sandbars. and lagoons. During the latter part of the Pleistocene Epoch (about 10,000 years ago), the level of the Atlantic Ocean rose and submerged parts of the Nigerian coast, drowning river mouths and creating islands and bays. The tidal drift of sands along the coast eventually built up the sandbar that now encloses Lagos Lagoon itself.

Topography

The federal territory includes a strip of the mainland on the western shore of the lagoon, several islands and marshes, and the western end of the sandbar. The main islands of Iddo, Lagos, Ikoyi, and Victoria are connected to each other and to the mainland by a system of bridges. They are subject to flooding after heavy rains, but the installation of a series of drainage canals has alleviated much of the problem. The smaller islands at the western entrance to Lagos Lagoon are, however, swampy and only sparsely inhabited. All of the territory is low-lying, and the highest point on Lagos Island is only 22 feet above sea level.

Lagos has a tropical climate with an average temperature of 80° F (27" C) and an annual rainfall average of 72 inches. There are two seasons; the rainy season lasts from May to September, with the largest amount of precipitation falling in July, and the dry season lasts from October to April. The dusty, dry, and cold harmattan wind blows from the northeast from December to February. The original vegetation and animal life of the area have virtually disappeared. There are, however, scattered palm trees and green tropical vegetation on the mainland, and the Lagos Zoo is popular.

Because the surrounding swamps and smaller lagoons were long used as refuse dumps, water pollution is a serious problem. Reclamation projects have added about 1,000 acres of land to the city area, reduced the offensive stench of the swamps, and decreased mosquito breeding



The harbour, business section, and old town, Lagos Island. In the background, Carter Bridge, which connects with Iddo Island on the northwest.

Marc and Evelyne Bernheirn—Rapho Guillumette

grounds. Though atmospheric pollution was not a problem in the early 1970s, it was feared that the rapid industrialization would make clean air a matter for concern.

Settlement pattern. Lagos developed from traditional Yoruba settlements on the north of the island and later from communities established by Europeans, as well as by repatriated people from the Americas and Sierra Leone, on the southern coast. With the industrial development and population increase of the 20th century, the land between these areas was built up, and residential suburbs were founded to the north on the mainland.

The original settlement on the northwestern tip of Lagos Island now forms the core area of Lagos City. A slum area, it is characterized by narrow streets, poor housing, overcrowding, and a lack of sanitation facilities. Its traditional-minded inhabitants strongly resisted the slum clearance program of the late 1960s. Other slum areas include the mainland suburbs of Shomolu, Mushin, and Ajegunle. Newer sections of Lagos, such as Victoria Island, Ikoyi, Ebute Metta, Yaba, Apapa, and Surulere, are well planned, with wide streets, open spaces, and adequate sanitation facilities.

Pre-eminent among the neighbourhoods of Lagos is Isale Eko, site of the palace of the *oba* of Lagos, the Iga Idungaran, which occupies about ten acres. Close to it are the Obun Edo and Ebute Ero markets.

The business district

The main business district, occupying Lagos Island's southwestern shore, contains an increasing number of multistoried buildings. Its restricted land area is intensively utilized, for this is the heart of the city, where its primary activities—commerce, finance, administration, and education—are performed. Slum clearance in the district made provisions for some expansion, but some businesses were relocating to the outskirts of the city. Industry is located on the mainland in the city areas of Apapa, Ijora, and Yaba, and in the suburbs of Ikeja and Mushin.

Transportation. Transportation problems have beset Lagos throughout the 20th century. A railway, with its headquarters in Ebute Metta and its terminal at Iddo, connects the mainland's residential and industrial centres during the morning and afternoon rush hours. Bus services operating throughout the city carry about 100,000 persons daily. The city is also well served by taxis, and there are so many private automobiles, motorcycles, and bicycles that traffic jams are constant. Bridges to the islands from the mainland are unable to handle the growing volume of traffic, and slow-moving rush hour traffic is further frustrated by a lack of sufficient parking facilities on Lagos Island. It is anticipated that by the mid-1970s motorists may be forced to leave their cars on the mainland and take buses to the islands.

The city is the terminus for the nation's road network. An airport at Ikeja receives most of Nigeria's visitors from overseas, is served by 18 international airlines, and is used by the Nigerian Air Force. Lagos is the headquarters of the Nigerian Airways Corporation and the focus for flights from the country's 28 other airports.

The port of Lagos

The port of Lagos consists of two quays—Customs Quay, on Lagos Island, and the more important Apapa Quay, on the mainland. Both are served by a channel that is protected by breakwaters to the east and west. Apapa, the main outlet for Nigeria's exports, is served by a railway and also has a small floating dock. The creeks and lagoons are plied by numerous canoes, engaged in fishing and carrying goods between Lagos and other coastal cities in Nigeria or in Dahomey. The port is administered and maintained by the Nigerian Ports Authority.

The people. With a metropolitan population of about 1,600,000 and a city proper of 945,000 inhabitants, Lagos is tropical Africa's second-largest city. Its size is attributable to its position as a leading industrial centre; the relative youth of the population—70 percent are between 20 and 29 years of age—is an indication of its attraction to labour. The crowded conditions of the city are at their worst in central Lagos, where densities were reported to exceed 14,000 persons per square mile in 1961. Over 80 percent of the population are Yorubas, but the city attracts goodly numbers of immigrants from other parts of the country and from abroad. Among other Nigerian ethnic groups the Hausa and Ibo are the largest; there are also Edo and Ijaw communities. Among those from other African countries, the largest number come from Dahomey. There are about 10,000 Europeans, Americans, and Asians in the city. The largest white groups originate from the United Kingdom, the United States, France, and Germany. Over half of the population are Christian, 42 percent are Muslim, and about 3 percent practice traditional religions.

Housing. The city's most pressing problem is housing. The severe overcrowding and lack of sanitation in central Lagos prompted the slum-clearance scheme that led to the redevelopment of 70 acres by the early 1970s. In the poor districts, houses lack piped water, toilets, and electricity. The buildings are composed of cement or mud and are roofed with corrugated iron. In the wealthier areas, relatively new houses are equipped with modern conveniences; they range in architecture from Brazilian-style storied buildings to English country and colonial styles. Population density is low because of the predominance of single-family dwellings. Population pressure in the rertricted metropolitan area was potentially relieved in 1967 by the creation of Lagos State; the extension of the city's political control on the mainland will permit the establishment of new industrial and residential suburbs to the north and west.

Manufacturing

The economy. ***Industry.*** Industrial development has been promoted by the tax concessions accorded to investors, by the large reserve of labour, and by the availability of transport facilities. Most establishments are small, the largest employing fewer than 1,000 persons. Apapa contains the largest industrial area, while the suburb of Ikeja is also growing in industrial importance. Manufacturing and processing operations include automobile and radio assembly, food and beverage processing, metalworks, and the production of paints and soap. Textile, cosmetic, and pharmaceutical manufacture is also important. The fishing industry possesses trawlers, laboratory facilities, and a research library.

Commerce. There are 16 licensed banks, owned by Nigerians, Europeans, Americans, or Asians. The Central Bank of Nigeria is the sole bank of issue for Nigerian currency. The city is the national centre for insurance, as well as the seat of the Lagos Stock Exchange.

Public utilities. The federal government is responsible for the daily supply of about 24,000,000 gallons of water to the city; the piped water supply is of good quality, is filtered, and is chlorinated. Electricity is provided by the Electricity Corporation of Nigeria, with the main generating station, powered by coal and diesel oil, located on Iddo Island.

Administration and social conditions. ***Governmental institutions.*** As a federal territory, Lagos is independent of the other regions and is the direct responsibility of the national government. Since 1967 it has been governed by the Lagos State government and by a City Council Caretaker Committee appointed by the military government of Lagos State. The city is also the seat of the Nigerian House of Representatives and of the Federal Supreme Court and the location of a number of foreign embassies and consulates.

Health and safety. Medical and health facilities—ranging from large specialized hospitals to small maternity clinics—include a teaching hospital of the University of Lagos, an orthopedic hospital, a tuberculosis sanatorium, and a mental institution. The most common diseases in the city are malaria, bronchitis, pneumonia, and dysentery. The narrow and crowded streets cause a large number of accidents, which lead to more hospital admissions than do major illnesses. The Municipal Board of Health has reduced the incidence of malaria through swamp reclamation and conducts periodic inspections of residences. Detachments of the Nigerian Police are stationed in Lagos, and there is a fire brigade. The city contains the headquarters of the Nigerian Red Cross.

Education. The Lagos State government is responsible for education in the federal territory. Most of the schools are run by religious organizations with governmental assistance. There are about 450 free primary schools, almost 80 secondary schools, five teacher-training colleges, and four technical and vocational colleges. Of the total student enrollment of more than 227,000, the overwhelming preponderance attend primary schools, where there are more male than female students. The University of Lagos, founded in 1962, has an enrollment of over 2,000.

Cultural life. Religious institutions include churches and mosques. Cultural associations, clubs, debating societies, and ethnic unions form an integral part of Lagos life, and the performing arts are popular. The National Museum houses excellent historical examples of Nigerian arts and crafts. Organizations include the Nigerian Institute of International Affairs and the Nigerian Institute of Management, as well as British, German, Italian, and American institutes. There are several city libraries, the National Library of Nigeria, and the library of the University of Lagos. There are many bookshops, and numerous cinemas offer varied programs that include American Westerns and Indian films on mystical themes.

Museums and cultural institutions

The media. Lagos is the centre of the country's international telecommunications system and has good postal services. Radio Nigeria and the Nigerian Television Service broadcast from Lagos in English and 15 or more African languages. Leading daily newspapers include ***The Daily Times, The Sunday Times, The Morning Post, The West African Pilot,*** and ***The Daily Sketch.*** There are also a large number of weeklies, monthlies, and specialized quarterly journals; most are published in English, but a few are in Yoruba.

Recreation. Popular sports include association football (soccer), boxing, wrestling, tennis, and horse racing. The lagoons are dotted with sailboats, and Victoria Island has a beautiful sand beach. There are walkways lined with coconut groves along the Marina, and the open spaces throughout the city are used for recreational purposes. At the many nightclubs, the music includes high life (a form of West African dance music), waltzes, rumbas, rock, soul, and rhythm and blues.

The future. Lagos faces such general urban problems as air and water pollution, traffic congestion, and the need for expansion. Suburban development on the mainland is being promoted, and continued land reclamation and slum clearance will add space for new construction. The creation of Lagos State will allow for the planning of new industrial districts that may reduce pollution dangers. To ameliorate the transportation problem, a monorail has been proposed that would link Lagos Island with the mainland.

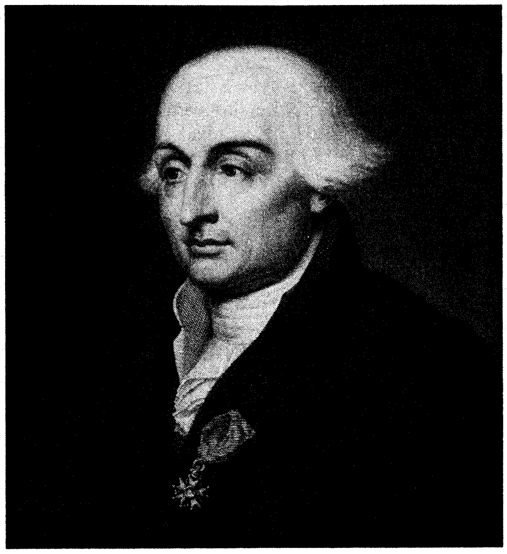
BIBLIOGRAPHY. The most recent authoritative study of Lagos is contained in A.L. MABOGUNJE, ***Urbanization of Nigeria*** (1968). J.B. LOSI, ***History of Lagos*** (1914, reprinted 1967); and J.B. WOOD, ***Historical Notices of Lagos,*** 2nd ed. (1933), are valuable for the early history of Lagos before it became the hub of Nigerian life. Conventional works that touch on the history of Lagos include A.C. BURNS, ***History of Nigeria,*** 7th ed. (1969); MICHAEL CROWDER, ***The Story of Nigeria*** (1962); and T.F. HALLIGEY, ***The Yoruba Country: Abeokuta and Lagos*** (1893). ROBERT S. SMITH, ***Kingdoms of the Yoruba*** (1969), contains valuable sections on Lagos. On the British occupation of Lagos, the best accounts are those by J.F.A. AJAYI, "The British Occupation of Lagos," ***Nigeria Magazine*** (1961); and W.D. McINTYRE, "Commander Glover and the Colony of Lagos, 1861–71," ***Journal of African History,*** 4:57–79 (1963). P.T. BAUER, ***West African Trade*** (1954); and B.W. HODDER, "The Growth of Trade at Lagos, Nigeria," ***Tijdschrift voor Economische en Sociale Geografie,*** 50:197–202 (1959), give a good picture of the economic life of Lagos. On the port of Lagos, the most reliable works are ***Lagos: The Federal Capital*** (Government Printer, 1955); "The Port of Lagos, Nigeria: Description of Apapa Wharf Extension," ***The Dock and Harbour Authority*** (1956); ***Metropolitan Lagos,*** United Nations Report (1964); and the official ***History of Lagos Harbour*** (Federal Ministry of Transport, 1956). PETER MARRIS, ***Family and Social Change in an African City: A Study of Rehousing in Lagos*** (1961), and "Slum Clearance and Family Life in Lagos," ***Human Organization,*** 19:123–128 (1960), are excellent studies. Demographic information is contained in some of the works mentioned above, but the ***Population Census of Lagos, 1950*** (1951), is a solid source; on architecture, A.L. MABOGUNJE (*op. cit.*) is the best account; see also G.J. AFOLABI OJO, ***Yoruba Palaces*** (1967); ***The Iga Idunganran: Palace of the Oba of Lagos*** (Federal Ministry of Information, 1966); and PIERRE VERGER, ***Flux et reflux de la traite des Negres entre le Golfe de Benin et Bahia de Todos os Santos du XVII***^{e} ***au XIX***^{e} ***siècle*** (1968), which contains many photographs illustrative of Lagos architecture over the years.

(B.I.O.)

Lagrange, Joseph-Louis, Comte de

A French–Italian mathematician, Joseph-Louis Lagrange was the intellectual disciple of Leonhard Euler, the leading 18th-century mathematician. He excelled in all fields of analysis and number theory, as well as in analytical and celestial mechanics. His most important book is ***Mkcanique analytique*** **(1788)**, the textbook on which all later work in this field is based.

By courtesy of the trustees of the British Museum; photograph, J.R. Freeman & Co. Ltd.



Lagrange, engraving by Robert Hart.

Early achievements

He was born January 25, 1736, in the Italian city of Turin, of a well-to-do family of French origin on his father's side. His father was treasurer to the King of Sardinia and lost his fortune in speculation. Lagrange later said, "If I had been rich, I probably would not have devoted myself to mathematics." His interest in mathematics was aroused by the chance reading of a memoir by the English astronomer Edmond Halley. At 19 (some say 16) he was teaching mathematics at the artillery school of Turin. He was to be instrumental in founding the Turin Academy of Sciences. His early publications, on the propagation of sound and on the concept of maxima and minima, were well received; Leonhard Euler, in Berlin, praised Lagrange's version of his theory of variations. The young mathematician continued to surprise his contemporaries with his discoveries.

By 1761 he was already recognized as one of the greatest living mathematicians. In 1764 he was awarded a prize offered by the Paris Academy of Sciences for an essay on the libration of the Moon, the apparent oscillation that causes the slight changes in position of lunar features on the face that the Moon presents to the Earth. In this essay he used the equations that now bear his name. His success encouraged the Academy in 1766 to propose, as a problem, the theory of the motions of the satellites of Jupiter. The prize was again awarded to Lagrange; and he won the same distinction in 1772, 1774, and 1778. In 1776, on the recommendation of Euler and the French mathematician Jean d'Alembert, Lagrange went to Berlin to fill a post at the academy vacated by Euler, at the invitation of Frederick the Great, who expressed the wish of "the greatest king in Europe" to have "the greatest mathematician in Europe" at his court.

Life in Berlin

Lagrange stayed in Berlin until 1787. His productivity in those years was prodigious: he published papers on the three-body problem, which concerns the evolution of three particles mutually attracted according to Newton's law of gravity; differential equations; prime number theory; the fundamentally important number-theoretic equation that has been identified (incorrectly by Euler) with John Pell's name; probability; mechanics; and the stability of the solar system. In his long paper "Réflexions sur la resolution algébrique des équations" (1770), he opened a new period in algebra and inspired Galois to his group theory.

A kind and quiet man, living only for science, Lagrange had little to do with the parties and intrigues around the King. When Frederick died, Lagrange preferred to accept Louis XVI's invitation to Paris. He was given apartments in the Louvre, was continually honoured, and was treated with respect throughout the Revolution. From the Louvre, he published his classical ***Mécanique analytique*** **(1788)**, a lucid synthesis of the hundred years of research in mechanics since Newton, based on his own calculus of variations, in which certain properties of a mechanistic system are inferred by considering the changes in a sum (or integral) that are due to conceptually possible (or virtual) displacements from the path that describes the actual history of the system. This led to independent coordinates that are necessary for the specifications of a system of a finite number of particles, or "generalized coordinates." It also led to the so-called Lagrangean equations for a classical mechanical system in which the kinetic energy of the system is related to the generalized coordinates, the corresponding generalized forces, and the time. The book was typically analytic; he stated in his preface that "one cannot find any figures in this work."

Work after the Revolution

The revolution that began in 1789 pressed Lagrange into work on the committee to reform the metric system and then into teaching. When the great chemist Antoine-Laurent Lavoisier was guillotined, Lagrange commented: "It required only a moment to sever that head, and perhaps a century will not be sufficient to produce another like it." When the École Polytechnique was opened in 1795, he became, with Gaspard Monge, the leading professor of mathematics. His lectures were published as ***Théorie des fonctions analytiques*** (1797) and ***Leçons sur le calcul des fonctions*** (1804) and were the first textbooks on real analytical functions. In them Lagrange, worried about the weak foundation of the calculus having to do with the ratios of small quantities and the limits of such ratios, namely, the differential calculus, tried to base it on algebra and thus eliminate the infinitesimal—a gallant but unsuccessful attempt. He also continued to work on his ***Mkcaniaue analvtique,*** but the new edition appeared only after his death.

Napoleon honoured the aging mathematician, making him a ***senateur*** and a *comte de l'Empire*, but he remained the quiet, unobtrusive academician, a venerable figure wrapped in his thoughts. He married twice. His second wife, the daughter of the astronomer Pierre-Charles Le Monnier, was much younger than he. When he died on April 10, 1813, he was interred in the Panthéon. The funeral orations were given by the astronomer Pierre-Simon Laplace and the naturalist Bernard Lacépède.

BIBLIOGRAPHY. A recent biography of Lagrange is F. BURZIO, *Lagrange* (1942), in Italian. The only other extensive biography is by Lagrange's friend, the astronomer J.B. DELAMBRE, ***Notice sur*** *Lagrange* (1816). E.T. BELL, ***Men of Mathematics*** (1953), devotes a chapter to Lagrange. An interesting essay is G. SARTON, "Lagrange's Personality," in ***Proceedings of the American Philosophical Society,*** 88:457–496 (1944). See also histories of mathematics such as C.B. BOYER, ***A History of Mathematics*** (1968). Lagrange's works have been published in 14 volumes, ***Oeuvres*** (1867–92); Delambre's biography is included in the first volume.

(D.J.S.)

Lahore

Lahore, the second largest city of Pakistan after Karāchi, is situated in a flourishing region of fertile alluvial soil in the upper Indus plain. It lies on the historic routes at the bridgehead over the Rāvi, a tributary of the Indus. It is important as an administrative, industrial, commercial, agricultural-marketing, and educational centre and as a military cantonment and also gives its name to a district and a division.

The city of Lahore was the capital of the Punjab for over 900 years; in 1955 it was selected as the capital of West Pakistan. After the dismemberment of the province of West Pakistan into four provinces in 1970, Lahore became the capital of a new province of the Punjab.

History. Information about the history of the settlement prior to the Muslim period is fragmentary and dubious. Hindu legend attributes the founding of Lahore to Lava, or Loh, son of Rāma, after whom it is said to have been named Lohawar. A Chinese traveller, Hsüan Tsang, however, made no mention of it in connection with his journey from Taki (northern Punjab) to Jullunder in 630.

Lahore figured prominently in the historical annals of Punjab from 1152 to 1849. It was the seat of government of the Ghaznavid dynasty from 1152 to 1186, during which time it was put to ransom by the Ghūrid Muʿizz-ud-Din Muhammad, in 1181, and later taken in 1186. From this time to 1238, Lahore was the centre of opposition to the authority at Delhi, while itself subject to incursions of the turbulent Khokhars. During the following century, the city lay at the mercy of recurrent Mongol incursions. In 1301 it was raided by Dua the Chagatai and in 1398 was destroyed by a detachment of the Central Asian conqueror Timur's army. It was rebuilt by Mubārak Shāh in 1422, and in 1451 came under Bahlūl Khān Lodī. The city seems to have enjoyed a period of peace under him and his lineage.

Mughal period

The city was captured by the Mughal Bābur's troops in 1524, and there ensued the golden period of Lahore's history under Mughal rule. Under the Mughals it occasionally became the place of royal residence. The city underwent an unprecedented expansion during the reign of Shāh Jahān (1628–57) but declined in importance during the reign of Aurangzeb.

From the death of Aurangzeb (1707) to the accession of Ranjit Singh (1798), the fate of Lahore was singularly unfortunate, since it was subjected to Sikh insurrections. With the invasion of Nāder Shāh, Lahore for some time became an outpost of his empire, but its history thereafter was associated with the rise of the Sikhs. Under the Sikh ruler Ranjit Singh (1799–1839) it once more became the seat of a powerful government but declined again under his successors. At the time of the British occupation in 1849, the "environs of Lahore were a mere expanse of crumbling ruins."

The contemporary city. The urban centre of Lahore comprises an area of 128 square miles (332 square kilometres). The city is sited on the left bank of the Rāvi on gently sloping terrain about 700 feet above sea level. The environs of Lahore contain three distinct alluvial landforms: active floodplain, a meander floodplain, and a "cover" floodplain inundated only occasionally. Most of the city spreads over the cover floodplain, the surface materials of which consist of deposits from repeated sheetflooding accompanied by vertical accretion.

Lahore has a subtropical continental type of climate, with hot summers and cool winters. The mean annual temperature and rainfall, respectively, are 74" F (23" C) and 20 inches (500 millimetres). June is the hottest month, with mean maximum and minimum temperatures of 106" F (41° C) and 79" F (26" C). January, the coldest month, registers a mean maximum temperature of 69" F (21" C) and a mean minimum of 40° F (4° C). July is the rainiest month (mean rainfall 5.5 inches, or 140 millimetres), and November the driest (mean rainfall 0.10 inches, or three millimetres). From July to September the weather, owing to a combination of high temperatures and humidity, is oppressive.

In terms of vegetation, Lahore is often eulogized as the "city of gardens," for, in comparison with other big urban centres of Pakistan, it contains much greenery in the form of roadside ornamental trees and sizable gardens.

Air pollution is increasing in the city because of an expanding number of automobiles and increasing industrialization. Water pollution is in evidence in ill-drained low-lying peripheral areas with high water tables.

Lahore presents a diversity of urban landscape and street patterns, resulting from its long and checkered history. Three main elements stand out:

Major sections of the city

1. An old walled city of amorphous growth, with narrow streets of uncertain width and direction, high intensity of development, great vertical expansion, inappropriate admixture of urban functions, and a congested appearance. This forms the hard core, or the inner region, of the city.

2. Anglicized districts and the cantonment, containing modern marketing areas and fashionable residential quarters. These districts, excluding the cantonment, form the intermediary zone of the city and were added during the late 19th and early 20th centuries. They manifest a lesser admixture of urban functions and moderately wide roads forming a loose rectangular pattern. The main functions are administrative, educational, commercial, and residential.

3. Postindependence residential localities, exhibiting some of the important traits of contemporary urban planning. These, together with the cantonment, encircle the inner city. This is quite extensive and contains planned residential localities, industrial sites, educational institutions, and agricultural land. The intensity of development here is least. The cantonment is a distinctive area because of its specialized function and the strict gridiron pattern of its streets. (Ed.)

The old city was formerly surrounded by a 30-foot-high brick wall and a moat, but these have been replaced, except in the north, by a garden. A circular road around the rampart gives access to the city by 13 gates. The architecture of Lahore cannot compare with that of Delhi, although there are some beautiful buildings in the fort dating from the time of Jahiingir and Shāh Jahān. The mosque of Wazīr Khān (1634) and the northern wall of the fort display magnificent examples of kashi or encaustic tile work. The Bādshāhī (Imperial) Mosque, built by Aurangzeb, is a lofty structure of red sandstone ornamented with marble tracing. The buildings of Ranjit Singh and his mausoleum are common and meretricious in style. He was, moreover, responsible for much of the despoiling of the earlier buildings, having stripped the Muslim tombs of their ornaments and sent them to Amritsar. Two of the chief features of the city lie outside its wall—Shāhdara and Shālīmār gardens. Shāhdara, about three miles to the north across the Rāvi, contains the tomb of the Mughal emperor Jahāngīr (1569–1627), surrounded by a garden. The tomb itself is a splendid marble monument in a building largely of red sandstone, with a profusion of marble ornaments arranged in elegant mosaics representing flowers, and inscriptions in Arabic and Persian. The Shālīmār gardens six miles to the east, which were laid out in 1641 by Shāh Jahān, are among the most magnificent in the world. (K.S.-U.A.)

The intracity road pattern is an irregular one with no defined geometrical design, except in some parts where it is dominantly rectangular. The pattern has resulted in a number of focal points in various parts of the city, the two most important of which are the Anārkalī-Mall Road area and the railway station. The road pattern is ill adapted to modern traffic needs, and the overall land-use pattern of the city is haphazard and poorly defined, except in the recently developed areas.

Railways link Lahore with all the important urban centres of Pakistan. The city lies on the Peshāwar–Karachi railway line and has, apart from the city railway station, several outer stations. Some 10,000,000 rail passengers are handled annually.

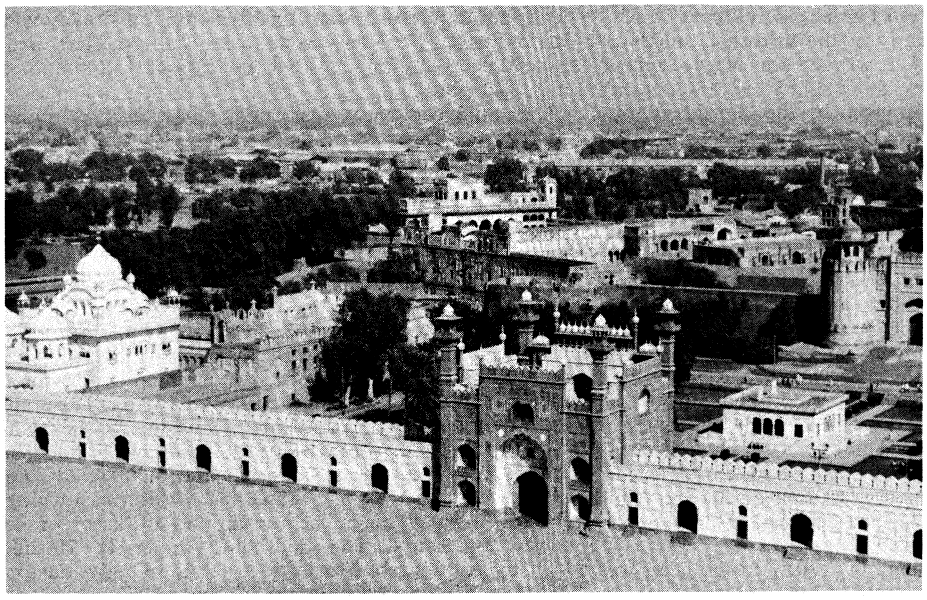
Motor transport of goods and passengers is also very important: the number of passengers carried on incoming and outgoing buses exceeds 2,000 per day, and a large number of trucks, exceeding 3,500, handle goods traffic to and from Lahore. By air, Lahore is directly connected with Rāwalpindi, Peshāwar, Lyallpur, Multān, Karāchi, and Quetta. Air traffic is increasing.

Demographic statistics

The estimated population of the city in 1971 was 1,985,800. The enumerated population (1961) was 1,296,477 (males 723,199; females 573,278). The overall density of population of the urban area works out to 10,129 persons per square mile. At the beginning of the 20th century, the population of Lahore was 202,964 (1901).

Like several other large urban agglomerations of Pakistan, Lahore suffers from overcrowding, with an average of 3.2 occupants per habitable room, which compares unfavourably with the statutory overcrowding limits of about two persons per room in countries such as Great Britain. In the poorer residential areas of the city, overcrowding is a serious menace.

Lahore presents a diversity of architectural features. The walled-city area and its outskirts contain a large number of buildings of the Mughal period, and a large number of government and public buildings, erected during the British period, particularly in the later part of the



Tomb of Ranjit Singh (left) and Lahore Fort (right) as seen from the turret of the Bādshāhī Mosque, Lahore.
Frederic Ohringer—Nancy Palmer Agency

19th century, are to be found. These generally suggest a blending of the Oriental and Western structural traits. The postindependence upper class residential colonies abound in beautiful modern structures.

Economic activity

Lahore contains 18 percent of the total industrial establishments of Pakistan. Manufacturing occupies some 30 percent of the total labour force of Lahore, textiles being the single most important industry. Trade and commerce provide employment for about 25 percent of the total labour force, and services (government, community, and personal) provide employment for about 26 percent. Lahore is the second largest centre, after Karāchi, of banking and insurance.

Public utilities include water supply, electricity, and sewerage. Water-supply arrangements, though quite elaborate, are inadequate. It is estimated that there is a deficiency of about 30 percent, or 7,000,000 to 10,000,000 gallons, per day.

The sewerage system is a semicombined one, designed to convey the domestic sewage and only a relatively small part of storm water. Overflows are provided to divert excessive storm water into drainage channels. Newly developed areas such as Gulberg have a waterborne sewage system, while older parts have open drains for sullage and conservancy crews for the collection of night soil.

Health facilities in the city include 18 hospitals, a large number of dispensaries and privately owned clinics, and over 700 private medical practitioners. It is estimated that there are only three hospital beds per 1,000 persons, a deficiency made more acute by the high incidence of disease. Being the provincial metropolis, Lahore also provides specialized medical services for its hinterland.

Lahore is an important educational centre, with two universities and numerous colleges and schools. The number of primary schools exceeds 360. According to the 1961 census, the city had a literacy rate of 32.7 percent of the total population.

Lahore has numerous and diversified cultural centres, ranging from city libraries to museums, theatres, clubs, and cultural societies. It is very deficient in public auditoriums, however, having only seven of limited capacity. Numerous English and Urdu newspapers and periodicals are published. The city has both radio and television stations.

Ranking high in the administrative hierarchy, Lahore has numerous political, governmental, and semi-governmental institutions.

The proportion of open recreational lands, including parks and playgrounds, is low in the urban centres of Pakistan. In Lahore, open recreational land accounts only for 4.6 percent of the total area, or 0.63 acre per 1,000 persons—while, according to generally accepted British and U.S. standards, it should be between five and ten acres.

Except in the planned residential districts, there is a paucity of neighbourhood parks and children's playgrounds. Lahore has, however, a few famous and large town parks.

BIBLIOGRAPHY. Published literature on Lahore consists mainly of archaeological and historical accounts. Such works include: ALEXANDER CUNNINGHAM, *Ancient Geography of India*, ed. by S.N. MAJUMDAR (1924); T.H. THORNTON, *Lahore: A Historical and Descriptive Note, Written in 1860* (1924); S.M. LATIF, *Lahore: Its History, Architectural Remains and Antiquities* (1892); H.K. TREVASKIS, *The Land of Five Rivers* (1928); R.E.M. WHEELER, *Five Thousand Years of Pakistan* (1950); and M. BAQIR, *Lahore: Past and Present* (1952).

(Ed.)

Lakes and Lake System

Definition and general occurrence

Lakes are bodies of slowly moving or standing water that occupy inland basins. Definitions that precisely distinguish lakes, ponds, swamps, and even rivers and other bodies of nonoceanic water are not well established. It may be said, however, that rivers and streams are relatively fast moving; marshes and swamps contain relatively large quantities of grasses, trees, or shrubs; and ponds are relatively small in comparison to lakes. Geologically defined, lakes are temporary bodies of water.

Within the global hydrologic cycle, freshwater lakes play a very small quantitative role, constituting only about 0.009 percent of all free water, which amounts to less than 0.4 percent of all continental freshwater. Saline lakes and inland seas contain another 0.0075 percent of all free water. Freshwater lakes, however, contain well over 98 percent of the important surface waters available for use. Apart from that contained in saline bodies, most other continental waters are tied up in glaciers and ice sheets and the remainder in groundwater.

Four-fifths of the 125,000 cubic kilometres (30,000 cubic miles) of lake waters occur in a small number of lakes, perhaps 40 in all. Lake Baikal, in central Asia, contains about 22,000 cubic kilometres (5,000 cubic miles) of water, and Lake Tanganyika (19,000 cubic kilometres [4,500 cubic miles]) and Lake Superior (12,000 cubic kilometres [3,000 cubic miles]) are the next largest. The Great Lakes of North America contain a total of about 25,000 cubic kilometres (6,000 cubic miles) of water and, together with other North American lakes larger than ten cubic kilometres (two cubic miles), constitute about one-fourth of the world's lake waters.

Although lakes are to be found throughout the world, the continents of North America, Africa, and Asia contain about 70 percent of the total lake water, the other continents being less generously endowed. A fourth of the total volume of lake water is spread throughout the world in uncounted numbers of small lakes. Anyone who has flown over much of the Canadian plains area cannot help but be struck by the seemingly endless skein of lakes and ponds covering the landscape below. Though the total volume of water involved is comparatively small, the surface area of lake water is substantial. The total surface area of all Canadian lakes has been estimated to exceed the total surface area of the province of Alberta. The state of Alaska has over 3,000,000 lakes with surface areas greater than eight hectares (20 acres).

The larger, deeper lakes are a significant factor within the cycle of water—from rain to surface water, ice, soil moisture, or groundwater and thence to water vapour. These lakes receive the drainage from vast tracts of land, store it, pass it on seaward, or lose it to the atmosphere by evaporation. On a local basis, even the smaller lakes play an important hydrologic role. The relatively high ratio of exposed surface area to the total water volume of these lakes accentuates their effectiveness as evaporators. In some cases the efficiency of lakes in losing water to the atmosphere is locally undesirable, because of demands for it by public and industrial requirements. In some basins, lakes are the terrestrial end point of the hydrologic cycle. With no outflow downstream toward the oceans, these closed lakes swell or recede according to the balance of local hydrologic conditions.

Importance to man

In today's industrial societies, requirements for water—much of which is derived from lakes—include its use for dilution and removal of municipal and industrial wastes, for cooling purposes, for irrigation, for power generation, and for local recreation and aesthetic displays. On a national level, in 1965 the United States used 160,000,000,000 gallons per day for industrial purposes, 22,000,000,000 gallons per day for all municipal reasons, and 141,000,000,000 gallons per day for irrigation. Obviously, these requirements vary considerably among regions, climates, and countries.

In another vein, it is convenient to use water to dilute liquid and some solid wastes to concentrations that are not intolerable to the elements of society that must be exposed to the effluent or wish to use it. The degree of dilution that may be acceptable varies from situation to situation and is often in dispute. In some cases, dilution is used purely to facilitate transport of the wastes to purification facilities. The water may then be made available for reuse.

Lake water is also used extensively for cooling purposes. Although this water may not be affected chemically, its change in thermal quality may be detrimental to the environment into which it is disposed, either directly, by affecting fish health or functions, or indirectly, by causing an excessive plant production and ultimate deoxygenation due to biological decay. Both fossil- and nuclear-fuelled power plants are major users of cooling water. Steel mills and various chemical plants also require large quantities. In Sweden an estimated 3.5 percent of the total national surface-water flow is presently required for cooling purposes, and it is estimated that this requirement will increase to 80 percent by the year 2000. For the United States these estimates are 10 and 50 percent.

Concern with thermal pollution of surface waters is concentrated principally on rivers and small lakes. With power requirements in modern societies increasing by about 7 percent per year, however, some apprehension has been expressed about the future thermal loading of even the largest lakes. It has been predicted that thermal inputs to each of the North American Great Lakes will increase by nearly 11 times by the year 2000. In terms of energy to be disposed in this fashion, the numbers are staggeringly large. These lakes have such large volumes, however, and such large surface areas (from which much of the heat goes into the atmosphere) that there is some question about the nature and magnitude of the actual effects.

The economic importance of waterways as communication links is enormous. In the earliest times, when travel by many societies was substantially by water, travel routes became established that resulted in relationships between cultural factors and surface hydrology networks. Today, river and lake systems serve as communication links and play an important role in shipping because of the large cargo capacities of merchant vessels and the still fairly uncongested condition of inland waterways. Oceanic shipping lanes play the major role, but river and lake systems, which link inland ports with the oceans, have been key factors in the rates of economic growth of many large inland ports.

In the Northern Hemisphere, icing on lakes precludes normal shipping commerce. Despite the limited shipping season, however, cargo tonnage on the Great Lakes for 1966 was about equal to Suez Canal tonnage and was more than twice the Panama Canal tonnage.

Commercial fisheries and other food industries reap great harvests from the major lakes of the world. The quality of the fish catch has steadily decreased in many lakes, however, with the more desirable species becoming less plentiful and the less desirable species gradually dominating the total wherever pollution and contamination of waters have occurred. Other commercial harvesting from lakes includes waterfowl, fur-bearing mammals, and some plant material, such as rice.

Abuse of lakes

Each of the uses described has associated with it the means for abuse of the very characteristics of lakes that make them desirable. Wise management of natural resources has never been man's forte. Municipalities and industries have polluted lakes chemically and thermally, the shipping that plies large inland water bodies leaves oil and other refuse in its wake, water used for irrigation often contains chemical residues from fertilizers and biocides when it is returned to lakes, and the populace that so desperately demands clean bodies of water for its recreation often ignores basic sanitary and antipollution practices, to the ultimate detriment of the waters enjoyed.

Among the major problems affecting the optimum utilization and conservation of lake waters are eutrophication (decay processes), chemical and biological poisoning, and decreases in water volumes. In the former case, discussed in more detail later, the enrichment of lakes with various nutrients supports biological productivity to an extent in which the ultimate death and decay of biological material places an excessive demand on the oxygen content, resulting in oxygen depletion in the worst cases. Phosphates and nitrates are two of the types of nutrients that are most important in this connection, particularly since they are often introduced in critical quantities in waste effluents from human sources. Other examples of chemical pollution of lakes include the introduction of DDT and other pesticides and heavy metals such as mercury. Bacteriological contamination of lake waters resulting in levels that constitute a hazard to health is another common result of man's inhumanity to his natural environment.

Water-quantity problems are complex, being related to natural vagaries of supply and levels of consumptive utilization of water. In the latter case, the percentage of water returned to the source after utilization varies with the use. The largest losses are due to actual water diversions and processes that result in evaporative losses. The use of large quantities of lake water for cooling purposes by industry and utilities, for example, may raise lake temperatures near the effluents sufficiently to cause increased evaporation. The use of certain types of cooling towers results in even larger losses. Some of the water evaporated will stay within the lake basin, but some will be lost from it.

Influence on local climate

Another example of this type of loss is connected with the possible application of weather-modification techniques to alleviate the heavy lake-effect snowfalls experienced along the lee shores of large lakes in intermediate latitudes. Redistribution of precipitation always raises the possibility of redistribution of water among various basins.

Lake-effect snowfall is just one example of the influence of lakes on local climate. The ability of large bodies of water to store heat during heating periods and to lose it more gradually than the adjacent landmasses during cooling periods results in a modifying influence on the climate. Because of this propensity, air passing over a lake's surface is cooled during summer and heated during winter. Consequently, the predominantly downwind side of a lake is more influenced by the ameliorating effects of a lake.

In most instances, moisture is also passed to the atmosphere. In summer, lake cooling serves to stabilize the air mass, but winter heating tends to decrease stability. The moisture-laden, unstable winter flows off lakes therefore produce so-called snow belts, which significantly affect downwind cities. The snow belts are usually of limited extent, often within a mile of the lake shore.

This article treats lake basins and sedimentation; the physical and chemical properties of lake waters; lake currents, waves, and tides; the hydrologic balance of lakes; and life in lakes and lake systems. For further information on related systems and environments, see SWAMPS, MARSHES, AND BOGS; RIVERS AND RIVER SYSTEMS; PLAYAS, PANS, AND SALINE FLATS; and ESTUARIES. The place of lakes within the hydrologic cycle is further dealt with in the article HYDROLOGIC CYCLE, and some aspects of lake sedimentation and water chemistry are treated in the articles SEDIMENT YIELD OF DRAINAGE SYSTEMS; RIVER DELTAS; and FRESHWATER, GEOCHEMICAL PROPERTIES OF. See also WATER WAVES; WATER RESOURCES; ICE IN RIVERS AND LAKES; and AQUATIC ECOSYSTEM for further detail on these topics, as they relate to lakes.

This article is divided into the following sections:

I. Lake basins

CLASSIFICATION OF BASINS

The name given to the study of lakes is limnology. Limnologists have used several criteria for the development of systems for classifying lakes and lake basins but have resorted particularly to the mechanisms that have produced lake basins. These have been summarized and examined in *A Treatise on Limnology*, by the American limnologist G.E. Hutchinson, which includes treatment of tectonism, volcanism, landslides, glaciation, solution, river action, wind action, coastline building, organic accumulation, animal activity, meteoritic impact, and human activity.

Basins formed by tectonism, volcanism, and landslides. Tectonism—or movements of the Earth's crust—have been responsible for the formation of very large basins. During the late Miocene (the Miocene Epoch includes the time interval from 26,000,000 to 7,000,000 years ago), broad, gentle earth movements resulted in the isolation of a vast inland sea across southern Asia and southeastern Europe. Through most of the Tertiary Period (from 65,000,000 to 2,500,000 years ago), subbasins developed that gradually were characterized by a great range of salinities. Resumption of communication with the oceans occurred later, and there is evidence of considerable variation in water levels. The present remnants of these inland bodies of water include the Caspian Sea and the Aral Sea, along with numerous smaller lakes. The Black Sea, which was also once part of this large inland basin, is now in direct communication with the oceans.

In some cases, elevated land areas may already contain depressions that eventually form lake basins. Lake Okeechobee, Florida, is cited as being such a basin, formed by uplift of the ocean floor.

Tectonic uplift may interfere with natural land-drainage patterns in such a way as to produce lake basins. The Great Basin of South Australia, some of the lakes in Central Africa (*e.g.*, Lakes Kioga and Kwania), and to some extent Lake Champlain, in the northeastern United States, are examples of this mechanism. Land subsidence, due to earthquake activity, also has resulted in the development of depressions in which lakes have evolved. Many such cases have been reported within the past 300 years.

The damming of valleys as a result of various tectonic phenomena has resulted in the formation of a few lake basins, but faulting, in its great variety of forms, has been responsible for the formation of many important lake basins. Abert Lake, in Oregon, lies in the depression formed by a tilted fault block against the higher block. Indeed, many lakes in the western United States are located in depressions formed through faulting, including Lake Tahoe, in the Sierra Nevada, California. Great Salt Lake, Utah, and other nearby salt lakes are remnants of a large Pleistocene lake, Lake Bonneville, which was formed at least partly by faulting activity (the Pleistocene Epoch includes the interval from about 2,500,000 to about 10,000 years ago).

In other parts of the world, faulting has also played an important role in basin formation. Lake Baikal and Lake Tanganyika, the two deepest lakes in the world, occupy basins formed by complexes of grabens (down-dropped faulted blocks). These lakes are among the oldest of modern lakes, as are other graben lakes, particularly those within the East African rift system which extends through the East African lake system and includes the Red Sea (see further RIFT VALLEYS).

Crater lakes and lava damming

Basins formed from volcanic activity are also greatly varied in type. The emanation of volcanic material from beneath the surface can be explosive, or it can issue in a gentle and regular manner. This range of activity and the variation of types of material which may be involved result in the possible development of many different types of basins.

One broad category includes those occupying the actual volcanic craters or their remnants. Crater lakes may occupy completely unmodified cinder cones, but these are rare. Craters caused by explosions or by the collapse of the roofs of underground magma (molten silica) chambers and those caused by explosion of new volcanic sources and that are built of nonvolcanic material are other examples. The latter are termed maars, following the local name for such forms in Germany. They are found, however, in several locations, including Iceland, Italy, and New Zealand. The maars of the volcanic district of Eifel, West Germany, are the most well-known of these formations.

The collapse of magma chambers and the development of very large surface craters called calderas is an important source of lake basins. Crater Lake, Oregon, is a typical example, exhibiting characteristically great depth and a high encircling rim. Some caldera basins evolved with gently sloping sides, however, due to the deposition of material from a series of explosions and a more gentle collapse of the structure. Secondary cones may develop within calderas, as shown by Wizard Island, in Crater Lake. The largest caldera in the world, which contains Lake Toba, in Sumatra, was formed through a combination of volcanic action and tectonic activity. Lake Toba's basin is contained in a strike-slip fault belt along the entire length of the Barisan Mountains of Sumatra. A vast, initial eruption of lava under gas pressure collapsed the magma reservoir, forming a depression that filled with water, producing the lake. Renewed volcanic activity subsequently led to the formation of an island in the centre, but a second collapse later cut it in two. Additional tectonic activity has further modified the lake's configuration.

Lake basins may also arise from the action of lava flows that emanate from volcanic fissures or craters. Lake Mývatyn, in Iceland, was formed in a basin arising from the collapse of the interior part of a large lava flow. Other basins have formed as the result of volcanic damming. This usually happens where a lava flow interrupts the existing drainage pattern.

Lake basins also may form following the blockage of a drainage depression by landslides. These may be temporary in nature because of the eroding action of the lake on the damming material. Lake Sārez in the Pamirs is stable, being dammed by a rockslide.

Basins formed by glaciation. The basin-forming mechanism responsible for the most abundant production of lakes, particularly in the Northern Hemisphere, is glaciation. The Pleistocene glaciers, which seem to have affected every continent, were especially effective in North America, Europe, and Asia. The retreat of ice sheets produced basins through mechanical action and through the damming effect of their ice masses at their boundaries (see further GLACIATION, LANDFORMS PRODUCED BY).

In some cases, lakes actually exist in basins made of ice. In other cases, water masses may form within ice masses. Such occurrences are rare and are not very stable. Damming by ice masses is a more common phenomenon but is also likely to be relatively temporary. Glacial moraine (heterogeneous sedimentary deposits at glacier margins) is also responsible for the occurrence of dammed lake basins. The Finger Lakes of New York State are dammed by an end-moraine.

Ice-scour lakes, cirque lakes, and fjords

Ice sheets moving over relatively level surfaces have produced large numbers of small lake basins through scouring in many areas. This type of glacial rock basin contains what are known as ice-scour lakes and is represented in North America by basins in parts of the high Sierras and in west central Canada (*e.g.*, near Great Slave Lake). Tens of thousands of these lakes are found in the ice-scoured regions of the world. Many of them are interconnected with short streams and may contain narrow inlets. Characteristically, they may be dotted with numerous islands and sprawling bays. Many are comparatively shallow. Where they are particularly abundant, they may cover up to 75 percent of the total surface, as in Quetico canoe country of Minnesota.

Glacier scouring associated with the freezing and thawing of névé (granular snow adjacent to glacier ice) at the head of a glaciated valley may produce a deepened circular basin termed a cirque. These are found in widely scattered mountain locations. The action of glaciers in valleys can produce a similar type of basin, often occurring in series and resembling a valley staircase. Ice movement from valleys through narrow openings has produced another type of rock basin, known as glint lake basins, particularly in Scandinavian regions.

Piedmont and fjord (*i.e.*, a river valley that has been "drowned" by a rise of sea level) lakes are found in basins formed by glacial action in long mountain valleys. Excellent examples are found in Norway, the English Lake District, the European Alps, and the Andes. In North America, several regions contain this type of lake basin. In British Columbia, many good examples exist, the largest of which are the Okanagan and Kootenay systems. These are long, narrow lakes of substantial depth. In northwestern Canada, some of the largest lakes, including Lake Athabaska, Great Slave Lake, and Great Bear Lake, are of this type, although they are not found in the same type of mountainous terrain. These lakes, as well as the North American Great Lakes, resulted from the movements of large ice sheets that deepened existing valleys.

The Wisconsin (latest stage of Pleistocene glaciation) ice sheet was responsible for shaping the present Great Lakes system, which drains mainly eastward to the Atlantic through the St. Lawrence River, during its retreat. The principal stages in the history of these lakes have received much study, and several stages of retreat and advance of the ice sheet have been identified. Behind the lobes of the ice sheet, ice lakes developed that drained according to the modifications of pre-existing valleys for glacial action. As the mass of ice retreated far to the north, glacial rebound (uplift of the Earth's crust in response to removal of the loading by ice) caused a general tilting of the land surface; the new lake basins also contributed to the subsequent changes through their own erosional action.

The material comprising glacial moraines or glacial outwash may provide dams that confine postglacial waters. The Finger Lakes, in New York State, constitute one interesting group of this type. These lakes were formed through glacial scouring of existing valleys, which were blocked at both the northern and southern ends by morainic deposits.

A variety of basin types have been formed in the different types of glacial drift-deposits, including basins in morainic material, kettle lakes, channels formed by water movement in tunnels beneath the ice masses, and lake basins formed by thawing in permafrost. An interesting example of glacial action is the formation of giant's kettles; these are glacial potholes in the form of deep cylindrical holes. Their origin is still uncertain. Sand gravel, or boulders, are sometimes found in their bottom. The kettles vary from a few inches to several feet in diameter. Good examples are found in the Alps, Germany, Norway, and in the United States.

Formation of natural dams

Basins formed by fluvial and marine processes. Fluvial action in several forms can produce lake basins; the most important processes include waterfall action, damming by sediment deposition from a tributary (fluviatile dams), sediment deposition in river deltas, damming by tidal transport of sediments upstream, changes in the configuration of river channels (*e.g.*, oxbow lakes and levee lakes), and solution of subsurface rocks by groundwater.

This last mechanism has produced the well-known formations in the Karst region, in Yugoslavia, which include subterranean and surface cavities and basins in limestone. The term karstic phenomena is applied to similar cases in many parts of the world (see further CAVES AND CAVE SYSTEMS). Solution lakes in Florida (*e.g.*, Deep Lake) are also of this origin, as are Liinersee and Seewlisee, in the Alps. Other rock types susceptible to solution basin formation include gypsum and halite. Mansfeldersee in Saxony formed in this manner.

In some coastal areas, longshore marine currents may deposit sufficient sediment to block river outflows. This damming action may be of varying intensity, and it may also occur in lake regions, where such current action causes sediment deposition that leads to the formation of multiple lakes. Accumulation of organic plant material can also result in structures that produce lake basins; Silver Lake, Nova Scotia, evolved from damming by plant material. Structural formations of coral are another potential cause of damming.

Basins formed by wind action, animal activity, and meteorites. Wind action may lead to dam or dune construction or erosion and thus can play a role in lake-basin formation. The latter case has been demonstrated in North America; a number of basins in Texas and northward, on the plains east from the Rocky Mountains, are thought to have originated from wind erosion—at least in part. Moses Lake in Washington state was formed by windblown sand that dammed the basin.

Mammals have constructed lake-forming dams: the American beaver is highly skilled at this, and his activities in this connection have established him as a symbol of industriousness. Man has also been busy in this regard and is fully capable of producing lakes that would rival the largest of the more natural variety. Plans once proposed for the damming of the Yukon River in Alaska would, if carried through, result in the formation of a lake larger than Lake Erie in surface area. Other human activities, such as quarrying and mining, also have produced cavities suitable for lake formation.

The last major mechanism of basin formation is that due to meteoritic impact. Meteorite craters (*q.v.*) are best preserved in arid climates and are often dry for this reason. A few lakes are known in craters, however, including Ungava Lake, in Quebec. In many other cases, it has not been possible to definitely confirm that basins that have the appearance of meteorite craters have, indeed, been produced by meteorite impact. Controversial ones include the bay lakes of southeast North America.

TOPOGRAPHY OF BASINS

Lakes meet with both the atmosphere and the underlying material of their terrestrial basins and interact with each. The topography and configuration of the lake bottom and the nature of the bottom materials vary considerably. They are of sufficient importance to most lake processes to warrant recognition as basic lake characteristics.

The surface area of a lake can easily be determined by cartographic techniques, but lake-volume determinations require knowledge of lake depths. Throughout the world, lakes important enough to warrant study have been sounded, and many nations have completed comprehensive programs to determine the bathymetry of large numbers of lakes. Lake sounding involves traversing a lake to collect either point or continuous measurements of depth until an accurate survey is made. Modern sounding devices measure the time taken for emitted sound to return after reflection from the bottom, relying on a knowledge of the speed of sound in water. The more sophisticated of these also provide for detection of the depths of stratification in sedimentary materials on the lake bottom. The employment of laser devices from aircraft is a recent development that is based on the transmission of light beams with wavelengths that will penetrate water.

For more practical purposes, lake morphology is a stable characteristic. Shore erosion, sediment deposition and transfer, and other processes, however, including dredging by man, may significantly alter a lake's bottom topography and thus affect navigation, currents, and ecological factors, such as fish spawning grounds.

SEDIMENTS AND SEDIMENTATION

Lake sediments are comprised mainly of clastic material (sediment of clay, silt, and sand sizes), organic debris, chemical precipitates, or combinations of these. The relative abundance of each depends upon the nature of the local drainage basin, the climate, and the relative age of a lake. The sediments of a lake in a glaciated basin, for example, will first receive coarse clastics, then finer clastics, chemical precipitates, and then increasingly large amounts of biological material, including peats and sedges.

Clues to lake history

Geologists can deduce much about a lake's history and the history of the lake basin and climate from the sedimentary records on its bottom. A sediment core contains such clues as ripple marks caused by current or wave action, carbonaceous layers, and alternations of strata that include cold- and warm-water species of fossils, pollen, and traces of chemicals of human derivation. These data provide the basis for extensive documentation of lake history (paleolimnology). Some well-known historical events, such as major volcanic eruptions, the clearing of North American forests by early settlers as revealed by pollen concentrations, the first extensive use of certain heavy metals by industry, and nuclear explosions, provide reference points in the sediment record.

Many of the materials that are detrimental to the ecology of a lake—*e.g.*, excessive quantities of nutrients, heavy metals, pesticides, oil, and certain bacteria—are deposited in lake sediments by chemical precipitation or the settling of particulate matter. These materials are potentially available for regeneration into the lake water and must be considered in any planning for measures to abate lake pollution. Within the uppermost lake sediments, large volumes of interstitial water are often present. This water may have high concentrations of nutrients and other constituents and enhance the exchange potential with the lake proper.

Clastic sediments. Waters draining into a lake carry with them much of the suspended sediment that is transported by rivers and streams from the local drainage basin. Current and wave action along the shoreline is responsible for additional erosion and sediment deposition, and some material may be introduced as a result of wind action. Rivers and streams transport material of many different sizes, the largest being rolled along the riverbed (the bed load). When river water enters a lake, its speed diminishes rapidly, bed-load transport ceases, and the suspended load begins to settle to the bottom, the largest sizes first. Lake outlets carry with them only those materials that are too small to have settled out from the inflows or those that have been introduced adjacent to the outflow. Because dynamic processes that keep materials suspended are generally more active near the shore, lake sediments are usually sorted by size; the rocks, pebbles, and coarse sands occur near shore, whereas the finer sands, silts, and muds are, in most cases, found offshore.

Clastic material over most of a lake basin consists principally of silts and clays, especially away from shores and river mouths, where larger material is deposited. Clays exist in a variety of colours, black clays containing large concentrations of organic matter or sulfides and whiter clays usually containing high calcium carbonate concentrations. Other colours, including reds and greens, are known to reflect particular chemical and biological influences.

Organic sediments are derived from plant and animal matter, Forna is recognizable plant and animal remains, *äfja* finely divided remains in colloidal suspension, and *gyttja* is a deposit formed from *äfja* that has been oxidized. Rapid accumulation of organic matter in still lakes is not uncommon; in the English Lake District, five metres (15 feet) of lake sediment of organic origin accumulated in 8,000 years. Pollen analysis has been used to accurately decipher climatic conditions of the lake in the past.

Varved deposits (*q.v.*) are the product of an annual cycle of sedimentation; seasonal changes are responsible for the information. Varves are a common feature in many areas and especially so where the land has received meltwaters from ice sheets and glaciers. The deposits consist of alternating layers of fine and coarse sediments.

Coarse clastic materials seldom are larger than boulders (25 centimetres [ten inches]), and the type of material in sizes larger than silt and clay often reveals its source. Materials along lakeshores can usually be traced back to a particular eroded source within the local drainage basin, and the distribution of this material provides evidence of the predominant current or wave patterns in the lake.

Volcanic ash is deposited downwind from its source. Ash from volcanic activity during the Pleistocene Epoch can often be dated and used as a stratigraphic marker. Lakes throughout the northwestern United States contain some of the best examples (the Mazama ash), and one deposit in the central United States, called the Pearlette ash deposit, occurs in beds as thick as three metres (ten feet).

Deposition of salts

Chemical precipitates. The major chemical precipitates in lake systems are calcium, sodium, and magnesium carbonates and dolomite, gypsum, halite, and sulfate salts. Calcium carbonate is deposited as either calcite or aragonite when a lake becomes saturated with calcium and bicarbonate ions. Photosynthesis can also generate precipitation of calcium carbonate, when plant material takes up carbon dioxide and bicarbonate and raises the pH above about 9 (the term pH is a measure of the acidity or alkalinity of water; acid waters have a pH of less than 7, and the pH of alkaline waters range from 7 to 14).

Dolomite deposition occurs in very alkaline lakes when calcium carbonate and magnesium carbonate combine. Recent dolomites have been found in Lake Balkhash, in the Soviet Union. In many saline lakes, gypsum deposition has occurred; Lake Eyre, Australia, is estimated to contain more than 4,000,000,000 tons of gypsum. For gypsum to be deposited, sulfate, calcium, and hydrogen sulfide must be present in particular concentrations. Hydrogen sulfide occurs in deoxygenated portions of lakes, usually following the depletion of oxygen resulting from decomposition of biological material. Bottom-dwelling organisms are usually absent.

Lakes that contain high concentrations of sodium sulfate are called bitter lakes, and those containing sodium carbonate are called alkali lakes. Soda Lake, California, is estimated to contain nearly 1,000,000 tons of anhydrous sulfate. Magnesium salts of these types are also quite common and can be found in the same sediments as the sodium salts. Other salts of importance occurring in lake sediments include borates, nitrates, and potash. Small quantities of borax are found in various lakes throughout the world. Mono Lake, in California, due to its high alkalinity is devoid of any life.

The gradual increase of sediment thickness through time may threaten the very existence of a lake. When a lake becomes shallow enough to support the growth of bottom-attached plants, these may accelerate the extinction of a lake. In several European countries, steps are being taken to restore lakes threatened by choking plant growth. Lake Hornborgesjon, Sweden, which has been prized as a national wildlife refuge, became the subject of an investigation in 1967 that may lead to a program involving the cutting and removal of plants that have seriously reduced the area of exposed water. It is expected that the lake level may be raised at least one metre above the present mean summer level. Lake Trummen, also in Sweden, has been treated by dredging its upper sediments. In Switzerland, Lake Wiler (Wilersee) has been treated by the removal of water just above the sediments during stagnation periods.

II. Lake waters

CHEMICAL COMPOSITION

Although the chemical composition of lakes varies considerably throughout the world, due to the varying chemistry of the erosion products of different lake basins, in most cases the principal constituents are similar. Human influences have also contributed greatly to the chemical makeup of lakes, and, although industrial effluents vary from lake to lake, many of the chemical effects of human activities are similar throughout the world. Another source in the chemical balance of lakes is the dissolved and suspended material contained in precipitation. Again, human activities have been responsible for steadily increasing concentrations of this input.

Concentration of salts in lake water

Salinity, nutrients, and oxygen. Salinity is the total concentration of the ions present in lake water and is usually computed from the sodium, potassium, magnesium, calcium, carbonate, silicate, and halide concentrations. Several important bodies of inland waters, often called inland seas, have very high salinities. Great Salt Lake, in Utah, has a salinity of about 200,000 milligrams per litre, as compared to Lake Superior's value of about 75 and an estimated mean for all rivers of 100 to 150. These ions are steadily introduced to lakes from rivers and rainwater, where they concentrate because of the evaporative loss of relatively pure water.

Where inflowing rivers erode igneous rocks, lake salinity values are relatively low, but, where soluble salts are available for erosion, salinities are relatively high. In general, it has been found that, of the cations (positively charged ions), calcium concentrations are highest, followed by magnesium, sodium, and potassium, in that order. Of the major anions (negatively charged ions), carbonate is usually the most abundant, followed by sulfate and chloride.

Other inorganic ions, though present in smaller concentrations, are of great importance. In particular, the nutrients (especially phosphate, nitrate, and silicate), heavy metals (*e.g.*, mercury, manganese, copper, lead), and polychlorinated hydrocarbons (DDT, for example) have attracted recent interest because of their role in ecological problems. Although sources of nutrients and mercury exist that are not directly related to human activities, budget studies and studies of the historical records available in sediment cores clearly reveal the great impact of human disposal of these constituents in lakes. Rainfall and dry fallout are small but significant chemical inputs to lakes. The release of gases and particulate matter into the atmosphere from factories and similar sources has increased dramatically in recent years, with consequent alterations in the chemistry of rainwater. It has been estimated, for example, that 16,000 tons of nitrogen, about 8 percent of the total from all sources, is introduced annually to Lake Erie from atmospheric action.

The substance of most interest in lakes is oxygen; once introduced to the lake water, its concentration is subject to factors within the water. Biological production (photosynthesis) releases oxygen into the water, while biological decay consumes it. Various chemical reactions within the lake system also affect the concentration of dissolved oxygen. The main source is the passage of oxygen through the air–water interface, which is affected principally by the lake temperatures; at low temperatures the partial pressure of dissolved oxygen in water is reduced. Consequently, during cold seasons, especially when vertical mixing is greatly enhanced due to lack of thermal structure and increased wind stirring, lakes are replenished with oxygen. In the warmer seasons, although surface waters may remain more or less saturated and even supersaturated, the concentrations are lower. Beneath the surface, oxygen consumption due to biological decay may cause serious depletion. Oxygen depletion also occurs near the bottom due to processes at the mud–water interface, many of which are still inadequately explained.

In winter months, a rapid formation of ice or the establishment of strong winter thermal stratification may significantly inhibit the replenishment of oxygen. Where ice cover lasts for long periods, a loss of oxygen at the mud–water interface may have repercussions for the whole lake, particularly if density currents cause significant vertical transport.

In tropical regions, where the winter replenishment mechanism (turnover) is absent, there is great reliance on the occasional occurrence of cold spells or on significant nighttime cooling to promote oxygen replenishment. Deep lakes in these regions are often anoxic (lacking in oxygen) in the deeper portions.

At any particular time, lake waters or waters entering a lake may have a biological or chemical potential for oxygen utilization. Measurements of this are termed BOD (biological oxygen demand) or COD (chemical oxygen demand). These concepts are used as partial indicators of the quality of waters being introduced to a lake.

Lakes that have a vertical salinity gradient strong enough to prevent winter turnover will usually be deoxygenated at depths where the vertical diffusion of oxygen is less than the oxygen demand. Such lakes are termed meromictic.

Carbon dioxide. Another gaseous substance of great

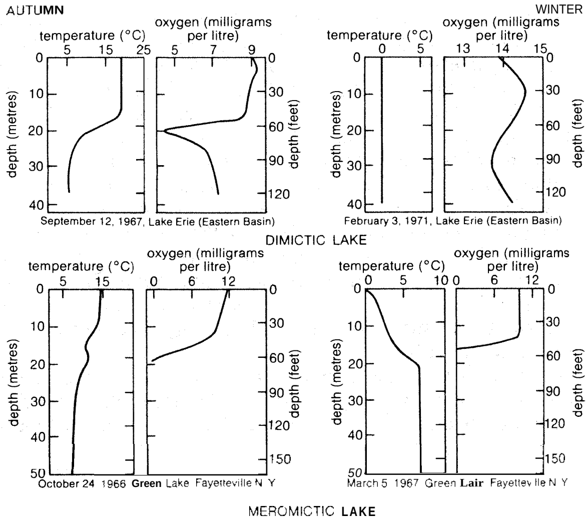


Figure 1: Seasonal variation of temperature and dissolved oxygen in dimictic (having two circulation periods annually) and meromictic (undergoing incomplete circulation at the fall overturn) lakes (see text).

importance that is exchanged with the atmosphere at the surface is carbon dioxide. Photosynthesis requires the presence of carbon dioxide, and it is released during biological breakdown.

Relation of carbon dioxide to pH and carbonate solubility

Carbon dioxide is very soluble in lake water; it forms carbonic acid, which dissociates and raises the concentration of hydrogen ions (lowering the pH). The relative proportions of bicarbonate, carbonate, and free carbon dioxide depend upon the pH. At high values of pH, carbonate ions will predominate; at low values, free carbon dioxide and carbonic acid will predominate.

Various carbonates (particularly sodium, calcium, potassium, and magnesium) are important to the carbon dioxide system. Increased pressure of carbon dioxide in the system increases the solubilities of these carbonates. In some cases, photosynthetic activity results in precipitation of certain carbonates. The entire carbon dioxide system and its behaviour at various pH values is very complex but can be interpreted from historical knowledge of lake sediments.

In waters that are neither very acidic (pH much less than 7) nor very basic (pH much greater than 7 but less than 14), the carbon dioxide system serves as a buffer, because, within limits, a change in pH will cause a shift within the system that ultimately serves to offset the pH change. Consequently, most lakes have a pH between 6 and 8. Some volcanic lakes are extremely acid, however, with pH values below 4, and some lakes with very high pH values, such as Lake Nakuru, Kenya, also occur in nature.

Sulfates, nitrates, and phosphates. Sulfate usually occurs as a principal ion in lake waters. Under anaerobic conditions in which bacteria persist in the oxidation of biological material, hydrogen sulfide is produced. When anoxic conditions exist in the deep waters just above the sediments, and the water is acidic enough to precipitate the iron present, hydrogen sulfide occurs. The characteristic and unpleasant odour of this gas is often popularly identified with the "death" of a lake. Big Soda Lake, Nevada, is extremely rich in this substance.

Nitrogen and its various compounds form another complex system in lakes, appearing as free nitrogen in solution, organic compounds, ammonia, nitrite, and nitrate. Sources of nitrogen compounds include influents to the lake (the most important source), fixation in the lake, and precipitation. Losses are experienced mainly through effluents but also by denitrification, sediment formation, and loss to the atmosphere.

Orthophosphate and various organic phosphates are the most important phosphorous compounds in lakes. Phosphates and nitrates are heavily consumed in the upper portion of lakes during periods of high productivity of phytoplankton. Increased concentrations occur in deeper portions due to decay of falling biological material and regeneration from the sediments, especially during anoxic conditions or stormy periods in shallow lakes. As limiting nutrients in many lake productivity cycles, phosphates and nitrates are often identified as controllable elements in situations where abatement is necessary to control eutrophication. Carbon is also a necessary constituent for production and in some cases can be the limiting component. Because carbon is less easily controlled and not often limiting, however, phosphates are most frequently named as substances to be reduced in effluents from industry and municipalities.

Silica also is present in lake waters, and, as with the other nutrients, it is introduced in influents and to some extent from the sediments. The production of diatom blooms is a major process for reducing silicate concentrations. Within this context, silica can also be regarded as a limiting nutrient.

THERMAL PROPERTIES

Basic physical data

Pure water freezes at 0° C (32" F), boils at 100" C (212" F), and has a latent heat of evaporation of 539.55 calories per gram, a latent heat of sublimation (ice) of 679 calories per gram, and a specific heat of 1.01 calories per gram, per ° C, at 0° C. The temperature of maximum density at atmospheric pressure occurs at 3.94" C (39.09" F). At the freezing point, ice has a lower density than water. For natural waters with high salinities, such as the oceans and inland seas, each of the values above is significantly altered. In most lakes, however, they are quite representative.

The density of water increases at pressures above one atmosphere (the pressure at sea level). Thus, pure water at 10" C (50" F) has a density of 0.9995 at one atmosphere and 1.0037 at the pressure existing at a lake depth of 1,000 metres (3,000 feet). Water raised from great depths to conditions of lower pressure experiences adiabatic cooling (without significant heat exchange with surrounding water), but there are very few lakes in which this factor can be of much significance.

Vertical mixing and overturn. It is useful to know how the temperature of maximum density changes with depth (*e.g.*, from 3.94" C at the surface to 3.39" C at 500 metres depth [38.10° F at 1,500 feet]). The fact that the temperature of maximum density of most lake waters is close to 4° C (39" F), whereas ice forms at temperatures close to 0° C, is of considerable importance. As lakes cool down to 4° C in response to surface cooling, vertical mixing takes place. When density increases with depth, the lake is said to be stable. Unstable conditions exist when density decreases with depth. Cooling at the surface to temperatures below 4° C establishes stability based on a negative thermocline (a positive thermocline is a vertical decrease in temperature with depth), because density will increase with depth. Ice then forms at the surface, enabling liquid water to exist beneath the ice in lakes, unless they are shallow enough to freeze to the bottom.

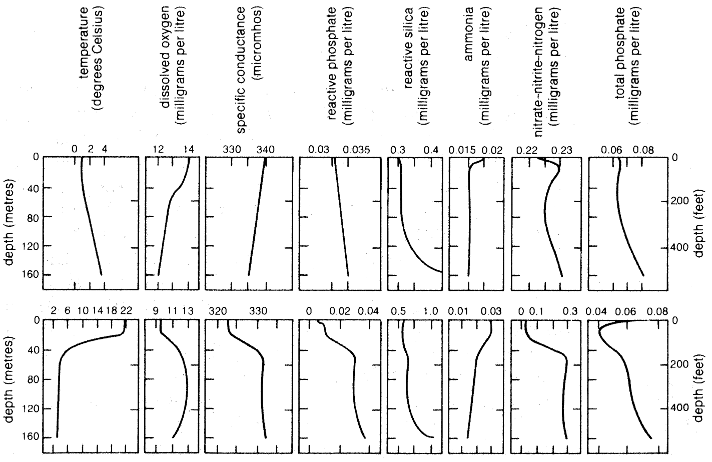


Figure 2: Distribution of some physical and chemical variables in central Lake Ontario in (bottom) July 1969 and (top) June 1970.

During the warming season, after ice has melted, heating increases the density of the surface waters, causing them to sink until stability is achieved. When surface heating proceeds above the temperature of maximum density, this process ceases, and the vertical thermal structure maintains and strengthens its stable condition, based on a positive thermocline. Turnovers tend to be seasonal.

Dimictic, monomictic, holomictic, and meromictic lakes

Mixing due to cooling or warming processes that increase the density of surface waters sufficiently to cause them to sink results in what is termed circulation, or overturn, of lake waters. Lakes that cool to below 4" C in winter experience two turnover periods, as just described, and are called dimictic lakes. Most lakes in temperate regions fall into this category. Lakes that do not cool to below 4" C undergo overturn only once per year and are called warm monomictic. Lakes that do not warm to above 4° C also experience only one overturn period per year and are called cold monomictic. There are many examples of the former, including lakes in the tropical regions and generally as far north as about 40". The cold monomictic type, however, is less common but can be found at high latitudes and at high aititudes (in the Alps, for example).

All the types described that circulate at least once throughout are called holomictic. It is possible, however, for lakes to be stable despite the thermal processes that normally induce overturn due to the existence of a positive salinity gradient with depth (chemocline). This type is called meromictic, and, in those cases where stability is permanent in at least part of the lake, the deep waters do not experience overturn and consequently are deoxygenated. Three principal origins of meromixis have been recognized. Ectogenic meromixis results from either the intrusion of seawater into a lake, as in the case of flooding from an unusually high sea level (*e.g.*, Hemmelsdorfersee, in Germany), or the introduction of freshwater through land drainage and precipitation to a saline lake (*e.g.*, Soda Lake, Nevada). Crenogenic meromixis is due to the introduction of saline water by springs, and biogenic meromixis is due to the uptake of salts from the lake sediments. North American examples include Lake Mary, Wisconsin. and Sodon Lake, Michigan.

A strong vertical salinity gradient that exists in the upper portion of a lake will affect the thermal structure by inhibiting the downward mixing of heat. In holomictic lakes, however, the downward mixing of heat due primarily to wind action usually compresses or concentrates the thermocline until it essentially separates an upper layer (epilimnion) from a lower layer (hypolimnion), each possessing weak or nonexistent vertical thermal gradients. The thermocline normally begins to grow at the beginning of the warming season. As summer passes and autumn commences, it intensifies and deepens. The onset of the cooling sees the beginning of the decay of the thermocline from above, although it usually continues to deepen until it is completely destroyed. The process just described is commonly found in lakes in temperate regions and is a seasonal phenomenon. During any period of strong warming, one or more shallower thermoclines may be observed to develop and move downward to the seasonal thermocline.

The heat budget of lakes. The heat budget of a lake includes several major factors: net incoming solar radiation, net exchange of long-wave radiation emitted by the lake surface and the atmosphere, transfer of sensible heat at the surface interface, and latent-heat processes. Those processes that are usually of much smaller importance include net inflow and outflow of heat advected by streamflow, precipitation, and groundwater flow, conduction from terrestrial heat flow, and dissipation of kinetic energy. In some cases, however, river inflow may be of more importance, such as where flow is from a nearby glacier or where the volume inflow is a significant fraction of the lake volume. Within a large lake the heat-budget considerations for a particular location must also take into account the local advection of heat within the lake by currents.

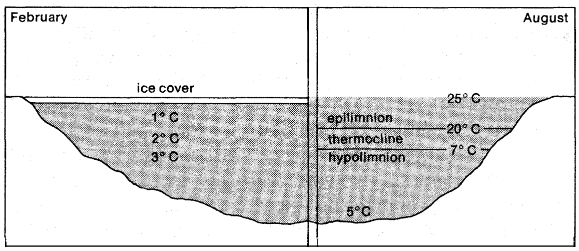


Figure 3: Temperature distribution in lakes in the middle latitudes of North America in winter and summer.

Incoming solar energy varies seasonally and with the latitude and is greatly influenced by cloud cover. The fraction that is reflected away from the lake surface depends upon the solar angle, the turbidity of the atmosphere, and the wave state, or surface roughness. In middle latitudes this ranges from about 6 percent in summer months to about 14 percent in winter.

The amount of radiation emitted by the lake surface is proportional to the fourth power of the surface temperature, whereas the radiation emitted by the clouds and atmosphere overlying the lake depends primarily upon the amount and height of the clouds and the temperature and moisture content of the atmosphere near the lake surface.

The fluxes of sensible heat and moisture at the lake surface are of great importance yet are still poorly understood. They depend upon the vertical gradients of temperature and vapour pressure above the water, respectively, and upon the factors that influence the transfer processes, such as wind and atmospheric stability. The transfer of sensible heat may be either into or out of the lake surface, usually on a seasonal basis but also sometimes on a diurnal basis. It is also possible but less likely for condensation to occur on a lake surface.

Heat flow through the bottom of lakes is normally of small significance, but exceptions exist. In a very deep lake where low rates of heating are important, such as Lake Baikal, Soviet Union, the results may be detectable. In some ice-covered lakes where other sources of heating are small, heat flux through sediments also has been shown to be significant.

The dissipation of wind energy that has been transferred to water movements is quite insignificant, as is the effect of heat transfer due to chemical and photosynthetic processes.

In latitudes and altitudes where ice is a factor, the latent heats of fusion and of evaporation of ice must also be considered within any heat-budget considerations. Heat-balance studies have been performed for lakes that are always ice covered. Solar radiation is often an important factor where ice thickness and consistency permits penetration. The heat balance of the ice is often difficult to assess, as long-wave radiation and evaporation factors are not easily measured and are very important. The exchange of sensible heat may not be large during summer months in these cases but is likely to be significant in the colder months. Several lakes that are ice covered have been shown to be meromictic; two examples are Lake Tuborg, Ellesmere Island and Lake Bonney, Antarctica.

Heat balance of two representative lakes

Heat-balance measurements or estimates have been made for many lakes throughout the world. Results show that the difference between the highest and lowest heat content for each lake varies from around 5,000 calories for high and low latitudes to around 45,000 calories for some midlatitude lakes.

The relative importance of each of the major terms of the heat budget is shown by data for two North American lakes: Lake Ontario, a large, deep, middle-latitude lake; and Lake Hefner, a relatively small, shallow lake in Oklahoma. The energy unit frequently employed is the langley (one gram calorie per square centimetre), and the figures given are approximate monthly means of langleys per day. Net solar radiation input to Lake Ontario varies from 80 to 600 (Lake Hefner varies from 200 to 600), midwinter to midsummer. Net losses due to long-wave radiation from Lake Ontario are nearly 100 throughout the year (Lake Hefner varies from **100** to **200).** Evaporation losses for Lake Ontario vary from **250** in midwinter to slightly negative values in early summer (Lake Hefner varies from **450** in late summer to **150** in spring). Conduction of heat from the surface of Lake Ontario varies from **250** in winter to about minus **100** in summer (Lake Hefner varies considerably from 80 to **—80** for the same time interval).

Heat added to a lake at the surface is usually mixed mechanically downward as a result of wind action. This process keeps the upper portion of a lake relatively uniform thermally. Consequently, a thermal gradient (thermocline) becomes established between the upper mixed layer (epilimnion) and the deep portion of the lake (hypolimnion). In shallow lakes or shallow portions of large lakes, the thermocline will eventually intercept the lake bottom so that no hypolimnion exists. Normally, as the heating season progresses, the thermocline intensifies and deepens. Secondary thermoclines may develop in the epilimnion, and these will migrate downward to the main seasonal thermocline. On very warm, still days, a thin surface layer may store heat before a mixing episode transfers heat downward. When the cooling season commences, the mixing that tends to destroy the thermocline is enhanced by vertical convection. If the cooling continues until the entire thermocline is eliminated, the lake becomes essentially isothermal and no longer exhibits the characteristics of a two-layered system.

When a lake is stratified, the most important process for downward transfer of heat to the hypolimnion is through eddy conduction. The coefficient of eddy conductivity is determined empirically and varies substantially from lake to lake. Mixing processes are generally more active in coastal areas, so that isotherms can be expected to slope downward toward shore. In large, relatively unprotected lakes, wind stress at the surface causes convergence or divergence or both of shallow waters along coastlines. Isotherms will slant upward toward the shore, and hypolimnion water may even become exposed at the surface. These occurrences are of great importance with regard to the distribution of heat within stratified lakes.

Thermal pollution in lakes

Heat introduced to lakes in large quantities, as a waste product of cooling processes in power-generating plants and other industrial concerns, is presently viewed with some concern as a pollutant, especially in small lakes. If the heat is injected at the surface it will spread initially according to the momentum of the influent and the speed and direction of ambient surface currents. When the initial momentum is sufficiently dissipated, the heat will spread mainly as a consequence of turbulent mixing processes. Throughout these events, substantial losses of heat to the atmosphere may occur, so that the full effects of the thermal input are not borne solely by the lake. Temperature values at the surface, adjacent to the influent-heat source, may be raised to a very high level—as much as several centigrade degrees. Under certain conditions fish-activity tolerances may be exceeded, and undesirable algae and plankton production may be stimulated.

If waste heat is not released at the surface but is diffused over a large depth range or injected at depth, the large local-surface-temperature problem is avoided. Losses to the atmosphere in this case, however, are also greatly reduced, and the net heat input to the lake as a whole is much greater. Over a long period, this may prove to be more detrimental to the general ecology than near-surface injection.

III. Lake hydraulics

CURRENTS

The principal forces acting to initiate water movements in lakes are those due to hydraulic gradients, wind stress, and factors that cause horizontal or vertical density gradients. Lakewater movement is usually classified as being turbulent.

Hydraulic effects are frequently the result of inflows and outflows of water. These may be substantial and continuous or weak and sporadic; in terms of the ratio of the volume of the inflow or outflow to the lake volume, the latter is the most frequently observed situation.

The stress of wind moving over the lake surface causes a transport of water within the lake, as well as the movement of energy downwind through the mechanism of surface waves. The wind is therefore one of the most important external forces on a lake. It can be relatively consistent in speed and direction, or it can be highly variable in either or both.

Pressure **gradients.** Water movements can occur as a result of internal pressure gradients and from density gradients caused by variations in temperature, sediment concentration, or the concentration of dissolved substances. Surface water in lakes can become more dense than underlying water either by cooling or heating, because the temperature of maximum density for pure lake water is about 4° C **(39" F).** Water entering a lake from rivers with a high concentration of dissolved substances will sink to a lake level of similar density. These movements are both horizontal and vertical, but the net effect is downward, if not vertical, motion.

Horizontal pressure gradients can result from many different processes that act to produce density gradients. One example is the situation of solar heating in a shallow nearshore region, where the heat is committed to the warming of a relatively small volume of water. This produces a water of lower density than the near-surface water of an adjacent deep region, where the heat is spread throughout a greater volume. Consequently, the pressure gradient force will act to move the warmer water offshore and to replace it from below with cooler water.

Influence of the Coriolis effect

Lake currents are the result of complex interactions of forces, but in many cases a small number of particular forces dominate. In the case of horizontal flow in the absence of horizontal pressure gradients, assuming no friction, water set in motion will curve to the right in the Northern Hemisphere because the Earth rotates from west to east. This effect is called the Coriolis force, and it will continue to influence water motion until there is a balance with the centrifugal force. This movement causes free-floating markers to move in an elliptical manner with a period that depends upon the latitude. In Lake Ontario, for example, it is about **17** hours. Where a dominating pressure gradient exists, the balance of the pressure-gradient force with the Coriolis force results in the so-called geostrophic flow, at right angles to the pressure gradient, with low pressure on the left (Northern Hemisphere). These conditions are most nearly realized only in very large lakes and in the oceans.

In those small lakes where hydraulic effects dominate, steady flow conditions may be achieved through balance with friction. This situation is commonly encountered in rivers, and relationships exist between mean current speed and the slope and mean depth of the river or narrow lake. These are called gradient currents and occur following situations where the wind or atmosphere pressure gradient causes a tilting of the lake surface (denivellation). In cases where the Coriolis force is a significant factor, the flow down a lake will tend to move toward the right (in the Northern Hemisphere). The development of a deeper countercurrent to the left will occur to compensate for the piling up of water on the right side.

Horizontal pressure gradients will be important in lakes where there are significant inflows of water with markedly different density from ambient lake density or where significant differential surface heating occurs.

Wind stress. Currents resulting from wind stress are the most common in lakes. Considerable research is still underway into the mechanism of transfer of wind momentum to water momentum. The stress on the lake is proportional to some power of wind speed, usually taken to be **2,** although it evidently varies with wind speed, wave conditions, and atmospheric stability. In large, deep lakes, away from the boundaries, where wind-stress effects may be balanced by Coriolis-force effects, theory suggests that the surface current will move in a direction **45°** to the right of the wind and that deeper currents are progressively weaker and directed further to the right. The depth at which flow is opposite to the wind direction is effectively the depth below which there is no influence from the wind. This depth, designated D, can theoretically occur at about 100 metres (300 feet) in large, deep, midlatitude lakes. Observations show varying degrees of fidelity to theory because of complications from coastal effects and thermal stratification.

In coastal regions, if water depth is a significant fraction of or greater than D, winds blowing parallel to the shore will transport water either onshore or offshore. In the latter case, where the coast is to the left of the wind flow (Northern Hemisphere), the water driven offshore is replaced by cooler, deeper water (upwelling).

Internal waves and Langmuir circulation. Under stratified conditions a strong thermocline will essentially separate a lake into two layers. Shearing forces that develop between these layers cause a motion, termed internal waves, that may serve to directly dissipate a substantial portion of a lake's kinetic energy and act as a coupling between motion in the epilimnion and hypolimnion. A great range of periodicities is observed in the oscillations of the thermoclines, particularly in large lakes. Internal seiches, which are responsible for relatively long-period internal waves, are discussed later.

A small-scale circulation phenomenon that has aroused considerable attention on lakes is Langmuir circulation. On windy days, parallel "streaks" can be observed to develop on the water surface and exhibit continuity for some distance.. These streaks may be caused by convergence zones where surface froth and debris collect. Langmuir circulation thus appears to be a relatively organized mixing mechanism wherein sinking occurs at the streaks and upwelling occurs between the streaks. Under favourable circumstances, this appears to be a key process for mixing heat downward in lakes.

SURFACE WAVES

Wind blowing over a calm lake surface first produces an effect that may appear as a widely varying and fluctuating ruffling of the surface. The first wave motion to develop is relatively regular, consisting of small, uniformly developed waves called capillary waves. These are quite transient, dissipating rapidly if the wind dies away or developing to the more commonly observed and more persistent gravity waves.

Energy will be continually fed to the waves by the frictional drag of the air moving over the water and by the direct force of the wind on the upwind face of the waves. The latter effect occurs only while the waves move more slowly than the wind. Pressure differences at the air–water interface also contribute energy to surface waves. Energy losses occur due mainly to turbulence in the water and, to a smaller extent, to the effects of viscosity.

Growth and movement of waves

Waves will continue to grow as long as there is a net addition of energy to them. Their height will increase as a function of wind speed and duration and the distance over which it blows (fetch). Most lakes are so small that fetch considerations are unimportant. Studies in larger lakes, however, have shown that the height of the highest waves are related to the fetch. In these lakes, waves as high as several metres are common, although waves of about seven metres (23 feet) are the highest to be expected. Wave heights in a given portion of a lake may vary considerably, due to interactions that suppress some waves and amplify others. As waves develop, their lengths increase, even after their height has stopped increasing. The phenomenon of swell, commonly observed in the oceans, is not truly realized, even in the largest lakes.

Waves travel in the same direction as the wind that generated them and at right angles to their crests. If they meet a solid object rather than a sloping beach, much of their energy will be reflected. If they enter shallow water obliquely, they are refracted. Wave speed, for waves longer than four times the depth of the water, is approximately equal to the square root of the product of the depth and the gravitational acceleration. For waves in relatively deep water, the wave speed is proportional to the square root of the wavelength.

As wave height increases, the sharpening of the wave crest may result in instability and a breaking off of the crest, a process hastened by the wind. This results in the familiar whitecaps. Waves that run ashore break up in surf. The wave height first decreases slightly, then increases, and the speed decreases, and eventually the wave form disappears as it crumbles into breakers. These can be plunging forms, in which the top curls right over the forward face, or of the spilling type, in which the crest spills down the forward face. A particular wave may break several times before reaching shore.

SEICHES

Cause and characteristics. If a denivellation, or tilting of a lake's surface, occurs as a result of a persistent wind stress or atmospheric pressure gradient, the cessation of the external forcing mechanism will result in a flow of water to restore the lake level. The flow would be periodic and uniform with depth, except for the damping effects of the lake-bottom friction and internal turbulence. Because of this, each successive tilt of the lake surface in the opposite direction occurs at a level slightly less than the previous one. The oscillation proceeds, moving the water back and forth until damping levels the water or until wind and pressure effect another tilt. This process is seiching; the lake oscillation is a seiche. The basic seiche has a single node, but harmonics of the oscillation occur, with several nodes being possible.

Seiche periods

The period of the uninodal seiche can be estimated from a formula that equates it to twice the length in the direction of the tilt, divided by the square root of the product of the mean lake depth and the gravitational acceleration.

Seiches have been noted, recorded, and studied for hundreds of years. Lake Geneva, Switzerland, was one of the first lakes to be studied in connection with seiching; it

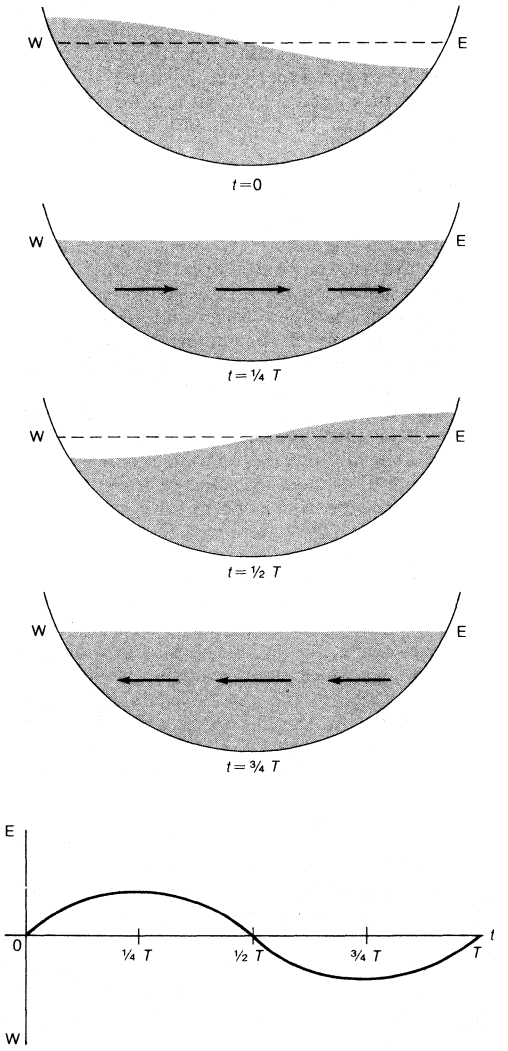


Figure 4: Four phases of elevations and currents. Below, time graph of current at centre of lake (*t* is time and T is seiche period; see text).

has an observed uninodal period of about 74 minutes and a binodal period of about 35 minutes. The observed uninodal periods of Loch Treig and Loch Earn, Scotland; Lago di Garda, Italy; Lake Vetter, Sweden; and Lake Erie, North America, are approximately nine, 14.5, 43, 179, and 880 minutes, respectively.

Long, relatively narrow lakes that are exposed to a predominance of wind flow along their major axes are most likely to exhibit so-called longitudinal seiches. Transverse seiching can occur across the narrower dimension of a lake; that observed in Lake Geneva, for example, has a period of about ten minutes.

The height of the denivellation depends upon the strength and duration of the forcing mechanism, as well as on the lake size and dimensions. In small lakes, level changes of a few centimetres are common, whereas, in the Great Lakes, intense storms can produce changes as great as two metres (seven feet). If the disturbance causing the tilting moves across the lake at close to the speed of the shallow-water wave speed, a profound amplification can occur, with possible disastrous consequences.

True tides that result from the gravitational effects of the Moon and Sun are rarely measurable in lakes, but small values of tidal components occasionally have been discerned.

Internal seiches. Internal seiching results from thermal stratification. The layers separated by the thermoclines oscillate relative to one another. Observed uninodal periods for Loch Earn, Lake Geneva, Lake Baikal, and Lake Cayuga (New York) are approximately 16, 96, 900 (binodal), and 65 hours, respectively.

Because hypolimnion water is very different from epilimnion water with regard to both thermal and biological characteristics, the massive movements of water and the turbulent exchanges that can occur during internal seiching are very important. Substantial portions of the bottom of shallow lakes can experience periodic alternation of exposure to hypolimnetic and epilimnetic water, and hypolimnetic water can be periodically exposed to the surface.

EFFECTS OF WAVE AND CURRENT ACTION

Changes in the coastline

Shore erosion and coastal features. In a lake's early stages of existence, its shore is most susceptible to changes from wave and current action. As these changes occur, there is a tendency over time to an equilibrium condition—a balance between form and processes that depends upon the nature of the materials present (*e.g.*, the size of sand and gravel present). The effectiveness of waves in the erosion process depends in part upon the depth and slope of the lake bottom. Where the shore consists of a shear cliff adjacent to deep water, wave energy will be reflected away without much erosional effect. The refraction of waves in zones of irregular coastline tends to concentrate wave energy at some locations and dilute it in others. Thus, features extended out into the lake will receive more wave energy, and the tendency is to smooth out an irregular coastline. Other net effects of shore erosion are an increase in the surface area of a lake and a reduction in its mean depth.

As erosion takes place, the distribution of erosion products results in transport of finer material offshore. The resulting terrace is called the beach in its above-water manifestation and the littoral shelf where it is below water. Landward, beyond the beach, a wave-cut cliff is usually found. The steeper slope that often separates the littoral shelf from the benthos (bottom) zone in the central part of the lake is called the step-off by some limnologists.

Water movement directed at an angle to the coastline will result in the generation of currents along the shore. Erosion products will then be transported down the coast and may be deposited in locations where transport energy is dissipated due to movement around a bend or past an obstruction. A buildup of such material is called a spit. If a bay becomes completely enclosed in this way, the spit is called a bar.

Water in very shallow lakes that are subjected to strong winds may be piled against the lee shore to such an extent that counter currents will develop from along the lee shore around each side of the lake. The cutting effects of these currents are known as end-current erosion and may characteristically alter the shape of a lake frequently subjected to winds from a particular direction.

Bottom morphology. The bottom morphology of a lake can be greatly influenced by deposition of sediment carried by inflowing rivers and streams. Although this process can be modified by wave and current action, most lakes are sufficiently quiet to permit the formation of substantial deltas. In very old lake basins the relief may become so extensively decreased due to the great buildup of deltaic deposits and the long-term effects of river widening, that deposition on the outer portions of a delta will fail to balance the effects of wave erosion. A delta, in these circumstances, will begin to shrink in size (see further RIVER DELTAS).

Problems with artificial lakes

It is very important to understand lake processes that affect the basin morphology and to be able to predict their trends and their impact on human activities. Increasingly, man is imposing his ability to change natural events in lakes, and he has often encountered problems by not anticipating a lake's reaction to his projects. The actual creation of a lake by damming a river is a major undertaking of this type. One recent example is Lake Diefenbaker, in Saskatchewan. In this region of prairie farmland, the banks of the new lake are extremely vulnerable to erosion, and planners have had to contend with the consequences of bank cutting and infilling of the basin. There are many examples of lesser engineering undertakings that have had to face the consequences of a lake's reaction. The building of jetties or breakwaters, for example, may interfere with natural circulation features. In some cases this has resulted in the reduction of flow past a harbour and increase in flow past a previously stable shoreline, with the result that the harbour has filled in or been blocked by sediment deposition, while the stable shoreline has become badly eroded (see further BEACHES).

IV. The hydrologic balance of the lakes

THE WATER BUDGET

Importance of water-budget studies

The role of lakes within the global hydrologic cycle has been described earlier. Lakes depend for their very existence upon a balance between their many sources of water and the losses that they experience. This so-called water budget of lakes is important enough to have warranted considerable study throughout the world, with each lake or lake system possessing its own hydrologic idiosyncracies. Aside from being of scientific interest, water-budget studies serve to reveal the dependence of each lake on particular hydrologic factors, thus enabling better management practices. These may include restrictions on water utilization during drought conditions, dike construction and evacuations prior to flooding, control of water levels to ensure efficient power production, and major decisions associated with diversions of watercourses in order to enhance water-quantity- and water-quality-management activities.

Often, man is able to react to predicted imbalances in the hydrologic budget, although he is usually unable to influence the basic natural factors that cause the imbalances. Precipitation and evaporation, for the most part, are uncontrollable, although some advances have been made in evaporation suppression from small lakes through the use of monomolecular surface films. Groundwater flow is not controllable, except where highly restricted flow can be tapped. Rivers and streams, however, can be subjected to regulation by well-established practices through the use of dams, storage reservoirs, and diversions. It is mainly through these controls that efforts are made to make the most efficient usage of water as a resource.

When man takes steps to alter elements of a basin's water budget, careful consideration must be given to the consequences of the hydrology and ecology of the entire watershed. Dredging operations for the purpose of harbour clearance or improvements to a navigable channel, for example, may increase the outflow from an upstream lake, increase shore erosion, or regenerate undesirable sedimentary constituents into the lake or river water. The damming of a river or a lake outlet to increase local water storage may also result in undesirable effects, such as an increased evaporation from the larger surface area, the restriction of fish movement, or changes in the thermal climate of the downstream flow. Diversions and dam-site construction may also result in flooding of important bird-breeding areas or a lowering of other lakes in the system, resulting in undesirable consequences.

Water input. The usual major input of water to a lake derives from streams and rivers, precipitation, and groundwater. In some cases inflow may come directly from glacier melt. The relative importance of each of the major sources varies from lake to lake.

Stream and river flow are usually seasonally variable, depending upon precipitation cycles and snowmelt. At low altitudes some rivers exhibit a peak during a high precipitation period in winter and then a second peak associated with a subsequent spring snowmelt that feeds the nearby high-altitude tributaries. In regions where precipitation can occur in great quantities at high rates, streams swell quickly, and water is delivered in relatively large volumes to downstream lakes.

A great deal of work has been done to improve the ability to measure and record streamflow. Consequently, it is usually the most accurately known of the inflow terms in the water budget. Most frequently, the height of the river level (stage) correlates well with the water discharge. In other cases, direct river-flow measurements are taken periodically with flow meters.

Precipitation over lake basins

Precipitation reaching a lake's surface directly may be the major input; this is true of Lake Victoria, Kenya–Tanganyika. In other cases, where the lake basin is large with well-developed drainage to a deep lake of small surface area, precipitation may be a small component. Precipitation that falls elsewhere in the lake basin may reach the lake through either surface or groundwater flow, or it may be lost due to evapotranspiration.

Measurements or estimates of precipitation for a basin are difficult to achieve. Even where elaborate networks of rain gauges exist or where these are supplemented by meteorological radar installations, total basin-precipitation data are still considered to be poor. Measurements of direct precipitation over lakes are exceedingly rare; this situation is especially serious in the case of a large lake for which nearby land data are not necessarily representative of conditions over the lake. Each climatic region throughout the world has its typical precipitation pattern, and the lakes within the regions are affected accordingly.

Groundwater controls on lake level

Groundwater reaches lakes either through general seepage or through fissures (springs). Groundwater is taken to be water in that zone of saturation that has as its surface the water table. The depth of the water table can be determined by digging a well into the saturated zone and noting the level of water—unless the water is under pressure, in which case it will rise in the well to a level above the water table. Clearly, it is possible for a lake level to coincide with the water table. In fact, unless impermeable material intervenes, the water table will drop to, rise to, or lie level with a lake surface. Groundwater that is lost from the saturated zone to a lake is termed groundwater discharge. Groundwater introduced to the saturated zone from a lake is termed recharge. The rate at which groundwater is exchanged between a lake and the saturated zone depends mainly upon the level of the water table and the pressure conditions within the saturated zone.

In permeable materials the zone above the water table is called the zone of aeration, and water within it is called soil moisture. Soil moisture is classified into three types: hygroscopic water adsorbed on the surface of soil particles; water held by surface tension in capillary spaces in the soil and moving in response to capillary forces; and water that drains through the soil under gravitational influence. The latter will most significantly contribute to groundwater recharge and to the water balance of a lake. The second category will generally be subject to loss due to transpiration by plants.

Water output. Lakes that have no outlets, either above or below surface, are termed closed lakes, whereas those from which water is lost through surface or groundwater flows are called open lakes. Closed lakes, therefore, lose water only through evaporation. In these cases, the loss of water that is less saline than the source water results in an increasing lake salinity.

Evaporation losses from lake surfaces

The process of evaporation results from a vertical gradient of vapour pressure over the water surface. Next to the water surface, saturation conditions exist that are a function of the temperature at the interface. The vapour pressure in the air above the surface is calculated from the temperature of the air and the wet-bulb temperature. The rate at which evaporation occurs also depends upon the factors that affect the removal of the saturated air above the surface (for example, wind speed and thermal convection).

Studies of evaporation must surely constitute a sizable proportion of all hydrological and oceanographic work. The principal categories of evaporation studies are water budget, energy budget, bulk aerodynamic techniqnes, and direct measurements of vapour flux (see further HYDROLOGIC CYCLE).

The so-called aerodynamic technique is based upon Dalton's formula, which correlates evaporation with the product of the vapour pressure gradient and the wind speed. Studies during the past 20 years have produced a host of variations of this equation, determined empirically using independent measurements of evaporation. One of the most often used of these was developed in a study of Lake Hefner, and even this work has been subsequently modified to suit other climates and conditions. Few workers are satisfied with the present state of the art in the use of the aerodynamic equations. Nevertheless, once an equation of this type is satisfactorily developed for a particular lake, having been checked with independent methods, it is attractive because it usually employs data that can be routinely observed.

The direct measurement of vapour fluxes is an extremely intricate proposition, as motions over a water surface are usually turbulent, and instruments capable of measuring rapidly changing vertical motions and humidities are required. Not the least of the difficulties is the likelihood that the kind of turbulence over large bodies far from land is significantly different from that over land. Recent advances in theoretical developments and instrumentation continue to encourage this type of study. In turn, successes in this field offer the opportunity for the refinement of empirical techniques more practically suited for general lake investigators.

In many lake studies, data from evaporation pans have been used to determine lake evaporation. Pans have even been developed for flotation on lakes. Pans cannot truly simulate lakes, however, as they constitute a different type of system (they are not exposed to the atmosphere in the same way, they exchange heat through their sides, and they do not store heat in the same way as lakes).

Some examples of evaporation estimates include annual totals of between 60 and 90 centimetres (two and three feet) for Lake Ontario (using different techniques and for different years); about 75 centimetres (2.5 feet) for Lake Mendota, Wisconsin; over 210 centimetres (seven feet) for Lake Mead, Arizona and Nevada; about 140 centimetres (4.5 feet) for Lake Hefner; about 660 millimetres for Lake Ijsselmeer, The Netherlands; and about 109 millimetres for Lake Baikal.

Water output from a lake in the form of surface-water outflow generally depends upon the lake level and the capacity of the effluent channel. Although lakes often have many surface inflows or at least several incoming streams or rivers, they generally have but one surface effluent.

Water-level fluctuations. The net water balance for a particular lake will vary according to the periodic and nonperiodic variations of the inputs and outputs and is reflected in the fluctuations of the lake level. Because the prime influencing factors are meteorological, the periodicity of seasonal events are often seen in water-level records.

Seasonal variations of level

Lake-level rises generally coincide with or closely follow seasons of high precipitation, and falls of level generally coincide with seasons of high evaporation. Complications are introduced by a variety of factors, however. The storage of heavy winter precipitation as snowpack is one example. The release of this water during the spring thaw may also be hampered by the presence of river ice, resulting in late-spring or summer peaks. In large drainage basins the full effects of heavy precipitation may not be immediately realized in the lake-water balance because of the time required for basin drainage. Where glacier melt is a major input to a lake, the level changes respond to seasonal heating as well as seasonal precipitation.

Although artificial controls, in the form of diversions, river dredging, and dams, affect the levels of the Great Lakes, the latter provide good examples of seasonal variations because of the lengthy record of levels available. The rivers draining to these large lakes are relatively stable; that is, the ratio of maximum to minimum flow is about 2 or 3 to 1, compared to 30 to 1 for the Mississippi River and 35 to 1 for the Columbia River. A 67-year average of lake levels by month shows that high water occurs, on the average, in September for Lake Superior and in June for Lake Ontario. Lows occur in March and December–January, respectively. The mean range in seasonal levels, for this period, is about 30 centimetres (one foot) for Lake Superior and about 45 centimetres (1.5 feet) for Lake Ontario. The pattern varies considerably from year to year, however, and periods of exceptional precipitation and drought are shown in the records. These events ultimately affect the downstream lakes, but, because of their relatively small discharge volumes, it takes 3.5 years for 60 percent of the full effect of a supply change to Lake Huron–Michigan to appear in the outflow from Lake Ontario.

The seasonal changes in a lake's level may be superimposed on longer term trends, which in some cases dominate. Several of the large lakes of the world have lengthy water-level records that illustrate long-term periods of relative abundance of water and drought. In Central Africa, Lakes Victoria, Albert, Tanganyika, and Nyasa exhibit substantial long-term features, some of which are consistent, suggesting that a common climatological factor is responsible for their existence (see further CLIMATIC CHANGE). Nevertheless, others of these features are not consistent within the lakes and have not been adequately explained.

The principal climatological factors that would most affect long-term lake-level variations have not been recorded for long periods at many locations. Regular precipitation observations were not made before about 1850. Some useful evidence is found in such natural records as tree rings and peat-bog stratigraphy.

On a worldwide basis, there is evidence of a period of low levels in the middle 19th century and near the end of the first quarter of the 20th century. Lake George, in Australia, the Caspian Sea, several lakes in western North America, and Pangong Lake, in Tibet, are examples that have exhibited these features.

LAKE EXTINCTION

Possible causes of lake extinction

The life history of a lake may take place over just a few days, in the case of one formed by a beaver dam, or, for the largest lakes, it may cover geological time periods. A lake may come to its end physically through loss of its water or through infilling by sediments and other materials. Reference has previously been made to the chemical–biological death of a lake, which is not necessarily the end of it as a physical entity but may in fact be its termination as a desirable body of water.

Geological processes involving the uplift and subsequent erosion of mountains and the advance and retreat of glaciers, establish lake basins and then proceed to destroy them through infilling. Lake basins may also lose their water through drought or through changes in the drainage pattern that result in depletion of water inflows or enhancement of outflows.

The chemical–biological changes within a lake's history offer a fine example of ecological succession. In the early stages a lake contains little organic material and has a poorly developed littoral zone. Particularly in temperate zones, such conditions favour a plentiful oxygen content, and the lake is said to be oligotrophic. As erosion progresses and as lake enrichment and organic content increase, the lake may become sufficiently productive to place an excessive demand upon the oxygen content. When periods of oxygen depletion occur, a lake is said to be eutrophic. An intermediate stage in this course of events is called mesotrophy. In the case of oligotrophy the vertical oxygen distribution is essentially uniform, or orthograde. Under eutrophic conditions, oxygen values decrease with depth, and the vertical distribution is called clinograde.

The limits of oligotrophic and eutrophic conditions have been set in terms of the rate at which oxygen is depleted from the hypolimnion. These limits are arbitrary but are approximately 0.03 and 0.05 milligrams per square centimetre per day as the upper limit of oligotrophy and the lower limit of eutrophy, respectively.

As eutrophic conditions develop, bottom sediments become enriched in organic material, and bottom plants spread throughout the littoral zone. As infilling proceeds, the plant-choked littoral zone spreads lakeward. Eventually, the littoral zone becomes a marsh, and the central part of the lake diminishes to a pond. When the lake finally ceases to exist, terrestrial vegetation may flourish, even to the extent of forestation.

(R.K.L.)

V. Life in standing water

STANDING WATER AS A BIOLOGICAL ENVIRONMENT

Standing waters—ranging from days-old ponds to million-year-old lakes—despite differences in size and in age, share certain factors in common; these determine the kind of organisms that can live in such communities. Among the most important of these factors are light, temperature, and dissolved nutrients; wind waves are important in large lakes.

The influence of light. If a lake is turbid, photosynthesis is possible only in the uppermost lighted layers. In fact, when the sun is high in the sky, the illumination in the upper few centimetres of even a murky lake is likely to be too high, rather than too low, for optimum photosynthesis. Deep in the lake, however, light is scarce. Attenuation of light is a logarithmic process; for example, if one-half of the incident sunlight reaches a depth of one metre (about three feet), only a quarter of it will reach two metres, an eighth of it three metres, and so on. Accordingly, the illumination and the amount of photosynthesis drop off extremely rapidly with increasing depth.

The food requirements of organisms, however, are nearly constant. Respiration of both plants and animals continues in the dark, so in the deeper levels of a lake respiration is likely to exceed photosynthesis. If the depth is great enough, plants will not be able to manufacture enough organic material to keep themselves alive, and the animals must depend upon food falling from the illuminated surface layers of the lake.

Lake stratification

It is convenient to think of a deep lake as divided into two layers, a food-generating, or trophogenic, zone, in which there is enough light for photosynthesis to exceed respiration over a 24-hour period, and a food-consuming, or tropholytic, zone, in which respiration exceeds photosynthesis. The boundary between trophogenic and tropholytic zones (the compensation level) commonly moves up and down with the seasons, particularly in polar lakes with their very pronounced seasonal pulse of incident sunlight.

Temperature effects. Temperature is also very important to life in lakes. It acts directly, of course; fluctuations in temperature are gradual and less extreme than they are in air. In the temperate zone during spring and summer and in the tropics during the dry season, solar heat is commonly delivered to the surface of a lake faster than the wind can stir it into the depths. Unless the lake is very shallow relative to its size, the warm surface water, being lighter, rides over the cool water beneath, and the lake is effectively divided into two compartments until either autumn or a rainy season cools the surface enough for complete mixing to occur.

The surface layer of a stratified lake is called the **epilimnion** and the deep layer the hypolimnion, and the boundary zone between them, where temperature changes rapidly with depth, is usually known as the thermocline. Epilimnion and hypolimnion are not completely sealed off from each other. There is a continual rain of debris, largely corpses and feces, through the thermocline and some upward movement of gas bubbles, mostly methane produced by bacterial decomposition of organic sediment. Organisms with a broad enough range of temperature tolerance may move freely through the thermocline in either direction. In general, however, movement through the thermocline is slight.

From Ecology *and* Field Biology by Robert L. Smith. Copyright © 1966 by Robert Leo Smith. By permission of Harper & Row. Publishers

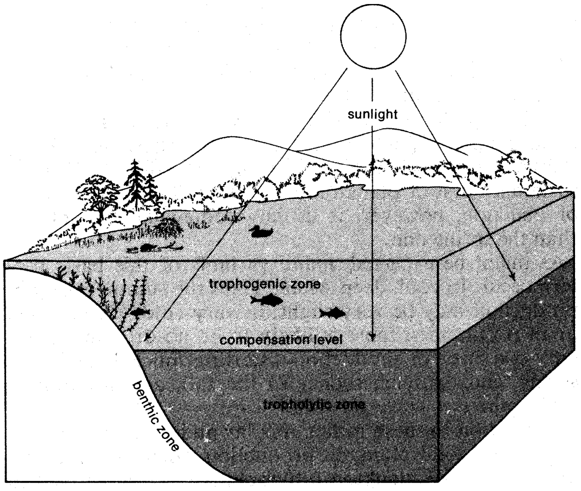


Figure 5: A lake in midsummer, showing zones of food production and consumption.

The relative extent of the **epilimnion** and the **trophogenic** (food-generating) zone and of the hypolimnion and the tropholytic (food-consuming) zone depends upon both the transparency of the water and the interplay of wind mixing and solar heating: but, in general the **epilimnion** (surface layer) tends to be strongly trophogenic, the hypolimnion (deep layer) tropholytic. During stagnation, respiration by organisms in the deep water may use a large part of the hypolimnetic store of dissolved oxygen, so that only animals capable of respiring without oxygen can survive there. If the hypolimnetic volume is very large or if it includes an appreciable part of the trophogenic zone, then an adequate supply of oxygen will be available throughout the period of stagnation, and the hypolimnion may provide a refuge for oxygen-demanding animals, such as lake trout or whitefish, which cannot tolerate summer surface temperatures.

Chemical stratification

In most lakes temperature differences are the most important cause of stratification; of special interest, however, is the minority that are stratified because of chemical differences between deep and shallow water. In this condition, called meromixis, a pocket of dense water (the monimolimnion), rich in dissolved salts, lies under the fresher upper water of the lake.

The warmest lakes known are meromictic ones of such high transparency that an appreciable amount of solar heat penetrates to the monimolimnion, where it accumulates. The density changes in such chemically stratified lakes are much greater than those resulting from temperature changes in freshwaters and occasionally lead to the formation of a lake—such as Soap Lake, in Washington—that, in wintertime may have a surface layer of ice although its hypolimnion is almost hot enough to poach eggs.

Mineral factors. Life in calcium bicarbonate lake waters is determined less by changes in the major constituents—calcium, magnesium, sodium, potassium, bicarbonate, sulfate, chloride, and silica—than by variations in biologically important substances that are commonly present in much smaller concentrations. The production of a lake commonly can be increased by adding phosphorus or compounds containing nitrogen; the addition of both increases production more than the addition of either alone. Changes in silica content have serious consequences for diatoms, algae with a siliceous cell wall; iron affects plant life directly and also because it chemically controls the availability of phosphorus.

In laboratory experiments it is possible to show that lake algae require numerous elements—nitrogen, phosphorus, sulfur, potassium, magnesium, silicon, sodium, calcium, iron, manganese, zinc, copper, boron, molybdenum, cobalt, and vanadium—and three vitamins. In most lakes, however, most of these materials are generally present in such adequate supply that it is usually difficult to demonstrate a deficiency.

Wind waves and seiche effects. Lunar tides are insignificant to life in lakes, being barely detectable on only a few of the world's largest lakes. Progressive wind waves, however, are a dominant influence on the plants and animals of the shore zone. In locales where the winds are strong and act over considerable surface area, waves

From Ecology *and Field* Biology by Robert L. Smith, Copyright © 1966 by Robert Leo Smith. By permission of Harper & Row. Publishers

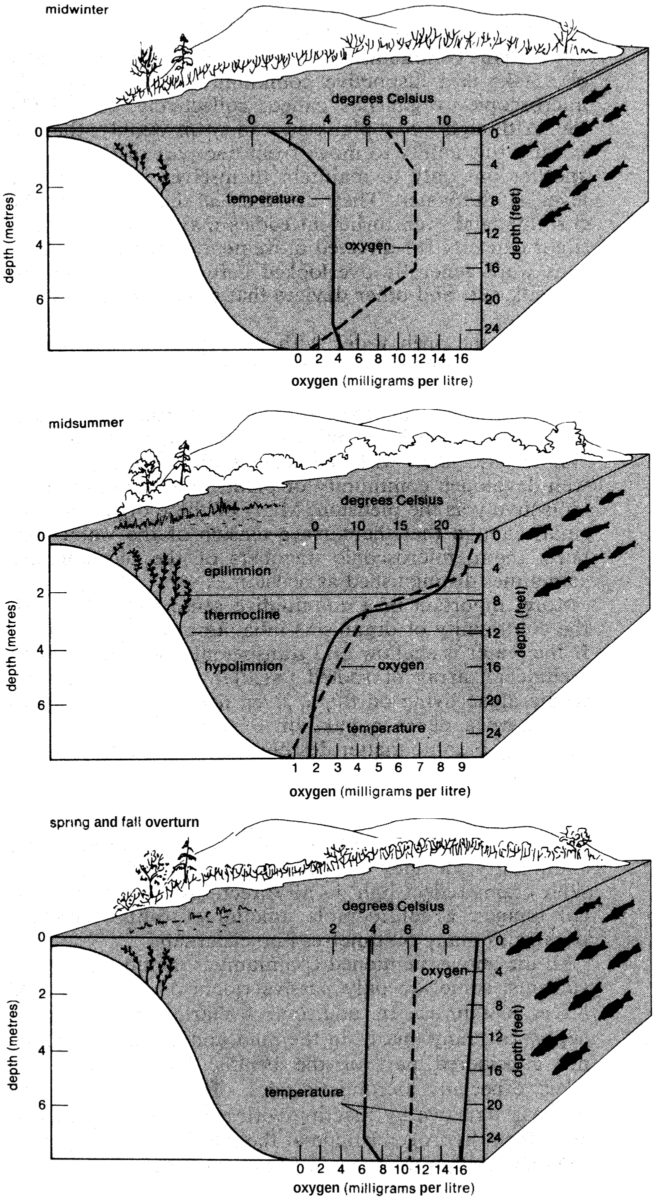


Figure 6: The influences of the seasons on lake stratification. The oxygen and temperature curves of the superimposed graphs are related to the depth of the water, as shown.

more than five metres (about 16 feet) high can be produced—high enough to be a hazard to navigation and to remove mud, clay, silt, and many organisms from beaches. Waves are probably the dominant influence on the distribution of shore plants and of the many smaller organisms of all sizes adapted to life among them.

The effect of an internal seiche

Less evident most of the time, though perhaps even more significant to life in many lakes, are standing waves called seiches, which are usually produced by the wind or by abrupt pressure disturbances. The initial disturbance tends to pile the water against one shore of the lake, leaving the opposite shore more exposed than normal. After the disturbance passes the lake, water rocks back and forth like soup in a bowl. Seiches only occasionally exceed a metre (three feet) in height, at which times they can cause considerable damage to shore structures and can sweep persons off low piers and docks. Biological significance is attached not so much to the surface seiche, however, as to the corresponding displacement of the thermocline, set up at the same time. Such an internal seiche is usually of greater amplitude than the surface seiche; in fact, all of the epilimnion may rock toward one shore, leaving hypolimnetic water exposed at the other. These large displacements of water are the main mechanism for generating turbulence in the hypolimnion of a stratified lake, and they play a vital role in the biological and chemical economy of such a lake.

CHARACTER OF LACUSTRINE POPULATIONS

Lacustrine communities. Suspended in the open water of a lake is a distinctive community of small—mostly microscopic—organisms called, collectively, the plankton. Although many plankters can swim weakly, they do not use this ability to move from one part of the lake to another but only to maintain themselves at a desirable level in the water. They are thus at the mercy of the currents, and even turbulent eddies may overpower their slight capacity for directed movement; planktonic organisms were generally overlooked before the invention of fine silk nets and other devices that separated them from the water.

Sharing the open water with the plankton are animals known collectively as the nekton; they are large, active swimmers able to move under their own power from one part of a lake to another. Fish, seals, turtles, hippopotamuses, and other large animals of lakes are all nektonic.

Often associated with the surface film of the lake is a well-developed community of plants and animals known collectively as the pleuston. Many of these organisms are small enough that the surface tension of the water supports them; microscopic members of the pleuston are sometimes distinguished as neuston.

More important in a quantitative sense is the benthos, the community of organisms inhabiting the lake bottom. If the water is shallow and transparent, the benthos may include an array of rooted vascular plants with microscopic algae living on them. Even deep or opaque lakes may have a dense population of benthic animals, sustained by organic matter descending from the trophogenic zone.

Lakes with waves adequate to maintain sand beaches are ringed with a psammon community, tiny organisms adapted to grow or move through the grains of sand. This community includes both plants and animals. Although sand grains reflect light in all directions and destroy visual images, a considerable amount of diffuse sunlight penetrates many centimetres into clean sand.

All the above-mentioned communities include micro-organisms; some are only passive riders on the bodies of larger creatures. In addition, bacteria are found to depths of many metres in the mud under a lake—depths not considered part of the benthic habitat. Stratified lakes, especially meromictic ones, often have a layer of bacteria that live by reacting with sulfate compounds so as to form hydrogen sulfide at the interface between oxygenated and nonoxygenated (reduced) zones.

Changing lake populations. In the Arctic and in the temperate zone the abundance of plants of both major producing communities, the plankton and the shallow-water benthos, fluctuates with the seasons. This presents no problem as far as the higher plants ringing the shore are concerned: like the land vegetation, they sprout in the spring, grow more slowly during the summer, and die back in the fall in response to the same climatic controls that influence trees or grasses.

Planktonic cycles

The fluctuation in abundance and activity of the planktonic algae, however, is more complex. A common pattern consists of two peaks of abundance and activity each year, one in the spring and one in the fall, with a period of relative quiescence during the summer. It appears that during the winter, when solar energy is scarce, photosynthesis is reduced, and many planktonic organisms die, releasing the nutrient salts of which they are composed. If the lake is covered with ice and snow, this process is so enhanced by the intensification of the darkness that oxygen may be sufficiently depleted to kill fish.

In spring the ice melts and freely circulating water, rich in nutrients, is exposed to bright sunlight. The result, predictably, is a spurt of photosynthetic activity and an increase in plankton abundance. During summer, thermal stratification occurs and with it a gradual depletion of nutrients in surface layers.

With the autumn overturn, however, nutrient-rich water is once again exposed to surface light intensities, which, though lower than those of spring, are still sufficient to sustain a burst of plankton production. The fall "bloom" of plankton, however, is usually somewhat less intense than the spring one.

As might be expected, many variants of this basic pattern occur. In cool, deep Alpine lakes the summer drop in production may be very slight; in very cold Alpine lakes or in Arctic ones. there is likely to be no drop at all but only one burst of production—starting much later in the spring, ending much sooner in the fall, and peaking in midsummer.

In addition to these rather regular and predictable fluctuations in total plankton production, there are changes in the plantlike plankton (phytoplankton), with first one species, then another, appearing in a fairly regular seasonal succession, repeated with minor variations from year to year. There is also a succession in the animal-like plankton (zooplankton), with some species remaining throughout the year but others appearing only during the winter or only during the summer.

These changes probably could be understood in terms of seasonally changing chemical and physical conditions, but the exact way in which the environment acts to control a single species throughout the year has not been worked out in a single lake as yet.

It seems likely that the appearance of blue-green algae in many lakes during midsummer, when dissolved nutrients are at their lowest ebb, has to do with their capacity to utilize atmospheric nitrogen, giving them an advantage over spring and fall species that must have their nitrogen presented to them in the form of nitrates or ammonia. It is also possible that biochemical factors may be important: perhaps the blue-greens require organic growth factors released by the dominant algae of spring.

ADAPTATIONS TO THE LACUSTRINE ENVIRONMENT

Compared with rivers and oceans, lakes are rather transitory features of the Earth. A lake is likely to fill in or be drained, to evaporate or freeze solid, in a time that is short compared to the rate of evolution. Many ponds are seasonal, some desert lakes (playas) hold water for only a few weeks after each rain, and Lake Eyre, in Australia, contains water so seldom that its flooding is usually recorded in the scientific literature.

Responses to the changes in lakes

Dormancy and emigration. Because of the ephemeral nature of lakes, many aquatic organisms have mechanisms of resting in a dormant state or for moving from one lake to another. Aquatic vascular plants are commonly provided with large, thick-coated seeds, with underground storage rhizomes (underground stems), or with overwintering buds that will keep the population alive while most of the active individuals succumb to drought or cold. *Daphnia*, the water flea, normally reproduces parthenogenetically (without the need of mating), but, when food gets scarce or the pond begins to dry up or some other unfavourable circumstances develop, males appear in the population and fertilize the females. The offspring of this sexual union develop inside a thick-walled resting egg, which is shed into the environment at molting, instead of being placed in the mother's brood pouch. The resting egg is capable of withstanding adverse conditions of cold or drought and does not hatch until favourable circumstances have returned. Similar resting eggs are produced by many other planktonic crustaceans and by bryozoans. In the fall, in temperate-zone lakes, large, globular, jelly-filled colonies of bryozoans are blackened by thousands of small, round resting eggs, each armed with a ring of anchor-like hooks that increase the egg's chances of being carried to a surface suitable for growth.

Many planktonic algae produce resistant spores capable of withstanding drought, and the larvae of bivalve mollusks are adapted to cling for dispersal to the gills of fish or to the feet of ducks. The short adult phase in the life of aquatic insects is also a device for moving from lake to lake. A very large part of the lacustrine (lake) biota seems to consist of species well-adapted to flee to more favourable areas should the lake in which they live become unsuitable. This is true even of such large, self-sufficient animals as the hippopotamus and the crocodile.

Planktonic adaptations. Plankters must stay suspended in the light water to survive. If they sink into the tropholytic zone, they die, either by poisoning or by starvation. Many plankters have overcome this threat by moving enough to keep themselves suspended. A few have developed gas bubbles or oil droplets to aid buoyancy. Other plankters have adapted by becoming as small as possible; thus, they sink very slowly through the water, relying on normal turbulence to hinder sinking further. If the growth of the population is greater than the slow loss to the depths, then the species can survive. This is probably the commonest way out of the dilemma and may explain why so many plankters are microscopic in size, especially in freshwater. In the sea, with its higher salinity, the sinking problem is less acute, and larger plankters are found.

Many of the planktonic organisms in lakes are surrounded either with a jelly coat or with long spines sticking out in all directions. Both are probably adaptations for increasing effective diameter without greatly increasing mass, thus reducing the sinking rate. Many diatoms are slightly asymmetrical—sigmoid or S-shaped, for example—a physical advantage that also reduces the rate of sinking.

Many species of lake plankton show one or another of two bizarre phenomena that may adapt them to open-water existence, although just how is still obscure. The first phenomenon is daily vertical migration: the plankters move up and down in the water, often considerable distances, according to the daily cycle of solar illumination. The second phenomenon is cyclic change of form. This occurs in animals such as ***Daphnia*** and in some plants, all of which reproduce asexually as they change in form from one generation to another. Some populations show only slight change during the course of each year; others go through such variation that the extremes in form would certainly be taken for different species if the intermediate types were not known. These form changes are induced largely by changes in temperature during development, although turbulence also plays a role. Of the many ideas that have been advanced for the adaptive value of the changes—that they are adjustments to the properties of warm water, that they keep the animals swimming in strata in which food is densest, that they are defense mechanisms against seasonal predators—none seems a convincing explanation for the general phenomenon.

Water-balance problems

Physiological adaptations. The ancestors of most animals that live in lakes were either terrestrial or marine. Adapted to a situation in which water was scarce, these ancestors had to be able to conserve water and to excrete extra salts and waste products with minimal water loss. The water in most lakes is much more dilute than is the body fluid of the animals; there is thus a strong tendency for the animals to gain water. This problem has been solved in two fundamentally different ways. One is typified by the freshwater mussels, which have an enormous area of mantle (an extension of the body wall) and gill exposed to the water; it might seem that they would take up water at a very high rate. The mussel, however, has developed extremely dilute body fluid—thus minimizing the tendency for water to cross its membranes—and a kidney that, because it is able to excrete very dilute urine, can excrete the excess water that does happen to enter the body. The problem has been solved in a different way by most insects in freshwater, which have a hard outer skeleton (or exoskeleton), covering as much as possible of the body; because the exoskeleton does not allow water to pass through (*i.e.*, it is impermeable), water uptake is reduced and can be handled adequately by an excretory apparatus designed for terrestrial performance.

Not all lake dwellers face the problem of an excess of water. The problem of those who live in waters many times more concentrated than the sea is to conserve water and get rid of excess salts. Relatively little is known about the excretory mechanisms of animals in such environments, however.

Water and salt balance are more easily studied in large nektonic and benthic animals than in microscopic organisms, but the latter, with their relatively thin walls and enormous surface-to-volume ratio, are exposed to severe stresses. In the case of single-celled protozoans, the problem is met by a structure called a contractile vacuole, by which dilute fluids are excreted.

Lake animals such as mussels, which have a large permeable surface exposed to the environment, are able to obtain their oxygen by diffusion from the water. Many insects, however, exploit the natural advantages of air as their respiratory medium.

Lake plants are commonly well provided with hollow tissues in their stems, through which they are able to pump oxygen down to roots and tubers buried deeply in mud. Mosquito larvae of the genus *Mansonia* have learned to tap into this gas supply with their respiratory organs and so free themselves from the usual repeated journeys to the surface to replenish their air.

Species interactions. Probably because most lakes are young and transitory, they have few of the elaborate and subtle interactions between species that are so conspicuous a part of the marine or terrestrial scene. A few cases are reminiscent of sea corals and the algae that live in them (zooxanthellae). Freshwater sponges, for example, are green because of the algae within them when they grow in the light; the same sponges are white when they grow in darkness. Some salamander eggs are usually green as a result of algae growing in the jelly masses.

Among the fishes of the oldest lakes are some interesting cases of egg predation; in one, a species of cichlid fish feeds exclusively on the eggs of mouth-brooding cichlids of other species. The predatory cichlid mimics the prey cichlid very closely in colour pattern and general appearance, persuading the mouth brooder to spit out its incubating eggs.

Parasitic relations between lake dwellers are elaborate, probably because the cycle has evolved in rivers and been secondarily extended to lakes. Many parasitic life cycles involve both lakes and terrestrial organisms; for example, the schistosomes, a group of blood flukes that are widely distributed in the lakes of the world, pass several larval stages in freshwater snails but spend their adult lives in the bloodstream of a terrestrial vertebrate. In most of the temperate zone the vertebrate host is usually a duck. Occasionally, certain of the larvae may penetrate the skin of humans, where they die, producing an annoying but usually short-lived dermatitis called swimmer's itch. In Japan and much of the tropics, however, the larvae of several species complete their development in man and cause a serious human disease.

BIOLOGICAL PRODUCTIVITY

Generally, the productivity of lakes is lowest in the Arctic, intermediate in the temperate zone, and highest in the

Correlations with latitude

tropics. The dark Arctic winter provides little solar energy for photosynthesis, and most of the springtime period of bright sunshine and long days occurs when the lakes are still covered by a thick layer of ice. In the tropics, on the other hand, where there is an adequate supply of solar energy all year and no ice cover to absorb it, high temperatures speed decomposition of plants and animals, so that nutrients of which they are composed are returned rapidly to the water and support new growth.

A second important influence on the productivity of lakes is the nature of the underlying bedrock. Rocks that are readily soluble and rich in nutrient elements, such as calcium, potassium, silica, and phosphorous, provide for a high and sustained rate of photosynthesis. Lakes resting on insoluble rocks, even though located in the favourable light and temperature environment of the tropics, have a productivity so low as to be unmeasurable by standard methods. It is in tropical regions of internal drainage, with much volcanic ash in the rocks—such as the rift valleys of tropical East Africa—that the most productive lakes in the world are found.

Presumably, an estimate of the total amount of each nutrient in the water and organisms of a lake would show a clear relation with productivity, but such an estimate is difficult to make. Even for phosphorus, which can be detected easily by radioisotope-tracer methods, the estimate is too cumbersome to be used on a widespread comparative basis.

There is, however, a fairly strong correlation between the bicarbonate–carbonate content of lakes and their primary productivity. This is likely to be at least partly the result of a correlation between bicarbonates and the generally unknown and unmeasurable total supply of nutrient salts, but it also may mean that plants in most lakes are chronically short of a convenient carbon source for photosynthesis.

If a lake is deep and transparent, it can sustain a respectable productivity per unit surface area while at the same time exhibiting a low productivity per unit volume of water. In a shallow lake, on the other hand, all the primary production takes place in a short water column, so that its effects, such as the production of algal blooms, are much more pronounced. This concentration of such growth is beneficial to organisms higher up on the food chain, such as zooplankton and fish, because large volumes of water are not required to obtain food. This is probably one reason the shore zone of a lake normally supports much denser populations of fish than does the open water.

The depth of mixing is important to the phytoplankton in another way. If it is so great that the phytoplankters spend an appreciable part of their time in the tropholytic zone, they may use up a very large part of the food they manufacture in their own respiration. This is probably why the very dense perennial blooms of highly productive tropical lakes develop only in shallow lakes or in meromictic ones, in which mixing is shallow and contained largely in the zone of trophogenesis.

Lake food chains

In general, only about 10 percent of the energy passing through one level in any food chain is consumed by the next trophic level above, which places a severe limit on the possible length of lake food chains. Few, if any, consist of more than six links, and in small lakes they may be only three links long. A high-level predator can increase its population by shortening the food chain on which it is dependent, just as human populations can (and do) by shifting from a diet rich in meat to one that is largely cereal. The resulting increase of food from a short food chain, however, is obtained at a considerable increase in the energy used to gather the food.

(D.A.L.)

BIBLIOGRAPHY. G.E. HUTCHINSON, *A Treatise on Limnology*, vol. 1, *Geography, Physics, and Chemistry* (1957) and vol. 2, *Introduction to Lake Biology and the Limnoplankton* (1967), contains a comprehensive treatment of the physical, chemical, and biological aspects of lakes and includes excellent bibliographies of previous work; basic books that contain useful bibliographies are F. RUTTNER, *Grundriss der Limnologie*, 2nd ed. (1952; Eng. trans., *Fundamentals of Limnology*, 3rd ed., 1963), a short introduction to lakes, with main emphasis on their biology; and D.G. FREY, *Limnology in North America* (1966). R.E. COKER, *Streams, Lakes, Ponds* (1954), is often used as a textbook on limnology; but for physical, hydrologic, and geologic aspects of lakes, the interested reader should see: J. PROUDMAN, *Dynamical Oceanography* (1953); B. KINSMAN, *Wind Waves: Their Generation and Propagation on the Ocean Surface* (1965); R.L. WIEGEL, *Oceanographical Engineering* (1964); and S.N. DAVIS and R.J.M. DEWEIST, *Hydrogeology* (1966). The subject of water use and planning is covered in R.J. CHORLEY (ed.), *Water, Earth and Man* (1969); J.G. NELSON and M.J. CHAMBERS (eds.), *Water* (1969); W.R.D. LEWELL and B.T. BOWER, *Forecasting the Demands for Water* (1968); UNITED STATES WATER RESOURCES COUNCIL, *The Nation's Water Resources: The First National Assessment* (1968); and F.E. MOSS, *The Water Crisis* (1967). Journals of particular interest include: *Limnology and Oceanography* (bimonthly), *Journal of the Fisheries Research Board of Canada* (monthly), *Proceedings* of the International Association for Great Lakes Research, and several oceanographic and meteorological journals. See also M.M. KOZHOV, *Lake Baikal and Its Life* (1963), a general account of the world's oldest lake and its rich biota; J.G. and P.R. NEEDHAM, *A Guide to the Study of Fresh-Water Biology*, 5th rev. ed. (1962), an introductory manual for the identification of lake and stream life; and W.R. EDMONDSON, *Fresh-Water Biology*, 2nd ed. (1959), a thorough and rather technical guide to the identification of most groups of lacustrine organisms.

(R.K.L./D.A.L.)

Lamarck, Jean-Baptiste de Monet, chevalier de

Jean-Baptiste de Lamarck, pioneer biologist in the formative era of that science, is remembered for bold speculations that all animals had evolved from primitive corpuscles through progressive improvements. His notion that evolution occurred as new characteristics were acquired through interaction with the environment and passed on to progeny is at odds with the Darwinian theory—now accepted without major qualification as the leading principle of evolutionary biology—of genetically determined variations selected by competition. Controversies generated by conflicts between these beliefs have tended to obscure Lamarck's substantial contributions to comparative anatomy and the study of the principles of classification and nomenclature of the invertebrates.

By courtesy of the Museum National D'Histoire Naturelle, Paris

Lamarck, engraving by W.H. Lizars (1788–1859).

Lamarck was born on August 1, 1744, in Picardy in northern France, the youngest of 11 children of a baron and lieutenant of infantry. Intended for the priesthood, he was sent to a Jesuit school at Amiens, but after his father died he took the opportunity to enlist in an infantry regiment, serving several years (1761–68). He became interested in plants while stationed on the Riviera and, following his resignation from the army, embarked upon the study first of medicine and then of botany, to which he soon devoted himself entirely under the French botanist Bernard de Jussieu at the Jardin du Roi (the royal botanical gardens) in Paris. Drawing on nine years of field study and collecting, Lamarck published a three-volume flora of France in 1778. Botany had become universally popular, and a wide public greeted his *Flore françoise* as a useful manual of identification. It did not adhere slavishly to the methods of the Swedish botanist Carolus Linnaeus and won for Lamarck appointment to the Académie des Sciences, which at that time was restricted to 42 members. Comte Georges de Buffon, the leading naturalist of the day, engaged him as tutor to his son during two years of travel in central Europe visiting botanical gardens and other learned institutions. He devoted the years following to voluminous botanical writings for the ***Encyclopédie me'thodique,*** successor of the famous ***Encyclope'die*** founded by Diderot, and to working as curator of the royal herbarium.

The revolution of 1789 was devoted to remaking institutions of intellect as well as of politics, and so the royal collection of natural history was discontinued. Lamarck addressed a memoir to the National Assembly condemning the random cabinets for display of curiosities built up by well-meaning amateurs and urged instead that collections be applied to the progress of science through the establishment of a great museum of natural history. Within such a collection objects "ought to be arranged in methodical or properly systematic order," not for display at random: each division of nature (animal, vegetable, and mineral) should be subdivided by classes, and those in turn by orders, and so to genera, with a written catalogue that would be the basis for systematic knowledge. Lamarck was one of the originators of the modern concept of the museum collection, an array of objects whose arrangement constitutes a classification under institutional sponsorship, maintained and kept up to date by knowledgeable specialists. When the Muséum National d'Histoire Naturelle was founded in 1793, Lamarck was placed in charge of the invertebrates, of which he had already made an important collection. He seems to have been the first to relate fossils to the living organisms to which they corresponded most closely, evincing a remarkable awareness of the great unities underlying the diverse external appearances of organisms.

Systematic order established in museum collections

By the end of the 18th century, enough had been learned in the sciences of chemistry and physiology to persuade the most acute inquirers that new understandings of great consequence might be attained through patient search for clues to fundamental relationships. Lamarck had been satisfied with and indeed excited by the looser, less critical notions of natural rhythm and the sense of cosmic unities entertained by 18th-century writers. It seemed to him that the new chemistry of Antoine Lavoisier led away from grand facts into a labyrinth of details. He feared that science would cease to be a coherent system whereby all men might understand the world and their place in it, becoming instead the confined domain of a few specialists. So he conceived a plan for a series of treatises, elaborating a unified view of physical processes and chemistry, geology, climate, and life. The first of these was a two-volume speculative treatment of matter and energy, ***Recherche~sur les causes des principaux faits physiques, et particulihrement sur celles de la combustion*** (1794), followed in 1796 by ***Réfutation de la the'orie pneumatique, ou de la nouvelle doctrine des chimistes modernes,*** in which he opposed his own theory of combustion to the views of Lavoisier and the French chemist Comte Antoine de Fourcroy. Neither of Lamarck's works was calculated to appeal to the mood of caution then coming to govern most serious scientific work, and Lamarck did not know how to dramatize his views for a wider public.

His ***Hydroge'ologie*** (1802) offered a history of the earth interpreted as a series of inundations by a global sea, each accompanied by organic deposits building up the continents. Among the insights that were highly advanced for his day was Lamarck's recognition that the type of fossil occurring in a deposit would permit inferences as to whether the deposit had been built up as deep marine sediments or coastal deposits. The book also revealed an extraordinary perception of the vastness of geologic time. "Time is insignificant and never a difficulty for Nature. It is always at her disposal and represents an unlimited power with which she accomplishes her greatest and smallest tasks." This treatise was also neglected, to Lamarck's deepening sorrow. Increasingly, science was being conducted through networks of mutual criticism in which evidence and data were employed to secure wide acceptance of essential facts before general theories were attempted. Scorning these procedures, Lamarck was transformed into a scientific outcast and gradually became an embittered solitary.

In 1800 he announced a revision of the classification of lower animals that had been left in a confused state by Linnaeus. He was able to penetrate superficial resemblances in form, as between certain worms and mollusks, through discriminating analysis of the functions and complexity of essential organs. This work he placed on an empirical foundation, "having at my disposal the magnificent collections of the Museum and another fairly rich, which I have myself made in the course of nearly thirty years' work." Published as ***Système des animaux sans vertèbres, ou table général des classes*** in 1801, Lamarck's first major work on the invertebrates reflected current research, most notably the anatomical studies of Cuvier, and established the basic arrangement for these animals that served as a guide to inquiry throughout the 19th century and is still largely accepted. These systematic studies of invertebrates were climaxed by the publication of his life's work, ***Histoire naturelle des animaux sans vertèbres,*** from 1815 to 1822, a complete vindication of his proposal to establish museum collections as the basis for revisionary work in systematic biology.

Guide to invertebrates

Lamarck imagined a vast sequence of life forms extending like a series of staircases from the simplest to the most complex. Impelled by "excitations" and "subtle and ever-moving fluids," the organs of animals became more complex and took their place on successively higher levels. This was the summary view of the relationship between physical energy and the overall organization of life set forth in ***Recherches sur l'Organisation des Corps vivans*** (1802) and the ***Philosophie Zoologique*** (1809). In the latter work he stated two "laws" that he held to govern the ascent of life to higher stages: first, that organs are improved with repeated use and weakened by disuse; second, that such environmentally determined acquisitions or losses of organs "are preserved by reproduction to the new individuals which arise." Thus, in a celebrated example, the forelegs and neck of giraffes have become lengthened through their habit of browsing. With the publication of Charles Darwin's ***Origin of Species*** 50 years later, these views of Lamarck became the centre of interest and controversy. As originally formulated, however, they were part of an elaborate surmise about processes for whose operation Lamarck had no direct evidence. To apply excerpts from so general a course of speculation to questions made much more precise through the application of Darwinian theory a century or more later—especially within the field of genetics, of which Lamarck had no conception—necessarily entails radical alterations of his meaning. From a lifelong, direct exposure to plants and animals Lamarck gained an intuitive sense of the dynamic quality of life, the close interdependence of physical and vital processes upon which the modern science of biology rests. Indeed Lamarck was the first to use the word biology, in 1802. But in the history of that science he may best be considered a forerunner rather than a founder, except in the systematic biology of the invertebrates, for which he established not only the best procedures of inquiry but also the kind of institution within which these inquiries have since been most successfully pursued.

Lamarck versus Darwin

Lamarck died blind and in poverty in Paris on December 18, 1829.

BIBLIOGRAPHY. There is no modern biography of Lamarck. Information on his work may be found in MARCEL LANDRIEU, "Lamarck, le fondateur du transformisme," ***Mémoires de la Société Zoologique de France,* vol.** *21* (1908), a

richly detailed chronological account of his major interests; H. BENTLEY GLASS, O. TEMKIN, and W.L. STRAUS (eds.), *Forerunners of Darwin: 1745–1859* (1959), an appraisal of Lamarck's contribution to evolutionary thought; and PHILIP C. RITTERBUSH, *Overtures to Biology: The Speculations of Eighteenth-Century Naturalists* (1964), a general account of speculative themes in 18th-century natural history. Two important translations of Lamarck's writings are by HUGH ELLIOTT, *Zoological Philosophy* (1914, reprinted 1963), his major speculative work; and by ALBERT V. CAROZZI, *Hydrogeology* (1964), an excellent modern edition of his work on the physical globe.

(P.C.R.)

Lamartine, Alphonse de

The earliest of the French Romantic poets, Alphonse de Lamartine also played a prominent, though less well-known, political role in the early 19th century. His spontaneous response to nature and his emotional sincerity rescued French verse from the abstractions and rigid structures of an exhausted classicism.

Giraudon

Lamartine, oil painting by François Gerard (1770–1837). In the Musée National de Versailles et des Trianons.

He was born at Mâcon on October 21, 1790. His father, an aristocrat, was imprisoned during the culminating phase of the French Revolution known as the Reign of Terror but was fortunate enough to escape the guillotine. Alphonse was educated at the college at Belley, which was maintained by the Jesuits though they were suppressed in France at this time.

Early military career

Lamartine had wanted to enter the army or the diplomatic corps, but because France was ruled by Napoleon, whom his faithful royalist parents regarded as the usurper, they would not allow him to serve. Thus, lacking a settled occupation. he remained idle until the Bourbon monarchy was restored in 1814, when he served in Louis XVIII's bodyguard. The following year, however, Napoleon returned from exile and attempted to rebuild his empire during the Hundred Days. Lamartine emigrated to Switzerland. After Napoleon's defeat at Waterloo and the Second Bourbon Restoration, he abandoned the military profession.

Attracted to literature, he wrote some tragedies in verse and a few elegies. By this time his health was not good, and he left for the spa of Aix-les-Bains, where, in October of 1816, on the shore of Lake Bourget, he met the brilliant but desperately ill Mme Julie Charles. Early in 1812 Lamartine had fallen deeply in love with a young working girl named Antoniella. In 1815 he had learned of her death, and later he was to recast her as Graziella in his prose "anecdote" of that name. He now became passionately attached to Madame Charles, who, because of her vast connections in Paris, was able to help him find a position. After her death in December 1817, Lamartine, who had already dedicated many strophes to her (notably "Le Lac"), devoted new verses to her memory (particularly "Le Crucifix").

In 1820 Lamartine married Maria Ann Birch, a young Englishwoman connected by marriage to the Churchills. The same year he published his first collection of poetry, *Méditations poétiques*, and finally joined the diplomatic corps, as secretary to the French embassy at Naples. *Meditations* was immensely successful because of its new romantic tone and sincerity of feeling. It brought to French poetry a new music; the themes were at the same time intimate and religious. If the vocabulary remained that of the somewhat faded rhetoric of the preceding century, the resonance of the sentences, the power of the rhythm, and the passion for life sharply contrasted with the often-withered poetry of the 18th century. The book was so successful that Lamartine attempted to extend it two years later with his *Nouvelles méditations poétiques* and his *Mort de Socrates*, in which his preoccupation with metaphysics first became evident. *Le Dernier Chant du pèlerinage d'Harold*, published in 1825, revealed the charm that the English poet Lord Byron exerted over him. Lamartine was elected to the French Academy in 1829, and the following year he published the two volumes of *Harmonies poétiques et religieuses*, a sort of alleluia, filled with deist—and even occasionally Christian ("L'Hymne au Christ")--enthusiasm.

Political career

That same year (1830), when Louis-Philippe acceded to the throne as constitutional monarch after the July Revolution, Lamartine abandoned his diplomatic career to enter politics. He refused to commit himself to the July Monarchy, however, and preserving his independence, he set out to draw attention to the social problem. After two unsuccessful attempts he was elected deputy in 1833. Yet he still wanted to write a poem, *Les Visions*, that he had been thinking about since 1821 and that he had conceived of as an "epoch of the soul." The symbolic theme was that of a fallen angel cast out of heaven for having chosen the love of a woman and condemned to successive reincarnations until the day on which he realized that he "preferred God." Lamartine wrote the last fragment of this immense adventure first, and it appeared in 1836 as *Jocelyn*. It is the story of a young man who intended to take up the religious life but, instead, when cast out of the seminary by the Revolution, falls in love with a young girl; recalled to the order by his dying bishop, he renounces his love and becomes a "man of God," a parish priest, consecrating his life to the service of his fellow men. In 1838 Lamartine published the first fragment of this vast metaphysical poem under the appropriate title *La Chute d'un ange* ("The Fall of an Angel"). In 1832–33 he travelled to Lebanon, Syria, and the Holy Land. He had by then definitively lost the Catholic faith he had tried to recover in 1820; a further blow was the death in Beirut, on December 7, 1832, of his only remaining child, Julia. A son born in Rome in 1821 had not survived infancy.

After a collection published in 1839 under the title *Recueillements poétiques* ("Poetic Recollections"), Lamartine interrupted his literary endeavours to become more active as a politician. He was convinced that the social question, which he himself called "the question of the proletariat," was the principal issue of his time; he deplored the inhumanity of the worker's plight; he denounced the trusts and their dominant influence on governmental politics, directing against them two discourses, one in 1838, another in 1846; he held that a working class revolution was inevitable and did not hesitate to hasten the hour, promising the authorities, in July 1847, a "revolution of scorn." In the same year he published his *Histoire des Girondins*, a history of the right, or moderate, Girondin Party during and after the French Revolution, which earned him immense popularity with the left-wing parties.

After the revolution of February 24, 1848, the Second Republic was proclaimed in Paris, and Lamartine became, in effect, head of the provisional government. The propertied classes, who were at first startled, pretended to accept the new circumstances, but they were unable to tolerate the fact that the working class possessed arms with which to defend themselves. In April, 1848, Lamartine was elected to the National Assembly by ten *départements*. The bourgeoisie, represented by the right-wing parties, thought they had elected in Lamartine a clever manipulator who could placate the proletariat, while military forces capable of establishing order, such as they conceived of it, were being reconstituted. The bourgeoisie was enraged to discover, however, that Lamartine was, indeed, as he had proclaimed himself to be, the spokesman of the working class. On June 24, 1848, he was thrown out of office and the revolt crushed.

Publications of last years

A broken man, Lamartine entered the "twilight" of his life. He was 60 years old in 1850, and his debts were enormous, not because he had been personally extravagant but because of the allowances he gave his sisters to compensate for the total property inheritance he had received as the only male in the Lamartine family. For 20 years he struggled desperately, though in vain, against bankruptcy, publishing book after book: *Raphael,* a transposed account of his love for Julie Charles, *Les Confidences* and *Nouvelles Confidences*, wherein he intermingled real and imaginary elements *(Graziella* is a fragment of it); novels: *Genevieve, Antoniella, Mémoires politiques,* the last work being of great historical interest; a periodical titled *Cours familiers de littérature* (1856–68/69), in which he published such poems as "La Vigne et la maison" and "Le Désert"; some historical works that remained unequalled, including *Histoire des Constituants, Histoire de la Restauration, Histoire de la Russie, Histoire de la Turquie*.

He died February 28, 1869, in Paris, nearly forgotten by his contemporaries and pursued by the hate of the bourgeoisie whom he had frightened in 1848. His stature remains that of a great poet, "a *voyant,* a seer, who was still inhibited by the old style of prosody" as the 19th-century symbolist poet Arthur Rimbaud would later say, a man haunted by the problem of God, and a courageous and lucid political combatant.

MAJOR WORKS

POETRY: *Méditations podtiques* (1820; *Mdditations,* 1920); *Nouvelles méditations poe'tiques* (1823); *La Mort de Socrates* (1823; *The Death of Socrates,* 1829); *Le Dernier Chant du pèlerinage d'Harold* (1825; *The Last Canto of Childe Harold's Pilgrimage,* 1827); *Les Visions* (critical edition by H. Guillemin, 1936); *Harmonies poe'tiques et religieuses* (1830); *Jocelyn,* 2 vol. (1836; Eng. trans., 1837); *La Chute d'un ange* (1838); *Recueillements poe'tiques* (1839).

HISTORICAL AND POLITICAL WRITINGS: *Histoire des Girondins,* 8 vol. (1847; *History of the Girondists,* 1847–48); *Histoire de la Re'volution de* 1848, 2 vol. (1849); *Histoire de la Restauration* (1851–52; *History of the Restoration of Monarchy in France,* trans. by the author and Capt. Rafter, 1851–53); *Histoire des Constituants* (1854; *History of the Constituant Assembly,* 1858); *Me'moires politiques* (1863); *La France parlementaire (1834–1851), oeuvres oratoires et e'crits politiques* (1864–65).

STORIES: *Raphael, pages de la vingtième année* (1849; *Raphael; or, Pages of the Book of Life at Twenty,* 1849); *Graziella* (1852; *Graziella; or, My First Sorrow,* 1871); *Geneviève* (1851; *Genevieve; or, the History of a Servant Girl,* 1851); *Le Tailleur de pierres de Saint-Point* (1851; *The Stonecutter of St. Point: A Village Tale* and *The Stonemason of Saint-Point: A Village Tale,* both 1851).

OTHER WORKS: *Souvenirs, impressions, pensées et paysages, pendant un voyage en Orient,* 1832–1833 (1835; *A Pilgrimage to the Holy Land, Comprising Recollections, Sketches, etc. Made During a Tour in the East in* 1832–3, 1835); *Les Confidences* (1849; *Confidential Disclosures* and *Memoirs of My Youth,* both 1849); *Nouvelles Confidences* (1851; *The Wanderer and His Home,* 1851); *Me'moires inédits* (1790–1813), published by his niece from 1870. Also general correspondence and numerous volumes of biography.

BIBLIOGRAPHY. A critical edition of the *Oeuvres poe'tiques complètes de Lamartine* now exists in the "Pléiade" collection (1963). For the prose works, it is necessary to refer to the edition furnished in 1863 by Lamartine himself, which includes neither *Antoniella* (1867) nor the *Me'rnoires posthumes* (1871). *Le Cours familier de litte'rature* (1856–69) has never been republished in full. The *Correspondence* is only complete—and imperfectly-for the years 1830–34 and the whole, very fragmentary, appears only in the volumes published in 1881 by the poet's niece. JEAN DES COGNETS, *La Vie intérieure de Lamartine* (1913), remains excellent; it was supplemented in 1916 by *Lamartine orateur by* LOUIS BARTHOU, and in 1925 by C. LATREILLE, *Les Dernières années de Lamartine,* 1852–1869. A book by H. GUILLEMIN, intended for students, *Lamartine, l'homme et l'oeuvre,* was published in 1940. By the same author are *Jocelyn, étude historique et critique* (1936), *Connaissance de Larnartine* (1942), and *Lamartine et la question sociale* (1946). Another important work is *Les Travaux et les jours d'Alphonse de Lamartine* by the MARQUIS DE LUPPE (1942); a popularized study (in French) is *Lnmartine* by MAURICE TOESCA (1969). *Le Livre du centerwire: etudes rec. et présentées par Paul Viallaneix* (1971) and *Actes du Congrès lamartinien* 111 (1971), should be consulted as well. Two English-language studies are A. PIRAZZINI, *The Influence of Italy on the Literary Career of Alphonse de Lamartine* (1917); and H.R. WHITEHOUSE, *Life of Lamartine, 2* vol. (1918, reprinted 1969).

(He.Gu.)

Lamiales

The order Lamiales includes four families of flowering plants: Verbenaceae (verbena family); Lamiaceae, sometimes called Labiatae (mint family); Phrymaceae (*Phryma*, or lopseed, family); and Callitrichaceae (water-starwort family). The family Lamiaceae, with over 3,500 species, is the largest, and is among the ten largest families of flowering plants.

GENERALFEATURES

Size range and diversity of structure. The order Lamiales shows a considerable diversity in size and habit, ranging from the slender, weak-stemmed aquatic and semiaquatic herbs of the Callitrichaceae to large trees of tropical forests in some genera of the Verbenaceae.

The teak tree

Of the many trees in the family Verbenaceae, the best known is teak *(Tectona grandis),* which ranges in height from 80 to 150 feet (20 to 50 metres) and has large leaves 18 to 24 inches (46 to 61 centimetres) long. Several genera of the family are entirely shrubby, a few are woody climbers, and others are annual and perennial herbs, small nonwoody plants that either complete their development and die within a single growth season, overwintering by means of seeds, or that live and produce seed from year to year, surviving winter or dry seasons by means of underground parts. Plants in the family Lamiaceae are mostly annual and perennial herbs, less often shrubs. Trees are rare and of small stature, generally not more than about 30 feet (nine metres) tall.

Distribution and abundance. The closely-related families Verbenaceae and Lamiaceae, when considered together, form one of the large natural alliances of flowering plants, worldwide in distribution. The Verbenaceae tends to be tropical and to occur in low-lying regions, in contrast to the Lamiaceae, which is predominantly temperate to subtropical and more often occurs in mountain or foothill regions.

The family Verbenaceae, with about 100 genera and 2,600 species, has its greatest concentration of genera and species in the warm parts of South America, Africa, and Madagascar. Lesser concentrations occur in continental Asia, Japan, and the islands of the Indian Ocean and the Pacific Ocean. The smallest concentrations are found in Europe, Asia Minor, and North America. In the family Lamiaceae, which has about 160 genera and 3,500 species, the primary centre of concentration is in the Old World from the Canary Islands to the Himalayas, with secondary outlying centres in Ethiopia, Madagascar, the Cape region of southern Africa, southern India, Sri Lanka (Ceylon), the Malaysian region, southern Australia, and Hawaii. In the New World a second principal centre of concentration lies in the mountain region extending from central Mexico southward to Bolivia and central Argentina; seondary centres include the northern plateau of Mexico, southwestern and southeastern United States, the highlands of Brazil and Guyana, the Caribbean region, and Chile.

Phryma leptostachya, the single species of the family Phrymaceae, has a distribution discontinuous between eastern North America and eastern Asia, a distributional phenomenon common to many plants and one that illustrates the floristic relationships between these two regions. This pattern is seen in such woody genera as *Mag-*

Split-distribution pattern

nolia (order Magnoliales), *Hamamelis* (Hamamelidales), and Wisteria (Fabales) but not in any other member of the Lamiales. In Callitriche, the single genus of the family Callitrichaceae, there are 25 or more species, which occur around the world, always associated with water.

Economic importance. The families Verbenaceae and Lamiaceae are of considerable economic importance, particularly the Lamiaceae.

From the Verbenaceae are obtained several useful wood products and some beautiful garden plants. Teak (from *Tectona* grandis), one of the most valuable of all known timbers, is native to India and Burma but is also cultivated for its timber in the Asiatic tropics. Other useful timbers are obtained from the genera Vitex, *Citharexylum*, *Peronema*, and *Premna*. The family Lamiaceae is an important source of oils, which are used in the manufacture of perfumes, deodorants, and toilet preparations and as flavourings for foods and beverages. Perfume oils are obtained commercially from basil (Ocimum), lavender (Lavandula), marjoram (*Origanum*), rosemary (*Rosmarinus*), thyme (*Thymus*), and sage (Salvia). Horehound (Marrubium vulgare) yields an essential oil used medicinally. Spices and flavouring oils are obtained from spearmint (Mentha spicata), peppermint (*Mentha piperata*), and balm, or melissa (Melissa *officinalis*). Spices are obtained from sage (Salvia *officinalis*); savory (*Satureja hortensis*); origanum, or European oregano (Origanum vulgare); thyme (*Thymus* vulgaris); Creten dittany (*Amaracus dictamnus*); sweet marjoram (*Majorana hortensis*); and basil (*Ocimum basilicum*).

Ornamental plants

Both the Lamiaceae and the Verbenaceae have many attractive ornamental plants. In the Lamiaceae most of the plants used as seasoning herbs are also used as garden ornamentals. Border and bedding plants come from several herbaceous (nonwoody) genera with brightly coloured and showy flowers, such as *Monarda*, Mentha, *Colquhounia*, *Phlomis*, Physostegia, and *Lavandula*. The shellflower (Molucella laevis) with large, saucer-like green calyxes (sepals—the outer series of petallike, usually green, structures in a flower) is popular for flower arrangements. *Coleus*, which has brightly coloured leaves, is one of the most popular houseplants. Shrubby ornamentals in the family Verbenaceae include the Chinese hat plant (Holmskioldia *sanguinea*), pigeonberry (*Duranta* erecta), the beauty-berries (Callicarpa species), the glory-bowers (Clerodendrum), the chaste tree (Vitex *agnus-castus*), lemon verbena (Aloysia *triphylla*), and the lantanas (*Lantana camara* and L. *montevidensis*).

Several of the mangroves of the genus *Avicennia* (Verbenaceae) have astringent bark that is used for tanning.

There are a few edible plants in the Verbenaceae and the Lamiaceae. The beauty-berry (Callicarpa *americana*) of the southeastern United States has juicy, slightly aromatic, sweet, fleshy fruits that are sometimes eaten. Several species of *Stachys* and Lycopus, in the eastern United States, and Coleus *esculentus*, of southern Africa, have tuberous roots that are used as root vegetables. The seeds of chia (Salvia *columbariae*), one of the sages widespread in the dry parts of the western United States, were once used extensively as food by the Indians. Members of both families have been used for various medicinal purposes. Bugleweed (*Ajuga reptans*) of the Lamiaceae, for example, has been used in treatment of bleeding of the lungs and as a safe and mild narcotic.

NATURAL HISTORY

Life cycle. *Pollination* and seed dispersal. Most of the Lamiales have brightly coloured, often bilaterally symmetrical flowers, which usually occur in showy clusters. These flowers are pollinated by insects and birds. Flowers with short corolla tubes are usually pollinated by bees; those with longer tubes, by butterflies and moths. The lobes of the corolla serve as a landing place or platform for the insect visitor, but in bird-pollinated flowers a platform is not usually present.

Pollination in the family Callitrichaceae takes place either on the surface of the water or in the water.

Mechanisms ensuring cross-pollination

The unisexual flowers of Callitriche are of necessity cross-pollinated, but cross-pollination is the rule in the order. The families with bisexual flowers are adapted to cross-pollination through the agency of insects and birds. In addition, there are several floral devices that further ensure cross-pollination. One is a time difference in the maturation of the stamens and stigmas. Flowers of the families Lamiaceae and Verbenaceae are largely protandrous; that is, the anthers (pollen sacs) mature earlier than the stigmas on a particular flower. This differential maturation time is sometimes accompanied by a change in position of the stamens, relative to the stigmas, during the life of the flower. In the large genus Clerodendrum (Verbenaceae), for example, the long stamens and style are at first parallel to each other; then the style bends down, away from the stamens as the anthers open. Following this the stamens curve downward, and the style returns to its original position before the stigmas become receptive to pollen.

In some genera of the Lamiaceae, there is a change from the bisexual condition to one in which there are two kinds of plants: some bisexual; others only female. This is called gynodioecy and occurs in Origanum, Salvia, *Mentha*, *Glechoma*, and Ziziphora. Sometimes, when this occurs, the plants with the female flowers are smaller than those with bisexual flowers. In a few instances size and sexual differences have not been recognized, and the two forms of one species have been described as separate species.

Fruits and seeds of the Lamiales show little diversity in structure in relation to their dispersal. The calyxes (sepals), except in *Callitriche*, remain attached much longer than in most flowers, usually enclose the fruits or seeds, and act as the vehicle for disseminating them.

Vivipary. *Avicennia*, a mangrove, is one of the remarkable genera of the Verbenaceae. It is a small genus of about 14 species found in coastal mud flats, tidal estuaries, or salt marshes of tropical or subtropical Africa, Asia, Australia, the Philippines, and North and South America. In these restricted habitats it occurs with *Rhizophora* and *Bruguiera* (family Rhizophoraceae) and *Sonneratia* (Sonneratiaceae) to form the vegetation type called mangrove forest. Mangrove plants are all similar in appearance in having prop roots or sucker shoots, and their seedlings are viviparous—that is, they germinate while still attached to the parent plant. In Avicennia, the seedling, when ready to be released from the parent plant, has a protruding rudimentary root and shoot that enable it to become embedded as soon as it falls into the mud. Vivipary is regarded as an adaptation to a saline environment. While growing on the parent plant, the seedlings store chloride, and their cell sap contains a very high salt content. Later, when the seedlings fall, they are able, because of their internal salt content, to take up water from the salty mud.

Root adaptations to obtain oxygen

Ecology. Some members of the Lamiales are adapted to specialized growth requirements. One of the most outstanding of these is seen in plants of the genus Avicennia. These plants have long cable-like roots that run horizontally through the mud and send up numerous erect suckers, or pneumatophores, the ends of which stand above the surface of the water and supply the underground roots with needed oxygen.

In the South American genus *Junellia* (Verbenaceae) are plants adapted to dry conditions that occur at high elevations from 10,000 to 14,000 feet (3,000 to 4,000 metres) in the Andes mountains and others adapted to dry desert regions of lower elevations. In the family Lamiaceae a number of species of *Scutellaria*, *Pogogyne*, *Stachys*, Lycopus, and *Mentha* grow in wet marshy ground. Several species of Salvia, which occur in the arid southwestern United States, are hard-leaved evergreen shrubs that are resistant to long dry spells.

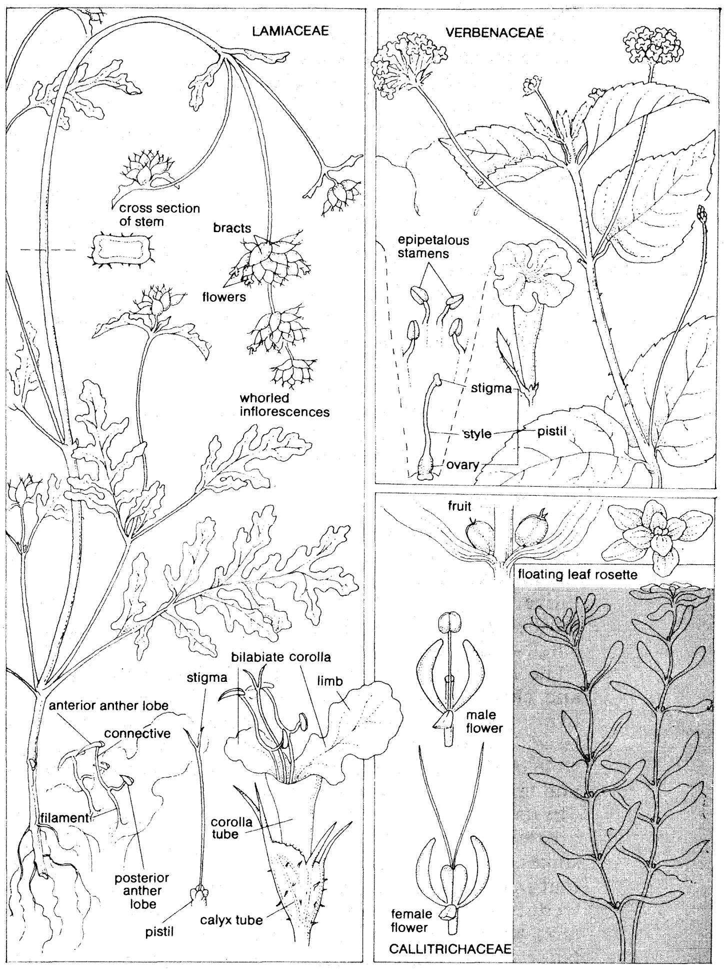
FORM AND FUNCTION

Vegetative characteristics. Members of the order Lamiales characteristically have opposite or whorled evergreen leaves. Square stems, popularly associated with the Lamiaceae, are lacking in some members of the family, particularly those with a shrubby habit. Square stems also occur in a few herbaceous members of the family Verbenaceae. The leaf blades of many of the Lamiaceae and the Verbenaceae and the single species of the Phrymaceae are entire—smooth margined—but in others the margins show all possible variations from smooth to variously toothed or cut; occasionally they are lobed. Compound leaves are uncommon. In some members of the family Lamiaceae, the stems, leaves, and inflorescences (flower clusters) are more or less hairy, sometimes with glandular hairs that secrete an aroma characteristic of the particular plant. In addition, there are frequently oil-secreting glands on the surfaces of the leaves and flowers. Except for the well-known lemon verbena (Aloysia *triphylla*; also called Lippia citriodora), scented foliage in the Verbenaceae is rare.

Aquatic plants

The family Callitrichaceae, well adapted to an aquatic or semi-aquatic habitat, is characterized by extreme reduction in its vegetative parts. The delicate and slender stem, which is either supported by water or lies prostrate, consists of a much-reduced vascular bundle and pith composed—in cross section—of only two or three cells.

Flower and fruit characteristics. The plants of the order Lamiales, with the exception of the family Callitrichaceae, have attractive, often brightly coloured flowers. They usually are not individually large but are generally clustered in various showy and often conspicuous inflorescences. These are usually racemes—flower clusters with a central axis and short side branches bearing flowers that mature from the bottom of the cluster upward—or whorls, and they have bracts (leaflike appendages below flowers) that are occasionally colourful or enlarged. The flowers are almost always bisexual (*i.e.*, structures of both sexes are found in one flower), with the basic parts, sepals and petals, five in number and more or less irregular or zygomorphic (bilaterally symmetrical). It is usually the corolla, the collection of petals, that is brightly coloured, mostly in shades of red or blue; white and yellow are rare. The calyx, or collection of sepals, is mostly tubular and five-parted or two-lipped at the apex and is usually green, smaller, and less conspicuous than the corolla. The corolla consists of a tube (the basal part) and a limb (the terminal showy part). The tube may be short or long in relation to the limb and is usually straight with parallel sides; it is rarely pouched. The limb shows variation in shape and size. It may be divided into five more or less irregular parts or into two unequal parts to form a two-lipped corolla. The five-parted corolla is more often seen in the family Verbenaceae, and the two-lipped one is typical of the Lamiaceae. The flowers in the family Phrymaceae have a tubular calyx and a two-lipped corolla. There may be four stamens (male pollen-producing structures) usually in two pairs of unequal length, sometimes two, or rarely five. They are attached to the inner side of the corolla tube and are contained within or sometimes extended beyond it. The ovary is superior; *i.e.*, it is located above the point where the sepals, petals, and stamens arise. It is composed of two carpels—simple ovaries or segments of a compound ovary—that usually appear to be four because of the division of each into two parts. Within each carpel are two ovules, seed precursors, which are usually erect but with the micropyles, tiny openings through which fertilization takes place, directed downward. The single style, the narrow upper part of the compound ovary or pistil, may or may not extend beyond the end of the corolla tube along with the stamens. The stigmatic surface, the pollen-receiving portion of the style tip, usually matures after the pollen is shed from the stamens of the same flower. The fruit in the family Lamiaceae usually consists of two to four hard, dry, one-seeded nutlets, each containing one seed. Sometimes, in the Verbenaceae, the fruit is a drupe, a stone-seeded fleshy fruit, which breaks up into one-seeded pyrenes, nutlet-like fruit segments.



Representative members of the order Lamiales.

Drawing by M. Pahl; from A. Enaler and K. Prantl, *Die Naturlichen Pflanzenfamilien, 2nd ed.*, vol. 19c (1931); Duncker & Humblot

Reduced flowers of Callitriche

The tiny flowers of Callitriche are lacking in showy parts—*i.e.*, sepals and petals—and are as much reduced as a flower can be and still function. They consist of only their essential organs, each flower having but one stamen and one pistil (female complex with stigma, style, and ovary). Each flower is solitary in a leaf axil—the upper angle between a leafstalk and the plant stem or branch—and is usually subtended by a pair of tiny, delicate bracts. Both staminate (male) and pistillate (female) flowers are on the same plant, so that the plants are monoecious, a condition not found elsewhere in the order Lamiales. The ovary of the female flower is two-celled and two-lobed; each cell is further divided into two parts as in the Lamiaceae and the Verbenaceae, but unlike them there are two separate styles. The small solitary fruits break up into four tiny one-seeded nutlets.

EVOLUTION

Fossil record. Fossil records show occasional occurrences of modern genera of both Lamiaceae and Verbenaceae. Perhaps the earliest record is that of *Vitex pentadactyla* (Verbenaceae) in the Cretaceous Period (65,000,000 to 136,000,000 years ago) of Bohemia. Nutlets of *Hancea* (Lamiaceae), a genus that now occurs in eastern Asia, have been found in the London clay flora of the Eocene Epoch (38,000,000 to 54,000,000 years ago). Records also have been found of a few present-day species. *Verbena officinalis*, now widespread in Europe, has been found in deposits of the Pliocene Epoch (2,500,000 to 7,000,000 years ago) of Belgium.

The fossil record, although small, indicates that some modern genera have been in existence since the Cretaceous and early Tertiary and suggests that during these times groups similar to those of the present day were being differentiated.

Phylogeny. The meagre fossil record is of no help in explaining relationships of the Lamiales to closely related families in other orders. Nor does it explain relationships among the families within the order. The characters most generally useful for establishing relationships among the flowering plants are in the flowers and fruits, and these seldom are found in the fossil record. As a result, the living members of these groups must provide the clues to relationships. In their floral structures, for example, the plants of the order Lamiales are considered to be among the most highly evolved groups of dicotyledons, the group of flowering plants with two "seed leaves" and net-veined leaves.

Relationship between mint and verbena families

The similarity between the families Lamiaceae and Verbenaceae has been evident for many years. The four-lobed ovary and gynobasic style—a style arising from near the base of the ovary—of the Lamiaceae separate it from the Verbenaceae in which the entire, rarely shallow, four-lobed ovary has a terminal style. Several genera of the Verbenaceae, however, have a nearly gynobasic style partially sunk between the four shallow lobes of the ovary. This and other intergrading characters make it difficult to separate all the members of one family from the other, but at the same time they emphasize the closeness of the relationship between the two families.

CLASSIFICATION

Distinguishing taxonomic features. Of the four families here comprising the Lamiales, the Callitrichaceae is the most easily distinguished by its greatly reduced vegetative and floral structure. The Phrymaceae is distinguished from the Lamiaceae and the Verbenaceae by its single-celled ovary.

The most useful characters for distinguishing the large and closely related Lamiaceae and Verbenaceae are the ovary and fruit structure, type of inflorescence, and, to a lesser extent, habit. The four-lobed ovary and gynobasic style of the Lamiaceae contrasts with the nonlobed ovary and terminal style of the Verbenaceae.

In its inflorescence types, the family Verbenaceae shows greater diversity than the Lamiaceae, with racemes, spikes, and compound cymes, complex-branched clusters that mature from the apex downward. The Lamiaceae infrequently has racemes and frequently has whorled clusters, which may be separated or crowded. In fact, the latter inflorescence occurs widely enough in the family to be considered characteristic of it.

In the Verbenaceae the woody habit predominates, but in the Lamiaceae the plants are mostly herbaceous and, less often, somewhat shrubby.

Annotated classification. The descriptions of the four families presented here are generalized to include most of their genera. In the large families there may be exceptions. Thus, the Lamiaceae is said to be bisexual, yet the family includes at least one genus, *Iboza*, with male and female flowers on separate plants and other genera with separate female flowers besides the bisexual ones. Exceptions to the characteristic two-lipped flowers are seen in Ajuga and *Teucrium* in which the upper lip of the corolla is lacking. In the Verbenaceae are several genera so unlike most of the family that from time to time they have been placed in separate families. One of these is *Avicennia*, which some botanists place in its own family, the Avicenniaceae, on the basis of its ovary structure.

ORDER LAMIALES

Herbs, shrubs, or trees; leaves nearly always opposite and simple; 5-parted flowers usually bisexual and with a gamopetalous (1-piece, of fused petals), irregular, or bilaterally symmetrical corolla; stamens 2, 4, or 5, arising from the inner surface of the tubular corolla; ovary superior (positioned above the attachment point of the calyx and corolla), usually with 2 carpels (ovule-bearing segments of the ovary) and 4 locules (chambers), ovules usually 1 in each locule; usually 1 style, forked at the apex, terminal on the ovary or arising from its base; fruit of 4 nutlets, a drupe that separates into 4 pyrenes, or rarely a capsule. Four families, about 260 genera, and over 6,000 species, worldwide distribution.

Family Callitrichaceae (water-starwort family)

Annual or perennial monoecious herbs with slender delicate stems; leaves opposite, entire; flowers solitary in leaf axils, subtended by two bracteoles (small leaflike appendages), unisexual, radially symmetrical; sepals and petals lacking; male flower consisting of 1 stamen with a 2-celled anther; female flower consisting of 1 pistil; ovary superior with 2 carpels and 4 locules, 1 ovule in each locule, pendulous, anatropous (reversed (the micropyle opening is close to the point of the funiculus, or stalk, attachment); styles 2, separate to the base; fruit composed of 4 nutlets, variously keeled or winged. One genus, Callitriche, with 25 or more aquatic or semi-aquatic species distributed on all continents.

Family Phrymaceae (lopseed family)

Perennial herbs; leaves opposite, entire; flowers solitary, in opposing axils, arranged in terminal elongated slender racemes, bilaterally symmetrical; calyx tubular, 2-lipped; corolla 2-lipped, lower lip 3-lobed, larger than upper lip; stamens 4 in 2 pairs, attached to corolla tube; ovary superior, 1-chambered with free central placentation; ovule 1, erect, orthotropous (straight—the micropyle opening is at the apex opposite the point of attachment); fruit a 1-seeded nutlet, enclosed in the persistent calyx. One genus and species (*Phryma* leptostachya), in damp woods of eastern North America and eastern Asia.

Family Verbenaceae (verbena family)

Herbs, shrubs, trees, or climbers; stems occasionally square in cross section; leaves usually opposite, sometimes whorled, mostly simple, rarely compound, without stipules; inflorescence variable, often large and showy; flowers irregular or bilaterally symmetrical, calyx mostly 5-lobed or 5-toothed, persistent; corolla of fused petals, irregularly 5-lobed or occasionally 2-lipped; stamens 4 in 2 pairs, rarely 5, attached to petals; ovary superior, sometimes shallowly 4-lobed, with 2 carpels, 4 locules, and ovule attachment along the central axis; 1 ovule in each locule, mostly erect, anatropous; 1 terminal style; stigma divided into 2 equal lobes; fruit generally a drupe separating into *2* to 4 pyrenes, or nutlets, or rarely (in *Avicennia*) a capsule; seed with a thin seed coat and straight embryo, lacking endosperm (nutritive tissue) except in *Avicennia*, *Stilbe*, and Chloanthes. About 100 genera and 2,600 species, predominantly in tropics and subtropics around the world.

Family Lamiaceae (mint family)

Mostly herbs, shrubs, or subshrubs, rarely trees or climbers; stems usually square in cross section; leaves opposite, simple, rarely pinnately compound (*i.e.*, with leaflets arranged on both sides of a central axis) or lobed, without stipules; flowers usually in a series of successive whorls, the whorls scattered or contiguous and spikelike, or flowers solitary at each axil; bracts usually present, sometimes showy and colourful; flowers bilaterally symmetrical or nearly so; calyx tubular, persistent, 5-lobed, often 2-lipped, usually with 5, 10, 13, or 15 conspicuous ribs; corolla 5-lobed, usually 2-lipped, typically the upper lip 2-lobed and erect and the lower 3-lobed, concave or spreading; stamens 2 or 4 in 2 pairs, attached to the petals, 1 pair usually longer, a staminode (sterile stamen) rarely present; anthers 2-celled, opening longitudinally, the connective tissues between the anther sacs sometimes well-developed and the anterior cell reduced or absent; nectar-secreting disk present at base of ovary; ovary superior, 4-lobed, with 2 carpels and 4 locules, ovule attachment basal, 4 ovules, anntropous, erect; style 1, forked at apex, usually arising from base of ovary; fruit typically of 4 nutlets, dry, rarely fleshy, separate or cohering in pairs, enclosed in persistent calyx. About 160 genera and 3,500 species distributed in temperate regions around the world.

Critical appraisal. The family Callitrichaceae with its tiny unisexual flowers may seem out of place in the Lamiales. Because of its flowers its relationship has been placed from time to time with other families that are largely unisexual and without flower petals, such as the Haloragaceae or the Euphorbiaceae. Recent embryological studies of the ovule, however, have shown features that are characteristic of the Verbenaceae, which tends to place the relationship of the Callitrichaceae with the Lamiales. Placing it in this order might still be questioned, however, since its well-developed endosperm and separate styles are not generally found in the Lamiales.

The families of the Lamiales were formerly placed in the large order Tubiflorae. This order contained about 20 families with bisexual flowers having tubular corollas. In their floral structures these families are among the highly evolved groups of dicotyledons.

BIBLIOGRAPHY. J. BRIQUET, "Verbenaceae," "Labiatae," and "Phrymaceae," in A. ENGLER and K. PRANTL (eds.), Die *Natürlichen Pflanzenfamilien*, vol. 4, pt. 3a, pp. 132–182, 183–375 and pt. 3b, pp. 361–362 (1897), a classic treatment in German (no comparable work in English) that describes the families, their vegetative and floral organs, fruit development, pollination, and geographical distribution; ARTHUR CRONQUIST, The Evolution and Classification of Flowering Plants (1968), a recent treatment, carefully worked out, that brings together new information and reference citations; K. FAEGRI and L. VAN DER PIJL, The Principles of Pollination Ecology (1966), a

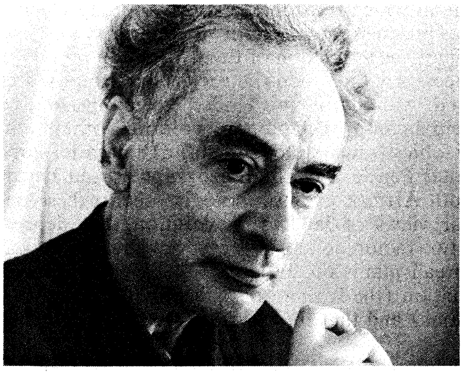
general discussion of pollination, the different kinds of pollination mechanisms in flowering plants, and pollinating agents (the chapter entitled "Case Histories" discusses, with examples, pollination within some of the Labiatae); HANS MELCHIOR in A. ENGLER, *Syllabus der Pflanzenfamilien*, 12th ed. vol. 2 (1964), an excellent summary treatment in German in which the families and their respective subfamilies, sections, genera, and geographic distributions are described; F. PAX, "Callitrichaceae," in ENGLER and PRANTI., *op. cit.*, vol. 3, pt. 5, pp. 120–123 (1897), a classic treatment in German of the family, with its single genus, *Callitriche*, that discusses vegetative and floral organs, pollination, and geographical distribution; ALFRED B. RENDLE, *The Classification of Flowering Plants*, 2nd ed., vol. 2, *Dicotyledons* (1959), brief family descriptions; J.C.T. UPHOF, *Dictionary of Economic Plants*, 2nd ed. rev. (1968), a standard reference of useful plants that includes genera of Verbenaceae and Labiatae used for food, shelter, and medicine; J.C. WILLIS, *A Dictionary of the Flowering Plants and Ferns*, 7th ed. rev. by H.K. AIRY SHAW (1966), a standard encyclopaedic reference in which all the families and genera of flowering plants, including those of the Lamiales, are described together with their geographical occurrence and numbers of species.

(E.McC.)

Landau, Lev Davidovich

Lev Davidovich Landau, one of the greatest theoretical physicists of the 20th century, made an outstanding contribution to the development of that subject in the U.S.S.R.

USSR Magazine—Sovfoto



Landau.

Education

Landau was born on January 22, 1908, in Baku, Azerbaijan, Russia. His father was an engineer working in the Baku oil industry and his mother a doctor who had at one time done physiological research. Landau graduated at 13 from the gymnasium and, because he was too young to go to the university, attended the Baku Economical Technical School. He matriculated in 1922 at Baku University, studying physics and chemistry, and transferred in 1924 to the Leningrad State University, which at that time was the centre of Soviet physics. Graduating in 1927, he continued research at the Leningrad Physico-Technical Institute. At that time there were practically no outstanding senior theoretical physicists in the Soviet Union, and since the younger men had to teach themselves and each other, it was important for them to go abroad and be in touch with the Western theoretical physics schools that were flourishing in such centres as Copenhagen and Munich. Landau got his first chance to go abroad in 1929, on a Soviet government travelling fellowship supplemented by a Rockefeller Fellowship. After brief stays in Gottingen and Leipzig, he went to Copenhagen to work in Niels Bohr's Institute for Theoretical Physics. It is probably no exaggeration to say that the development of present-day theoretical physics owes more to Bohr's Institute than to any other place in the world. Almost all of the leading theoretical physicists of the 1920s and 1930s spent some period at this institute. Landau always considered himself a pupil of Bohr's, and his attitude to physics was greatly influenced by Bohr's example. After his stay in Copenhagen he visited Cambridge and Ziirich before returning to the U.S.S.R. Apart from short visits to Copenhagen in 1933 and 1934, Landau spent the remainder of his life in his own country.

In 1932 Landau went to Kharkov to become the head of the Theoretical Division of the Ukrainian Physico-Technical Institute, which position he combined in 1935 with that of head of the Department of General Physics at the Kharkov A.M. Gorky State University. In Kharkov Landau began to build a Soviet school of theoretical physics, so that Kharkov soon became the centre of theoretical physics in the U.S.S.R. It was also in Kharkov that, with his friend and former student, E.M. Lifshits, he started to write the well-known *Course of Theoretical Physics*, a set of nine volumes that together span the whole of the subject. His great interest in the teaching of physics is also shown in his plans for a "Course of General Physics" and even a series "Physics for Everybody."

Landau required that his students master all necessary mathematical techniques before coming to him. After that he expected them to master the so-called theoretical minimum, which included a basic knowledge of all the domains of theoretical physics. Only the ablest of the students were able to pass this minimum. In this way his students became proper physicists, rather than narrow specialists.

Work on liquid helium

In 1937 Pyotr Leonidovich Kapitsa, a low-temperature experimentalist, persuaded Landau to move to Moscow and to head the Theory Division of the S.I. Vavilov Institute of Physical Problems, which had been created by the U.S.S.R. Academy of Sciences. There, Landau's close interest in experimental physics led to his development of the theory of liquid helium, the last of the elements to be liquefied and the most remarkable of all liquids. Kapitsa had found that liquid helium was superfluid—that is, that it had less resistance against moving through a tube than any other known liquid. Landau's theory to explain this peculiar behaviour was the work for which he was awarded the Nobel Prize for Physics for 1962.

Landau's attitude to physics and physicists was critical; he did not suffer fools gladly. While always willing to help anybody, he hated pomposity. People either adored him or were his bitter enemies, and he was imprisoned during the Stalin era, in 1938, and only a personal intervention by Kapitsa freed him.

In 1937 Landau married K.T. Drobanzeva, and in 1946 they had a son, Igor, who became an experimental physicist.

In Moscow Landau continued to make significant contributions to almost all parts of physics. The topics he covered range from low-temperature to nuclear physics, from the theory of metals to stellar energy, from cosmic rays to plasmas, from hydrodynamics to atomic physics. Landau's contributions are partly reflected in such terms as Landau diamagnetism and Landau levels in solid-state physics, Landau damping in plasma physics, the Landau energy spectrum in low-temperature physics, or Landau cuts in high-energy physics.

On January 7, 1962, Landau was involved in a car accident. He was unconscious for six weeks and was several times declared clinically dead, but he somehow revived. Distinguished specialists from several countries helped to save his life. After Landau had regained consciousness his faculties slowly returned to him, but he was no longer able to perform creative work. His physical condition never returned to normal, and he died April 1, 1968.

Apart from the Nobel Prize, Landau received many other honours. In the U.S.S.R. he was directly elected a member of the Academy of Sciences, was given the title of Hero of Socialist Effort, and was awarded three State Prizes, as well as a Lenin Prize. He was a foreign member of the Royal Society of London and of the academies of The Netherlands, Denmark, and the United States, as well as a recipient of the Max Planck Medal and the Fritz London Prize.

BIBLIOGRAPHY. A popular account of Landau's life, and especially of his last six years after his accident, may be found in A. DOROZYNSKI, *The Man They Wouldn't Let Die* (1966). His scientific papers were published in D. TER HAAR (ed.), *Collected Papers of L.D. Landau* (1965). *All* his books have been published in English translation.

(D.t.H.)

Landform Evolution

Landform evolution is a descriptive term that may be applied to all changes in the earth's surface features throughout time. In the broad sense, such changes are a consequence of the interaction of two basic processes: relative uplift on the earth's surface and downwearing by weathering and erosion. The eternal struggle between these opposing processes is reflected at any particular point in time by the landforms, or scenery, that appear on earth. The casual visitor to Victoria Falls, the Alps, Lake Baikal, or elsewhere is, in effect, a witness to this contest, for no landform or feature on the earth's surface is unchanging. At some point in the geological past it did not exist, and at some future date it will have vanished once again.

The transient character of landforms is both inescapable and obvious when time spans of great duration are involved. But when the focus is on a somewhat shorter time period, and when landform evolution is examined more closely, a number of difficult questions arise. In a general way, it might be asked how a given landform originated, how it changed in the past, and how it will change in the future. In more precise terms, however, the inquiry really centres on whether the processes of uplift or downwearing dominate at a given point in time and throughout time, on the precise nature of the interaction between these processes, and on the effect of this interaction on the configuration of landforms at the earth's surface. The latter question involves a host of physical, chemical, and biological factors and processes, many of which are imperfectly understood. In one sense it is adding the query **why** to the previously asked *how*; in another, it is asking the ultimate question, namely, whether landform evolution accords with some natural law that causes change to occur in a predictable manner.

Early theories of landform evolution

These questions have long been of interest to man and many views have been set forth. The Greek authors Herodotus and Strabo and the Persian Avicenna perceived correctly that the landforms about them were acted upon and molded by physical processes. Cogent arguments along these lines were provided in the 15th and 16th centuries by Leonardo da Vinci, Agricola, and others who recognized that stream erosion and fluvial processes were responsible for the creation of mountain valleys and the transportation of sediment.

This promising beginning toward understanding was halted abruptly by theological constraints on the time available for the occurrence of landform evolution. Archbishop Ussher of Ireland, for example, concluded from biblical studies in 1654 that earth and man had been created in 4004 BC, on October 26, at 9:00 A.M. His pronouncement, accepted literally, forced the conclusion that all aspects of the earth's surface features must have developed in less than 6,000 years. There thus arose a catastrophic school of thought, so called because its adherents interpreted the evolution of all landforms in terms of catastrophic events that presumably had occurred during extremely brief periods of time.

Abraham Werner (1750–1817), German mineralogist, was one of the chief proponents of this view. During a long tenure in Germany he produced an entire generation of student disciples who went so far as to champion the notion that all rocks of the earth's crust had been rapidly precipitated from a single "universal sea." At that time there was no knowledge of the several advances and retreats of glacial ice during the Pleistocene Epoch, and the enormity of geological time was still unsuspected. Combined with the theological constraints imposed on the earth's age, it is not surprising that the valleys of the Swiss Alps and the glacially derived debris on their floors commonly were explained by recourse to the advance and retreat of the biblical flood. Georges Cuvier (1769–1832) the French paleontologist, gave the same explanation for disrupted strata and unconformities in sedimentary deposits in France. As recently as the early part of the 19th century, in fact, there was conflict between supporters of a relatively modern view of nature and those who wished to interpret landforms in terms of the universal deluge that had raised the Ark.

Present theories of landform evolution

Controversy about the nature of landform evolution has persisted to the present day. There are, presumably, no advocates of Ussher's chronology or of a universal rush of waters. Moreover, since quantitative data for the retreat of the chalk cliffs of England, the canyon walls along the Colorado River, the escarpment at Niagara Falls, and similar features, became available, none has doubted that landforms tend to change gradually, at observable rates in most instances. With the replacement of the theory of catastrophism by that of gradualism, several distinct concepts of landform evolution have emerged. They are uniformitarianism, the cycle of erosion, regional morphogenesis, dynamic equilibrium, and entropy-probability. Briefly, these ideas may be described as follows.

Uniformitarianism. According to uniformitarianism, the processes operative on the earth's surface today are similar in kind to those that were operative in the past. Hence, if runoff from a storm is seen to produce gullies in hillslopes today, it is argued that knowledge of this modern hillslope erosion, and of the transportation and deposition of the removed sediment, can be extrapolated in time to interpret hillslopes and sedimentary accumulations of the past.

Cycle of erosion. The cycle of erosion, also called the geographic cycle by some, is a quasi-anthropomorphic explanation of landform evolution set forth by William Morris Davis (1850–1934), the American geographer, in the latter part of the 19th century and the early part of the 20th century. Its proponents assumed that landforms traverse a cycle in which hillslopes, valleys, mountains, and river systems progress from "youth" to "maturity" to "old age," principally as a function of time, although rock type and the processes acting upon the rocks were supposed to be considered. Although Davis tended to emphasize time as the principal causative factor in evolution, others emphasized tectonics (crustal movements) and parallel slope retreat. Walter Penck, a German geographer, and Lester King, a South African geomorphologist, set forth somewhat similar views of landform evolution that depend upon these two factors. King, in fact, drew the conclusion that widespread plains on the continents were produced by pediplanation (the formation of pediments under humid conditions) and that they are correlative surfaces of denudation.

Morphogeneric regions. According to the concept of regional morphogenesis, similarities of landforms are thought to result primarily from similarities of climatic regimes in the regions involved. Thus, in a gross way, the angularity of desert landscapes and the gentler expression of temperate and humid landscapes are taken as examples of morphogenetic regions. Considered in detail, this concept, like that of the cycle of erosion, falters because it requires too much dependence on a single factor, or single group of factors, and the denigration of others of equal importance.

Interestingly, the predominant Russian view of landform evolution tends to parallel the regional morphogenetic system, though instead of emphasizing climate Soviet workers emphasize tectonics and speak of "morphostructural units" as a basic conceptual framework. Simply stated, this means that a relationship between form and uplift is thought to exist, such that the genesis or history of a given landform or region can be deduced from its geographic aspect.

Dynamic equilibrium. The idea of dynamic equilibrium can best be explained in terms of high school chemistry: If reversible reactions of substances A and B yield substances C and D and if any of these four substances is added after equilibrium has been attained, there will then occur a reaction in a direction that will tend to minimize the changed conditions, and a new balance point or state of equilibrium will be established. Analagous examples from the elementary laws of motion in physics and from other disciplines can be given. In terms of landform evolution, the conception of dynamic equilibrium means that a balance tends to be established between the resistance of the rocks on the one hand and the processes of downwearing and erosion on the other. This "balance" may be expressed as a particular hillslope or stream bed profile.

Concept of balance between rock resistance and erosion

Any change in conditions, such as renewed uplift or climatic change, will, like the addition of one of the substances in the chemical example, cause a concomitant change in the hillslope or stream bed profile toward a new state of equilibrium. Because landform systems are not static, equilibrium is called dynamic to indicate its constant change or fluctuation about a mean position that is termed the steady state. An important consequence of the equilibrium concept is that landforms in equilibrium must be considered timeless.

The idea of dynamic equilibrium is in direct contrast to that of the erosion cycle, which is the notion that the history of a landform is revealed by its stage or aspect within some fixed development sequence. Dynamic equilibrium is, in fact, in contrast to nearly all other views of landform evolution because they set forth *chronologies* of denudation regardless of the procedural element (for example, tectonics, climate) that is being emphasized. Thus, landforms are time bound within other frameworks of thought.

Entropy-probability. The last concept to be considered, entropy-probability, is not applied to landforms solely in terms of its inverse thermodynamic relation to free energy but, rather, in terms of probabilities. Entropy in landform evolution is related to the most probable distribution of energy and to the most probable configurations of stream profiles, drainage networks, and the like. It has been found, for example, that the plans and profiles of river systems and the distribution of branches and other aspects of drainage networks accord well with comparable features that have been deduced theoretically or that have been generated by random-walk statistical procedures. The body of knowledge that has accrued in this area suggests that hillslopes, river systems, and other features of the surface of the earth generally accord with the most probable configuration of their class or type, unless physically constrained in some manner. This most probable configuration represents both the most probable distribution of energy within a given system and the most probable point of fluctuation of dynamic equilibrium. The entropy-probability concept is in one sense, therefore, an extension of that of dynamic equilibrium. The additional notion of randomness in nature is important, however, because it suggests that certain aspects of landform evolution may be indeterminate. If drainage networks evolve according to the laws of chance, for example, and if their aspect in nature reflects only a distribution about the most probable network configuration, then it is clear that although a great deal can be said about networks as a class, treatment of a single, specific drainage network rests upon ground of far lesser strength.

Because landform evolution is a subject that is intimately associated with the concepts outlined above, this article will treat these concepts in detail, and classes or kinds of specific landforms only incidentally. The interested reader should refer to such articles as COASTAL FEATURES; DESERTS; PEDIMENTS; SAND SHEETS AND SAND DUNES; JUNGLES AND RAIN FORESTS; GLACIATION, LANDFORMS PRODUCED BY; and RIVER DELTAS for treatment of landforms as such. For more detailed treatment of some of the components of landform systems that are germane here see HILLSLOPES; SOILS; SEDIMENT YIELD OF DRAINAGE SYSTEMS; ROCKS, PHYSICAL PROPERTIES OF; RIVERS AND RIVER SYSTEMS; FLUVIAL PROCESSES: WEATHERING; CLIMATE; CLIMATIC CHANGE; and PHYSIOGRAPHIC EFFECTS OF TECTONISM. For further information about geomorphology, the discipline concerned with landform evolution, see GEOLOGICAL SCIENCES and EARTH SCIENCES; for the relevance of the topic to aspects of earth history, see EARTH, GEOLOGICAL HISTORY OF; and for information on the geological time periods of greatest pertinence to this article, see PLEISTOCENE EPOCH and HOLOCENE EPOCH.

UNIFORMITARIANISM

The concept of uniformitarianism is often equated to the phrase "the present is the key to the past." The allusion made is to present and past processes that are operative upon and within the earth. James Hutton (1726–97), a Scottish naturalist, was the first to recognize this truth; and his views were expanded by his friend John Playfair (1748–1819), a mathematician at Edinburgh, and ensconced in geological thought for all time by Sir Charles Lyell, the English geologist, in his *Principles of Geology*, published between 1830 and 1832.

Methodological and substantive uniformitarianism

Uniformity in nature is undoubted, and it is a commonplace that rocks weather today as they did in the past, that coasts are beaten by waves and tides as in the past, and that the dunes that move in response to wind action today must also have migrated under this impetus in the past. These are the dictates of common sense, and their denial would appear to be akin to repeal of the fundamental laws of nature, but the concept of uniformitarianism has not been viewed with unity by all observers. The rub involves the question of whether past processes are similar to modern processes when magnitudes, intensities, and rates are considered. In the light of modern data, most authorities today would probably agree that uniformitarianism is a dual concept: substantive uniformitarianism, representative of the original view, holds that present and past processes are similar in kind and rate, whereas methodological uniformitarianism asserts that although process rates may have varied throughout time, the concept of uniformitarianism properly pertains only to the principle of invariance of natural laws in space and time. The truth of this proposition is manifest from plotted graphs of mountain building, petroleum and coal reserves, extinctions of life forms, extent of glacial ice, and many other variables against time. In each case the frequency of occurrence of processes, or their magnitude, or both, fails to accord with a uniform distribution throughout time. This demonstrates that although present and past processes are similar in kind, process rates must have been variable.

This is a fact of considerable importance with respect to landform evolution. Central to the controversy between the adherents of Davis' anthropomorphic theory and proponents of dynamic equilibrium, for example, is the question of whether a given landform reflects the historical imprint of time; that is, whether the effects of former processes are discernible, or whether the landform is fully adjusted to present processes. Though the main question will be discussed later, some aspects of methodological uniformitarianism are worthy of mention at this point.

Earth-Moon relationships

Earth-Moon relationships serve as an excellent example because the laws of motion that apply to these bodies are representative of invariant natural laws. Despite this fact, their relationship through time has been such that certain process rates have changed.

Both observational and theoretical data, for example, indicate that the Moon has accelerated in its orbit through time. Because the angular momentum of the two-boldy system must be conserved, the increase in the angular velocity of the Moon has to he associated with an increase of the Earth-Moon separation distance and a decrease in the rate of rotation of the Earth. A number of models can be postulated to account for this situation; but these are only speculative because neither the initial separation distance nor the initial time involved is known. The physical limitation on this point, regardless of the origin of the Moon, is the Roche limit, named for its discoverer Edouard Roche in 1850, which posits a minimum possible separation distance for the Earth and its satellite at 2.89 Earth radii. Closer proximity would result in destruction of the Moon.

Within the context of methodological uniformitarianism two points, above others, need to be stressed. First, the tide-generating potential on Earth is proportional to M/R^3, in which M is the mass of the Moon and R the Earth-Moon separation distance. This means that, aside from the host of complexities that enter tidal computations, it is beyond dispute that greater tides were experienced on Earth in the geological past because the separation distance was less than it is today. This may be related to such problems as the uneven distribution of blanket sands (sand-size sediments of great areal extent), cyclothems (rhythmic and alternating strata that include sandstones, limestones, and coal beds), or other features in the geological record, but the basic point is that past tidal processes must have differed in magnitude and frequency from those of today. Second, because the rate of rotation of the Earth has decreased, so fundamental a quantity as the length of day has also varied through time.

Based on a calculation of tidal dissipation in shallow seas and in the form of bodily or earth tides, it has been proposed that the length of the day on earth has increased by 1.8 milliseconds per century for about 1.5×10^9 years. Corroboration from a fascinating and independent source was provided in 1963 by the discovery of banding in Devonian (345 to 395,000,000 years ago) corals that represent increments of growth. A simple counting of fine bands within the annual bands provides the number of days per year. Counts from several specimens yielded an average of approximately 400 days per year, in middle Devonian time. These results differ from tidal-dissipation calculations of the increase in the length of day by only 0.7×10^{-10} sec per day (see also DATING, RELATIVE AND ABSOLUTE).

Changes in the length of day and the Earth's rotation rate through time bear upon a considerable number of processes and factors of interest. These include the duration and frequency of incoming solar radiation at a given point on Earth, oceanic current systems, and a wide variety of atmospheric phenomena that affect the weather and climate.

Effects on climate and erosion rates

The climate, in turn, is of considerable importance because it has both indirect and direct influence on the erosion of the earth's surface. Indirectly, climate tends to control the type and density of vegetation as well as the development of soils. The presence or absence of soils and vegetation will in turn cause considerable variation of the sediment yield from given areas, under identical precipitation conditions. The relation between annual sediment yield and effective precipitation that was developed in 1958 is still generally valid. Basically, the relationship shows that maximum yields are correlative with semi-arid conditions. As precipitation decreases, the sediment yield diminishes drastically because the lack of sufficient runoff under increasingly arid conditions compensates for the absence of vegetation. As precipitation increases above semi-arid levels, sediment yield again diminishes. In this instance, however, it is because greater abundance of vegetation inhibits erosion despite the greater amounts of runoff. When effective precipitation attains a value of 1,000 millimetres (about 40 inches) per year, the curve of relation flattens; higher annual values do not produce greater sediment yields.

With respect to the role of vegetation in inhibiting erosion rates, it is necessary to bear in mind that the earth's present plant cover has not always existed. Plants as ground cover were totally absent during Precambrian and early Paleozoic time, and on these grounds alone it is clear that erosion rates must have varied through geologic time. Present rates have been computed by a number of authorities, but the values vary because bed-load transport (coarse material in transit along the bed of a stream), for example, is unmeasured in most streams and because few data on the dissolved loads of rivers are available. The computation involves converting the annual tonnage of suspended sediment in streams into denudation rates, which are expressed as rates of uniform lowering of the land surface over large areas. The present rate for the continental United States has been given as 2.5–7.5 centimetres per 1,000 years, which is considered by many to be abnormally high when geological time is considered.

Magnitude and frequency of geomorphic processes

It is well established that climate has changed through geologic time and that, in a general way, this can be linked to changes in rates of denudation. The precise relationship between the magnitude and frequency of natural geomorphic processes, such as fluvial runoff, and the rate of accomplishment of work (*e.g.*, erosion) is still elusive. It has been argued that it is the natural event of moderate magnitude and relatively frequent occurrence that accomplishes most of the work on the surface of the earth. It must be noted, however, that this view applies best to depositional landforms in temperate regions.

The retreat of cliffs, for example, tends to be a process characterized by great magnitude (*e.g.*, landslides, avalanches) and infrequent occurrence; it can be termed a step-function process for this reason. The same is true of stream-channel erosion in many situations; the largest boulders will be transported only by extreme events (*e.g.*, the thousand-year flood). An examination of Coffee Creek, California, following such an extreme event, for example, has suggested that the role of infrequent events of large magnitude deserves renewed consideration in landform evolution. And an analysis of high mountain streams in the Sangre de Cristo Mountains of New Mexico by J.P. Miller (1923–62), the American geomorphologist, led him to conclude that a certain fraction of the bed material present today may have moved frequently only under Pleistocene runoff conditions that would have been extreme in terms of modern phenomena. Runoff events under arid conditions clearly belong to this same category. As the mean annual precipitation falls below 250 mm (about ten inches) per year, the variability of its distribution increases markedly, and the relatively infrequent flash flood becomes the most significant event in terms of work performed.

Thus, with respect to the uniformitarian view, the present is the key to the past; but at every hand there is abundant evidence of diverse nature which indicates that the magnitudes, intensities, and rates of processes have varied throughout time. The fact that the kinds of operative processes have been similar throughout time, however, permits rational examination of the relative merits of the several concepts of landform evolution. And the fact that process rates have varied throughout time provides an avenue for investigation of the extent to which the imprint of time is present on a given landform and under a given climatic regime.

CYCLE OF EROSION

The concept that landforms are subjected to a cycle of erosion, also called by some the geographic or geomorphic cycle, was set forth in more than 500 published works of William Morris Davis between 1884 and 1934. Particularly important among these works was a paper on the rivers and valleys of Pennsylvania, the features that he interpreted in terms of the concept. Following its publication in 1889, an entire generation of geographers and geomorphologists devoted the bulk of their productive years to the expansion, elaboration, or further illustration of the cycle of erosion. Despite the advent of quantitative methods in the 1950s and of alternative general syntheses of landform evolution, the views expressed by Davis are still accepted by some geologists as a correct conceptual framework of reference. Clearly they must, therefore, be considered in any general discussion of landforms and their change through time.

The Davisian concept of stage and age:

Davis believed that landforms evolve from an initial state to an ultimate state, and that this is accomplished by progressing through definite and characteristic stages after the initial regional uplift has occurred. The initial state may be represented by a relatively flat plain, perhaps one that has emerged from the sea. Uplift transforms such an area into fold mountains or block mountains, leading to the stage which Davis termed youth. Next, dissection of these massifs by fluvial processes causes the establishment of drainage systems and maximum relief during the stage termed maturity. Ultimately, and in Davis' view, inexorably, the streams then continue to transport debris to the sea until the entire region is reduced to sea level, or nearly so. This is the stage of "old age"; the surface is termed a peneplain, and the anthropomorphic cycle of erosion, which was to some extent inspired by debate about organic evolution in the 19th century, has run its course.

This easily grasped deductive scheme is depicted in Figure 1. The last stage shown, namely rejuvenation, was thought to occur when uplift again raised the peneplain of old age. Davis argued that the streams of such a region would cut downward with renewed vigour, essentially starting a second cycle of erosion. His interpretation of the folded mountain region of the Appalachians in terms of multiple erosion cycles was hailed as a triumph of geomorphic analysis at that time. Because it is still praised today, its possible flaws are worthy of some analysis.

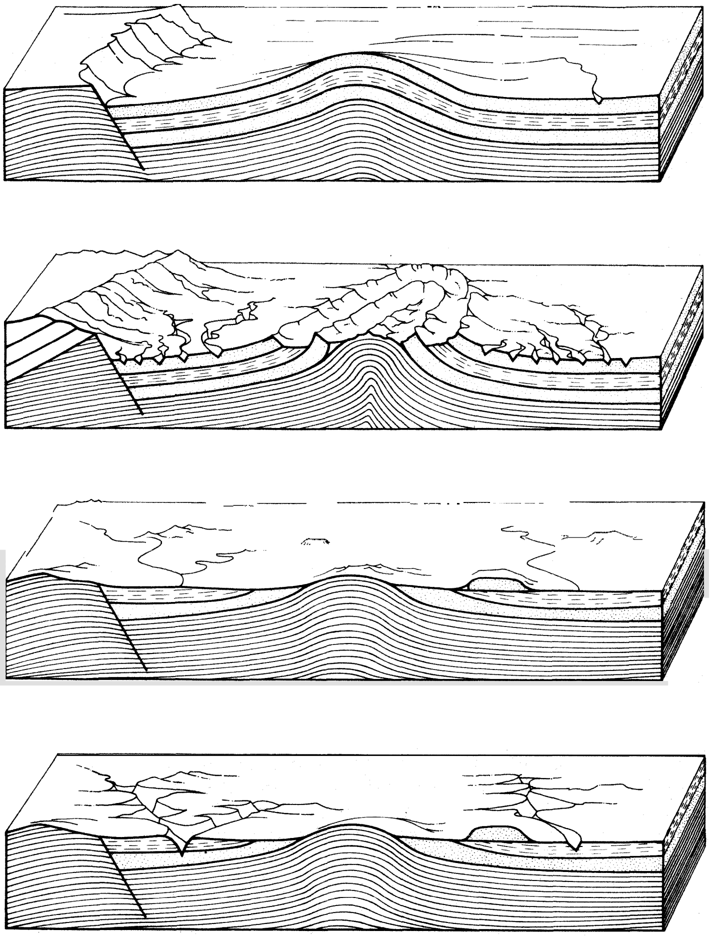


Figure 1: Stages in geographic cycle of a humid region: (from top to bottom) youth, maturity, old age, and rejuvenation.

From *Geomorphology* by A. Lobeck, © 1939 by McGraw-Hill Book Company. Used with permission of McGraw-Hill Book Company.

First, it should be noted that the stage names of the stages adopted by Davis lack the degree of precision that is generally required of a definition. Second, the invocation and widespread employment of the stage names is symptomatic of the fact that among the several important factors in landform evolution, time was vastly overemphasized. Davis was fond of using the notation

$$\text{landforms} = f\ (\text{structure} + \text{process} + \text{stage}).$$

He did so for pedagogic reasons and although this pseudo-quantitative expression is quite correct (landforms are, in general, a function of rock type, the processes that are operative upon the rocks, and time), Davis basically ignored structure (rock type) and processes. He emphasized stage, or time, and because of this literally thousands of papers have been devoted to the "matching" of landscapes from all parts of the world to block diagrams such as those depicted in Figure 1. Once a "stage" within the cycle of erosion was established, the history of the region could be written directly by using block diagrams.

Drainage systems and peneplains

Aside from these points, there is considerable difficulty involving drainage systems and peneplains. Davis thought that like the landforms, and, in fact, with them the streams of a region progressed, through an evolutionary sequence of stages. Wildly dashing rivers in steep mountain areas were "youthful"; meandering rivers that flowed across gently sloping plains, on the other hand, were "sluggish," "tired," and representative of the final stages of old age. This remarkable view flies in the face of present knowledge of the velocity of flow and stream discharge near the mouth of the Congo, the Amazon, the Mississippi, or any other great river of the world. The velocity of flow of most rivers tends to remain the same or to increase slightly in a downstream direction; discharge will increase greatly between headwaters and mouth. Meandering bears no relation to time, and it is ludicrous to discuss the "age" of the average river. Moreover, the assignment of stage names to entire river systems was basically incorrect on other grounds. Even in Davis' terms it can be said that along its course, between the divide area and the river mouth, any single river system in any large region exhibits most of the supposed attributes of each of the successive stages.

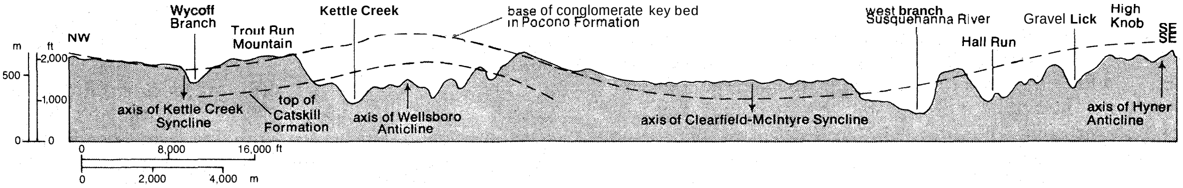
With respect to peneplains, it can be argued that, like river systems, no large area of perhaps subcontinental dimensions can truly represent a single stage or point in time It seems likely that the interior portion of anv such surface will be somewhat higher in-elevation than the peripheral parts. This alone might ultimately lead to an interpretation of the occurrence of two cycles of erosion, in the classical or Davisian view. Also, the degree of departure from a planar surface that can be tolerated or expected has not been specified; relativelv flat surfaces can be attributed to causes other than erosion cycles; and unequivocable examples of undissected "peneplains" do not readily come to mind. Finally, and most important, the evidence of peneplanation that was offered by Davis in support of his Appalachian interpretation consisted principally of: (1) the statement that accordant summit elevations occurred in areas of discordant (dissimilar) structure (rock type) and (2) that the courses of certain streams that flowed across the northeastward-trending regional structure could best be explained in terms of superposition from a pre-existing peneplain.

Inter-pretation of Appala-chian history

These two points are most instructive, because the search for supporting evidence has led to the contradictory view that the principle of dynamic equilibrium best explains the landscape in Davis' classical study area. This will be discussed later, but the general thrust may be indicated here. First, the advent of modern mapping and quantitative observation have shown that the "accordant summits" were, in part, a mirage; a plane cannot be well fitted to the ridge-top remnants. In addition, the work in the Appalachian area suggests that the "restored" peneplain would necessarily be as irregular as is the present landscape. In other words, it would seem that ridges and valleys were in existence in this region at the time when the erosion cycle concept dictates the presence of a nearly featureless plain. The close relationship of topography and structure is shown in Figure 2.

Still more telling, however, is the argument that even if accordant summit levels on hills of different rock type did exist today, these hilltops would not represent erosional remnants of a former peneplain but, rather, remnants of hills of different age. This follows directly from knowledge that different rock types weather at different rates: if a hill of sandstone and another of shale exist side by side today and exhibit the same summit elevation, then their original elevations (or time of landform formation) must have been quite different. This is clearly borne out by studies of the relationship between altitudes and rock types in Connecticut (Figure 3).

BY courtesv of the U.S. Geological Survey


Wycoff Branch
Kettle Creek
Trout Run Mountain
base of conglomerate key bed in Pocono Formation
west branch Susquehanna River
Gravel Lick
High Knob
NW
SE
SE
Hall Run
m
ft
2,000
500
1,000
0
0
axis of Kettle Creek Syncline
top of Catskill Formation
axis of Wellsboro Anticline
axis of Clearfield-McIntyre Syncline
axis of Hyner Anticline
0
8,000
16,000 ft
0
2,000
4,000 m

e 2: Relation of upland surfaces to rock structure shown [illegible] cross section across Renovo and Hammersley Fork quadrangles Potter [illegible]

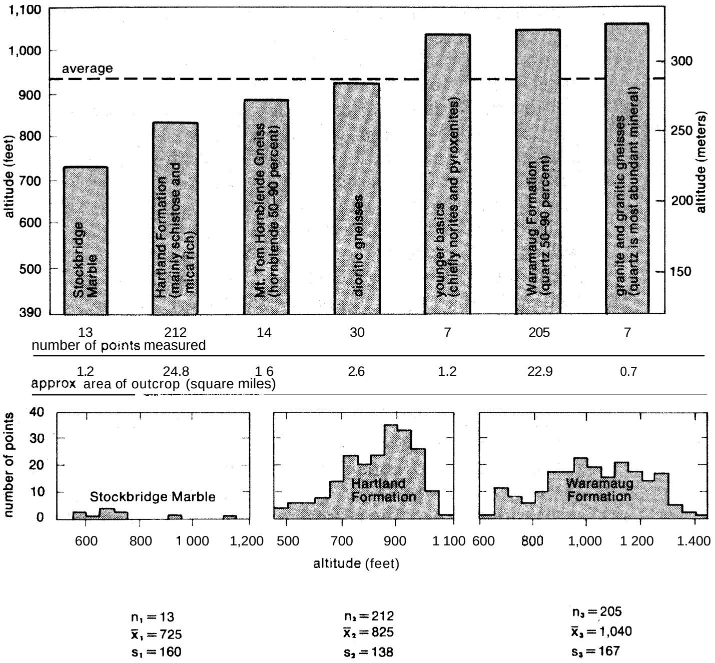


Figure 3: **(Top)** Mean altitudes of rock units in New Preston, Connecticut, quadrangle. (Bottom) Frequency distribution for three rock units.

From R. Flint, *Journal* of *Geology*, © 1963 by The University of Chicago Press

The course of streams in the Appalachian area has also been explained without recourse to the cycle of erosion concept. Available evidence strongly suggests that most river-course attributes are a function of differences in the resistance of existing rock types. There is no need to postulate the lowering of rivers by thousands of feet from hypothetical planar surfaces. A map of part of the Shenandoah Valley (Figure **8**) clearly shows, for example, the accentuated meanders of the North Fork where it flows over the Martinsburg Shale. The structure of this rock formation is such that the path of least resistance is to the northwest, and for this reason extreme sinuosity of river path exists. Hence on a larger scale it may be reasonable to suppose that stream trends across the major ridges can be accounted for by such preferred orientation or structural controls.

The cycle of erosion was challenged during Davis' lifetime, notably by Walter Penck, who believed that landform configuration was a function of the balance, or lack of balance, between rates of uplift and rates of denudation through time. In this view, greater rates of uplift would yield convex hillslopes, whereas greater rates of denudation would produce concave slopes. Straight slopes were said to reflect equality of the two rates. Most authorities today believe that Penck was more nearly correct than Davis, with respect to hillslope aspects, but that he too based his argument on deductive reasoning in the face of a dearth of data. Also, he ignored the effects of rock type and climate in his evolutionary scheme.

Open versus closed systems

In summary, the cycle of erosion as the first great synthesis of landform evolution theory, was widely accepted as the explanation of landforms and their origins in the past. Its chief virtue was its simplicity; once the inherent assumptions were accepted, any region of the earth's surface could be compared to some suitable block diagram and the history of the region set forth straightaway. The chief question that has arisen in the minds of modern geologists is whether this history is correct. Although dispute on the point continues, the dominant view today is that landform evolution should be considered within the framework of general systems theory. Within this framework, the Davisian cycle of erosion reflects a closed system in which the initial conditions of the system suffice to determine the final conditions and end products. The factor of overriding importance in the erosion cycle scheme is initial uplift; once this occurs, the outcome or ultimate fate of the region is fixed. It must become a peneplain after intermediate stages are traversed; rock type and processes are largely irrelevant. No allowance is truly made for the possible addition and subtraction of materials and energy, which is characteristic of an open system, the other type of general system. Likewise, the concept of attainment of equilibrium, after the initial state but prior to the ultimate state, was not envisaged by Davis and is not permitted in a closed-system framework of reference. Each of these conclusions seems erroneous in the light of modern knowledge.

Before turning to the evidence that appears to support dynamic equilibrium and the view that landforms are best considered with reference to open systems, some views on the role of climate in landform evolution must be considered. Like the time factor emphasized by Davis, climate is of undoubted significance—but it is not of overriding significance.

MORPHOGENETIC REGIONS

Climatic factor in landform evolution

The role of climate in landform evolution was acknowledged only indirectly by Davis and his many adherents. In their view, the cycle of erosion, or some modification of it, explained landforms everywhere; but largely because Davis' initial investigations were in Pennsylvania and New England, the temperate-humid regions were taken as "normal," and, in fact, the impress of glaciation or aridity was termed a climatic accident by C.A. Cotton, the New Zealand geographer, years ago, and only recently, A.L. Bloom, a coastal specialist, has discussed the ways in which such accidents alter the orderly rate of progression of the erosion cycle. From the meteorological vantage, of course, the term accident is most unfortunate. In general it is argued by all proponents of the concept that different landforms result from different climatic regimes and, as a consequence, that there are regions of the Earth that can be defined on morphological grounds because these regions are distinctive on climatic grounds. Such regions are termed morphogenetic, meaning essentially that the morphology of the forms present supposedly has a common genesis—which is climatic. In turn, the geomorphic processes will yield characteristics distinguishing the landscape from other areas developed under other climatic conditions.

In a gross way, it is generally agreed that climate is extremely important in the weathering of rocks and that it exerts an influence on the dynamics of processes that, in turn, will leave their own impress on the landforms of a region. Julius Biidel, the German geographer, has argued that for such reasons climatic variance must take prece-

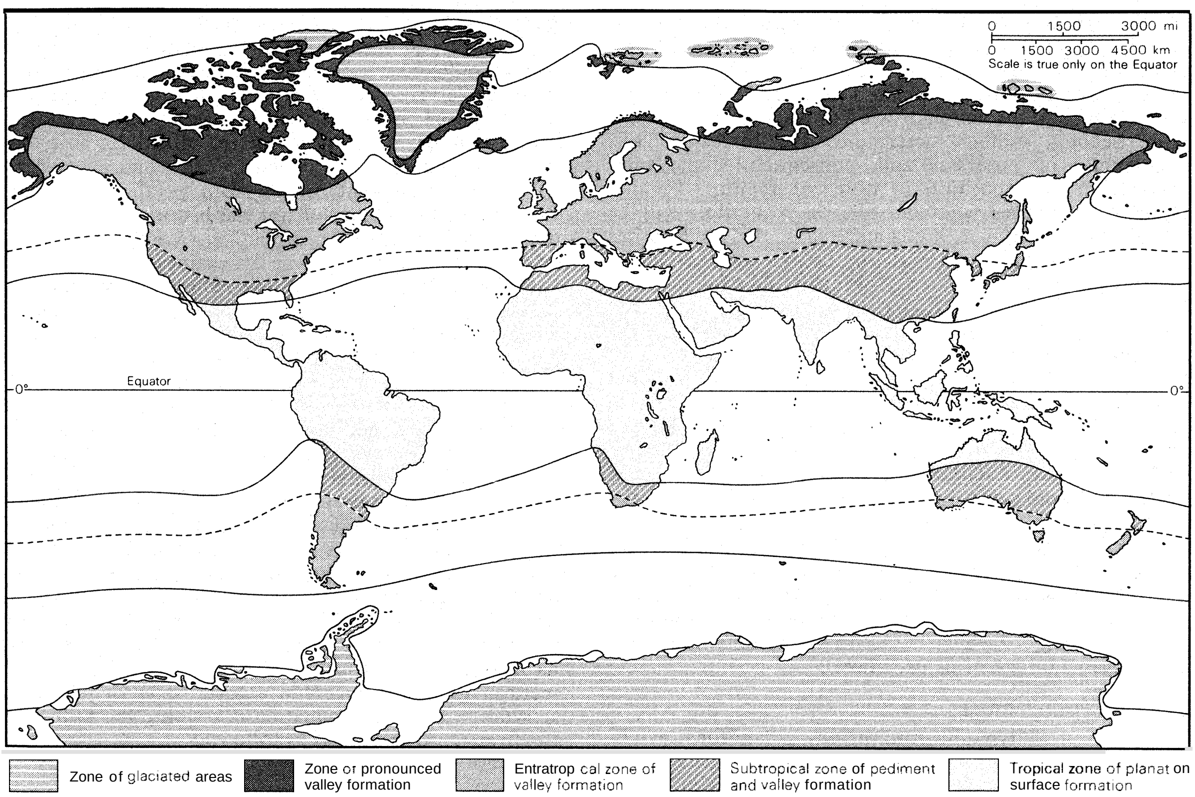

0 1500 3000 mi
0 1500 3000 4500 km
Scale is true only on the Equator
Equator
0°
0°
Zone of glaciated areas
Zone of pronounced valley formation
Extratropical zone of valley formation
Subtropical zone of pediment and valley formation
Tropical zone of planation surface formation

Figure 4: Morphogenetic regions of the earth.
Adapted from *Encyclopedia of Geomorphology*

dence over all structural factors. This is another way of saying that climate and climatic differences overshadow rock type as a controlling variable in landform configuration. The morphogenetic regions of the earth according to Biidel are shown in Figure 4. Considered in detail this map offers a few surprises, because maps showing the world distribution of deserts, for example, would reveal that at least two of Büdel's zones, and in some instances three, coincide with these arid areas; one would anticipate the existence of uniformity of forms in these regions according to the thesis as presented. Moreover, R. Common, who has studied slopes and processes in Ireland, Greece, and elsewhere, has indicated that there are ten zones in which temperature differences are likely to be of geomorphic significance and six zones in which important precipitation distinctions can be made. This difference in the number of zones of climatic distinction and, hence, of zones of supposed morphological distinction, varies as a function of the investigator involved. The fact that any variation exists demonstrates that the determination of morphogenetic regions is somewhat subjective; the qualitative criteria for recognition of such regions lead to disparate numbers, kinds, and locations of regions. The regions are primarily structural, topographical, and erosional and depositional. All are influenced by climatic controls.

Climate versus rock type

Proponents of the concept would say that climatic controls outweigh rock type as a landform factor because the significance of a given rock type and its resistance to the forces of erosion is to some extent a function of the climate in which it exists. Granite, for example, is a relatively hard rock that comprises many mountain ranges in temperate and other regions. In tropical zones, however, where intense chemical weathering has occurred, granites are weathered to great depths and commonly underlie plains and pediment-like areas. With respect to carbonate rocks, it is argued that these are "soft" in areas of frost action where they are readily shattered, and "hard" (in the sense of being cliff-forming rocks) in more humid regions. This is because the high degree of solubility and gross structure of carbonate rocks tend to permit cavern formation and the development of subterranean channels and does not lead to surficial erosion of any magnitude.

It is also true that processes such as frost wedging, wind action, oxidation, reduction, and others are indeed climatically linked. Despite this there is some reason to question the general thesis of the concept of morphogenetic regions. What has best been distinguished to date are the extreme "regional types." Opponents of the theory ask first if any of the supporters of regional morphogenesis have in fact gone much beyond the notion that tropical, glacial, arid, and temperate environments are generally recognizable. All protagonists would acknowledge that Arabia is inherently distinctive when compared to equatorial Africa or that Iceland and Florida are in some ways quite different, but they would insist that even this degree of acquiescence is deceptive because the real question is not whether the climates of these contrasting areas differ but whether their landforms differ. They would conclude that the answer may seem to be affirmative at a distance —but only at a distance.

It is commonly said, for example, that desert landforms are distinctive by reason of their high degree of angularity, particularly if such terrain is compared to the "rolling forms" of an area such as New England, where the climatic regime is quite different. The impression of rolling forms is, of course, largely conveyed by the soil cover or by the structure (*e.g.*, folded mountains) of certain areas. Where the Hudson River flows along the Palisades diabase sill (a tabular body of intrusive crystalline rock), however, this igneous rock type clearly serves as an angular cliff former. The hillslopes to be found in basaltic terrane along the Columbia River plateau, or in any temperate humid region, will differ hardly at all from hillslopes of the Black Haruj or similar volcanic terranes of the Sahara; many analyses of hillslopes in the U.S., Africa, Australia, and elsewhere have indicated that there is a basic interdependence of slope angle and particle-size distribution, which is governed by the lithology (or rock type) in question. This relationship is nearly ubiquitous and may suggest that, contrary to the claims of adherents of the morphogenetic regions concept, climatic controls have complemented but not transcended the basic importance of lithology in landform evolution. Aridity, however, has a great effect on the rate of erosion and evolution of landforms.

In the light of present knowledge on the subject, it would appear that climate is an important variable in landform evolution for the following reasons: first, climate determines those regions in which mechanical or chemical weathering and the action of wind or water will be important. Second, climate determines the intensities and frequencies of processes and, consequently, the rates of erosion or deposition in different regions. Finally, climate determines whether soils or vegetation will be present and, if so, their kinds and extent. This indeed has much to do with the general impression one receives of the morphology of an area. The available evidence also indicates, however, that hillslopes are strongly influenced by rock type. If this is true, then it would appear that the concept of morphogenetic regions is somewhat overblown. In the final analysis, landforms must be considered to be a response to multivariate controls; it is an error to assume that one factor subsumes the others, regardless of whether one wishes to make the case for time, for climate, or for rock type. Landforms tend to reflect the interaction of the materials of which they are composed and the applied stresses of processes that act upon them. As indicated by the Soviet geographer I.S. Shchukin in a recent critique, climate, rock properties, and physical processes are intimately intertwined in landform formation, and there is but little basis for the creation of a "climatic geomorphology" (which is essentially synonymous with morphogenetic regions as used here).

DYNAMIC EQUILIBRIUM

Gilbert's view of the relation between form and process

The concept of dynamic equilibrium in landform evolution was introduced by G.K. Gilbert, the American geologist, in his classic reports on the geology of the Henry Mountains, Utah, in 1877 and 1880. To explain the assemblage of landforms in this area, Gilbert applied several fundamental principles of fluvial erosion which today seem to have been intuitive and self-evident but which reflect his relatively sophisticated grasp of fluid mechanics and hydraulics at that time. Gilbert recognized that steep slopes erode more rapidly than gentle ones, that erosion is most rapid where rocks are least resistant, and that both the steepness of slope and water discharge affect sediment transport to a disproportionate degree. These principles led him to two important conclusions, namely, a stream tends to equalize the work done over each part of its course, and when the ratio of erosion as a function of slope becomes equal to resistance as a function of rock character, there is an "equality of action."

Gilbert did not give a name to his conceptual views, and his conclusions were not truly measured against the assumptions of the cycle of erosion. The Davisian bandwagon gathered momentum, and not for a generation was Gilbert recognized widely as a founder of this field of inquiry. It seems clear today, however, that Gilbert's conclusions are basic to, if not indeed synonymous with, the concept of dynamic equilibrium. This is the concept that landforms tend toward and attain a configuration that reflects a balance between rock resistance and the erosive action of natural processes upon them. Stated differently, landforms can be considered as part of an open system in a steady state of balance, in which every slope element is adjusted to its neighbours.

Both definitions imply an anticipation that changes in climate and relief, or, in general, changes in materials and energy, will affect the steady state and will produce new equilibriums between landform configurations and the environment. It is for this reason that equilibrium is termed dynamic. Topographic change is acknowledged to occur in response to changed conditions; therefore, the landscape will tend to shift toward some new equilibrium condition.

It should be noted again that the concept of dynamic equilibrium differs drastically from that of the cycle of erosion. No definite and inevitable sequence of landforms through time is postulated. Equilibrium states and configurations are recognized as highly probable and frequent conditions in nature, whereas only the old-age peneplain at base level is viewed as an equilibrium form in the Davisian scheme. And finally, a landform that is in a steady state of balance is considered timeless from the viewpoint of equilibrium; its history cannot be deduced from its configuration or other outward manifestations.

If the cycle of erosion is accepted as a correct conceptual framework, however, then landform configuration is at once indicative of history, or the effects of time, because any landform is assigned to a specific part of an evolutionary scale that ranges from youth to old age.

The answer to whether or not the historical imprint of time can be discerned from landform configuration can best be considered in terms of an example. If it be supposed that a steep cliff consisting of jointed sandstone is acted upon by weathering, gravity, and fluvial processes in such a manner that each year a certain number of fragments and small blocks are dislodged and tumble to the base of the cliff, then as long as the nature of the rock and the intensity of the processes acting upon it do not change, the rate of addition of material to this talus pile will remain fairly constant. Some material will, however, be removed from the talus pile by processes acting upon it, and ultimately the rate of removal will equal the rate of addition. At that time the talus pile will be in a state of dynamic equilibrium. Its volume will not change unless some factors that can alter the rates of addition or removal are changed.

C.C. Nikiforoff, the soils scientist, discussed soil formation in terms of dynamic equilibrium in 1942 and his work was adapted to landform evolution by J.T. Hack, the American geomorphologist, in a basic series of papers in the 1960s. The general equation describing the talus-pile case, from these sources, is as follows:

$$S_n = A\left[\frac{(1-r).[1-(1-r)^n]}{r}\right] + A,$$

in which S_n is the total volume of talus after n years, A is the amount of talus that accumulates in one year, and r is the rate of removal, expressed as a decimal proportion of the amount present. In the limit, with the passage of time, the equation reduces to:

$$S_n = A\left(\frac{1-r}{r}\right) + A, \text{ or } = \frac{A}{r}.$$

This equation shows that when the rate of removal r produces an absolute amount equal to the accumulation A, the volume of talus S_n will be a constant. A graphical expression of this relationship is shown in Figure 5, assuming that A = 1 and r = 0.3, or that for each unit increment of talus derived from the cliff, 30 percent of the amount present is removed from the pile. The limiting value in this case is 3⅓ times the amount added per unit of time, and it can be seen from the above graph that the limit is asymptotically approached after approximately 12 units of time.

By courtesy of the U.S. Geological Survey

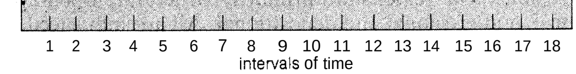


Figure 5: Nikiforoff's equation for growth of a deposit toward steady state in which two opposing processes are involved in formation of deposit.

Assuming that the talus pile attains a steady state, that S_n is constant, and that an investigator approaches the base of the cliff at any future time, it is interesting to ask what history of the talus-cliff configuration could be deduced? To question whether the system is at the stage of maturity or what might be concluded about the initial conditions, whether the volume of talus was larger, smaller, or zero 1,000,000 years earlier, would be somewhat specious. It seems evident from the example given, and from what has been said thus far, that the eradication of initial systemic conditions is a by-product of the attainment of steady-state conditions. This would lead to speculation as to whether all landforms are in fact at or near steady-state conditions today. Before turning to this question, field evidence for dynamic equilibrium must be considered. This is of three principal kinds: the equilibrium of slopes, the equilibrium of river systems, and the equilibrium of regions.

Altitudes, rock type, and slope equilibria

The equilibrium of slopes is well established in the sense that the logic of Gilbert's observations is overwhelming and that detailed field observations sustain expected relationships between slopes and rock character, other factors being equal. The histograms of altitude and rock type in Connecticut (Figure 3, bottom) serve as one example of differential erosion that is attributable solely to rock type. The Stockbridge Marble would be predicted to be the most easily erodible rock type because of its high solubility, whereas quartz-rich rocks would be expected to be at the other end of the erosional spectrum. This is the relationship found; lowest mean altitudes occur on terrain underlain by marble, and the greatest average altitudes coincide with outcrops of quartzose rocks.

The relationship of channel slope and stream length in rocks of different lithology is known for some streams in Virginia and Maryland. There is a clear distinction between relationships in sandstone and those in shale; for a given stream length, streams in sandstone areas flow over channel beds that are seven times steeper than those in shale areas. A curve of relationship for carbonate rocks has not been constructed because of the relatively great degree of scatter of the data. This scatter is caused by the chert content of the limestone, which is abundant in some stream channels and tends to produce increased gradients.

Profiles of hillslopes on similar rock types, as shown from Hack's studies in the same area, also show unsurprising relationships. The steepest profile prevails in terrain that predominantly consists of sandstone, the gentlest on hills composed of shale beds, and intermediate profiles are associated with carbonate rocks.

Hydraulic geometry of river systems

Other examples could be cited from all parts of the world, as mentioned above in the discussion of morphogenetic regions, but still more telling are the available data on equilibriums in river systems. The hydraulic geometry of stream channels, as it was termed by Luna Bergere Leopold and Thomas Maddock, Jr., both hydrologists with the U.S. Geological Survey, in a basic exposition of the subject in 1953, is based on the observation that fundamental relationships exist between water discharge and the width, depth, and velocity of flow, among other variables. At a given river cross section, for example:

$$
\begin{aligned}
w &= a\,Q^{b},\\
d &= c\,Q\text{, and}\\
v &= k\,Q^{m},
\end{aligned}
$$

in which w, d, and v are width of water surface, depth of flow, and velocity of flow, respectively; Q is water discharge of a given frequency of occurrence; and a, c, k and b, f, m are numerical constants and exponents in the power-function relationships shown. From the law of continuity (water input along a given river reach equals water output):

$$Q = Av,$$

in which Q is again water discharge, A is cross sectional area, and v is velocity of flow. Because area is equal to the product of width and depth,

$$Q = wdv,$$

and substituting from above:

$$Q = a\,Q^{b} \times c\,Q \;\times k\,Q^{m} = ackQ^{b+f+m}$$

This identity indicates that:

$$a \times c \times k = 1\text{, and } b + f + m = 1$$

The last is an important result of this analysis because b, f, and m are the slopes of lines that relate water discharge and width, depth, and velocity, and the average values of these exponents provide a means of characterizing and comparing river systems. The at-a-station (a given cross section of observation) relationships vary for different rivers and for different sections along their courses; but in a downstream direction it has been shown that $b = 0.5$, $f = 0.4$, and $m = 0.1$ for many rivers. This indicates that width, depth, and velocity of flow all tend to increase regularly with water discharge in a downstream direction and that channel width adjusts most rapidly.

Similarly, there are relationships between water discharge and suspended sediment load, channel roughness, and channel slope, and each of these can be expressed as power functions. For example:

$$l = p\,Q^{j}$$

in which l is suspended sediment load, Q is water discharge, p is a numerical constant, and j is the characterizing exponent, equal to about 0.8 in a downstream direction for many streams.

Generalized graphs of each of these relationships are shown in Figure 6; the solid lines give changes in a downstream direction with discharge, and the dashed lines show at-a-station results. Only channel roughness (a measure of particle size, among other factors) and channel slope decrease in a downstream direction, as would be expected. The block sections on the right half of the diagram permit visualization of the at-a-station and downstream relationships for different magnitudes of discharge; these correspond to different frequencies of occurrence of water discharge.

By courtesy of U.S. Geological Survey Professional Paper 252, U S. Government Printing Office

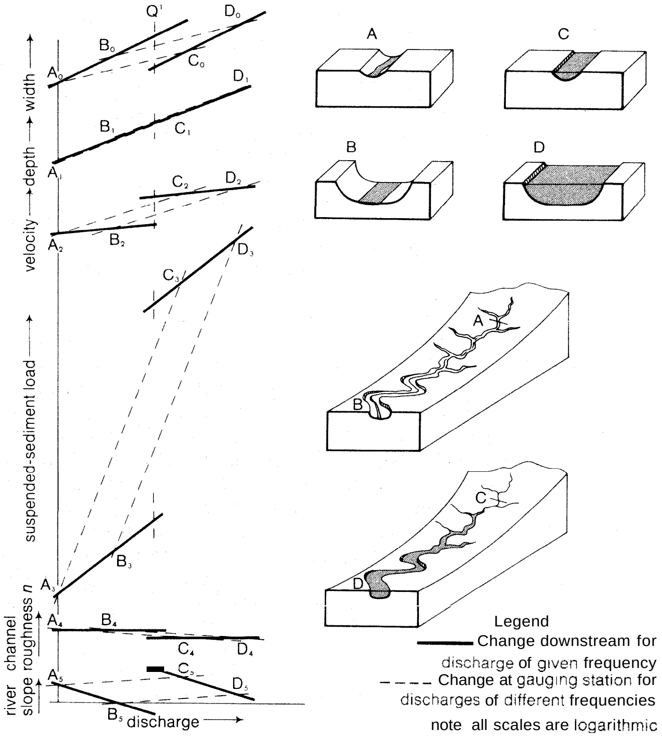


Figure 6: Average hydraulic geometry of river channels expressed by relations of width, depth, velocity, suspended-sediment load, roughness, and slope to discharge at a gauging station and downstream.

River systems and dynamic equilibrium

With respect to the magnitude and frequency of water discharge, it has been found that the channel-forming discharge, or that magnitude of flow that best relates to the dimensions and characteristics of a given channel, is the bankfull discharge. As implied by the name, bankfull discharge is a magnitude of flow that fills a channel; it has a recurrence interval (frequency of occurrence) of one to two years in humid temperate regions and an interval of as much as hundreds of years in extremely arid areas.

Typical relationships in humid-temperate regions are indicated by the map and cross section of Stevens Creek, South Carolina (Figure 7). This stream was studied in conjunction with long-term records of water discharge in an attempt to learn whether some physical aspects of a channel did in fact correspond to the **bankfull** stage. They found, in general, that water discharge with a recurrence interval of one to two years was related to the bench level adjacent to channel levees (predominant bench **3** in the cross section of Figure 7). The highest benches bordering the channel were correlative with a magnitude of discharge that occurred as infrequently as once in 25 to once in 90 years.

By courtesy of the U.S. Geological Survey

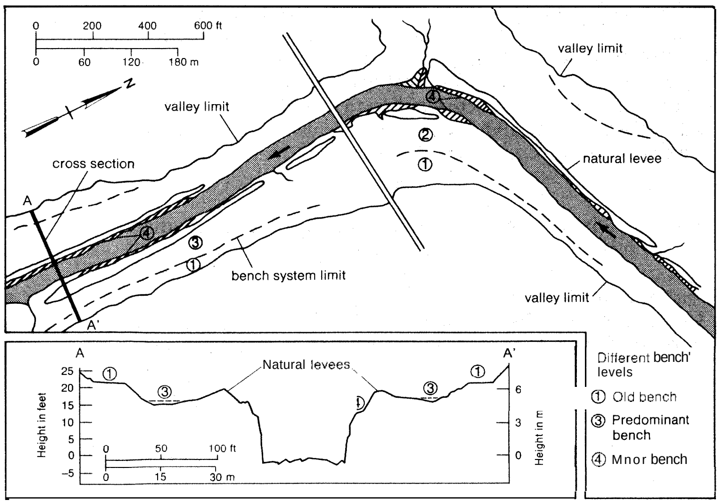


Figure 7: Map and cross section of Stevens Creek, South Carolina, showing the several benches associated with flow stages of different frequency of occurrence.

With respect to landform evolution and the merits of the dynamic equilibrium view, it must be concluded that evidence derived from study of river systems is both pervasive and persuasive. **A** river adjusts its profile and cross section to accord with its bankfull discharge and the requirements of its sediment load. In alluvial channels this adjustment is rapidly accomplished and the hydraulic geometry has been demonstrated to hold in a wide variety of physical environments and for rivers of all sizes. If climatic change occurs, or if uplift alters channel slope, then a river will adjust its channel geometry and load to correspond. In other words, most river systems are in a state of dynamic equilibrium. Rivers provide an example of the adjustment of forms to process and although a stratigraphic record is available in certain alluvial deposits, it is absurd to speak of the age of a river; once again, time and adjustment have obscured initial systemic conditions. These conditions change as the river goes through a series of stages in its life history.

The third line of evidence supporting the concept of dynamic equilibrium is evidence of the equilibrium condition of regions. In a sense this line of argument is somewhat superfluous; if hillslopes are in equilibrium and if drainage systems are in equilibrium, then it follows that many regions must also be in a state of equilibrium because they consist of these two elements. Nevertheless, Hack's intensive study of the Shenandoah Valley (Figure 8) is worthy of mention because similar work has not yet been done elsewhere and because it is part of an area long interpreted in terms of the erosion cycle. In this region the streams accord with the general hydraulic geometry; their longitudinal profiles are adjusted to differential rock resistance; the meanders of the North Fork are confined principally to the Martinsburg Shale, in which they parallel the rock structure; and the Potomac River enters and leaves the valley where narrow outcrops of weaker rock occur. **As** indicated by Hack, it is extremely unlikely that such close relationships between drainage and rock resistance and structure would prevail in this region if the streams were superposed.

Similarly, hillslopes reflect rock resistance to erosive stresses, and major topographic features accord with the generalization that lowlands and valleys are correlative with soft rocks (carbonates and shales), whereas the greatest altitudes and most persistent ridges are correlative with the most resistant rocks. Climatic conditions also modify the hillslope erosion.

Interpretation of landform history

Granting that equilibrium can and does exist in nature, it is fair to ask what can be said of landform evolution; is there some historical imprint of time that will yield upon analysis? It must be emphasized that the concept of dynamic equilibrium is a framework of thought that provides a rational basis for the study of landforms. It is a standard against which landscape features can be tested; studies of process intensities and rates of change can, in many instances, lead to conclusions about those features that are not in equilibrium. This narrows the search for time-dependent landforms and can lead to important conclusions.

For example, G.H. Dury, a British geomorphologist, observed that the wavelength of valley meanders clearly differed from the wavelength of stream meanders. The plots of wavelength for the Dnestr and Don rivers, and for rivers in France, England, and the U.S., sustain this view. Because it was known that the wavelength of stream meanders bears a consistent relationship to water discharge (or drainage area, which is closely correlative), it was possible for Dury to argue that the valley meanders represented misfit or nonequilibrium forms, which were related to water discharge during more humid, Pleistocene time.

There are many landforms that reflect the historical imprint of time in some way. Most of these are Pleistocene relicts: large cirques and V-shaped valleys in formerly glaciated areas; beach ridges around modern playas (dry lakes); and solution caverns and stream channels that reflect adjustment to moister hydrologic regimes in modern deserts. These and other landforms consequently reflect equilibriums with past processes rather than present processes. Given sufficient time, they will attain new steady-state conditions and their history will be erased. In the interim, however, interpretation is possible, and it can be said that one of the chief virtues of the equilibrium concept is that it permits discrimination between the time-independent and relict landforms in nature.

By courtesy of the U.S. Geological Survey

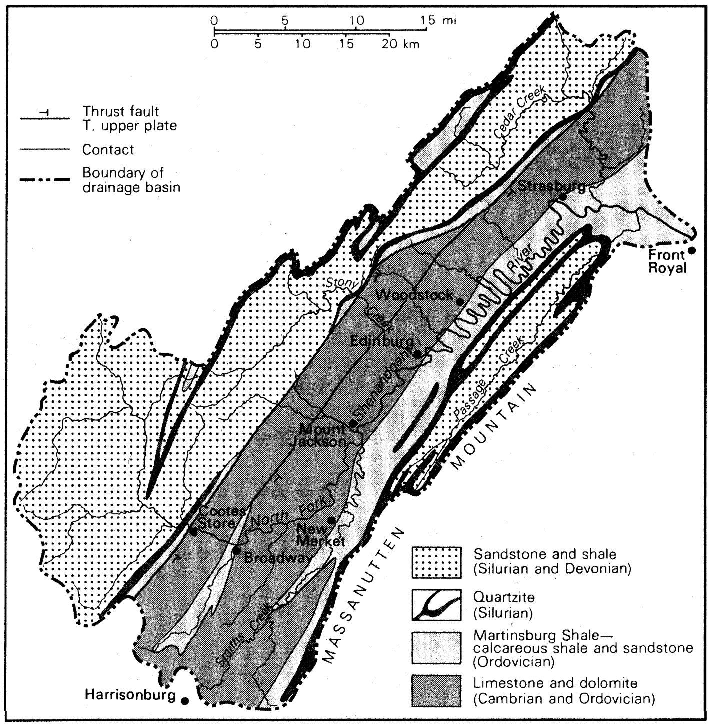


Figure 8: Geological map of the North Fork, Shenandoah River drainage area.

In addition, as summarized so well by Richard J. Chorley, the British geomorphologist, the equilibrium concept fosters recognition of the multivariate nature of geomorphic phenomena and directs attention to the entire landscape rather than only to those elements of supposed cyclic significance.

ENTROPY—PROBABILITY

Entropy and the work of river systems

The entropy-probability concept in landform evolution is a logical outgrowth of the dynamic equilibrium known to prevail in river systems. The hydraulic geometry of stream channels concerns the interrelationship and adjustment of channel morphology and the several hydraulic variables. In discussing this adjustment, however, nothing was said of the distribution of work done and energy available in relation to fluvial processes or of whether river systems reflect some basic natural law as they follow a course from highlands to the sea. These questions will be considered here, but a brief exposition of the thermodynamic and probabilistic connotations of entropy is required first.

In thermodynamics, entropy is an expression of the reduction of free energy in a system and, therefore, of a reduction in the ability of a system to perform work. This relationship can be expressed simply as:

$$\Delta F = \Delta H - T \Delta S,$$

in which ΔF, ΔH, and ΔS represent the changes of free energy, enthalpy (heat content), and entropy of a system that passes from one state to another, and T is absolute temperature. The essential point is that an increase of entropy (ΔS) must be associated with a decrease in free energy (ΔF). The second law of thermodynamics specifies this relationship for natural processes and provides as well the explanation for the lack of success of any perpetual motion apparatus.

A river system obviously involves natural processes; each droplet of water in the headwaters of the system contains a certain amount of potential energy that is a function of its elevation above sea level. With the passage of time, as the stream transports sediment to the sea and the landmass is reduced, if no other changes affect the system then it must suffer an energy decrease and a corresponding rise of entropy. Therefore, overall entropy will increase as work is done by the river system under these conditions.

But entropy also relates to states of order-disorder and to probability. A mass of gas in a container at room temperature consists of individual gas molecules, each of which follows random paths and engages in random collisions with its fellows. This may be contrasted with a situation in which the same amount of gas is compressed into a small volume so that the motions of the molecules are greatly restricted. The first state could be designated as one of great disorder and high entropy, whereas the second would be characterized by a high degree of order and low entropy.

In this example, work would need to be done in order to produce the second state, but there is some finite probability that all the gas molecules in a given closed container will cluster at one end through chance alone. The probability that this will occur through chance alone is small indeed, but it does serve to illustrate the basic point, namely, that there are several possible alternate states in all natural systems and that each such state is able to be characterized by a certain probability of occurrence, a certain degree of order or disorder, and a certain potential for work. In this statistical sense the entropy of a system can be defined as:

$$S = \sum_{i=1}^{i=n} \log pi$$

in which S is the entropy of the system and p_i represents the probabilities of each of the possible alternate states ($p_1, p_2, p_3, \ldots p_n$).

When the sum of the logarithms of the probabilities of all possible states is a maximum, the most probable energy distribution in the system will be achieved. This will occur when the probabilities of all alternative states are equal ($p_1 = p_2 = p_3 \ldots = p_n$). It has been argued that river systems are open systems in which the hydraulic geometry adjusts to achieve the most probable energy distribution and, further, that there is a tendency toward minimization of work along a stream's course. It can be shown on theoretical grounds that the most probable series of energy losses along a river's course is coincident with a series of entropy increases that are equal in magnitude for each unit length of river. Most important is the fact that when this occurs the longitudinal profile of a river will be best represented by an exponential curve. The basic relationship is:

$$\frac{dS}{dt} \cdot \frac{1}{Q} = \frac{dh}{dx} \cdot \frac{1}{h},$$

in which the left side of the equation is the rate of change of entropy (S) per unit discharge (Q) and the right side is the loss of elevation (h = head) with distance of flow (x) along a river's course. This fact is important because available data on the natural channels of streams in all types of environments have indicated that the preferred profile in nature is indeed an exponential, concave-upward path.

Stream profiles and the most probable path

If the exponential profile of streams is theoretically predicted to be the most probable flow path, then it follows that it should be reproducible by chance alone. For this reason random-walk procedures, generated by manual or computer methods, can be employed to test the theory. Figure 9 shows the mean longitudinal profile generated by two random-walk trials, the condition of the trials being that no constraints be imposed on length of path. It can be seen that the resulting curve is exponential in form and is asymptotic at base level (elevation = 0). A much greater number of trials yields the same result. If constraints on length are imposed upon the model, then a variety of curves is obtained, but these are of the same general form.

By courtesy of the U.S. Geological Survey

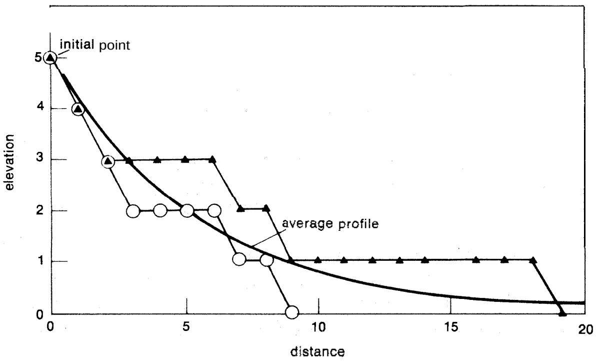


Figure 9: Stream profile generated by two random-walk trials.

The resulting form is related to the rate of work done by a given stream. Least work expenditure is required of streams with profiles of maximum concavity because the greatest fall or head loss in such cases is in the headwaters, where the water discharge is least. The general relationship of least-work, equal-work-distribution, and natural-river profiles is shown in Figure 10, which is based on the relationship

$$s \propto Q^z,$$

in which s is slope of channel, Q is water discharge, and z is the characterizing exponent.

The plan views of rivers have not been neglected in the search for underlying principles. Typical drainage nets have been generated by random-walk procedures. Aside from their obvious superficial resemblance to natural drainage nets, it has been shown that such random, and therefore most probable, drainage nets accord with the laws of stream order, stream number, bifurcation, and other empirical properties of networks set forth by R.E. Horton, an American hydraulic engineer, in a basic paper on this topic. A test of Horton's laws, as they are termed, was devised by R.L. Shreve who concluded that randomly generated channel networks exhibited all the properties required by Horton's laws and that the population of such networks conformed statistically to the attributes of 172 published sets of data on natural networks. The general conclusion that can be drawn from all such work must be that drainage networks, like channel profiles, can be reproduced by random-generating methods. This means that they too reflect the most probable network patterns, tend toward least-work paths, and are associated with an entropy distribution that leads to these results.

From *American Journal of Science*, vol. 262 (1964)

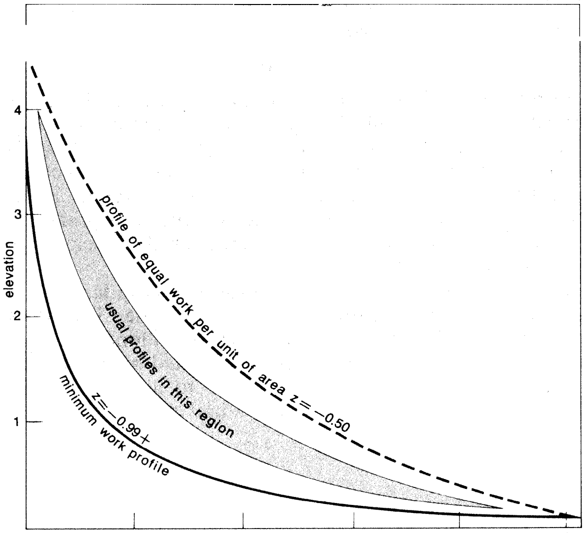


Figure 10: River profiles for a given relief in relative units.

Randomness in nature as a factor in landform configuration

Stochastic models, or those involving the element of randomness, are applicable to a wide variety of natural phenomena. These range from patterns of sedimentation revealed by the stratigraphic record, to the branching nature of rivers, trees, and lightning, and include the landform configurations in all natural environments. From the vantage of the entropy-probability concept it can be said that nature tends to produce the most probable patterns and configurations unless constrained in some way in specific cases. It may be noted that the logic of the position that hillslopes, drainage networks, and river channels in nature present the aspects they do simply because these aspects are most probable has much to commend it. Clearly, one would not wish to attempt to sustain the alternative proposition that these aspects are most improbable.

In summary, the entropy-probability concept serves to explain why dynamic equilibrium holds for most landforms in nature, that is, because the equilibrium condition reflects a nice distribution of available energy relative to work done, which is the most probable distribution. The previously indicated virtue of the equilibrium concept, namely that it served to segregate the time-independent from the time-bound landforms, still holds true. With respect to the entropy-probability view, this might be altered to say that the study of landform evolution must be considered a search for constraints, these constraints being the conditions that lead to departure from equilibrium forms. In this sense, however, modern studies of landform evolution still can treat the interpretation of landforms in terms of the historical imprint of time, but such studies must be entirely restricted to the nonequilibrium landforms.

In addition, the element of indeterminacy in landform evolution must be recognized. It is clear that individual landforms, like individual stream profiles, are members of populations that statistically cluster about mean values. Because these are the most probable values in nature, and commonly reflect steady-state values, studies of individual forms that fluctuate about the mean cannot provide a valid basis for extrapolation of class attributes. Given a thousand hypothetical talus piles it is possible that nearly half of these would exhibit slight increases of volume through time, whereas the remainder would exhibit diminution of volume. This possibility implies that general rules governing talus piles as a class may be indeterminate if the rules are derived from data based on the study of only one or a few members of that class. Stated differently, the principle of indeterminacy in nature is a logical consequence of the operation of randomness. The latter forces consideration of landforms and their evolution into the statistical realm.

BIBLIOGRAPHY. There are few books that treat all aspects of landform evolution, and the interested reader will need to investigate some specialized literature and journals to pursue this topic. Several books that are worthy of citation include R.W. FAIRBRIDGE (ed.), *The Encyclopedia of Geomorphology* (1968), articles on specific landforms and their presumed evolution; G.H. DURY (ed.), *Essays in Geomorphology* (1966), several articles of relevance on morphogenetic regions, the concept of grade or equilibrium, and statistical methods; L.B. LEOPOLD, M.G. WOLMAN, and J.P. MILLER, *Fluvial Processes in Geomorphology* (1964), a basic coverage of the principles of geomorphology that is derived from much of the specialized and fundamental literature on landform evolution; H.F. GARNER, *The Origin of Landscapes* (1974), and C.H. CRICKMAY, *The Work of the River* (1974), both of which treat fluvial processes in geomorphology; L.C. KING, *Morphology of the Earth,* 2nd ed. (1967), an exhaustive summary of the author's views on intercontinental denudation chronology (resulting from pediplain formation and parallel slope retreat on a vast scale); W.M. DAVIS, *Geographical Essays* (1909; reprinted 1954, ed. by D.W. JOHNSON), a reissue of some of Davis' basic writings on the cycle of erosion, including the article on the rivers and valleys of Pennsylvania; R.J. CHORLEY, A.J. DUNN, and R.P. BECKINSALE, *The History of the Study of Landforms,* vol. 1 (1964), lively and enjoyable account of ideas on earth history (from antiquity to 1893), can be read with profit by any curious lay reader; C.C. ALBRITTON, JR. (ed.), *The Fabric of Geology* (1963), a collection of essays designed to probe the philosophical aspects of geology. Of particular interest, with respect to landform evolution, are articles on rational and empirical methods of investigation and on association and indeterminacy in geomorphology: WALTHER PENCK, *Die morphologische Analyse* (1924; Eng. trans., *Morphological Analysis of Land Forms,* 1953), summarizes Penck's views on parallel slope retreat and landform evolution, largely ignored or overlooked by Davis and his adherents; Y. ABUL-HAGGAG, A *Contribution to the Physiography of Northern Ethiopia* (1961), and E.H. BROWN, *The Relief and Drainage of Wales* (1960), two relatively modern examples of the interpretation of regions within the framework of the cycle of erosion.

The specialized literature on landforms and their evolution is vast, and reference cannot be made here to the many fundamental contributions on the development of ideas. The serial publications involved would include minimally, however, the *Zeitschrift für Geomorphologie* (papers are in French, German, and English), *Revue de Géomorphologie Dynamique* (papers are in French), *Soviet Geography: Review and Translation* (occasionally contains useful surveys of ideas on landform evolution in the Soviet Union), and *Geographical Abstracts A: Geomorphology* (contains abstracts of world literature). Additionally, journals that principally treat geology may contain important papers on this topic (*e.g., Journal of Geology, American Journal of Science,* and *Geological Magazine).* The *U.S. Geological Survey Professional Paper* series includes papers on the hydraulic geometry of stream channels in several environments, on dynamic equilibrium and the entropy-probability concept, and on general systems theory in geomorphology. In this same vein, see *Benchmark Papers in Geology* (1971–), which present results of geomorphology symposia held at the State University of New York, Binghamton.

(L.K.L.)

Land Reform and Tenure

Land reform, in its many historical manifestations, has varied according to the functions performed by land itself. Land serves as a factor of production, a store of value or wealth, a status symbol, and a source of social and political influence. Land values reflect the relative scarcity of land, which usually depends on the ratio between the area of usable land and the size of the population dependent on it. As the per capita land area declines, the relative value of land rises; and land becomes increasingly a source of conflict among economic and social groups in the community.

The patterns of wealth and income distribution, and of social and political influence, that arise from land ownership are determined by the laws governing land tenure. These laws specify which forms of tenure are acceptable, and the privileges and responsibilities that go with them. They define the right to the land: whether the owner has full power to dispose of it as he wishes or whether it must be kept within the same family. For example, if private ownership is permitted, class differentiation is unavoidable; if only public or collective ownership is permitted, distinctions based on land ownership will be eliminated. Similarly, if land is held in hereditary tenure, the benefits will accrue only in the form of income; in contrast, absolute ownership may confer social and political advantages in addition to income.

The purposes of land reform

Historically and until very recent decades, land reform meant reform of the tenure system, or a plan for rapid improvements in the pattern of distribution of rights to the land and of the benefits and responsibilities derived from ownership of it. This idea of reform has been modified in recent years in response to the growing interest in guiding or planning national economic development and a recognition of the strategic role land and agriculture can play in economic development. It has therefore become more accurate to speak of land reform in terms of such things as planned, rapid improvement of the land tenure system; the pattern of cultivation; the scale of operation of farms; terms of tenancy, rural credit, marketing, and education; the state of technology; or any combination of these. This broader concept has characterized all recent reform movements, regardless of the political or ideological orientation of the reformers.

OBJECTIVES OF REFORM

Reform is usually introduced by governments, at their own initiative or in response to internal and external pressures, to resolve or avoid an economic, social, or political crisis. The objectives of reform may be only indirectly associated with the crisis, since reform programs may be introduced for their apparent merits, quite apart from the impending crisis.

The distinction between the real motives for reform and the objectives proclaimed by the reformers may be especially significant if the proclaimed objectives are imposed upon the reformers by political circumstances and do not have their full support. For example, the reformers may proclaim certain objectives (without any intention of enforcing them) to appease the peasants, to undermine the opposition, to gain international good will, or to comply with instructions from outside forces. Given the difficulty of isolating the real from the proclaimed purposes of land reform, the latter will be taken as a point of departure in this article.

Political objectives. The most common objective of land reform is to abolish feudal forms, which, in practice, usually means overthrowing the landlord class controlling the land and depriving it of power in favour of a reforming elite. If "foreigners" happen to be among the landlord class, the objectives are reformulated to call for the defeat of imperialism, and the end of foreign exploitation.

Another objective is to free the peasants and make them active citizens by ending their subjugation to and dependence on the exploiters. This objective assumes a moral and humanitarian character as a means of restoring to the peasants what assertedly had been taken away from them.

A third objective is to create democracy—a stated purpose of both non-Communist and Communist reformers. Most non-Communist reforms are based on the assumption that individual private ownership in the form of family farms will promote and sustain democratic institutions by increasing the number of independent family farmers.

Communist reformers have usually aimed at overthrowing both feudalism and capitalism on the assumption that private ownership of the means of production, including land, inherently breeds exploitation. In practice they have favoured radical land reforms with the stated goal of "returning land to the tillers" and creating a classless, democratic society. One of the more or less open political aims of Communist reformers is to rally the peasants in support of the new order. Once this has been done they seek to move the peasants toward the next step of socialized farming.

Finally, reform may be undertaken as a pragmatic step to resolve a crisis or avoid a revolution when the peasants become restless enough to threaten the government or the political order in general. Under these circumstances, the regime in power or a new one introduces and implements just enough reform to appease the peasants and contain the conflict, especially when the reformers are in sympathy with the landlord class or with the established order or when they do not feel strong enough to undertake a more radical reform.

These political objectives overlap and are frequently proclaimed in different combinations. All of them, however, are subject to two conditions: they tend to undergo change during the period of implementation, and they are usually kept vague enough to permit flexibility and modification as conditions change.

All land reforms tend to emphasize improvements of the peasants' social conditions and status through such measures as the alleviation of poverty and the redistribution of income and wealth in their favour.

Reformers attempt to raise the social level of the community by improving the conditions of the individual. In addition to the creation of employment opportunities, they focus attention on social improvements such as expansion of education and health services, the benefits of which accrue to the community at large. Given the long-term and lasting nature of social objectives, reformers usually aim at the younger generations.

Economic objectives. Economic development has become a major objective of governments and parties in recent decades. In pursuit of this, efforts have been made to encourage agricultural progress. The peasant who does not own his land or whose share of the crop is relatively small may have little incentive to invest capital or expend effort to improve the land and raise productivity. Agrarian reform is presented as a correction of such defects.

Another objective is to encourage more intensive cultivation, that is, the application of more labour and other resources to a given unit of land. This assumes that large landowners often use their land wastefully, even when abundant labour is available to farm it more intensively.

An equally important economic objective is to promote optimum-scale operation of the farm. Both excessively large farms (latifundia) and excessively small farms (minifundia) tend to be inefficient in the sense that output per unit of input is less than it can be under more favourable circumstances. Reform aims at creating farms of the correct size for full and efficient utilization of resources under given technological conditions.

Finally, reform aims at coordinating agricultural production with the rest of the economy. In their quest for economic development and industrialization, reformers hope to make the rural sector more responsive in various ways to the needs of the industrial sector—to use it as a source of new labour, of expanded food supplies, or of industrial raw materials. The rural sector may also be seen as a source of funds to invest in industry or of commodities that can be exported to strengthen the balance of payments. These functions are often expected to be performed simultaneously.

TYPES OF REFORM

The variety of reform objectives implies different types and measures of reform. They can be classed in five main categories according to whether they deal with land title and terms of holding, with land distribution, with the scale of operation, with the pattern of cultivation, or with the supplementary measures necessary for success of the reform.

Reforms that are concerned with the title to land and the terms upon which it is held mainly represent a transition from tradition-bound to contractual systems of landholding. Their implementation involves property surveys, re-

Title settlement and terms of holding land

cording of titles, and provisions freeing the landholder from restrictions or obligations imposed by tradition. Property surveys are introduced where land is held by a tribe or clan or where reallocation of cultivable land still follows tradition. The landholder in these situations lacks security of tenure and frequently does not even know the boundaries of his holdings. Though he cultivates the land, the right of disposal belongs to the tribe, clan, or feudal lord, as in medieval Europe and in modern Africa and the South Pacific islands. Reform therefore involves surveying the land area, devising a formula for its redistribution among the occupants, and recording the titles accordingly. In most cases this kind of reform tends toward establishment of private ownership or some form of individual family holding. Title settlement touches on at least three aspects of landholding: it may increase security of tenure and hence incentives; it may reorganize the system of inheritance in favour of offspring; and it tends to bring land onto the market so that land transactions become possible. Reform of this kind has little effect on the scale of operation until some land concentration or fragmentation takes place.

In areas of the world where the terms of holding and tenancy are regulated by tradition, frequently to the advantage of the landlord, reform may involve seeking to convert tenancy into a contractual agreement that offers some protection to the tenant, as has been the case in India and Pakistan. The terms of tenancy are made flexible and the tenant is given more incentive to improve the land and advance technology.

The redistribution of land

The most common type of reform involves the redistribution of titles from one individual to another, from individuals to a group or community at large, or from a group to individuals, regardless of how the land is finally held for purposes of cultivation. For example, the land of one landlord may be redistributed to many individuals as in Egypt, Iran, or Ireland. Or the land of individuals may be reallocated in favour of the community at large by abolishing private ownership as in Soviet Russia and China. Or, again, public land may be distributed to individuals as in various parts of Latin America.

The impact of redistribution on the scale of operation and on marketability of the land depends on the form it takes and the restrictions attached to it. Ownership may change hands without affecting the scale of operation. For example, if the redistributed farm was previously operated as a unit, its division means fragmentation and reduction of scale; however, if it was operated in smaller segments by individual tenants, transfer of title to these tenants would not affect the scale of operation. The final results depend on the measures taken to prevent adverse effects.

Efforts to improve the scale of operation

Land reform is, of course, able to improve the scale of farm operations by enlarging the farm unit or by reducing it. Enlargement applies when the holding is increased in size either by adding to it or by consolidating its fragmented parts into one intact unit. Farm consolidation involves reallocation of the total farm land within a region among holders through a system of land evaluation and exchange, possibly including compensation so that no one loses and all gain by increasing efficiency. The scale of operations may also be increased by pooling resources as in farm cooperatives that offer facilities otherwise inaccessible at the level of the relatively small farm. Collectives aim at the same objective, although overexpansion has often led to negative rather than to positive economic effects.

An equally common approach is to divide large, extensively cultivated farms into smaller and more intensively cultivable units. Reduction of the scale, however, has potential problems since it may be confused with redistribution of income and employment. In such cases, it is not uncommon that extra weight may be given to the social impact so that the division may result in excessively small units or in the break up of efficiently run farms. Reduction of the scale may also inhibit changes in technology, capital investment, and diversification, all of which require relatively large units.

Changing the pattern of cultivation

Changes in the pattern of cultivation relate directly to the important question of increasing productivity in agriculture. While other types of reform may influence productivity indirectly by their influence on security of tenure, or scale of operation, improvement of the pattern of cultivation affects productivity directly: through advances in technology, improved irrigation, and the application of fertilizers and pesticides.

Technological advance usually implies mechanization, although this need not always be the case since it may be possible to improve agricultural techniques by introducing crop rotation, investigating the quality of the soil, or by better use of available technology. Such changes, however, have limited effects unless new technology has been introduced. The state of technology largely determines the level of productivity or the ratio between outputs and inputs, with more advanced technology permitting the cultivation of more land per unit of labour. It also makes possible deeper plowing, more timely cultivation and harvesting, reclamation of areas previously inaccessible, and possibly wider diversification of the crops than previously attainable. By easing the physical burden of farm work, it helps to conserve human energy; and it may even be the most direct way to modify tradition without an open confrontation with traditional forces. Mechanization may also have some adverse effects, such as the release of labour that may be difficult to employ elsewhere or the consumption of capital badly needed in other sectors.

Improvements in irrigation include not only increasing the water supply but draining swampy land and regulating the quantity and quality of water flow. This enables more intensive cultivation and expands the variety of crops that can be raised. Irrigation has commonly been a public responsibility since it often involves large investments and infringes on tenure rights.

Both irrigation and technology are closely related to the use of fertilizer and other chemicals to restore fertility, improve the crops, and eliminate weeds and pests. Chemicals may be difficult to apply without irrigation, and neither may be practical unless technology has advanced beyond relatively primitive methods. Despite the potential benefits, improvement of the pattern of cultivation may be delayed by traditional attitudes in rural areas, or by the lack of skills to operate more advanced technology, or by scarcity of capital. Another difficulty is that changes in the pattern of cultivation are usually long-term investments that may be too slow to satisfy immediate pressures for reform.

Measures that supplement land reform

Many things need to be done outside the immediate sphere of agriculture if reform is to be successful. Probably the most important involve credit, marketing, and education. Unless the farmer is able and ready to take advantage of new opportunities and unless his product can be marketed profitably, reform efforts may be of no use. Costly or inaccessible credit, and the excessive charges of middlemen, increase the relative costs of farming. Supervised credit, subsidies, and low interest loans help to remove the need for traditional sources of credit; credit cooperatives help to ease the problems of financing. Similarly, marketing cooperatives and public supervision that reallocate the middleman's share back to the farmer help him realize greater benefits from the market.

Finally, improvements in general education are essential in any reform that involves the modernization of agriculture. These include literacy programs, the teaching of home economics, and vocational training, which is especially important in helping the young and unemployed and in providing skilled labour for industry.

EVALUATION AND CRITERIA OF SUCCESS

The difficulties of appraising agricultural reforms

Agricultural reform is a complex process of directed change, and its effects touch society in many ways. Any specific reform program may be difficult to evaluate because its various social, political, and economic objectives may be inconsistent; even its supporters may have different ideas about it. Moreover, there are no generally accepted criteria for determining the success of such a program, nor adequate tools for measuring its progress.

Economic criteria. Economic indicators are the easiest to use since the conditions of economic development are more readily identified than the conditions of social and political development. Economic development may be defined as a sustained increase in per capita real income. To be sustained, the rise of per capita real income must be accompanied by changes in the economic and social structures of society. From an economic standpoint, development requires an increase in total investment (capital formation), higher productivity, and full employment.

Capital formation, or investment, in agriculture implies that more resources will be put at the disposal of the farmer in the form of technology, fertilizers, and irrigation facilities, all of which contribute to his productive capacity and productivity. Since capital formation depends greatly on domestic saving and income, however, a higher rate of capital formation may thus be an indicator of the success of the reform in aiding economic development.

Another indicator is change in agricultural productivity. A rise in productivity implies higher efficiency, better use of resources, and an advance in the state of technology. It also suggests an increase in the level of income and a higher potential for saving and investment. In the final analysis, change in productivity may be the most important single indicator of the contribution made by reform to economic development. Change in the level of rural employment or unemployment provides another indicator.

Finally, an important indicator may be the change in agriculture's responsiveness to the demands of the nonagricultural sectors. The ability of agriculture to provide labour, food, industrial crops, and a market for industrial products is a significant measure of its contribution to industrialization.

Social criteria. Social and political accomplishments are more difficult to measure. One of the important indicators may be the degree of peasant participation in society, relative to other groups, in matters such as voting, representation, and decision-making. Change in social and political stability is another indicator. Stability is defined in this context as the ability of a government to change by constitutional and nonviolent means and to maintain peace and continuity of the social and political order without resort to force.

Attempts to measure changes

But these various indicators can only suggest that change has taken place. Evaluation of the results depends on the measurement and interpretation of change. Although it is usually tempting to describe a reform as successful or unsuccessful in general terms, it is more realistic to evaluate it according to each indicator separately, since the results are not necessarily additive. A negative impact on economic development may or may not be offset by higher stability, depending on the system of evaluation. Nor is it easy to measure change; some results may be intangible, and in some cases data on prereform conditions may not be accessible and hence the change not measurable. Assuming that measurement is feasible, however, evaluation of the results may follow one of two approaches: that of maximization or that of relative gain.

The maximization approach considers reform successful only if maximum benefits have been derived in the area under consideration. For example, a reform would be successful in increasing capital formation if any further effort would result in a decline of net investment. This approach has been applied to political and social results as well, such as the maximization of votes, equality, and social participation. The maximum is achieved when the cost of additional results is equal to the additional benefits derived.

The alternative approach is to measure success by the amount of gain or improvement over the prereform conditions. Thus, while the maximization approach asks how close to the ideal the results have been, the relative gain approach asks how much improvement has taken place. By this approach a reform would be successful to the extent to which capital formation has been encouraged and net investment increased. This is difficult to apply because of differences in underlying conditions. A 20 percent rise in capital formation in a capital-poor economy may be more significant than a 30 percent rise in a capital intensive economy. Similarly a 20 percent decline in the number of dissatisfied peasants in a peaceful or democratic society may be more conducive to stability and harmony than a 30 percent decline in a radical or violent society. For this reason the relative gain approach can be meaningful only if a critical minimum achievement is specified as a criterion of success in each specific situation. For instance, if the critical minimum is a 20 percent increase in per capita income in one economy and 30 percent in another, an increase of 20 percent in the first and 30 percent in the second would have equal significance as measures of success.

Whichever of these approaches is applied, it is still necessary to consider the subjective evaluation of reform programs. The degree to which the actual results approximate the expectations of the parties concerned will determine the success of a reform in promoting political stability. The reform may be considered successful if the peasants are satisfied that they achieved as much as possible and the landlords that they did not lose more than they had to. Thus, political stability may be another indicator of the success of a reform.

Paradoxes of reform

Inadequacies of the indicators. Whether the criteria used to measure the success of reform are the objective economic indicators or the more problematical social and political effects, there are inherent contradictions that tend to render the indicators vague and inconclusive. Several of these paradoxes are worth noting.

First, land redistribution often implies withdrawal of the landlord's contribution to the functioning of the farm. His managerial ability is no longer available, and his contributions to investment, credit, and general farm security are lost. Therefore, unless substitute sources of credit and investment capital are provided, land redistribution may retard rather than advance economic development. Further, while redistribution creates security, it may inhibit mobility of the new owners. A stake in the land tends to reduce the incentive to seek opportunity elsewhere, even though the new farm may be below optimum size and the farm family underemployed.

Second, reformers often must choose between breaking up efficiently run farms or reducing the land available for redistribution. Some efficiently run farms are exempted from redistribution; in other cases they are taken over as state farms, and the landless farmers are hired permanently as wage workers. Or the estate is broken up and the smaller units are pooled together in the form of a cooperative. Each of the alternatives has its own weaknesses—*e.g.*, the problem of how to weigh the economic gains of cooperative or state farming against loss of independence and freedom of decision making or the hope of becoming an owner.

Third, redistribution is often accompanied by compensation of landlords for expropriated property. Compensation tends to reduce the equalizing effects of reform since the relative position of the compensated landlord is little affected. Compensation payments may be lower than the market value or their real value may be depreciated by inflation, but these are uncertain results and may be difficult to account for in evaluating the reform.

Fourth, though it is usually hoped that reform will encourage saving, it is a fact that the farmers tend to spend the extra income on consumption, which may be unavoidable so long as farmers' incomes are relatively low and they prefer consumption to saving.

Fifth, reform tends to invite social and political participation prematurely. Participation, particularly after many years of backwardness, requires a degree of readiness and awareness that is slow in developing. Yet the farmers are placed in a position where immediate participation is expected. Hence the dilemma: delayed participation could be interpreted as reform failure, and premature participation could spell chaos. Shortcut measures to speed up participation often aggravate the problem.

Finally, land redistribution without preventive measures to preclude future concentration of ownership might simply prepare the way for future crises. Yet, preventive measures such as restrictions of land division, a ceiling on holding size, or regulation of land transfer infringe on the rights of ownership and may also inhibit the efficient owner from expanding his operations. These paradoxes have sometimes been resolved by subjecting private interests to public interest. But this depends on the ideology of the reformers and their ability to enforce the regulations.

HISTORY OF LAND REFORM

The ideas and principles discussed so far may be illustrated by a selective survey of the history of land reform movements, emphasizing the objectives and processes of reform, the difficulty of evaluating it, and the evolution of reform.

Ancient reforms. The recorded history of reform begins with the Greeks of the 6th century BC. Land in ancient Athens was held in perpetuity by the tribe or clan, with individual holdings periodically reallocated according to family size and soil fertility. Population increase, expansion of trade, growth of a money economy, and the opening up of new business opportunities eventually made financial transactions in land an economic necessity. Land itself continued to be inalienable, but the right to use the land could be mortgaged. Thus, peasants could secure loans by surrendering their rights to the product of the land, as "sale with the option of redemption." Lacking other employment, the debtor continued to cultivate the land as *hektdmor*, or sixth parter, delivering five-sixths of the product to the creditor and retaining the rest for himself. Mortgaged land was marked by *horoi*, or mortgage stones, which served as symbols of land enserfment. When Solon was elected archon or chief magistrate, *c.* 594 BC, his main objective was to free the land and destroy the *horoi*. His reform law, known as the *seisachtheia* or "shaking-off the burdens," cancelled all the debts, freed the *hektēmoroi*, destroyed the *horoi*, and restored land to its constitutional holders. Solon also prohibited the mortgaging of land or of personal freedom on account of debt.

The impact of the reform was extensive but of short duration. The *hektēmoroi* were freed, but since no alternative sources of support or credit were provided and creditors were uncompensated, dissatisfaction and instability continued to prevail. Two decades of anarchy were followed by a revolution, *c.* 561 BC, that brought Peisistratus to power, who enforced the reform and distributed lands of his adversaries (who were killed or in exile) among the small holders. He also extended loans to aid cultivation and prevent migration to the city and expanded silver mining to create employment. Although the amount of land redistributed is unknown, Peisistratus was apparently able to satisfy the peasantry, secure their loyalty, and stay in power for life. The reform seems to have been an effective mechanism to sustain tyranny or nonconstitutional rule, but the economic effects are too vague to evaluate.

The reforms of the Gracchi

The Roman reform by Tiberius and Gaius Gracchus came between 133 and 121 BC. The land reform law, or *lex agraria*, of Tiberius was passed by popular support against serious resistance by the nobility. It applied only to former public land, *ager publicus*, which had been usurped and concentrated in the hands of large landholders. Land concentration reduced the number of owners and hence the number of citizens and those eligible to serve in the army. In addition, such concentration was accompanied by a shift from cultivation to grazing, which reduced employment and increased the poverty of the peasants, producing a crisis. The motives of the reformers have been debated, but it would appear that sympathy with the poor and concern for political stability were the factors that led the Gracchi to attack usurpation and return the public domain to small cultivators.

The *lex agraria* specified minimum and maximum individual land holdings, with an allowance for male children of the family. Excess land would be expropriated and compensation paid for improvements. A standing *collegium*, or commission, was to enforce the law, but implementation was delayed because Tiberius was killed the same year of its passage. When Gaius was elected tribune about a decade later, he revived the reform and went even farther. He colonized new land and abolished rent on small holdings to put them in the same position as large holdings on which rent had been suspended as compensation for expropriation. However, Gaius was killed in 121 BC and within a decade the reform was reversed: private acquisition of public land was legalized, the land commission dissolved, rent on public land abolished, all holdings declared private property, and squatting on public land prohibited. Even colonization was ended and colonies established by Gaius were broken up. The era of reform came to an end and another period of land concentration was inaugurated.

Modern European reforms. The French Revolution brought a new era in the history of land reform. Reform meant dealing with survivals of the medieval tenures that had left a common heritage in most European countries and, through them, in the colonies. The measures and approaches varied from place to place and period to period.

The French Revolution and after

On the eve of the Revolution, French society was polarized, with the nobility and clergy on one side and the rising business class on the other. The middle class was relatively small, especially in the rural areas. The majority of the peasants were hereditary tenants, either *censiers*, who paid a fixed money rent, or *mainmortables*, or serfs, who paid rent in the form of labour services, *corvée*, of about three days a week. The peasants paid various other burdensome feudal dues and taxes, from which the nobility and clergy were exempted. Peasant poverty, which was intolerable enough, was exacerbated by their servile status. The Revolution overthrew the *ancien régime* and the feudal order, and introduced land reform.

The reform repealed feudal tenures, freed all persons from serfdom, abolished feudal courts, and cancelled all payments not based on real property, including tithes. Rents based on real property were redeemable. Once the law had been passed, however, the peasants seized the land and refused to pay any rents or redemption, and in 1792 all payments were finally cancelled. Land of the clergy and political emigrants was confiscated and sold at auction, together with common land. The terms of sale, however, often favoured the wealthy, which may explain the rise of a new class of large landowners among the supporters of Napoleon.

The social and political objectives of the reformers were fully realized. The *censiers* and serfs became owners. Feudalism was destroyed, and the new regime won peasant support. The economic effects, however, were limited. Incentives could not be increased substantially since the peasants already had full security of tenure prior to the reform. The scale of operations was not changed; and no facilities for credit, marketing, or capital formation were created. The major achievement was perpetuation of the small family farm as a basis of democracy, a contribution of the Enlightenment and the *Code Napoléon*. The small family farm has characterized French agriculture ever since.

There were other reforms in most European countries. England resolved its land problems by the enclosure movement that drove the small peasants into the towns, the consolidation of holdings, large scale operation, and private ownership. Sweden and Denmark pioneered between 1827 and*1830 by peacefully abolishing village compulsion, or imposed labour service, and the strip system of cultivation and by consolidating the land and dividing the commons among the peasants. Though influenced by the French Revolution, only after the 1848 revolutions did Germany, Italy, and Spain free the peasants and redistribute the land. Reform in Ireland took a whole century before substantive results were achieved, in the mid-1930s, after Ireland was divided into Northern Ireland and the Irish Free State. The tenants were converted into owners by subsidized purchase of the land.

Russian reforms

The first major Russian reform was the emancipation of the serfs in 1861. At the time of emancipation about 45 percent of the land was private property and the remainder was held as allotment land, cultivated in units averaging 9.5 acres by the peasant serfs against rent in kind and labour, payable to feudal lords. In contrast, fewer than 1,000 noble families owned about 175,000,000 acres and received rent therefrom. Conflict between such extremes of poverty and wealth, the expansion of trade and commerce, and the pressure of population caused restlessness among the peasants and rendered reform inevitable. As Tsar Alexander II put it: "It is better to abolish serfdom from above than to await the day when it will begin to abolish itself from below."

The Emancipation Act of 1861 abolished serfdom and distributed allotment land among the peasants. The homestead became hereditary property of the individual; but the field land was vested in the village *mir* as a whole. The peasant paid redemption through the village authority, while the landlord received state bonds as compensation equal to 75–80 percent of the land market value. Though legally freed, the private serf had to ransom his freedom by surrendering a part of the allotment land. In contrast, serfs belonging to the royal family were emancipated in 1863 and received the maximum amount of land fixed by law. Serfs belonging to the state were emancipated in 1866 and allowed to keep the land they occupied in perpetual tenure against money rent. The Cossacks received two thirds of the land, to be held in common, but instead of redemption they had to serve 20 years in the army. The serfs in mines and households were freed but received no economic assets.

Redemption payments, however. soon proved too burdensome, village restrictions were tight, and the allotment land area declined; all of which led to renewed restlessness and disturbances. Following the revolt of 1905, the government, under Pyotr Stolypin, tried to create middle class, independent farmers by replacing the village tenure with private ownership, consolidating holdings, and encouraging land purchase by individuals; but the time was too short for effective implementation. The Soviet Revolution overthrew the tsarist regime and introduced the concepts of public ownership and collectivization.

By decree in 1918, the Soviets abolished private ownership of land, made farming the sole basis of landholding, and declared collectivization a major objective of policy. Marketing of agricultural products became a state monopoly. In 1929 Stalin was able to embark on a full course of collectivization, which he pursued by various means and at high cost. By 1938 collective farms occupied 85.6 percent of the land and state farms 9.1 percent. Credit facilities and tractor stations supplemented collectivization, while agricultural production was integrated in the national plan for industrialization and development.

The costs of Soviet reform included the destruction of capital and the death of large numbers of peasants. Nevertheless, farm operation was enlarged, sometimes to the point of inefficiency; agriculture was coordinated with nonagricultural sectors; and mechanization was advanced sufficiently to release labour for the industrial sector. Total output and productivity increased; and capital formation was made possible through forced saving, taxes, and regulated prices. The peasant received health care, education, and better labour conditions. The objectives of the decree of 1918 have been fully realized.

Reform in eastern Europe

Reform in eastern Europe was complicated by the fact that most of the east European countries remained under foreign rule until the middle of the 19th century or later. In Hungary, the Decree of 1853 abolished the *robot,* or forced labour and feudal dues, freed the serfs, liberalized land transaction, and encouraged consolidation. The Romanian reform of 1864 freed the serfs and distributed both the land and the redemption payments in proportion to the number of cows or oxen each peasant had. Formal emancipation in Bulgaria was introduced by the Turkish government in the 1850s; but actual reform came in 1880, after independence. Each peasant, including sharecroppers and wage workers, who had worked the land for ten years without interruption, was entitled to the land he had cultivated. With the exception of Bulgaria, the distribution of ownership throughout most of eastern Europe remained highly uneven. Political instability reached a dangerous point between the two world wars. Following World War II, the east European countries acquired Communist governments with a strong tendency toward collectivized and mechanized agriculture.

Mexico. The Mexican reform of 1915 followed a revolution and dealt mainly with lands of Indian villages that had been illegally absorbed by neighbouring *haciendas* ("plantations"). Legally there was no serfdom; but the Indian wage workers, *peones,* were reduced to virtual serfdom through indebtedness. Thus, the landlords were masters of the land and of the *peones.* The immediate aim of reform was to restore the land to its legal owners, settle the title, and use public land to reconstruct Indian villages. The motives were mainly to reduce poverty and inequality and to secure political stability, which was then in the balance. A decree of 1915 voided all land alienations that had taken place illegally since 1856 and provided for extracting land from *haciendas* to re-establish the collective Indian villages, ejidos. The 1917 constitution reaffirmed those provisions but also guaranteed protection of private property, including *haciendas.* Nevertheless, a combination of loopholes, litigation, and reactionary forces slowed implementation; and effective reform came only after passage of the Agrarian Code of 1934 and the sympathetic efforts of President Lázaro Cárdenas.

The reform restored many villages and freed the *peones,* but land concentration and poverty continued. In 1950, more than 31 percent of the private cropland was owned by fewer than half of one percent of the owners. Small-scale operation was retained or encouraged, a fact explaining the decline of output in the early years. More recently, efficiently run farms have been exempted from distribution.

The social and political impact was more positive. The peasants acquired more land and liberty, and control by landlords was reduced, although it was replaced by village restrictions. At least legally, farming became the basis of landholding. Some have seen in land reform the reason for Mexico's political stability, although there have been sporadic peasant uprisings as recently as the mid-1960s.

REFORMS SINCE WORLD WAR II

Reforms for the purpose of economic development

Recent decades have witnessed widespread, comprehensive land reform programs in a number of countries. Reforms have been oriented toward economic development, with guidance and aid from the United Nations and various outside countries. The trend can be only selectively illustrated.

Japan. The Japanese reform came after the war at the insistence of the Allied Occupation Army. The reform was designed by experts to suit the uniquely high literacy rate and advanced industrialization of the country. Although the Meiji government had formally abolished feudalism and declared the land to be the property of the peasants, usurpation of land by the rich and by moneylenders had created classes of perpetual tenants and absentee landlords. In 1943, 66 percent of the land was operated by tenants against rent in kind that averaged 48 percent of the farmers' product. The trend was strengthened by population pressure on scarce land resources, which also resulted in fragmentation of holdings. The social class structure was closely tied to tenure, the owners in each village being at the top of the structure. Conflict between landlords and peasants was widespread.

After the war, the crisis was revived by food shortages, the breakdown of the urban economy, and the return of absentee landlords to the land. The Occupation Army insisted on reform, presumably to democratize the society and rehabilitate the economy. The reform law of 1946 established a ceiling on individual holdings and provided for expropriation and resale of excess land to the tenants against long-term payments. The government compensated the landlords in cash and bonds redeemable in 30 years. Tenants who remained were protected by contract, and rents were reduced to a maximum of 25 percent of the product. The redistributed land was made inalienable, though this restriction was relaxed four years later. The program also provided for marketing and credit cooperatives. An important supplementary measure was the Local Autonomy Law of 1947; this reform law decentralized the power structure and put village affairs in the hands of the villagers.

Within two years tenancy declined by more than 80 percent. Rent control and land distribution helped to equalize incomes in the villages and restore the sociopolitical status of the peasants. Crop yields per unit of land increased, but despite improved techniques the yields per man declined. In general the reform seemed to realize the objectives of the reformers and the peasants, although smallness of scale, low per capita incomes, underemployment, and insufficient mechanization have persisted. Even black market rents have developed. These problems were tolerable in view of the upsurge of the urban economy and the ability of the Japanese farmer to supplement his income from nonagricultural employment.

Egypt. The Egyptian reform of 1952 followed the revolution that overthrew the monarchy and brought young middleclass leaders to the helm. It was the most comprehensive reform outside the Communist countries. Though affecting only about 12 percent of the arable land, it was applied thoroughly and touched all aspects of rural life. Egypt had two main forms of tenure: private ownership and *waqf* or land held in trust under a special ministry and dedicated to charitable or educational purposes. *Waqf* land was inalienable, but private land was subject to speculation and concentration. In 1950, 1 percent of the owners had more than 20 percent of the private land, and seven percent had more than two-thirds. The operating unit was small, with 77 percent of all the holdings occupying less than one acre each. Tenancy was widespread and rents exorbitant. The peasants were exploited by middlemen who sublet the land to tenants, mediated between them and the market, and extended credit at high rates of interest.

The revolutionary reformers aimed at abolishing feudalism, recruiting peasant support, promoting economic development, and bringing the villagers back into the stream of national life. The Agrarian Reform Law of 1952 put a ceiling on individual holdings at 200 *faddān* (1 *faddān* = 1.038 acres), later reduced to 100 *faddān*, with special allowance for male children. The excess land was expropriated and distributed to the peasants in parcels not exceeding five *faddān*. Compensation was given in bonds, while land recipients had to repay in annual installments. The new owners were obligated to join cooperatives for production, marketing, and credit. Tenancy conditions were also regulated, with contract replacing traditional terms; rent could not exceed 50 percent of the product nor could a tenant hold more than 50 acres, to avoid subletting. An interesting feature of the reform was the special attention given to college graduates by allowing them up to 20 *faddān* parcels.

The Reform Law of 1952

The reform was enforced quickly and had a great impact on the morale of the peasants. The economic effects, however, were minor since agriculture had already been intensive and land yield high. Producer cooperatives served only to offset the impact of distribution on the scale of operation. Some increases in yield have been claimed, but the evidence is insufficient. Furthermore, little capital was redirected into productive investment since the compensation bonds were not negotiable. Peasant savings remained limited, income increments being spent mostly on consumption. Finally, underemployment in agriculture has remained widespread.

The social and political effects, however, were far reaching. Redistribution and regulation of rent raised the incomes of small owners and tenants. Cooperatives replaced the middleman and captured his share for the farmer. The peasant gained social status and enjoyed a higher level of political participation, mostly in support of the revolutionary regime. These effects, however, can be easily exaggerated. The peasants became dependent on the cooperatives whether they liked them or not. Great differences in landholding continued to exist, and peasant incomes remained low relative to urban incomes. Black market rents appeared. The example of Egypt suggests that successful land reform requires an upsurge in the industrial sector of the economy.

Southeast Asia. The model of Japan's reform has been followed in Southeast Asia, especially in Taiwan and South Vietnam. Both were influenced by American experts and by the anti-Communism of the governments. The objectives in both have been to sustain the political order, raise living standards, and promote some degree of economic development. Both reforms began with regulation of tenancy, restriction of rent, and the institution of written contract for leases, following which tenants were to be-transformed into owners. Taiwan's reform was implemented between 1949 and 1953, in three stages. First, rents, which had sometimes reached 70 percent of the product, were reduced to 37.5 percent. Next, tenant-farmed public land was sold to the tenants. When this was done, tenant-farmed private land was bought by the government and resold to the tenants.

Vietnamese reform

The Vietnamese reform was introduced in 1955. Rents were reduced to a maximum of 25 percent of the product. However, a ceiling of 100 hectares was put on individual holdings, and only the excess land was subject to redistribution in parcels of three to five hectares to the tenants. Reforms in both Taiwan and South Vietnam have been formally supplemented by other measures described as community development (more thoroughly implemented in Taiwan), including adult education, credit facilities, improved technology, and other social services. Some consolidation was attempted in Taiwan, but in both countries the scale of operation was little affected. The main effect seems to have been the regulation of tenancy and the redistribution of rent incomes. An innovation of Taiwan's reform was the partial compensation of landlords with industrial shares in public enterprises.

Taiwan's reform has been hailed as a major success, both in economic and political terms. Some observers, however, are unwilling to reach such a conclusion until the peasants have a free choice of tenure and farm organization. As in Vietnam, however, such a choice may be a long way off, given the political conflict in the region.

The Philippines introduced a reform program in 1963 with emphasis on economic development. The reform aimed primarily at replacing share tenancy with lease contracts and eventually with ownership, and at revitalizing agriculture through extension services. The reform was intended to increase credit facilities, create cooperatives, and establish social service clubs. The main effects initially were seed improvement, greater use of fertilizers, and an increase in contractual tenancy.

Various other reforms have been introduced in Southeast Asia. In Malaysia, Thailand, and Indonesia the emphasis has been on settlement of landless people on virgin public land. In Indonesia attempts have been made to shift people from relatively crowded to sparsely populated islands. In Thailand various approaches to settlement have been attempted, including cooperative and government-sponsored programs, although self-help has been emphasized. Fewer than 25,000 persons were settled between 1956 and 1965.

Reform in Malaysia

In contrast, the program in Malaysia has been highly organized and oriented toward economic development. It combines social and economic objectives with emphasis on the production of rubber and palm oil for export. The program gradually transforms the landless into hereditary tenants. A typical project covers 4,500 to 5,000 acres of jungle land and absorbs about 400 families. The land is cleared and planted by contract, and a village is constructed, with all the necessary services, before the settlers arrive. Each house has a quarter of an acre for a household garden. Crop land is divided in blocks of 120 to 200 acres, to be worked by a team of 15 to 25 people until the plants have matured. Upon maturation, each settler receives a share by lottery and a lease title for 99 years. This tenure arrangement precludes alienation, subdivision, or subleasing; it thus sidesteps the Islamic laws of inheritance and opens the door for institutional change in society at large.

The settler is responsible for the cost of clearing and planting, but the government pays the administrative costs. The settler is guaranteed supplementary employment to earn subsistence income pending maturity of the plants, and cultivation is supervised and guided by experts. The rate of settlement is determined by the overall economic plan. During the First Five Year Plan, 1957–60, 3,300 families were settled. During the second, 1961–5, 9,500 families, or 57,000 people, were settled. Although it is too early to measure the success or failure of the reform, it is already clear that land holding has become tied to cultivation; fragmentation and diseconomies of scale have been avoided, and cultivation has become a rational economic operation. The Malaysian program has much in common with the cooperative settlements of Israel, and also resembles the Gezira Scheme in Sudan.

Latin America. Except for the early example of Mexico, reform in Latin America has been recent, and appears to have come only in response to the threat of social and political instability and mounting international pressures. It must be seen against a background of rapidly increasing population; extreme contrasts between plantation economies and small units; high concentration of land ownership, income, and power; relatively backward cultivation methods; and extensive foreign ownership of land. Latin American society is also complicated by its ethnic mixtures and its high dependence on staple trade items such as sugar, tobacco, cocoa, coffee, and beef cattle.

As reform in Latin America has come in one country after another, each program has reflected the differing ideologies and objectives of the national leaders. Brazil has had several attempts at reform, the latest and most comprehensive of which was initiated in 1964. The measures have been indirect and relatively mild, the most important being taxation of idle land and large plantations. The rates of tax are progressively higher as the size of the farm increases, on the assumption that wealth redistribution will follow. Actual redistribution of land has received a low priority. Provisions for credit, tenancy protection, and colonization were made; but as of 1970 they had not been strongly implemented.

Reform in Cuba

At the other end of the Latin-American spectrum is the Cuban reform that followed the revolution of 1958. Cuba retained private ownership, but reduced it effectively in favour of the public sector. As proclaimed a few months before the overthrow of the old regime, the reform aimed at the elimination of *latifundia* tenure, expropriation of land owned by foreign companies, higher standards of living for the peasantry, and national economic development. It began by setting a ceiling of 30 *caballerias* (1 *caballaria* = 13.4 hectares) on individual holdings, with a maximum of 100 *caballerias* if economic operations required such a scale. All foreign-owned land was nationalized. Public land on which rice and cattle were raised was converted into state farms, and the peasants became permanent wage workers on these farms. Sugar plantations were converted into cooperatives to avoid their subdivision into small uneconomical units. Before long the ceiling on individual holdings was lowered to five *caballerias* and all such holdings became private family farms. The rest were nationalized, and the expropriated owners compensated with a pension for life. The reform was supplemented by the organization of national farmer associations; people's stores; credit, housing, and educational facilities; and the supply of machinery and fertilizers. In 1963 a major reorganization of state farms took place; they were subdivided on the basis of crop specialization into smaller operational units of about 469 *caballerias.*

Effects of the reform were comprehensive and immediate. The tenure institutions were radically changed in favour of public ownership, while *minifundia* and tenancies were abolished. Socially and politically, the reform realized the objectives of the reformers. Economically, the government claimed higher yields of sugar cane, vegetables, and fruit; but this was disputed by foreign observers.

Other Latin American reforms have fallen in a range between those of Brazil and Cuba. For example, that of Costa Rica has overlooked land concentration and income inequality and concentrated on the squatters or *parasitos,* who in 1961 numbered between 12,000 and 16,000 people. The reform aimed at legalizing existing squatter holdings, preventing further squatting, and conserving virgin land. Even this modest program was implemented very slowly. Colombia has had reform programs for at least 30 years; but concentration of ownership, fragmented small holdings, and backward methods of cultivation are still the rule rather than the exception.

Chile

Chile undertook various reform programs before achieving concrete results. In 1962 a program was enacted to encourage settlement of new land, but only about 1,000 families were settled. A comprehensive agrarian reform was introduced in 1965 with three main objectives: to make the agricultural workers owners of the land they had cultivated previously, to increase agricultural and livestock production, and to facilitate social mobility and peasant participation in political life. The Chilean reform is unique in its method of implementation. Once the plantation has been designated for expropriation and the prospective owners selected, they are organized into ***asentamientos,*** or settlement groups. The group elects a committee to take charge of settlement. The members cultivate the land as a team for three to five years. Meanwhile they receive training and guidance in social participation, decision making, and modern farming. Upon completion of the transition period, the land is divided among those who have shown promise, to be held outright and without restriction. All new owners are obligated to join cooperatives, the form of these being determined by the members. Between 1965 and 1969 more than 15,000 families were settled on 2,000,000 hectares expropriated by the reformers.

Observers of the Latin American scene have been pessimistic regarding the adequacy of these land reform programs. With the exception of Cuba, capital formation in agriculture has not substantially increased; the pattern of land distribution has undergone little change; and social and political stability has remained in question. The agrarian structure is still considered defective both from the point of view of the peasants and for purposes of economic development.

India and Pakistan

Other recent reforms. The tendency to reform the agrarian structure has been manifest in nearly all countries. India and Pakistan have concentrated on abolishing intermediaries who existed as survivals of traditional and feudal tenures. In India the tenants have become hereditary holders, with the title vested in the state. India has left reform to the individual states and emphasized peaceful and compensatory methods; hence the results have varied from one state to another. Pakistan, following the revolution of 1958, enacted a reform that made most of the tenants owners. In both countries, however, small-scale farming has persisted, and Pakistan has tolerated and protected owners of up to 500 acres. In neither country has fragmentation been effectively reduced or capital formation and modern cultivation methods significantly advanced.

China has eliminated private ownership and organized the peasants in village communes. Extensive supplementary measures have been spoken of, but the role and organization of the commune have varied according to the pressures upon the economy. Iran's reform made owners of most of the former sharecroppers. The scale of operation and the methods of cultivation, however, were not affected. Capital formation among the Iranian peasants has been limited since the small increases in their incomes have been spent mainly on consumption. Iran's reform has apparently succeeded in sustaining the existing social order.

Reform was introduced in Syria and Iraq following their revolutions, but the status of these reforms remains indeterminate. Reforms in Africa have been concerned mainly with title settlement and colonization of new land. Many African countries have tended toward some form of socialism in tenure arrangements. In most countries where partial reforms have been carried out the agrarian structure still suffers from defects that obstruct economic development and nourish social and political instability.

BIBLIOGRAPHY. The following references form a point of departure for further study; each of these, in turn, provides additional bibliographies.

A. WHITNEY GRISWOLD, *Farming and Democracy,* new ed. (1952), a lucid scholarly analysis of the political significance of various farm structures and a discussion of the family farm as a cornerstone of democracy; FOLKE DOVRING, *Land and Labor in Europe, 1900–1950: A Comparative Survey of Recent Agrarian History* (1956), an examination of the political and ideological bases of land policy since the French Revolu-

tion; D. WARRINER, *Land Reform in Principle and Practice* (1969), a provocative discussion of the evolution of reform and the conflict between theory and practice; E.H. JACOBY, *Evaluation of Agrarian Structures and Agrarian Reform Programs* (1966), a checklist of administrative and organizational changes that contribute to reform success; E.H. TUMA, *Twenty-Six Centuries of Agrarian Reform: A Comparative Analysis* (1965), a synthesis of theory and history with an attempt to formulate a general theory of agrarian reform; UNITED NATIONS DEPARTMENT OF ECONOMIC AFFAIRS, *Land Reform: Defects in Agrarian Structure As Obstacles to Economic Development* (1951), a brief but comprehensive survey of the relation between reform and development.

The ancient reforms may be approached through: I.M. LINFORTH, *Solon the Athenian* (1919); W.J. WOODHOUSE, *Solon the Liberator: A Study of the Agrarian Problems in Attica in the Seventh Century* (1938); W. WARDE-FOWLER, "Notes on Gaius Gracchus," *English Historical Review,* 20:209–227, 417–433 (1905); and E.G. HARDY, "Were the Lex Thoria of 118 B.C. and the Lex Agraria of 111 B.C. Reactionary Laws?," *Journal of Philology,* 31:268–286 (1910).

Reforms following the French Revolution may be approached through the following references: G. LEFEBVRE, *Quatre-vingt-neuf* (1939; Eng. trans., *The Coming of the French Revolution, 1789,* 1947); M.L.B. BLOCH, *Les caractères originaux de l'histoire rurale française, 2* vol. (1952–56; Eng. trans., *French Rural History: An Essay on its Basic Characteristics,* 1966); G.T. ROBINSON, *Rural Russia Under the Old Regime: A History of the Landlord-Peasant World and a Prologue to the Peasant Revolution of 1917* (1949); M.H. DOBB, *Soviet Economic Development Since 1917* (1948); N. JASNY, *The Socialized Agriculture of the U.S.S.R.: Plans and Performance* (1949); F. TANNENBAWM, *Mexico: The Struggle for Peace and Bread* (1950); C.O. SENIOR, *Land Reform and Democracy* (1958), mainly a study of Mexico's reform, but including discussions of concepts and theory; and E.H. TUMA, *Economic History of Europe* (forthcoming).

Any study of recent decades must begin with the UNITED NATIONS publications, especially *Progress in Land Reform,* Reports 1–4, and papers delivered at the World Land Reform Conference in Rome, 1966. These provide the latest information on the largest number of countries. An excellent reference which forms a background for recent reform is K.H. PARSONS, P.M. RAUP, and R.J. PENN (eds.), *Land Tenure: Proceedings of the International Conference on Land Tenure and Related Problems in World Agriculture* (1956).

(E.H.T.)

Language

Many definitions of language have been proposed. Henry Sweet, an English phonetician and language scholar, stated: "Language is the expression of ideas by means of speech-sounds combined into words. Words are combined into sentences, this combination answering to that of ideas into thoughts." The U.S. linguists Bernard Bloch and George L. Trager formulated the following definition in their *Outline of Linguistic Analysis* (1942): "A language is a system of arbitrary vocal symbols by means of which a social group cooperates." Definitions like these and, indeed, any succinct definition make a number of presuppositions and beg a number of questions. The first, for example, puts excessive weight on "thought," and the second uses "arbitrary" in a specialized, though legitimate, way (see below).

A number of considerations enter into a proper understanding of language as a subject:

1. Every physiologically and mentally normal person acquires in childhood the ability to make use, as both speaker and hearer, of a system of vocal communication that comprises a circumscribed set of noises resulting from movements of certain organs within his throat and mouth. By means of these he is able to impart information, to express feelings and emotions, to influence the activities of others, and to comport himself with varying degrees of friendliness or hostility toward persons who make use of substantially the same set of noises.

2. Different systems of vocal communication constitute different languages; the degree of difference needed to establish a different language cannot be stated exactly. No two people speak exactly alike; hence, one is able to recognize the voices of friends over the telephone and to keep distinct a number of different unseen speakers in a radio broadcast. Yet, clearly, no one would say that, for that reason, they speak different languages. Generally, systems of vocal communication are recognized as different languages if they cannot be understood without specific learning by both parties, though the precise limits of mutual intelligibility are hard to draw and belong on a scale rather than on either side of a definite dividing line. Substantially different systems of communication that may impede but do not prevent mutual comprehension are referred to as dialects of a language. In order to describe in detail the actual different speech patterns of individuals, the term idiolect, meaning the speech habits of a single person, has been coined.

Languages, dialects, and idiolects

3. Normally, people acquire a single language initially—their first language, or mother tongue, the language spoken by their parents or by those with whom they are brought up from infancy. Subsequent "second" languages are learned to different degrees of competence under various conditions, but the majority of the world's population remains largely monolingual. Complete mastery of two languages is designated as bilingualism; in a few special cases—such as upbringing by parents speaking different languages at home—speakers grow up as bilinguals, but ordinarily the learning, to any extent, of a second or other language is an activity superimposed on the prior mastery of one's first language and is a different process intellectually.

4. Language, as described above, is species-specific to man. Other members of the animal kingdom have the ability to communicate, through vocal noises or by other means, but the most important single feature characterizing human language (that is, every individual language), against every known mode of animal communication, is its infinite productivity and creativity. Human beings are unrestricted in what they can talk about; no area of experience is accepted as necessarily incommunicable, though it may be necessary to adapt one's language in order to cope with new discoveries or new modes of thought.

Animal communication systems are by contrast very tightly circumscribed in what may be communicated. Indeed, displaced reference, the ability to communicate about things outside immediate temporal and spatial contiguity, which is fundamental to speech, is found elsewhere only in the so-called language of bees. Bees are able, by carrying out various conventionalized movements (referred to as bee dances) in or near the hive, to indicate to others the locations and strengths of nectar sources. But nectar sources are the only known theme of this communication system. Surprisingly, however, this system, nearest to human language in function, belongs to a species remote from man in the animal kingdom and is achieved by very different physiological activities from those involved in speech. On the other hand, the animal performance superficially most like human speech, the mimicry of parrots and of some other birds that have been kept in the company of humans, is wholly derivative and serves no independent communicative function. Man's nearest relatives among the primates, though possessing a vocal physiology very similar to that of humans, have not developed anything like a spoken language.

Interaction of language with society

Language interacts with every other aspect of human life in society, and it can be understood only if it is considered in relation to society. This article attempts to survey language (both spoken and written) in this light and to consider its various functions and the purposes it can and has been made to serve. Because each language is both a working system of communication in the period and in the community wherein it is used and also the product of its past history and the source of its future development, any account of language must consider it from both these points of view.

The science of language is known as linguistics. It includes what are generally distinguished as descriptive linguistics and historical linguistics. Linguistics is now a highly technical subject; it embraces, both descriptively and historically, such major divisions as phonetics, grammar, and semantics, dealing in detail with these various aspects of language.

For a full account of the theory and methods of linguistic science, see the article LINGUISTICS. The articles PHONETICS; GRAMMAR; and SEMANTICS provide further information. For a general survey of known living and dead languages, consult the article LANGUAGES OF THE WORLD. Treatments of language families and individual languages, such as ROMANCE LANGUAGES; GERMANIC LANGUAGES; and BASQUE LANGUAGE, are also provided.

This article is divided into the following sections:

ATTITUDES TOWARD LANGUAGE

The practice of naming

As is evident from the introduction above, human life in its present form would be impossible and inconceivable without the use of language. People have long recognized the force and significance of language. Naming—applying a word to pick out and refer to a fellow human being, an animal, an object, or a class of such beings or objects—is only one part of the use of language, but it is an essential and prominent part. In many cultures men have seen in the ability to name an ability to control or to possess; this explains the reluctance, in several primitive and other communities, with which names are revealed to strangers and the taboo restrictions found in several parts of the world on using the names of persons recently dead. Lest it be thought that attitudes like this have died out in modern civilized communities, it is instructive to consider the widespread and perhaps universal taboos on naming directly things considered obscene, blasphemous, or very fearful. Indeed, use of euphemistic substitutes for words referring to death and to certain diseases actually seems to be increasing in some civilized areas.

Not surprisingly, therefore, several independent traditions ascribe a divine or at least a supernatural origin to language or to the language of a particular community. The biblical account, representing ancient Jewish beliefs, of Adam's naming the creatures of the Earth under God's guidance is well-known:

> So out of the ground the Lord God formed every beast of the field and every bird of the air, and brought them to the man to see what he would call them; and whatever the man called every living creature, that was its name (Gen. 2:19).

Norse mythology preserves a similar story of divine participation in the creation of language, and in India the god Indra is said to have invented articulate speech. In the much more sophisticated debate on the nature and origin of language given in Plato's Socratic dialogue *Cratylus,* Socrates is made to speak of the gods as those responsible for first fixing the names of things in the proper way.

A similar divine aura pervades early accounts of the origin of writing. The Norse god Odin was held responsible for the invention of the runic alphabet. The inspired stroke of genius whereby the ancient Greeks adapted a variety of the Phoenician consonantal script so as to represent the distinctive consonant and vowel sounds of Greek, thus producing the first alphabet such as is known today, was linked with the mythological figure Cadmus, who, coming from Phoenicia, was said to have founded Thebes and introduced writing into Greece. The Arabs had a traditional account of their script, together with the language itself, being given to Adam by God.

Biblical traditions concerning language

The later biblical tradition of the Tower of Babel (Gen. 11:1–9), whatever its exact place in ancient Near Eastern history, exemplifies three aspects of early thought about language: (1) divine interest in and control over its use and development, (2) a recognition of the power it gives to man in relation to his environment, and (3) an explanation of linguistic diversity, of the fact that people in adjacent communities speak different and mutually unintelligible languages, together with a survey of the various speech communities of the world known at the time to the Hebrews.

The origin of language has never failed to provide a subject for speculation, and its inaccessibility adds to its fascination. Informed investigations of the probable conditions under which language might have originated and developed are seen in the late-18th-century essay of the German philosopher Johann Gottfried von Herder, "Abhandlung über den Ursprung der Sprache" ("Essay on the Origin of Language"), and in numerous other treatments. But people have tried to go further, to discover or to reconstruct something like the actual forms and structure of man's first language. This lies forever beyond the reach of science, in that spoken language in some form is almost certainly coeval with *Homo* sapiens. The earliest records of written language, the only linguistic fossils man can hope to have, go back no more than about 4,000 or 5,000 years. Attempts to derive human speech from imitations of the cries of animals and birds or from mere ejaculations of joy and grief, as if onomatopoeia were the essence of language, were ridiculed for their inadequacy by the Oxford philologist F. Max Miiller in the 19th century and have been dubbed the bowwow and pooh-pooh theories.

On several occasions attempts have been made to identify one particular existing language as representing the original or oldest tongue of mankind, but, in fact, the universal process of linguistic change rules out any such hopes from the start. The Greek historian Herodotus told a story that King Psammetichus of Egypt caused a child to be brought up without ever hearing a word spoken in its presence. On one occasion it ran up to its guardian as he brought it some bread, calling out "bekos, bekos"; this, being said to be the Phrygian word for bread, proved that Phrygian was the oldest language of mankind. The naïveté and absurdity of such an account have not prevented its repetition elsewhere and at other times.

In Christian Europe the position of Hebrew as the language of the Old Testament gave valid grounds through many centuries for regarding Hebrew, the language in which God addressed Adam, as the parent language of all mankind. Such a view continued to be expressed even well into the 19th century. Only since the mid-1800s has linguistic science made sufficient progress finally to clarify the impracticability of speculation along these lines.

Language and thinking

When people have begun to reflect on language, its relation to thinking becomes a central concern. Several cultures have independently viewed the main function of language as the expression of thought. Ancient Indian grammarians speak of the soul apprehending things with the intellect and inspiring the mind with a desire to speak; and in the Greek intellectual tradition Aristotle declared, "Speech is the representation of the experiences of the mind" (*On Interpretation*). Such an attitude passed into Latin theory and thence into medieval doctrine. Medieval grammarians envisaged three stages in the speaking process: things in the world exhibit properties; these properties are understood by the mind of man; and, in the manner in which they have been understood, so they are communicated to others by the resources of language.

Rationalist writers on language in the 17th century gave essentially a similar account: speaking is expressing thoughts by signs invented for the purpose, and words of different classes (the different parts of speech) came into being to correspond to the different aspects of thinking.

Such a view of language continued to be accepted as generally adequate and gave rise to the sort of definition proposed by Henry Sweet and quoted above. The main objection to it is that it either gives so wide an interpretation to thought as virtually to empty the word of any specific content or gives such a narrow interpretation of language as to exclude a great deal of normal usage. A recognition of the part played by speaking and writing in social cooperation in everyday life has highlighted the many and varied functions of language in all cultures, apart from the functions strictly involved in the communication of thought, which had been the main focus of attention for those who approached language from the standpoint of the philosopher. To allow for the full range of language used by speakers, more comprehensive definitions of language have been proposed in recent years on the lines of the second one quoted above (*i.e.*, "A language is a system of arbitrary vocal symbols by means of which a social group cooperates").

A rather different criticism of accepted views on language began to be made in the 18th century, most notably by the French philosopher Étienne Bonnot de Condillac in "Essai sur l'origine des connaissances humaines" (1746; "Essay on the Origin of Human Knowledge") and by Johann Gottfried von Herder. These men were concerned with the origin and development of language in relation to thought in a way that earlier students had not been. The medieval and rationalist views implied that man as a rational, thinking creature invented language to express his thoughts, fitting words to an already developed structure of intellectual competence. With the examination of the actual and the probable historical relations between thinking and speaking, it became more plausible to say that language emerged not as the means of expressing already formulated judgments, questions, and the like but as the means of thought itself, and that man's rationality developed together with the development of his capacity for speaking.

Language as a means of thought itself

The relations between thought and speech are certainly not fully explained today, and it is clear that it is a great oversimplification to define thought as subvocal speech, in the manner of some behaviourists. But it is no less clear that propositions and other alleged logical structures cannot be wholly separated from the language structures said to express them. Even the symbolizations of modern formal logic are ultimately derived from statements made in some natural language and are interpreted in that light.

The intimate connection between language and thought, as opposed to the earlier assumed unilateral dependence of language on thought, opened the way to a recognition of the possibility that different language structures might in part favour or even determine different ways of understanding and thinking about the world. Obviously, all people inhabit a broadly similar world, or they would be unable to translate from one language to another; but, equally obviously, they do not all inhabit a world exactly the same in all particulars, and translation is not merely a matter of substituting different but equivalent labels for the contents of the same inventory. From this stem the notorious difficulties in translation, especially when the systematizations of science, law, morals, social structure, and so on are involved. The extent of the interdependence of language and thought—linguistic relativity, as it has been termed—is still a matter of debate, but the fact of such interdependence can hardly fail to be acknowledged.

THE STRUCTURE AND COMPOSITION OF LANGUAGE

Languages are immensely complicated structures. One soon realizes how complicated any language is when trying to learn it as a second language. If one tries to frame an exhaustive description of all the rules embodied in one's language—the rules by means of which a native speaker is able to produce and to understand an infinite number of correct, well-formed sentences—one can easily appreciate the complexity of the knowledge acquired by a child in mastering his mother tongue. The descriptions of languages written so far are in most cases excellent as far as they go, but they still omit more than they contain of an explicit account of a native speaker's competence in his language, by virtue of which one calls him a speaker of English, French, Swedish, or Swahili. The most recent developments in the study of language have served to reveal just how much more there is to do to bring palpable fact within systematic statement.

A detailed treatment of the science of linguistics is found elsewhere (see LINGUISTICS). Here it is proposed simply to give a brief outline of the way language or languages can be considered and described from different points of view, or at different levels, each contributing something essential and unique to a full understanding of the subject.

Phonetics and phonology. The most obvious aspect of language is speech. Speech is not essential to the definition of an infinitely productive communication system, such as is constituted by a language. But, in fact, speech is the universal material of human language, and the conditions of speaking and hearing have, throughout human history, shaped and determined its development. The study of speech sounds and of the physiology of speaking is called phonetics; this subject is dealt with further below, as well as in the article PHONETICS. Articulatory phonetics relates to the physiology of speech and acoustic phonetics to the physics of sound waves, their transmission and reception.

Speech, the universal material of human language

Phonetics covers much of the ground loosely referred to in language study as pronunciation. But, from a rather different point of view, speech sounds are also studied in phonology. Every language makes use of a very wide range of the articulations and resultant sounds that are available within the human vocal and auditory resources. Each language uses a somewhat different range, and this is partly responsible for the difficulty of learning to speak a foreign language and for speaking it "with an accent." But mere repertoires of sounds are not all that is involved. Far fewer general classes of sounds are distinctive (carry meaning differences) in any language than the number of sounds that are actually phonetically different. The English *t* sounds at the beginning and end of "tot" and in the two places in "stouter" are all different, though these differences are not readily noticed by English speakers; and, rightly, the same letter is used for them all. Similar statements could be made about most or all of the other consonant and vowel sounds in English.

What is distinctive in one language may not be distinctive in another or may be used in a different way; this is an additional difficulty to be overcome in learning to speak and understand a foreign language. In Chinese and in several other languages loosely called tone languages, the pitch, or tone, on which a syllable is said helps to distinguish one word from another: ma in northern Chinese on a level tone means "mother," on a rising tone means "hemp," and on a falling tone means "to curse." In English and in most of the languages of Europe (though not all—Swedish and Norwegian are exceptions) pitch differences do not distinguish one word from another, but form part of the intonation tunes that contribute to the structure and structural meaning of spoken sentences.

Languages differ in the ways in which consonant and vowel sounds can be grouped into syllables in words. English and German tolerate several consonants before and after a single vowel: "strengths" has three consonant sounds before and three after a single vowel sound (*ng* and *th* stand for one sound each). Italian does not have such complex syllables, and in Japanese and Swahili, for example, the ratio of consonant and vowel sounds in syllables and in words is much more even. Speakers of such languages find English words of the sort just mentioned very hard to pronounce, though to an Englishman they are perfectly "natural," "natural" in this context meaning "within the sounds and sound sequences whose mastery is acquired in early childhood as part of one's mother tongue."

All these considerations relating to the use of speech sounds in particular languages fall under the general heading of phonology; phonology is often regarded as one component of language structure.

Relations between words in sentences

Grammar. The other component is grammar. There is more to language than sounds, and words are not to be regarded as merely sequences of syllables. The concept of the word is a grammatical concept; in speech, words are not separated by pauses, but they are recognized as recurrent units that make up sentences. Very generally, grammar is concerned with the relations between words in sentences. Classes of words, or parts of speech, as they are often called, are distinguished because they occupy different places in sentence structure, and in most languages some of them appear in different forms according to their function (English "man," "men"; "walk," "walked"; "I," "me"; and so on). Languages differ in the extent to which word-form variation is used in their grammar; Classical Chinese had almost none, English does not have much, and Latin and Greek had quite a lot. Conversely, English makes much more use of word order in grammar than did Latin or Greek.

Traditionally, grammar has been divided into syntax and morphology, syntax dealing with the relations between words in sentence structure, morphology with the internal grammatical structure of words. The relation between "boy" and "boys" and the relationship (irregular) between "man" and "men" would be part of morphology; the relation of concord between "the boy [or "man"] is here" and "the boys [or "men"] are here" would be part of syntax. It must, however, be emphasized that the distinction between the two is not as clear-cut as this brief illustration might suggest. This is a matter for debate among linguists of different persuasions; some would deny the relevance of distinguishing morphology from syntax at all, referring to grammatical structure as a whole under the term syntax.

Grammar is different from phonology, though the word grammar is often used comprehensively to cover both aspects of language structure. Categories such as plural, past tense, and genitive case are not phonological categories. In spoken language they are, like everything else, expressed in speech sounds, but within a language these may be very different for one and the same category. In English noun plurals, the added *-s* in "cats," the vowel changes in "man, men" and in "goose, geese," and the *-en* in "oxen" are quite different phonologically; so are the past-tense formatives such as *-ed* in "guarded," *-t* in "burnt," vowel change in "take, took," and vowel and consonant change in "bring, brought." In Latin the genitive case can be represented in singular nouns by *-ī*, *-is*, *-ae*, *-ūs*, and *-eī*. The phonological difference does not matter, provided only that the category distinction is somehow expressed.

The same is true of the orthographic representation of grammatical differences, and the examples just given illustrate both cases. This is why the grammar of written language can be dealt with separately. In the case of dead languages, known with certainty only in their written forms, this must necessarily be done; insofar as the somewhat different grammar of their spoken forms made use of sound features not represented in writing (*e.g.*, stress differences), this can, at best, only be inferred or reconstructed.

Meaning expressed by grammatical forms, vocabulary, and intonation

Grammatical forms and grammatical structures are part of the communicative apparatus of languages, and along with vocabulary, or lexicon (the stock of individual words in a language), they serve to express all the meanings required. Spoken language has, in addition, resources such as emphatic stressing and intonation (see below). This is not to say, however, that grammatical categories can be everywhere directly related to specific meanings. Plural and past tense are fairly clear as regards meaning in English, but even here there are difficulties; in "if I knew his address I would tell you," the past-tense form "knew" refers not to the past but to an unfulfilled condition in the present. In some other languages greater problems arise. The gender distinctions of French, German, and Latin are very much part of the grammar of these languages, but only in a small number of words do masculine, feminine, and neuter genders correspond with differences of sex, or with any other category of meaning in relation to the external world (see also GRAMMAR).

Semantics. Language exists to be meaningful; the study of meaning, both in general theoretical terms and in reference to a specific language, is known as semantics. It embraces the meaningful functions of phonological features, such as intonation, and of grammatical structures and the meanings of individual words. Once again, it must be stressed that questions arising from the relations between grammar and meaning and between grammar and phonology are the subjects of continuing controversy today (see also SEMANTICS).

VARIETIES OF LANGUAGE

The word language contains a multiplicity of different designations. Two senses have already been distinguished: language as a universal species-specific capability of mankind, and languages as the various manifestations of that capability, as with English, French, Latin, Swahili, Malay, and so on. There is, of course, no observable universal language over and above the various languages that have been or are spoken or written; but one may choose to concentrate on the general and even the universal features, characteristics, and components of different languages and on the ways in which the same sets of descriptive procedures and explanatory theories may be applied to different languages. In so doing one may refer to language (in general) as one's object of study. This is what is done by linguists, or linguistic scientists, persons devoting themselves to the scientific study of languages (as opposed to the popular sense of polyglots, persons having a command of several different languages).

Intercomprehensibility of dialects

Languages and dialects. It has already been pointed out that no two persons speak exactly alike, and within the area of all but the smallest speech communities (groups of people speaking the same language) there are subdivisions of recognizably different types of language, called dialects, that do not, however, render intercommunication impossible nor markedly difficult. Because intercomprehensibility lies along a scale, the degree required for two or more forms of speech to qualify as dialects of a single language, instead of being regarded as separate languages, is not easy to quantify or to lay down in advance, and the actual cutoff point must in the last resort be arbitrary. In practice, however, the terms dialect and language can be used with reasonable agreement. One speaks of different dialects of English (Southern British English, Northern British English, Scottish English, Midwest American English, New England American English, Australian English, and so on, with, of course, many more delicately distinguished subdialects within these very general categories), but no one would speak of Welsh and English or of Irish and English as dialects of a single language, although they are spoken within the same areas and often by people living in the same villages as each other.

Certain problems arise over the use of the term dialect. The criterion of intercomprehensibility may link a chain of adjacent dialects together, such that at either end of the chain no intercomprehensibility is found. Thus, in a territorial sequence A, B, C, D, E, F, G . . . , speakers of A and B, B and C, C and D, and so on may be able to understand each other and, to a lesser extent, speakers of A and C, B and D, and so on, but a speaker of A and a speaker of F or a speaker of B and a speaker of G may find each other virtually unintelligible. Some such situation is said to obtain in the chain of spoken dialects extending across the French–Italian frontier, between the extremes of Parisian French and Roman Italian. In cases such as this the decision to allot one set to one language and the other to another may have to be taken on different grounds, such as political divisions.

Dialects and political boundaries

Because the frontiers of the modern nation-states of Europe tend to correspond more or less to language areas, with certain notable exceptions, such as Switzerland and Belgium, there is a general tendency to group European dialects within political frontiers as dialects of the national language of the country. This practice is reinforced under modern social conditions, as children of most countries in Europe are taught literacy in the "standard language," or standard dialect, of the country they live in. Swedish, Norwegian, and Danish are, in varying degrees, intercomprehensible for many speakers of any one of the three, and it is quite possible that without the political division of Scandinavia the existing dialects of each would be regarded as dialects of a single language or at any rate that the grouping of them within separate languages would be different.

A converse of the situation just referred to is the peculiar use of the term dialect in relation to Chinese, whereby all varieties of Chinese are loosely called dialects. In fact, several major divisions of Chinese speech—*e.g.*, Pekingese (Mandarin) in the north and Cantonese in the south—are mutually unintelligible as spoken languages. They do, however, employ the same writing system (the so-called Chinese characters), and in general the same written texts can be read and understood by speakers of either, though they would pronounce them differently. By the usual criterion of mutual intelligibility, there are several Chinese languages, each in turn subdivided into a number of dialects.

Yet a third use or, rather, misuse of the term would separate "the language" from "the dialects," as if a particular form of a language was in some way more representative of the language as a whole. Thus, Classical Arabic is sometimes contrasted with the spoken Arabic "dialects" of the contemporary Arabic-speaking world, and "standard English," at least in some parts of Great Britain, is contrasted with "the English dialects."

This leads to a further important observation about variations within a language. Speakers often differ in their dialect not only regionally but also socially, though these divisions are less easily represented graphically on the dialect maps of linguistic geographers. Sometimes, socially preferred dialects are themselves regional; for example, Parisian French, as the form of the language used in the capital, is regarded as standard. In most cases, however, nonregional differences are also involved: standard French is the French of educated Parisians, and standard has been loosely defined in this context as the speech of educated inhabitants of the capital city.

Prestige dialects and literary dialects

In a socially stratified community, the speech habits, like the other habits, of a group that enjoys general respect, political and social prestige, and economic power (or, at any rate, is thought to do so) tend to be preferred as models to imitate in the interests of social advancement. When literacy is the privilege of such a group, as was the case in most of Europe before the middle of the 19th century, such a dialect comes to be regarded as the language of education and of the educated and becomes the variety of the language directly represented in the writing system. Such dialects (for that is what they are) often correspond to the type of language in which national literatures are written. In English literature, though recognized authors of prose and poetry may speak regional varieties of the language (as the poet Alfred, Lord Tennyson is reported to have spoken Lincolnshire English), there are only a few pieces specifically written to be read in a nonstandard dialect. In Britain the poems of William Barnes of Dorset, the few Lincolnshire poems of Tennyson (*e.g.*, "Northern Farmer"), and above all the Scottish English poetry of Robert Burns are notable examples. In American literature one may cite the Uncle Remus tales of Joel Chandler Harris. The social prestige enjoyed by some Scottish dialects as compared with the nonstandard dialects of English spoken in England and Wales may have derived in part from the educational and literary reputation of Edinburgh, especially in the 18th century, but it was probably the eminence of Burns's Scottish English poetry that played the major role.

Standard British English is historically derived from a variety of southern English, the language of the Tudor court and of court circles. Though regional in origin and still having the largest concentration of speakers in London and the southeast of England, it is less locally rooted than other dialects, in part as the result of the greater mobility of well-educated and economically successful people. It is the type of English that foreign speakers are taught and are expected to aim at. The same holds true in large measure for educated Parisian French, which represents the French language as taught in other countries.

Influence of Dante on Italian language

In some languages the importance of literature or of a particular author in the development and emergence of the recognized standard language (*i.e.*, the standard dialect) has been considerable. Particularly in linguistic areas in which political unity came late—and, as a consequence, a court or capital city influence was not a strong factor—a widely read and respected body of literature has been found to have promoted one dialect above others as the representative of the language for literary and didactic purposes. Martin Luther's translation of the Bible into German in 1534 is regarded as having been a main influence in establishing Modern High German as the literary standard for the unified Germany of a later century and as the standard for teaching abroad. In Italy, like Germany a country that achieved political unity only in the 19th century, Dante's influence in the formation of a recognized Italian literary language, based on his native Florentine Tuscan, was overwhelming. This result, indeed, was deliberate; in his essay "De vulgari eloquentia" ("On the Vernacular Tongue"), Dante pleaded for the cultivation of some one dialect to represent Italian as a whole and to supply a focus of loyalty comparable to that exercised by the court in other countries.

Within an already recognized standard language and as part of its development, the effects of major works of literature may be very extensive. The number of expressions regularly used in English today that have their origin in Shakespeare or in the King James Bible is far greater than is often realized, and many such expressions are used without further reference to their Shakespearean or biblical source: "There's the rub" (*Hamlet*, Act III, scene 1), "A rose by any other name" (*Romeo and* Juliet, Act II, scene 2), "The time is out of joint" (Hamlet, Act I, scene 5), "Screw your courage to the sticking-place" (*Macbeth*, Act I, scene 7); "Jacob's ladder" (Gen. 28:12), "Noah's Ark" (Gen. 6:14), "A land flowing with milk and honey" (Ex. 3:8), "The shadow of death (Psalms 23:4), "The patience of Job" (James 5:11), and so on. Other authors have contributed to the English language in similar ways, but none are so significant in this respect as Shakespeare and the translators of the King James Bible.

It must be emphasized that, considered simply as modes of communication, nonstandard dialects are neither less "correct" nor more "correct" than standard dialects. There may be social reasons for preferring the use of standard dialects, but these do not involve inherent merits or demerits of a linguistic sort. Dialects differ at all levels of structure—in phonetics, in grammar, and in vocabulary—but there are no grounds for saying that one dialect is more orderly, systematic, or expressive than another. Each is capable of meeting the requirements that its speakers need to impose on it. Nor is it true, as is sometimes maintained, that nonstandard dialects are later degenerations of the original "purer" language. In British English, in fact, several nonstandard dialects preserve older features now lost in the standard; for example, a postvocalic *r* sound exists in the West-Country pronunciation of words such as "farm" and "cart," and a distinct 2nd-person singular pronoun is used in parts of the northern English counties, taking various phonetic forms and corresponding to "thou" and "thee" of earlier written English (see also DIALECTS).

Specialized language forms: slang, argot, and others. The above-mentioned social and regional dialect variations within languages are more or less natural; that is, they arise largely from the conditions of language use and language transmission: without deliberate intent on the part of speakers. There are, however, some deliberately created variations within languages. The socially unifying force of a single language or dialect and the divisive force of language and dialect differences have always been apparent. Language was a main inspiration in 19th-century European nationalism and is one of the factors in nationalist movements in the world today. Not surprisingly, groups within a society that set a special value on group identity and group consciousness deliberately develop and foster private dialects that are known to insiders but are mysterious and baffling to those not belonging to the group. The various underworld jargons and special trade argots can be cited, such as Loucherbème in Paris (from *boucher* "butcher," by systematic deformation), the private slang vocabulary of many schools and colleges, and the systematic and regular alteration of certain words in some private languages of schoolboys and students—American "pig Latin" and the former "Oxford and British public-school "-agger" talk ("nogger," meaning agnostic; "wagger pagger bagger," meaning wastepaper basket), developed from universally recognized forms such as "rugger," Rugby football. From all over the world the transient but zealously cherished speech styles of teen-agers can be noted, whereby, along with particular fashions in clothes, they maintain and assert their identity over against the childhood they have left and the adult world they do not yet recognize (see also SLANG).

Group identity and private dialects

Sometimes, as in the case of criminal argots, part of the function of special languages is deliberately to mislead and obstruct the rest of society and the authorities in particular; they may even become wholly impenetrable to outsiders. But this is not the sole or main purpose of most specialized varieties of language. Professions whose members value their standing in society and are eager to render their services to the public foster their own vocabulary and usage, partly to enhance the dignity of their profession and the skills they represent but partly also to increase their efficiency. An example of this is the language of the law and of lawyers.

The cultivation and maintenance of specialized types of language by certain professions should not be regarded as trivially or superficially motivated. In general usage, languages are necessarily imprecise, or they would lack the flexibility and infinite extensibility demanded of them. But for certain purposes in restricted situations much greater precision is required, and part of the function of the particular style and vocabulary of legal language is the avoidance, so far as may be possible, of all ambiguity and the explicit statement of all necessary distinctions. This is why legal texts, when read out of their context, seem so absurdly pedantic and are an easy target for ridicule. Similar provision for detail and clarity characterizes the specialist jargons of medicine and of the sciences in general and also of philosophy. Indeed, one might regard the formulas of modern symbolic logic as the result of a consciously developed and specialized written language for making precise the relations of implication and inference between statements that, when couched in everyday language, are inexact and open to misinterpretation. Some would go as far as to say that traditional metaphysics is no more than the result of misunderstanding everyday discourse and that the main purpose of philosophy is to resolve the puzzles that arise from such misunderstandings.

Stereotyped language of games

The use of specialized types of language in fostering unity is also evidenced in the stereotyped forms of vocabulary employed in the playing of certain games. Tennis scores use the sequence "love, 15, 30, 40, and game"; cricketers verbally appeal to the umpire when a batsman may be out by calling "How's that?" and the ways of being out are designated by stereotypes, "run out," "leg before wicket," "stumped," and so forth. The esoteric language of horse racing and its associated wagering of money is well-known, though not readily understood by outsiders.

The ancient but persistent recognition of the power of language is apparent in the respect for correctness in the use of language in any sphere of life having supernatural connections. Those credited with such connections employ special formulas and rigidly prescribed modes of diction; examples of the language of magic and of magicians are widespread, ranging from the usages of shamans and witch doctors to the ritual "abracadabra" of the mock magic displayed by conjurors at children's parties.

The efficacy of religious worship and of prayers is frequently associated with the strict maintenance of correct forms of language, taught by priests to their successors, lest the ritual become invalid. In ancient India the preservation in all its supposed purity of the language used in the performance of certain religious rituals (Sanskrit) gave rise to one of the world's most important schools of linguistics and phonetics. In the Christian churches one can observe the value placed by Roman Catholics on the Latin of the mass and by Church of England and Episcopalian churchmen on the formal English of the Authorized Version of the Bible and of *The* Book of *Common* Prayer, despite recent attempts at replacing these ritual forms of language by forms taken from modern spoken vernaculars.

Pidgins and creoles. Some specialized languages were developed to keep the outsider at bay. In other circumstances, languages have been deliberately created to facilitate communication with outsiders. This happens when people speaking two different languages have to work together, usually in some form of trade relation or administrative routine. In such situations the so-called pidgins arise, more or less purposively made up of vocabulary items from each language, with mutual abandonment of grammatical complexities that would cause confusion to either party. Pidgins have been particularly associated with areas settled by European traders; examples have been Chinook Jargon, a lingua franca based on an American Indian language and English and formerly used in Washington and Oregon, and Beach-la-mar, an English-based pidgin of parts of the South Seas.

Sometimes, as the result of relatively permanent settlement and the intermixture of two speech communities, a pidgin becomes the first language, or mother tongue, of later generations, ultimately displacing both the original languages. First languages arising in this way from artificially created pidgins are called creoles. Notable among creoles is the language of Haiti, Haitian Creole, built up from the French of the settlers and the African language of the former slaves; it shows lexical and grammatical features of both sources.

Creolization of a pidgin language

Creoles differ from pidgins in that, as first languages, they are subject to the natural processes of change like any other language (see below *Linguistic* change); and, despite the deliberately simplified form of the original pidgin, in the course of generations creoles develop their own complexities. The reason is plain to see. The restricted uses to which pidgins were first put and for which they were devised did not require any great flexibility. Once such a language becomes the first or only language of many people, it must perforce acquire the resources (*i.e.*, the complexity) to respond adequately to all the requirements of a natural language (see also PIDGIN).

Nonverbal language. Speech and writing are, indeed, the fundamental faculties and activities referred to by the term language. There are, however, areas of human behaviour for which the term is used in a peripheral and derivative sense.

When individuals speak, they do not normally confine themselves to the mere emission of speech sounds. Because speaking usually involves at least two parties in sight of each other, a great deal of meaning is conveyed by facial expression, tone of voice, and movements and postures of the whole body but especially of the hands; these are collectively known as gestures. The contribution of bodily gestures to the total meaning of a conversation is in part culturally determined and differs in different communities. Just how important these visual symbols are may be seen when one considers how much less effective telephone conversation is as compared with conversation face to face; the experience of involuntarily smiling at the telephone receiver and immediately realizing that this will convey nothing to the hearer at the other end of the line is common. Again, the part played in emotional contact and in the expression of feelings by facial expressions and tone of voice, quite independently of the words used, has been shown in tests in which subjects have been asked to react to sentences that appear as friendly and inviting when read but are spoken angrily and, conversely, to sentences that appear as hostile but are spoken with friendly facial expressions. It is found that it is the visual accompaniments and tone of voice that elicit the main emotional response. A good deal of what goes under the heading of sarcasm exploits these contrasts.

Expression by vocal, non-linguistic utterances

Just as there are paralinguistic activities such as facial expressions and bodily gestures integrated with and assisting the communicative function of spoken language, so there are vocally produced noises that cannot be regarded as part of any language, though they help in communication and in the expression of feeling. These include laughter, shouts and screams of joy, fear, pain, and so forth, and conventional expressions of disgust, triumph, and so on, traditionally spelled "ugh!," "ha ha!," etc., in English. Such nonlexical ejaculations differ in important respects from language: they are much more similar in form and meaning throughout mankind as a whole, in contrast to the great diversity of languages; they are far less arbitrary than most of the lexical components of language; and they are much nearer the cries of animals produced under similar circumstances and, as far as is known, serve similar expressive and communicative purposes. As noted above, some people have tried to trace the origin of language itself to them.

A language is a symbol system. It may be regarded, because of its infinite flexibility and productivity, as the symbol system ***par excellence.*** But there are other symbol systems recognized and institutionalized in the different cultures of mankind. Examples of these exist on maps and blueprints and in the conventions of representational art (*e.g.*, the golden halos around the heads of saints in religious paintings). Other symbol systems are musical notation and dance notation, wherein graphic symbols designate musical pitches and other features of musical performance and the movements of formalized dances. More loosely, because music itself can convey and arouse emotions and certain musical forms and structures are often associated with certain types of feeling, one frequently reads of the "language of music" or even of "the grammar of music." The terms language and grammar are here being used metaphorically, however, if only because no symbol system other than language has the same potential of infinite productivity, extension, and precision.

Languages are used by human beings to talk and write to other human beings. Derivatively, bits of languages may be used by humans to control machinery, as when different buttons and switches are marked with words or phrases designating their functions. A recent and specialized development of man–machine language is seen in the various "computer languages" now in use; *e.g.*, Cobol, Algol, and Fortran. These are referred to as programming languages, and they provide the means whereby sets of "instructions" and data of various kinds can be supplied to computers in forms acceptable to these machines. Various types of such languages are in use for different purposes. The development and use of computer languages must now be regarded as a distinct science in itself (for more information, see COMPUTERS).

PHYSIOLOGICAL AND PHYSICAL BASIS OF SPEECH

Primacy of speech in language

For an adequate understanding of human language it is necessary to keep in mind the absolute primacy of speech. In societies in which literacy is all but universal and language teaching at school begins with reading and writing in the mother tongue, one is apt to think of language as a writing system that may be pronounced. In point of fact, language is a system of spoken communication that may be represented in various ways in writing.

Man has almost certainly been in some sense a speaking animal from early in the emergence of Homo *sapiens* as a recognizably distinct species. The earliest known systems of writing go back perhaps some 5,000 years. This means that for many hundreds of thousands of years human languages were transmitted from generation to generation and were developed entirely as spoken means of communication. Moreover, in the world as it is today, literacy is still the privilege of a minority in many language communities. Even when literacy is widespread, some languages remain unwritten if they are not economically or culturally important enough to justify creating an alphabet for them and teaching them; then literacy is acquired in a second language learned at school. Such is the case with many speakers of South American Indian languages, who become literate in Spanish or Portuguese. A similar situation prevails in some parts of Africa, where reading and writing are taught in languages spoken over relatively wide areas. In all communities, speaking is learned by children before writing, and all people act as speakers and hearers much more than as writers and readers.

It is, moreover, a total fallacy to suppose that the languages of illiterate or so-called primitive peoples are less structured, less rich in vocabulary, and less efficient than the languages of literate civilizations. The lexical content of languages varies, of course, according to the culture and the needs of their speakers, but observation bears out the statement of the U.S. anthropological linguist Edward Sapir made in 1921: "When it comes to linguistic form, Plato walks with the Macedonian swineherd, Confucius with the head-hunting savage of Assam."

All this means that the structure and composition of language and of all languages have been conditioned by the requirements of speech, not those of writing. Languages are what they are by virtue of their spoken, not their written, manifestations. The study of language must be based on a knowledge of the physiological and physical nature of speaking and hearing. The details of these aspects of language are covered in other articles, especially PHONETICS and SPEECH, PHYSIOLOGY OF; only the essentials are given here.

Speech production. Speaking is in essence the by-product of a necessary bodily process, the expulsion from the lungs of air charged with carbon dioxide after it has fulfilled its function in respiration. Most of the time one breathes out silently; but it is possible, by adopting various postures and by making various movements within the vocal tract, to interfere with the egressive airstream so as to generate noises of different sorts. This is what speech is made of.

Vocal tract and speech organs

The vocal tract comprises the passage from the trachea (windpipe) to the orifices of the mouth and nose; all the organs used in speaking lie in this passage. Conventionally, these are called the organs of speech, and the use in several languages of the same word for the tongue as a part of the body and for language shows the awareness people have of the role played by this part of the mouth in speaking. But few if any of the major organs of speech are exclusively or even mainly concerned with speaking. The lips, the tongue, and the teeth all have essential functions in man's bodily economy, quite apart from talking; to think, for example, of the tongue as an organ of speech in the same way that the stomach is regarded as the organ of digestion is fallacious. Speaking is a function superimposed on these organs, and the material of speech is a waste product, spent air, exploited to produce perhaps the most wonderful by-product ever created.

Relatively few types of speech sounds are produced by other sources of air movement; the clicks in some South African languages are examples, and so is the fringe linguistic sound used in English to express disapproval, conventionally spelled "tut." In all languages, however, the great majority of speech sounds have their origin in air expelled through the contraction of the lungs. Air forced through a narrow passage or momentarily blocked and then released creates noise, and characteristic components of speech sounds are types of noise produced by blockage or narrowing of the passage at different places.

If the vocal cords (really more like two curtains) are held taut as the air passes through them, the resultant regular vibrations in the larynx produce what is technically called voice, or voicing. These vibrations can be readily observed by contrasting the sounds of f and v or of *s* and *z* as usually pronounced; "five" and "size" each begin and end with voiceless and voiced sounds, respectively, which are otherwise formed alike, with the tongue and the lips in the same position. Most consonant sounds and all vowel sounds in English and in the majority of languages are voiced, and voice, in this sense, is the basis of singing and of the rise and fall in speaking that is called intonation, as well as of the tone distinctions in tone languages. The vocal cords may be drawn together more or less tightly, and the vibrations will be correspondingly more or less frequent. A rise in frequency causes a rise in perceived vocal pitch. Speech in which voice is completely excluded is called whispering.

Formation of vowels

Above the larynx, places of articulation in frequent use are between the back of the tongue and the soft palate, between the blade of the tongue and the ridge just behind the upper front teeth, and between the lips. Stoppage and release (technically, plosion) at these places form the k (often written as *c*, "cat"), *t*, and *p* sounds in English and, when voicing is also present, the g (as in "gay"), d, and b sounds. Obstruction at these and other places sufficient to cause noise gives rise to what are called fricative sounds; in English these include the normal pronunciations of *s*, z, f, and v and the *th* sounds in "thin" and "then." A vowel is characterized as the product of the shape of the entire tract between the lips and larynx, without local obstruction though usually with voicing from the vocal cords. It is contrasted with a consonant, though the exact division between these two categories of speech sound is not always easy to draw. Different shaping of the tract produces the different vowel sounds of languages.

The soft palate may be raised or lowered. It is lowered in normal breathing and allows air to pass in and out through the nose. In the utterance of most speech sounds it is raised, so that air passing through the mouth alone forms the sound; if it is lowered, air passes additionally or alternately through the nose, thus producing the nasal sounds. All but a very few languages have nasal consonants (the English sounds m, n, and *ng* as in "sing") and some, such as French, have nasalized vowels as well. A few speakers regularly allow air to pass through their nasal passages all the time while they are speaking; such persons are said to "speak through the nose."

All articulatory movements, including the initial expulsion of air from the lungs, may be made with greater or less vigour, giving rise to louder or softer speech as a whole or to greater loudness on one part of what is said, for emphasis or contrast.

Every different configuration and movement of the vocal tract creates corresponding differences in the air vibrations that comprise and transmit sound. These vibrations, like those of all noises, extend outward in all directions from the source, gradually decreasing to zero or to below the threshold of audibility. They are called sound waves, and they consist of rapid rises and falls in air pressure. The speed at which pressure rises and falls is the frequency. Speech sounds involve complex waves containing vibrations at a number of different frequencies, the lowest being the voice pitch of singing and intonation, produced by the vocal cords in voiced sounds.

The eardrum responds to the different frequencies of speech, provided they retain enough energy, or amplitude (*i.e.*, are still audible). The perceptibly different speech sounds that comprise the spoken utterances of any language are the result of the different impacts on one's ears made by the different complexes of frequencies in the waves produced by different articulatory processes. As the result of careful and detailed observation of the movements of the vocal organs in speaking, aided by various instruments to supplement the naked eye, a great deal is now known about the processes of articulation. In recent years an array of other instruments has provided much information about the nature of the sound waves produced by articulation. Speech sounds may be and have been described and classified both from an articulatory viewpoint, in terms of how they are produced, and from an acoustic viewpoint, by reference to the resulting sound waves (their frequencies, amplitudes, and so forth). Articulatory descriptions are more readily understood, being couched in terms such as nasal, bilabial, lip-rounded, and so on. Acoustic terminology requires a knowledge of the technicalities involved for its comprehension. In that almost every person is a speaker and a hearer, it is clear that both sorts of description and classification are important, and each has its particular value for certain parts of the scientific study of language.

Articulatory and acoustic descriptions of speech

Language acquisition. As far as the production of speech sounds is concerned, all human beings are physiologically alike. People have differently shaped faces, as much as they differ in other aspects of bodily build, but it has been shown time and again that a child learns to speak the language of those who bring him up from infancy. In most cases these are his biological parents, especially his mother, but one's first language is acquired from environment and learning, not from physiological inheritance. Adopted infants, whatever their race or physical type and whatever the language of their actual parents, acquire the language of the adoptive parents who raise them just as if they had been their own children.

Different shapes of lips, throat, and other parts of the vocal tract have an effect on the voice quality of people's speech, and this is part of the individuality of each person's voice referred to above. Physiological differences, including size of throat and larynx, both overall and in relation to the rest of the vocal tract, are largely responsible for the different pitch ranges characteristic of men's, women's, and children's speech. But these individual differences do not affect one's ability or aptitude to acquire and speak any particular language.

Speech is species-specific to mankind. Physiologically, animal communications systems are of all sorts. The type that is in some ways functionally the nearest to human language, the system of bee dances, is entirely remote from human speech and does not make use of sound at all. The animal sounds superficially most resembling speech, the imitative cries of parrots and some other birds, are produced by very different physiological means: birds have no teeth or lips but vocalize by means of the syrinx, a modification of the windpipe above the lungs. Almost all mammals and many other animal species make vocal noises and evince feelings thereby and keep in contact with each other through a rudimentary sort of communication, but those members of the animal kingdom nearest to man genetically, the primates, have proved highly resistant to the acquisition of speech.

Advantages of speech as a means of communication

Man's development of speech has been linked to his upright posture and the freeing of his vocal cords from the frequent need to "hold one's breath in using the arms for locomotion. Certainly, speaking and hearing—as man's primary means of communication—have a number of striking advantages: speech does not depend on daylight or on mutual visibility, it can operate in all directions over reasonably wide areas, and it can be adjusted in loudness to cope with distance. As is seen in crowded rooms, it is possible to pick out some one person's voice despite a good deal of other noise and in the midst of other voices speaking the same language. In addition, the physical energy required in speaking is extremely small in relation to the immense power wielded by speech in man's life, and scarcely any other activity, such as running, walking, or tool using, interferes seriously with the process (even eating and drinking can be carried on simultaneously with talking, and the reluctance one may feel for "speaking with one's mouth full" is more a matter of cultural convention and good manners than of physiological difficulty).

The characteristics just outlined pertain to all of the world's languages, including those of the allegedly primitive peoples. What is more a matter of controversy is the extent to which man's biological inheritance is involved in language acquisition and language use. The fact that language is species-specific to man argues an essential cerebral or mental component, and in the last century certain aspects of speech control and use were located in a particular part of the human brain (Broca's convolution).

No one inherits the ability to speak a particular language, but every normal human child is born with the ability and the drive to acquire a language, namely, the one to which he is predominantly exposed from infancy. He brings to this task a very considerable innate ability, because his exposure is largely to a random selection of utterances (apart from any attempts at systematic teaching that he may encounter) occuring within his earshot or addressed to him. Yet by late childhood he has, through progressive stages, acquired the central or basic vocabulary of the language, together with its phonological and grammatical structure. Observation shows that this is substantially the same situation the world over, among literate and illiterate communities, and that much the same number of years of childhood is taken up by the process. Thus, it would appear that, objectively considered, all languages are roughly equal in complexity and in difficulty of mastery.

Innate human capacity for language acquisition

It is, therefore, clear that all normal humans bring into the world an innate faculty for language acquisition, language use, and grammar construction. The last phrase refers to the internalization of the rules of the grammar of one's first language from a more or less random exposure to utterances in it. The human child is very soon able to construct new, grammatically acceptable sentences from material he has already heard; unlike the parrot in human society, he is not limited to the mere repetition of whole utterances.

What is currently under debate is the part played by this innate ability and its exact nature. Until the 1950s scholars considered language acquisition to be carried out largely by analogical creation from observed patterns of sentences occurring in utterances heard and understood by the child. Such a view, much favoured by persons inclined to a behaviourist interpretation of human learning processes (*e.g.*, the U.S. linguist Leonard Bloomfield), stressed the very evident differences between the structures of different languages, particularly on the surface. Since the late 1950s, a number of linguists have been placing much more emphasis on the inherent grammar-building disposition and competence of the human brain, which is activated by exposure to utterances in a language, especially during childhood, in such a way that it fits the utterances into predetermined general categories and structures. Such linguists, inheritors of the 17th- and 18th-century interest in "universal grammar," put their stress on the underlying similarities of all languages, more especially in the deeper areas of grammatical analysis (for the distinction between deep structure and surface structure in grammar, see the article LINGUISTICS).

MEANING, VOCABULARY, AND STYLE IN LANGUAGE

The whole object and purpose of language is to be meaningful. Languages have developed and are constituted in their present forms in order to meet the needs of communication in all its aspects.

It is because the needs of human communication are so various and so multifarious that the study of meaning is probably the most difficult and baffling part of the serious study of language. Traditionally, language has been defined, as in the definition quoted above, as the expression of thought, but, as was seen, this involves far too narrow an interpretation of language or far too wide a view of thought to be serviceable. The expression of thought is just one among the many functions performed by language in certain contexts.

Distinguishing meaning by word order or grammatical form

Structural and lexical meaning. First, one must recognize that the meaning of any sentence comprises two parts, the meanings of the words it contains and the structural or grammatical meaning carried by the sentence itself. In English "the dog chased the cat" and "the boy chased the cat" differ in meaning because "dog" and "boy" are different words with different word meanings; the same applies to equivalent sentences in other languages. The two sentences "the dog chased the cat" and "the cat chased the dog," though containing exactly the same words, are different in meaning because the different word orders distinguish what are conventionally called subject and object. In Latin the two corresponding sentences would be distinguished not by word order, which is grammatically indifferent and largely a matter of style, but by different shapes in the lexical equivalents of "dog" and "cat." In Japanese the grammatical distinction of subject and object, normally marked by the word order subject–object–verb, can be reinforced by a subject particle after the first word and an object particle after the second.

The formal resources of any language for making distinctions in the structural meanings of sentences are limited by two things: the linear (time) dimension of speaking and the limited memory span of the human brain. Writing copies the time stream of speech with the linear flow of scripts. Diagrams and pictures employ two dimensions, and models employ three; but writing is partially relieved of memory-span restrictions by the permanence of visual marks. Because written texts are almost entirely divorced from oral pronunciation, sentence length and sentence complexity can be carried to extremes, as may be observed in some legal and legislative documents that are virtually unintelligible if read aloud.

Within these linear restrictions, distinctions corresponding to the main uses of language can be made. All languages can employ different sentence structures to state facts (declarative), to ask questions (interrogative), and to enjoin or forbid some course of action (imperative). More delicate means exist to soften or modify these basic distinctions: *e.g.*, "It's cold today, isn't it?"; "Isn't it still raining?"; "Shut the door, would you mind"; "Don't be long, will you?" Languages use their resources differently for these purposes, but, generally speaking, each seems to be equally flexible structurally. The principal resources are word order, word form, syntactic structure, and, in speech, pitch and stress placement. In English, as an example, a word or phrase can be highlighted by being placed first in the sentence when it would not normally occur there: compare "he can't bear loud noises" with "loud noises he can't bear" or "loud noises, he can't bear them." The object noun or noun phrase can also be put first by making the sentence passive; this allows the original subject to be omitted if one does not know or does not want to refer to an agent: "the town was destroyed (by the revolutionaries)." Within and together with all these possibilities, almost any word can be made contrastively prominent by being stressed (spoken more loudly) or by being uttered on a higher pitch, and very often these two are combined: "I asked you for red roses (not yellow)"; "I meant it for you (not her)"; "I know nothing about it (someone else may)." Prominence is especially associated with intonation, itself an important carrier of structural meaning in speech. One may state facts, ask questions, and give instructions with a variety of intonations indicating, along with visible gestures, different attitudes, feelings, and social and personal relations between speaker and hearer.

The possibilities of expressing structural meanings are a most important part of any language. They are acquired along with the rest of one's first language in childhood and are learned more slowly and with more difficulty in mastering a second or later language. Scholars are still only at the beginning of a full formal analysis of these resources, as far as most languages are concerned, and are still further from an adequate understanding of all the semantic functions performed by means of these resources.

Meanings of words

The other component of sentence meaning is word meaning, the individual meanings of the words in a sentence, as lexical items. The concept of word meaning is a familiar one. Dictionaries list words and in one way or another state their meanings. It is regarded as a sensible question to ask of any word in a language, "What does it mean?" This question, like many others about language, is easier to ask than to answer.

It is through lexical resources that languages maintain the flexibility their open-ended commitments demand. Every language has a vocabulary of many thousands of words, though not all are in active use, and some are known only to relatively few speakers. Perhaps the commonest delusion in considering vocabularies is the assumption that the words of different languages, or at least their nouns, verbs, and adjectives, label the same inventory of things, processes, and qualities in the world but unfortunately label them with different labels from language to language. If this were so, translation would be easier than it is; but the fact that translation, though often difficult, is possible indicates that people are talking about similar worlds of experience in their various languages.

Languages in part create the world in which men live. Of course, many words do name existing bits and pieces of earth and heaven: "stone," "tree," "dog," "woman," "star," "cloud," and so on. Others, however, do not so much pick out what is there as classify it and organize one's relations with it and with each other with regard to it. A range of living creatures are mammals or are vertebrates, because people classify them in these ways, among others, by applying selected criteria and so determining the denotation of the words mammal and vertebrate. Plants are vegetables or weeds according as groups of people classify them, and different plants are included and excluded by such classifications in different languages and different cultures.

Time and its associated vocabulary ("year." "month," "day," "hour," "minute," "yesterday," "tomorrow," and so on) do not refer to discrete sections of reality but enable people to impose some sort of order, in agreement with others, on the processes of change observed in the world. Personal pronouns pick out the persons speaking, spoken to, and spoken about; but some languages make different distinctions in their pronouns from those made in English. For example, in Malay, *kita*, which means "we," including the person addressed, is distinct from *kami*, a form for "we" that includes the speaker and a third person or persons but excludes the person addressed. In Japanese and in several other languages, a variety of words denoting the 1st and 2nd persons indicate additionally the observed or intended social relationship of those involved.

Other word meanings are even more language and culture bound, and in consequence harder to translate. "Right" and "wrong," "theft," "inheritance," "property," "debt," "sin," and "crime" (as different sorts of wrongdoing) are just a few of the words regulating one's conduct and relations with one's fellows in a particular culture. Translation becomes progressively harder as one moves to languages of more remote cultures, and it has been said that it requires "a unification of cultural context." Insofar as a person's understanding of the universe and of the relations between himself and other people is closely linked with the language he speaks, it must be assumed, and the evidence confirms this assumption, that the child progressively acquires such understanding along with his language.

Onomatopoeic words

The great majority of word shapes bear no direct relation to their lexical meanings. If they did, languages would be more alike. What are called onomatopoeic words are rather similar in shape through different languages: French *coucou*, English "cuckoo," and German *Kuckuck* directly mimic the call of the bird. English "dingdong" and German *bim-bam* share several sound features in common that partially resemble the clanging of bells. More abstractly, some direct "sound symbolism" has been seen between certain sound types and visual or tactile shapes. Most people agree that the made-up word "oomboolu" would better designate a round, bulbous object than a spiky one. In addition, the appropriateness of the vowel sound represented by *ee* in English "wee" and *i* in French *petit* "small" and Italian *piccolo* "small" for expressing things of small size has been traced in several languages.

Vocabulary. All this, however, is a very small part of the vocabulary of any language. For by far the largest number of words in a language there is no direct association between sound and meaning. English "horse," German Pferd, French *cheval,* Latin *equus,* and Greek *hippos* are all unrelated to the animal so named, except that these words are so used in the languages concerned. This is what is meant by the term arbitrary in the second definition of language quoted at the beginning of this article. Vocabulary has to be largely arbitrary, because the greater part of the world and of man's experience is not directly associated with any kind of noise, and it is a contingent, though universal, fact of history and biology that sound and not the material of some other sense is the basis of human language.

The relations between sentence structure and structural meanings are also largely arbitrary and tacitly conventional. Though loudness and stress for emphasis and certain linguistic indications of anger, excitement, and the like are more closely akin to nonlinguistic ejaculations and are somewhat similar across language divisions, actual intonations and features such as word order, word inflection, and grammatical particles, used in maintaining distinctions in structural meaning, differ markedly in different languages.

Not only are word meanings somewhat different in different languages; they are not fixed for all time in any one language. Semantic changes take place all along (see below), and at any moment the semantic area covered by a word is indeterminately bordered and differs from context to context. This is a further aspect and condition of the inherent and necessary flexibility of language. A person can be as precise or as imprecise as he needs or wishes lo be. In general, words are fairly imprecise; yet for particular purposes their meanings can be tightened up, usually by bringing in more words or phrases to divide up a given field in more detail. "Good" contrasts generally with "bad"; but one can, for example, grade students as "first-class," "excellent," "very good," "good," "fair," "poor," and "failed" (or "bad"). In this case, "good" now covers a restricted and relatively low place in a field of associated terms. Colour words get their meanings from their mutual contrasts. The field of visually discriminable hues is very large and goes far beyond the resources of any vocabulary as it is normally used. Children learn the central or basic colour words of their language fairly early and at the same time; such terms as red and green are normally learned before subdivisions such as crimson and scarlet or chartreuse. It is well-known that languages make their primary divisions of the spectrum of colours in different places; Japanese ***aoi*** covers many of the hues referred to in English by "green" and "blue," while "blue" covers much of the range of the two Russian words *goluboy* and *siny.* While the actual colour vocabularies of languages differ, however, recent research by Brent Berlin and Paul Kay has tried to show that "there exist universally for humans eleven basic perceptual color categories" that serve as reference points for the colour words of a language, whatever number may be regularly employed at any time.

Precision and imprecision of words

Ordinarily, considerable areas of indeterminate designation in colour vocabulary and in other fields are tolerated; between "red" and "purple" and between "purple" and "blue" there are hues that one would hesitate to assign firmly to one or the other and on which there would be considerable personal disagreement. When greater precision than normal is required—as, for example, in listing paint or textile colours—all kinds of additional terms can be brought into service to supplement the usual vocabulary: "off-white," "light cream," "lemon," "blush pink," and so on.

The vocabulary of kinship terms varies from language to language, reflecting cultural differences. English distinguishes the nearer kinsfolk by sex: "mother, father"; "sister, brother"; "aunt, uncle"; and others. Other languages, such as Malay, make a lexical distinction of age the primary one, with separate words for elder brother or sister and younger brother or sister. Still other languages—for example, some American Indian ones—use different words for the sister of a man and for the sister of a woman. But beyond this any language can be as precise as the situation demands in kin designation. When it is necessary, English speakers can specify "elder sister" and "female cousin," and within the overall category it is possible to distinguish "first and second cousins" and "cousins once removed," distinctions that it is ordinarily pedantic to make.

The best example of infinite precision available from a

Vocabulary of mathematics

strictly limited lexical stock is in the field of arithmetic. Between any two whole numbers a further fractional or decimal number may always be inserted, and this may go on indefinitely: between 10 and 11, 10% (10.5), 10¼ (10.25), 10% (10.125), and so on. Thus, the mathematician or the physical scientist is able to achieve any desired degree of quantitative precision appropriate to his purposes; hence the importance of quantitative statements in the sciences—any thermometric scale contains far more distinctions of temperature than are reasonably available in the vocabulary of a language ("hot," "warm," "cool," "tepid," "cold," and so on). For this reason mathematics has been described as the ideal use of language, but for many purposes in everyday life the very imprecision of natural languages is the source of their strength and adaptability.

Every living language can readily be adapted to meet changes occurring in the life and culture of its speakers, and the main weight of such changes falls on vocabulary. Grammatical and phonological structures are relatively stable and change noticeably over centuries rather than decades (see below); but vocabularies can change very quickly both in word stock and in word meanings. Consider as an example the changes wrought by modern technology in the vocabularies of all European languages since 1945. Before that date "transistor" and "cosmonaut" did not exist, and "nuclear disarmament" would scarcely have had any clear meaning.

Every language can alter its vocabulary very easily, which means that every speaker can without effort adopt new words, accept or invent new meanings for existing words, and of course, cease to use some words or cease to use them in certain meanings. Dictionaries list some words and some meanings as "obsolete" or "obsolescent" to indicate this process. No two speakers share precisely the same vocabulary of words readily used and readily understood, though they may speak the same dialect. They will, however, naturally have the great majority of words in their vocabularies in common.

Languages have various resources for effecting changes in vocabulary. Meanings of existing words may change. With the virtual disappearance of falconry as a sport in England, "lure" has lost its original meaning of a bunch of feathers on a string by which hawks were recalled to their handler and is used now mainly in its metaphorical sense of enticement. The additional meaning of "nuclear" has already been mentioned; one may list it with words such as computer and jet, which acquired new ranges of meaning in the mid-20th century.

Creation of new words

All languages have the means of creating new words to bear new meanings. These can be new creations; "Kodak" is one such, invented at the end of the 19th century by George Eastman; "chortle," now in general use, was a jocular creation of the English writer and mathematician Lewis Carroll (creator of *Alice in Wonderland),* and "gas" was formed in the 17th century by the Belgian chemist and physician Jan Baptist van Helmont as a technical term in chemistry, loosely modelled on the Greek *chaos* ("formless void"). But mostly languages follow definite patterns in their innovations. Words can be made up without limit from existing words or from parts of words; the sources of "railroad," "railway," and "aircraft" are obvious, and so are the sources of "disestablishment," first cited in 1806 and thereafter used with particular reference to the status of the Church of England. The controversy over the relations between church and state in the 19th and early 20th centuries gave rise to a chain of new words as the debate proceeded: "disestablishmentarian," "antidisestablishmentarian," "antidisestablishmentarianism." Usually, the bits and pieces of words used in this way are those found in other such combinations, but this is not always so. The technical term permafrost (terrain that never thaws, as in the Arctic) contains a bit of "permanent" probably not hitherto found in any other word.

A particular source of technical neologisms in European languages has been the words and word elements of Latin and Greek. This is part of the cultural history of western Europe, in so many ways the continuation of Greco-Roman civilization. "Microbiology" and "dolichocephalic" are words well formed according to the rules of Greek as they would be taken over into English, but no records survive of *mikrobiologia* and *dolichokephalikos* ever having been used in Ancient Greek. The same is true of Latinate creations such as "reinvestment" and "longiverbosity." The long tradition of looking to Latin and, since the Renaissance, to Greek also as the languages of European civilization, keeps alive the continuing formation of learned and scientific vocabulary in English and other European languages from these sources. The dependence on the classical languages in Europe is matched by a similar use of Sanskrit words for certain parts of learned vocabulary in some modern Indian languages (Sanskrit being the classical language of India). Such phenomena are examples of loanwords, one of the readiest sources for vocabulary extension.

Transmission of loanwords

Loanwords are words taken into a language from another language (the term borrowing is used for the process). Most obviously, this occurs when new things come into speakers' experiences as the result of contacts with speakers of other languages. This is part of the history of every language, except for one spoken by an impossibly isolated community. "Tea" from Chinese, "coffee" from Arabic, and "tomato," "potato," and "tobacco" from American Indian languages are familiar examples of loanwords designating new products that have been added to the vocabulary of English. In more abstract areas, several modern languages of India and Pakistan contain many words that relate to government, industry, and current technology taken in from English. This is the result of British role in these countries up to independence and the worldwide use of English as a language of international science since then.

In general, loanwords are rapidly and completely assimilated to the prevailing grammatical and phonological patterns of the borrowing language. The German word *Kindergarten,* literally "children's garden," was borrowed into English in the middle of the 19th century to designate an informal school for young children. It is now regularly pronounced as an English word, and the plural is kindergartens (not *Kindergarten,* as in German). Occasionally, however, some loanwords retain marks of their foreign origin: examples include Latin plurals such as cacti and narcissi (as contrasted with native patterns such as cactuses and narcissuses).

Languages differ in their acceptance of loanwords. An alternative way of extending vocabulary to cope with new products is to create a descriptive compound from within one's own language. English "aircraft" and "aeroplane" are, respectively, examples of a native compound and a Greek loan creation for the same thing. English "potato" is a loan; French *pomme de terre* (literally, "apple of the earth") is a descriptive compound. Chinese is particularly resistant to loans; "aircraft," "railway," and "telephone" are translated by newly formed compounds meaning literally "fly machine," "fire vehicle," and "lightning (electricity) language."

Language and conceptualization. The ability to speak and the ability to conceptualize are very closely linked, and the child learns both these skills together at the same time. This is not to say that thinking is no more than subvocal speech, as some behaviourists have proposed; most people can think pictorially and in simple diagrams, some to a greater degree than others, and one has the experience of responding rationally to external stimuli without intervening verbalization. But, as 18th-century thinkers saw, man's rationality developed and still goes hand in hand with his use of language, and a good deal of the flexibility of languages has been exploited in man's progressive understanding and conceptualizing of the world he lives in and of his relations with other men. Different cultures and different periods have seen this process differently developed. The anthropological linguist Edward Sapir put it well: "The 'real world' is to a large extent unconsciously built up on the language habits of the group."

Much of this lies in the irrecoverable prehistory of languages. The idea that there are still some primitive, al-

Language and thought patterns

most "fossil" languages, embodying a very low level of conceptualization, is a vain one. All that can be said is that languages are different and that, in part, the world is seen differently through the eyes of speakers of different languages. But, in some cases, part of the lexical adaptation of a language to developing thought patterns can be followed through. Ancient Greece saw a wholly unique growth and flowering of civilization in the 1st millennium BC, which has put virtually the entire civilized world in its debt ever since. In Greek, along with the emergence of certain abstract concepts and ways of thinking, one can follow some of the changes of word meanings and the coining of new words that accompanied this. As an example, the word *dikē* originally meant “way” or "manner"; thereafter, it acquired the meaning of the right way of doing something, the right way of behaving, and finally abstract right. Its derivative *dikaiosynē*, traditionally translated "justice," became the subject of philosophical debate and analysis by the Greek philosophers and covered almost the whole range of moral obligation involved in the relations of one person with others in society. Similar debate and refinement of key terms in the various branches of thought covered by Greek philosophy can be followed through; indeed, the term philosophy is directly taken from Greek *philosophia*, a compound formed not later than the 5th century BC from *philo-* (compare ***philein*** "to love") and ***sophia*** "wisdom" to refer to abstract speculation and debate of a fundamental nature about the world and man's place in it.

More recently, the development of the lexical resources of the languages of civilization can be observed, in one way or another, as they keep up with the scientific progress that dominates contemporary life.

An examination of the lexical structure of languages throws some light on the relations among various aspects of man's conceptualization. Spatial relations and their expression seem to lie very deep in the content of vocabulary. Words referring to time are drawn metaphorically from spatial words with great frequency: "a long/short time," "the near future," "far ahead/separated in time." Although time is a continuum, people readily divide it up into bits and record it rather as they do materials extended in space: "five years," "three months," “six seconds." This last use of vocabulary may be a particular trait of European languages and some others. An American Indian language is reported not to do this nearly so readily; it uses cardinal numbers only for discrete, countable objects. A separate class of words aligns the vocabulary of sequential time with that of intensity, so that repetition of the same activity again and again (to a European) is rather the intensification of a single activity. Certain differences in cultural attitudes and world outlook are said to accompany this kind of linguistic difference.

Expression of abstract relationships

Spatial terms are also freely used in the expression of other, more abstract relationships: "higher temperature," "higher quality," "lower expectations," “summit of a career," "far removed from any sensible course of action," "a distant relationship," "close friends," "over and above what had been said." It has been theorized that the linguistic forms most closely associated semantically with the expression of relations—case inflections in languages exhibiting this category—are originally and basically spatial in meaning. This “localist” theory, as it has been called, has been debated since the beginning of the 19th century and probably cannot be accepted as it stands, but the fact that it can be proposed and argued shows the dominant position that spatial relations hold in the conceptualization and verbalization of relations in other realms of thought.

It has been maintained that the human brain has a preference for binary oppositions, or polarities. If this is so, it will help explain the numerous pairs of related antonyms that are found: "good, bad"; "hot, cold"; "high, low"; "right, wrong"; "dark, light"; and so on. For finer discriminations, these terms can be put into more narrowly specified fields containing more than two terms taken together, but their most general use is in binary contrasts. Here, however, one term seems to represent the fundamental semantic category in question. In asking about size, one asks "How big is it?"; about weight, "How heavy is it?"; and about evaluation, "How good is it?" It is possible to ask how small, how light, or how bad something is, but such questions presuppose that the thing in mind has already been graded on the small side, on the light side, or on the bad side.

Style. The capacity for conceptualization possessed and developed by languages is by no means the only purpose language serves. A person's speech, supplemented by facial expression and gesture when speaker and hearer are mutually in sight, indicates and is intended to indicate a great deal more than factual information, inquiries, and requests. The fact that some of these other functions are performed by parts of a language usually mastered later by foreign learners gives rise to misinterpretation and often makes foreign speakers appear rude or insensitive when they are, in actuality, simply deploying fewer resources in the language.

Expression of emotional attitudes

Within the range of the structural and lexical possibilities of a language, speakers are able to convey their emotional attitudes and feelings toward the person or persons they are addressing and toward the subject matter of what they are saying. They are also able to conceal such feelings as one form of linguistic deception, though this is usually a harder task. These same resources are also exploited to arouse appropriate feelings and responses in others, again independently of any factual content. This is the chosen field of the propagandist, the preacher, the orator, the barrister, and the advertiser. All languages make use of intonation and voice qualities in these different ways; a person can produce and recognize the intonation and type of voice employed in coaxing, in pleading, in browbeating, and in threatening, in pleasure, and in anger, as well as those appropriate for matter-of-fact statements and the exposition of details about which the speaker has little or no emotional involvement. To describe exactly which phonetic features are brought into play is quite another matter, involving advanced competence in phonetic discrimination and analysis. This is one of the areas of speech about which all too little is currently known. Grammar and vocabulary are equally involved, though differently in each language. English speakers know the difference between "Come and give me a hand!" and "Could you possibly come and help me?'; "He's got the gift of gab” and "He is undoubtedly a fluent and persuasive speaker" are each appropriate for different occasions. By greetings and leave-takings a great deal of intended interpretation of the social relations between individuals can be expressed. Much of this is the “good manners" taught to children and expected of adults; these aspects of language behaviour vary from culture to culture, but in none are they wholly absent. It is, of course, equally possible to be deliberately bad mannered or deliberately to flout a linguistic convention or expectation, but this can be done only by knowing what is expected in the situation. The refinements of rudeness, like the refinements of politeness, insofar as the use of language is involved, require a very good knowledge of a language if it is other than one's mother tongue.

Written language is no less adapted to conveying more than just factual information, asking factual questions, and giving instructions. Intonation and tone of voice are clearly not reproducible in existing orthographic systems, but part of the skill of a novelist or a reporter is to convey these features of speech in his descriptions. Additionally, grammatical and lexical choices are available to the writer, as reading the examples above will show, and everyone knows the special artistry and techniques involved in composing written memorandums or letters if they are to achieve precisely the purpose for which they are intended.

Styles: variations within a language

These variations, written and spoken, within a language or within any dialect of a language, may be referred to as styles. Each time a person speaks or writes he does so in one or another style, deliberately chosen with the sort of considerations in mind that have just been mentioned, even though in speech the choice may often be routine. Sometimes style, especially in literature, is contrasted with "plain, everyday language." In using such plain, unmarked types of speaking or writing, however, one is no less choosing a particular style, even though it is the most commonly used one and the most neutral in that it conveys and arouses the least emotional involvement or personal feelings.

Stylistic differences are available to all mature native speakers and in literate communities to all writers, as well as to foreigners who know a second language really well. But there is undoubtedly a considerable range of skills in exploiting all the resources of a language, and, whereas all normal adults are expected to speak correctly and, if literate, to write correctly, communities have always recognized and usually respected certain individuals as preeminently skilled in particular styles, as orators, storytellers, preachers, poets, scribes, belletrists, and so forth. This is the material of literature. Once it is realized that oral literature is just as much literature as the more familiar written literature, it can be understood that there is no language devoid of its own literature.

In all languages certain forms of utterance have been considered worthy of preservation, study, and cultivation. In writing, the nature of written surfaces makes this fairly easy, though not all written material is deliberately preserved; much of it is deliberately destroyed, and, although the chance survival of inscriptions on stone or clay is of the greatest value to the archaeologist and historian, a good deal of such material was never intended to survive. Literature, on the other hand, is essentially regarded as of permanent worth. Printing and, in earlier days, the copying of manuscripts are the means of preserving written literature. In illiterate communities certain persons memorize narratives, poems, songs, prayers, ritual texts, and the like, and these are passed on, with new creations in such styles, to succeeding generations. Such skills, preservative as well as creative, are likely to be lost along with much of the surrounding culture under the impact of literacy. Here, modern technology in the guise of the tape recorder has come to the rescue, and many workers in the field of unwritten languages are recording specimens of oral literatures with transcriptions and translations while speakers having the requisite knowledge and skills are still available. A great amount of such material, however, must have been irretrievably lost from illiterate cultures before the 20th century.

Culture and literature

All languages have a literature, but different types of literature flourish in different languages and in different cultures. A warrior caste or a general respect for martial prowess fosters heroic verse or prose tales; strongly developed magical and mystery cults favour ritualistic types of oral or written literature; urban yearnings for the supposed joys of country life encourage the development of pastoral poetry, itself an outgrowth of the songs of shepherds and rural workers; and the same sense of the jadedness of city life is the best ground for the cultivation of satirical verse and prose, a form of literature probably confined largely to urban civilizations. Every language has the resources to meet these and other cultural requirements in its literature as the occasions arise, but some literary forms are more deeply involved in the structure of the language itself; this is made clear by the relative difficulty of translating certain types of literature and literary styles from one language to another. Poetry, in particular, is closely bound to the structure of the language in which it is composed, and poetry is notoriously difficult to translate from one language into another.

The special vocabularies and linguistic forms used in several games have already been mentioned. Here one may point to the widespread existence of verbal games themselves, based on the accidental features of a particular language. English-speaking children are accustomed to riddles, puns, and spelling games: "I spy with my little eye something beginning with *p*" (notice the regular formula with which this opens). These and similar word games have been found all over the world. Homer records the punning use by Odysseus of No-man (Greek *Outis*) as his name when he was about to attack Cyclops, who then roared out "No-man is killing me!" and so failed to attract any help (Odyssey 9:366–408). In some languages that make use of lexically distinctive tones, tone puns (words alike but for having different tones) are a form of word play.

As an intellectual challenge, the crossword puzzle in all its varieties, originally an American development early in the 20th century, has maintained and indeed greatly increased its popularity over much of the literate world that employs the Latin (Roman) alphabet. Crossword-puzzle solvers rely heavily on the relative probabilities of letter sequences in written words to suggest an answer to a partly filled line; and, depending on the particular style of the originator, crossword clues make use of many sorts of formal features in the language, among them spelling puns, spoken puns, and accidental letter sequences in words and phrases. To be able to solve a crossword puzzle in a second language shows a high degree of skill and knowledge therein.

LANGUAGE AND CULTURE

It has been seen that language is much more than the external expression and communication of internal thoughts formulated independently of their verbalization. In demonstrating the inadequacy and inappropriateness of such a view of language, attention has already been drawn to the ways in which one's mother tongue is intimately and in all sorts of details related to the rest of one's life in a community and to smaller groups within that community. This is true of all peoples and all languages; it is a universal fact about language.

Language as a part of culture

Anthropologists speak of the relations between language and culture. It is, indeed, more in accordance with reality to consider language as a part of culture. "Culture" is here being used, as it is throughout this article, in the anthropological sense, to refer to all aspects of human life insofar as they are determined or conditioned by membership in a society. The fact that a man eats or drinks is not in itself cultural; it is a biological necessity that he does so for the preservation of life. That he eats particular foods and refrains from eating other substances, though they may be perfectly edible and nourishing, and that he eats and drinks at particular times of day and in certain places are matters of culture, something "acquired by man as a member of society," according to the now-classic definition of culture by the English anthropologist Sir Edward Burnett Tylor. As thus defined and envisaged, culture covers a very wide area of human life and behaviour; and language is manifestly a part, probably the most important part, of it.

Although the faculty of language acquisition and language use is innate and inherited, and there is legitimate debate over the extent of this innateness, every individual's language is "acquired by man as a member of society," along with and at the same time as other aspects of that society's culture in which he is brought up. Society and language are mutually indispensable. Language can have developed only in a social setting, however this may have been structured, and human society in any form even remotely resembling what is known today or is recorded in history could be maintained only among people speaking and understanding a language in common use.

Transmission of language. Language is transmitted culturally; that is, it is learned. To a lesser extent it is taught, when parents deliberately encourage their children to talk and to respond to talk, correct their mistakes, and enlarge their vocabulary. But it must be emphasized that children very largely acquire their mother tongue (*i.e.*, their first language) by "grammar construction" from exposure to a random collection of utterances that they encounter. What is classed as language teaching in school either relates to second-language acquisition or, insofar as it concerns the pupils' first language, is in the main directed at reading and writing, the study of literature, formal grammar, and alleged standards of correctness, which may not be those of all the pupils' regional or social dialects. All of what goes under the title of language teaching at school presupposes and relies on the prior knowledge of a first language in its basic vocabulary and essential structure, acquired before school age.

Transmission of culture through language

If language is transmitted as part of culture, it is no less true that culture as a whole is transmitted very largely through language, insofar as it is explicitly taught. The fact that mankind has a history in the sense that animals do not is entirely the result of language. So far as researchers can tell, animals learn through spontaneous imitation or through imitation taught by other animals. This does not exclude the performance of quite complex and substantial pieces of cooperative physical work, such as a beaver's dam or an ants' nest, nor does it preclude the intricate social organization of some species, such as bees. But it does mean that changes in organization and work will be the gradual result of mutation cumulatively reinforced by survival value; those groups whose behaviour altered in any way that increased their security from predators or from famine would survive in greater numbers than others. This would be an extremely slow process, comparable to the evolution of the different species themselves.

There is no reason to believe that animal behaviour has materially altered during the period available for the study of human history, say the last 5,000 years or so, except, of course, when man's intervention by domestication or other forms of interference has itself brought about such alterations. Nor do members of the same species differ markedly in behaviour over widely scattered areas, again apart from differences resulting from human interference. Bird songs are reported to differ somewhat from place to place within species, but there is little other evidence for areal divergence. By contrast with this unity of animal behaviour, human cultures are as divergent as are human languages over the world, and they can and do change all the time, sometimes with great rapidity, as among the industrialized nations of the 20th century.

The processes of linguistic change and its consequences will be treated below. Here, cultural change in general and its relation to language will be considered. By far the greatest part of learned behaviour, which is what culture involves, is transmitted by vocal instruction, not by imitation. Some imitation is clearly involved, especially in infancy, in the learning process, but proportionately this is hardly significant.

Diffusion of knowledge through writing and printing

Through the use of language, any skills, techniques, products, modes of social control, and so on can be explained, and the end results of anyone's inventiveness can be made available to anyone else with the intellectual ability to grasp what is being said. Spoken language alone would thus vastly extend the amount of usable information in any human community and speed up the acquisition of new skills and the adaptation of techniques to changed circumstances or new environments. With the invention and diffusion of writing, this process widened immediately, and the relative permanence of writing made the diffusion of information still easier. Printing and the increase in literacy only further intensified this process. Modern techniques for almost instantaneous transmission of the written and spoken word all over the globe, together with the rapid translation services now available between the major languages of the world, have made it possible for usable knowledge of all sorts to be made accessible to people almost anywhere in the world in a very short time. This accounts for the great rapidity of scientific, technological, political, and social change in the contemporary world. All of this, whether ultimately for the good or ill of mankind, must be attributed to the dominant role of language in the transmission of culture.

Variations within languages. The part played by variations within a language in differentiating social and occupational groups in a society has already been referred to above. In language transmission this tends to be self-perpetuating unless deliberately interfered with. Children are in general brought up within the social group to which their parents and immediate family circle belong, and they learn the dialect and speaking styles of that group along with the rest of the subculture and behavioral traits and attitudes that are characteristic of it. This is a largely unconscious and involuntary process of acculturation, but the importance of the linguistic manifestations of social status and of social hierarchies is not lost on aspirants for personal advancement in stratified societies. The deliberate cultivation of an appropriate dialect, in its lexical, grammatical, and phonetic features, has been the self-imposed task of many persons wishing "to better themselves" and the butt of unkind ridicule on the part of persons already feeling themselves secure in their social status or unwilling to attempt any change in it. Much of the comedy in George Bernard Shaw's *Pygmalion* turns on Eliza's need to unlearn her native Cockney if she is to rise in the social scale. Conversely, it is readily apparent today that middle class people, mostly adolescents, who for some reason want to "opt out" of the social group of their parents make every effort to abandon the distinctive aspects of the social dialect that would mark them, along with dress and general behaviour, as members of a group whose mores they are, at least temporarily, affecting to reject. Culturally and subculturally determined taboos play a part in all this, and persons desirous of moving up or down in the social scale have to learn what words to use and what words to avoid if they are to be accepted and to "belong" in their new position. All through the ages, a good part of the material for "comedies of manners" has come from the social role of language variation within a society.

The same considerations apply to changing one's language as to changing one's dialect. Language changing is harder for the individual and is generally a rarer occurrence, but it is likely to be widespread in any mass immigration movement. In the 19th and early 20th centuries, the eagerness with which immigrants and the children of immigrants from continental Europe living in the United States learned and insisted on speaking English is an illustration of their realization that English was the linguistic badge of full membership in their new homeland, at the time when the country was proud to consider itself as the melting pot in which people of diverse linguistic and cultural origins would become citizens of a unified community.

Special vocabularies of games, trades, professions

The same sort of self-perpetuation, in the absence of deliberate rejection, operates in the special languages of games and of trades and professions (these are in the main concerned with special vocabularies). Game learners, apprentices, and professional students learn the locutions together with the rest of the game or the job. The specific words and phrases occur in the teaching process and are observed in use, and the novice is only too eager to display an easy competence with such phraseology as a mark of his full membership of the group; *e.g.*, golfers are keen to talk of birdies, fairways, and slicing.

Languages and variations within languages play both a unifying and a diversifying role in human society as a whole. Language is a part of culture, but culture is a complex totality containing many different features, and the boundaries between cultural features are not clear-cut, nor do they all coincide. Physical barriers such as oceans, high mountains, and wide rivers constitute impediments to human intercourse and to culture contacts, though modern technology in the fields of travel and communications make such geographical factors of less and less account. More potent today are political restrictions on the movement of people and of ideas, such as divide western Europe from Communist eastern Europe; the frontiers between these two political blocs represent much more of a cultural dividing line than any other European frontiers.

The distribution of the various components of cultures differs, and the distribution of languages may differ from that of nonlinguistic cultural features. This results from the varying ease and rapidity with which changes may be acquired or enforced and from the historical circumstances responsible for these changes. In contemporary Europe, as the result of World War II, a major political and cultural division cuts across an area of relative linguistic unity in East and West Germany. It is significant, however, that differences of vocabulary and usage are already noticeable in the German speech from each side, overlying earlier differences attributed to regional dialects; one may surmise that, if the present political situation endures for several more generations, the East–West frontier will come to mark a definite dialect boundary within the German language as well.

Language distribution. Language, no less than other aspects of human behaviour, is subject to purposive interference. When people with different languages need to communicate, various expedients are open to them, the most obvious being second-language learning and teaching. This takes time, effort, and organization, and, when more than two languages are involved, the time and effort are that much greater. Most people are monolingual, and those with a working knowledge of three or four languages are much fewer than those with a competence in just one second language. Other expedients may also be applied. Ad hoc pidgins for the restricted purposes of trade and administration were mentioned above. Tacit or deliberate agreements have been reached whereby one language is chosen for international purposes when speakers of several different languages are involved. In the Roman Empire, broadly, the western half used Latin as a lingua franca, and the eastern half used Greek. In western Europe during the Middle Ages, Latin continued as the international language of educated people, and Latin was the second language taught in schools. Later, the cultural, diplomatic, and military reputation of France made French the language of European diplomacy. This use of French as the language of international relations persisted until the present century. At important conferences among representatives of different nations, it is usually agreed which languages shall be officially recognized for registering the decisions reached; and the provisions of treaties are interpreted in the light of texts in a limited number of languages, those of the major participants.

International and diplomatic languages

Since World War II the dominance of the English-speaking peoples in science and technology and in international commerce has led to the recognition of English as the major international language in the world of practical affairs, with more and more countries making English the first foreign language to be taught and thus producing a vast expansion of English-language-teaching programs all over the world. Those whose native language is English do not sufficiently realize the amount of effort, by teacher and learner alike, that is put into the acquisition of a working knowledge of English by educated first speakers of other languages.

As an alternative to the recognition of particular natural languages as international in status, attempts have been made to invent and propagate new and genuinely international languages, devised for the purpose. Of these, Esperanto, invented by the Polish–Russian doctor L.L. Zamenhof in the 19th century, is the best known. Such languages are generally built up from parts of the vocabulary and grammatical apparatus of the better known existing languages of the world. The relationship between the written letter and its pronunciation is more systematic than with many existing orthographies (English spelling is notoriously unreliable as an indication of pronunciation), and care is taken to avoid the grammatical irregularities to which all natural languages are subject and also to avoid sounds found difficult by many speakers (*e.g.*, the English ***th*** sounds, which most Europeans, apart from English speakers, dislike). Despite the high hopes of their inventors, these artificial languages have not made much progress, though an international society of Esperanto speakers does exist (for more information, see also INTERNATIONAL LANGUAGE).

Deliberate interference with the natural course of linguistic changes and the distribution of languages is not confined to the facilitating of international intercourse and cooperation. Language as a cohesive force for nation-states and for linguistic groups within nation-states has for long been manipulated for political ends. Multilingual states can exist and prosper; Switzerland is a good example. But linguistic rivalry and strife can be disruptive. Language riots have occurred in Belgium between French and Flemish speakers and in parts of India between rival vernacular communities. A language can become or be made a focus of loyalty for a minority community that thinks itself suppressed, persecuted, or subjected to discrimination. The French language in Canada in the mid-20th century is an example. In the 19th and early 20th centuries Irish Gaelic came to symbolize Irish patriotism and Irish independence from Great Britain. Since independence, government policy continues to insist on the equal status of English and Irish in public notices and official documents, but, despite such encouragement and the official teaching of Irish in the state schools, a main motivation for its use and study has disappeared, and the language is giving ground to English under the international pressures referred to above.

Language as a cohesive political force

For the same reasons, a language may be a target for attack or suppression, if the authorities associate it with what they consider a disaffected or rebellious group or even just a culturally inferior one. There have been periods when American Indian children were forbidden to speak a language other than English at school and when pupils were not allowed to speak Welsh in British state schools in Wales. Both these prohibitions have been abandoned. Since the Spanish Civil War of the 1930s Basque speakers have been discouraged from using their language in public, as a consequence of the strong support given by the Basques to the republican forces. Interestingly, on the other side of the Franco-Spanish frontier, French Basques are positively encouraged to keep their language in use, if only as an object of touristic interest and consequent economic benefit to the area.

Translation. So far, some of the relatively large-scale effects of culture contacts on languages and on dialects within languages have been surveyed. A continuous concomitant of contact between two mutually incomprehensible tongues and one that does not lead either to suppression or extension of either is translation. As soon as two speakers of different languages need to converse, translation is necessary, either through a third party or directly.

Before the invention and diffusion of writing, translation was instantaneous and oral; persons professionally specializing in such work were called interpreters. In predominantly or wholly literate communities, translation is thought of as the conversion of a written text in one language into a written text in another, though the modern emergence of the simultaneous translator or professional interpreter at international conferences keeps the oral side of translation very much alive.

Oral and written translation

The tasks of the translator are the same whether the material is oral or written, but, of course, translation between written texts allows more time for stylistic adjustment and technical expertise. The main problems have been recognized since antiquity and were expressed by St. Jerome, translator of the famed Latin Bible, the Vulgate, from the Hebrew and Greek originals. Semantically, these problems relate to the adjustment of the literal and the literary and the conflicts that so often occur between an exact translation of each word, as far as this is possible, and the production of a whole sentence or even a whole text that conveys as much of the meaning of the original as can be managed. These problems and conflicts arise because of factors already noticed in the use and functioning of language: languages do not operate in isolation but within and as part of cultures, and cultures differ from each other in various ways. Even between the languages of communities whose cultures are fairly closely allied, there is by no means a one-to-one relation of exact lexical equivalence between the items of their vocabularies.

In their lexical meanings, words acquire various overtones and associations that are not shared by the nearest corresponding words in other languages; this may vitiate a literal translation. The English author and theologian Ronald Knox has pointed to the historical connections of the Greek ***skandalon*** "stumbling block, trap, or snare," inadequately rendered by "offense," its usual New Testament translation. In modern times translators of the Bible into the languages of peoples culturally remote from Europe are well aware of the difficulties of finding a lexical equivalent for "lamb," when the intended readers, even if they have seen sheep and lambs, have no tradition of blood sacrifice for expiation nor long-hallowed associations of lambs with lovableness, innocence, and apparent helplessness. The English word uncle has, for various reasons, a cozy and slightly comic set of associations. The Latin poet Virgil uses the words *avunculus Hector* in a solemn heroic passage of the Aeneid (Book III, line 343); to translate this by "uncle Hector" gives an entirely unsuitable flavour to the text.

Poetry translation

The translation of poetry, especially into poetry, presents very special difficulties, and the better the original poem, the harder the translator's task. This is because poetry is, in the first instance, carefully contrived to express exactly what the poet wants to say. Second, to achieve this end, the poet calls forth all the resources of the language in which he is writing, matching the choice of words, the order of words, and grammatical constructions, as well as phonological features peculiar to the language in metre, perhaps supplemented by rhyme, assonance, and alliteration. The available resources differ from language to language; English and German rely on stress-marked metres, but Latin and Greek used quantitative metres, contrasting long and short syllables, while French places approximately equal stress and length on each syllable. The translator must try to match the stylistic exploitation of the particular resources in the original language with comparable resources from his own. Because lexical, grammatical, and metrical considerations are all interrelated and interwoven in poetry, a satisfactory literary translation is usually very far from a literal word for word rendering. The more the poet relies on language form, the more embedded his verses are in that particular language, and the harder they are to translate adequately. This is especially true with lyrical poetry in several languages, with its wordplay, complex rhymes, and frequent assonances.

At the other end of the translator's spectrum, technical prose dealing with internationally agreed scientific subjects is probably the easiest type of material to translate, because cultural unification (in this respect), lexical correspondences, and stylistic similarity already exist in this type of usage in the languages most commonly involved, to a higher degree than in other fields of discourse.

Significantly, it is this last aspect of translation to which mechanical and computerized techniques are being applied with some prospects of limited success. Machine translation, whereby, ultimately, a text in one language could be fed into a machine to produce an accurate translation in another language without further human intervention, has been largely concentrated on the language of science and technology, with its restricted vocabulary and overall likeness of style, for both linguistic and economic reasons. Attempts at machine translation of literature have been made, but success in this field, more especially in the translation of poetry, seems very remote at present.

Translation as an art

Translation on the whole is an art, not a science. Guidance can be given and general principles can be taught, but after that it must be left to the individual's own feeling for the two languages concerned. Almost inevitably, in a translation of a work of literature something of the author's original intent must be lost; in those cases in which the translation is said to be a better work than the original, an opinion sometimes expressed about the English writer Edward Fitzgerald's "translation" of The *Rubdiydt of Omar Khayyám*, one is dealing with a new, though derived, work, not just a translation. The Italian epigram remains justified: *Traduttore traditore* "The translator is a traitor."

Messages and codes. Translation serves to extend the communicative value of a text. Sometimes people want to restrict it. Confidential messages, spoken and written, require for their efficacy that they be known to and understood by only the single person or the few persons to whom they are addressed. Such are diplomatic exchanges, operational messages in wartime, and some transmissions of commercial information. Protection of written messages from interception has been practiced for many centuries. Recent developments in telegraphy and telephony have made protection against unauthorized reception more urgent, whether of texts transmitted as speech or as series of letters of the alphabet. Scrambling of telephony is a common expedient; the wave frequencies through which the sounds are to be transmitted are altered at the source so as to be unrecognizable and then reconverted by the intended recipient's receiver. Codes and ciphers (cryptography) are of much longer standing in the concealment of written messages, though their techniques are being constantly developed. Such gains are, of course, countered by developments in the techniques of decipherment and decoding (as distinct from getting hold of the key to the system in use). An important by-product of such techniques has been the reading and interpretation of inscriptions written in otherwise unknown languages or unknown writing systems for which no translation exists. The recent, very significant decipherment of the Linear B script and its recognition as Mycenaean Greek, an early Greek dialect written in a form of orthography quite distinct from the later classical Greek alphabet, was first achieved by the application of cryptographic "code cracking" methods (see also CRYPTOLOGY).

LANGUAGE LEARNING: BILINGUALISM AND LITERACY

Role of the brain in language learning

Every physiologically and mentally normal person has learned the main structure and basic vocabulary of his mother tongue by the end of childhood. It has been pointed out that the process of first-language acquisition as a spoken medium of communication is largely achieved from random exposure. There is legitimate controversy, however, over the nature and extent of the positive contribution that the human brain brings to the activity of grammar construction, the activity by which the child develops an indefinitely creative competence from the finite data that make up his actual experience of the language. Creativity is what must be stressed as the product of first-language acquisition. By far the greater number of all the sentences anyone hears and utters during his lifetime are new; that is, they have not occurred before in his personal experience. But individuals find no difficulty at all in understanding at once almost everything they hear nor for the most part in producing sentences to suit the requirements of every situation. This very ease of creativity in man's linguistic competence makes it hard to realize its extent. The only regularly reproduced sentences in most speakers' experience are the stereotyped forms of greeting and leave-taking and certain formalized responses to recurrent situations, such as shopping, cooperative activities in repetitive jobs, the stylized parts of church services, and the like.

Yet, despite this really immense achievement that the progressive mastery of one's first language constitutes, it arouses no comment and attracts no credit. It is simply part of what is expected of one in growing up. Different people may be singled out for praise in certain uses of their language, as good public speakers, authors, poets, tellers of tales, and solvers of puzzles, but not just as speakers. The credit that some individuals acquire in certain communities for "speaking correctly" is a different matter, usually the result of speaking as one's mother tongue a prestigious standard dialect among people most of whom speak another, less favoured one.

Bilingualism. The learning of a second and of any subsequently acquired language is quite a different matter. Except for one form of bilingualism (see below), it is a deliberate activity undertaken when one has already nearly or fully acquired the basic structure and vocabulary of one's first language. Of course, many people never do master significantly more than their own first language. It is only in encountering a second language that one realizes how complex language is and how much effort must be devoted to subsequent acquisition. It has been said that the principal obstacle to learning a language is knowing one already, and it may also be that the faculty of grammar construction exhibited in childhood is one that is gradually lost as childhood recedes.

Whereas every normal person masters his mother tongue with unconscious ease, people vary in their ability to learn additional languages, just as they vary in other intellectual activities. Situational motivation, however, Ability to learn additional languages

appears to be by far the strongest influence on the speed and apparent ease of this learning. The greatest difficulty is experienced by those who learn because they are told to or are expected to, without supporting reasons that they can justify. Given a motive other than external compulsion or expectation, the task is achieved much more easily (this, of course, is an observation in no way confined to language learning). In Welsh schools it is found that English children make slower progress in Welsh when their only apparent reason for learning Welsh is that there are Welsh classes. Welsh children, on the other hand, make rapid progress in English, the language of most further education, the newspapers, most television and radio, most of the better paid jobs, and of any job outside Welsh-speaking areas. Similar differences in motivation have accounted for the excellent standard of English, French, and German acquired by educated persons in the Scandinavian countries and in Holland, small countries whose languages, being spoken by relatively few foreigners, are of little use in international communication. This attainment may be compared with the much poorer showing in second-language acquisition among comparably educated persons in England and America, who have for long been able to rely on foreigners accommodating to their ignorance by speaking and understanding English.

It is often held that children brought up bilingually in places in which two languages are regularly in use are slower in schoolwork than comparable monolingual children, as a greater amount of mental effort has to be expended in the mastery of two languages. This is by no means proved; and, because much of a child's language acquisition takes place in infancy and in the preschool years, it does not represent an effort in the way that consciously learning a language in school does, and indeed it probably occupies a separate part of the child's mental equipment. The question of speed of general learning by bilinguals and monolinguals must be left open. It is quite a separate matter from the job of learning, by teaching at home or in school, to read and write in two languages; this undoubtedly is more of a labour than the acquisition of monolingual literacy.

Types of bilingualism

Two types of bilingualism have been distinguished, according to whether the two languages were acquired from the simultaneous experience of the use of both in the same circumstances and settings or from exposure to each language used in different settings (an example of the latter is the experience of English children living in India during the period of British ascendancy there, learning English from their parents and an Indian language from their nurses and family servants). However acquired, bilingualism leads to mutual interference between the two languages; extensive bilingualism within a community is sometimes held partly responsible for linguistic change (see below). Interference may take place in pronunciation, in grammar, and in the meanings of words. Bilinguals often speak their two languages each with "an accent"; *i.e.*, they carry into each certain pronunciation features from the other. The German word order in "He comes tomorrow home" has been reported as an example of grammatical interference; and in Canadian French the verb introduire has acquired from English the additional meaning "introduce, make acquainted" (which in metropolitan French is *présenter*).

Literacy. The acquisition of literacy is something very different from the acquisition of one's spoken mother tongue, even when the same language is involved, as it usually is. Both skills, speaking and writing, are learned skills, but there the resemblance ends. The child learns his first language at the start involuntarily and mostly unconsciously from random exposure, even if no attempts at teaching are made. Literacy is deliberately taught and consciously and deliberately learned. There is current debate on the best methods and techniques for teaching literacy in various social and linguistic settings. Literacy is learned through speech, by a person already possessed of the basic structure and vocabulary of his language.

Such facts should be very obvious, but the now-accepted, though fairly recent, standard of near-universal literacy in technologically advanced countries, along with the fact that in second-language learning one usually acquires speech and writing skills at the same time, tends to bring these two parts of language learning under one head. Literacy is manifestly a desirable attainment for all communities, though not necessarily in all languages. It must be borne in mind that there are many distinct languages spoken in the world today by fewer than 1,000 or 500 or even 50 persons. The capital investment in literacy, including teaching resources, teacher time and training, printing, publications, and so forth, is vast, and it can be economically and socially justified only when applied to languages spoken and likely to continue to be spoken by substantial numbers over a wide area.

Effects of literacy

Literacy is in no way necessary for the maintenance of linguistic structure or vocabulary, though it does enable people to add words from the common written stock in dictionaries to their personal vocabulary very easily. It is worth emphasizing that until relatively recently in human history all languages were spoken by illiterate speakers and that there is no essential difference as regards pronunciation, structure, and complexity of vocabulary between spoken languages that have writing systems used by all or nearly all their speakers and the languages of illiterate communities.

Literacy has many effects on the uses to which language may be put; storage, retrieval, and dissemination of information are greatly facilitated, and some uses of language, such as philosophical system building and the keeping of detailed historical records, would scarcely be possible in a totally illiterate community. In these respects the lexical content of a language is affected, for example, by the creation of sets of technical terms for philosophical writing and debate. Because the permanence of writing overcomes the limitations of auditory memory span imposed on speech, sentences of greater length can easily occur in writing, especially in types of written language that are not normally read aloud and that do not directly represent what would be spoken. An examination of some kinds of oral literature, however, reveals the ability of the human brain to receive and interpret spoken sentences of considerable grammatical complexity.

In relation to pronunciation, writing does not prevent the historical changes that occur in all languages. Part of the apparent irrationality of English spelling, such as is found also in some other orthographies, lies just in the fact that letter sequences have remained constant while the sounds represented by them have changed. For example, the *gh* of "light" once stood for a consonant sound, as it still does in the word as pronounced in some Scots dialects; and the k of "knave" and "knight" likewise stood for an initial k sound (compare the related German words *Knabe* and *Knecht*). A few relatively uncommon words, including some proper names, are reformed phonetically, specifically to bring their pronunciation more in line with their spelling. Spelling pronunciations, as these are called, are a product of general literacy. In London, the pronunciation of "St. Mary Axe" as if it were spelled "Simmery Axe" is now decidedly old-fashioned. "St. John" and "St. Clair" survive as proper names with their old pronunciations, in the latter case helped by the presence of the alternative spelling "Sinclair."

WRITTEN LANGUAGE

Historically, culturally, and in the individual's life, writing is subsequent to speech and presupposes it. Aristotle expressed the relation thus: "Speech is the representation of the experiences of the mind, and writing is the representation of speech" (On Interpretation). But it is not as simple as this would suggest. Alphabetic writing, in which, broadly, consonant and vowel sounds are indicated by letters in sequence, is the most widespread system in use today, and it is the means by which literacy will be disseminated, but it is not the only system, nor is it the earliest.

Evolution of writing systems. Writing appears to have been evolved from an extension of picture signs: signs that directly and iconically represented some thing or action and then the word that bore that meaning. Other

Character scripts

words or word elements not readily represented pictorially could be assigned picture signs already standing for a word of the same or nearly the same pronunciation, perhaps with some additional mark to keep the two signs apart. This sort of device is used in children's word puzzles, as when the picture of a berry is used to represent, say, the second half of the name Canterbury. This opens the way for what is called a character script, such as that of Chinese, in which each word is graphically represented by a separate individual symbol or character or by a sequence of two or more such characters. Writing systems of this sort have appeared independently in different parts of the world.

Chinese character writing has for many centuries been stylized, but it still bears marks of the pictorial origin of some characters. Chinese characters and the characters of similar writing systems are sometimes called ideograms, as if they directly represented thoughts or ideas. This is not so. Chinese characters stand for Chinese words or, particularly as in modern Chinese, bits of words; they are the symbolization of a particular language, not a potentially universal representation of thought. The & sign, standing for "and in English printing, is a good isolated example of a character used in an alphabetic writing system.

Character writing is laborious to learn and imposes a burden on the memory. Alternatives to it, in addition to alphabetic writing, include scripts that employ separate symbols for the syllable sequences of consonants and vowels in a language, with graphic devices to indicate consonants not followed by a vowel. The Devanagari script, in which classical Sanskrit and modern Hindi are written, is of this type, and the Mycenaean writing system, a form of Greek writing in use in the 2nd millennium BC and quite independent of the later Greek alphabet, was syllabic in structure. Japanese employs a mixed system, broadly representing the roots of words by Chinese characters (the Japanese learned writing from the Chinese in and after the 5th century AD) and the inflectional endings by syllable signs. These syllable signs are an illustration of the way in which a syllabic script can develop from a character script: certain Chinese characters were selected for their sound values alone and, reduced in size and complexity, have been standardized as signs of a particular consonant and vowel sequence or of a single vowel sound.

Greek alphabet

The Greek alphabet came from the Phoenician script, a syllabic-type writing system that indicated the consonant sounds. By a stroke of genius, a Greek community decided to employ certain consonantal signs to which no consonant sound corresponded in Greek as independent vowel signs, thus producing an alphabet, a set of letters standing for consonants and vowels. The Greek alphabet spread over the ancient Greek world, undergoing minor changes. From a Western version sprang the Latin (Roman) alphabet. Also derived from the Greek alphabet, the Cyrillic alphabet was devised in the 9th century AD by a Greek missionary, St. Cyril, for writing the Slavic languages.

Spelling. Alphabetic writing is not and cannot be an exact representation of the sequence of sounds or even of the sequence of distinctive sounds in the spoken forms of words and sentences. "Consonant" and "vowel" mean different things when applied to letters and to sounds, though there is, of course, much overlap. The y at the beginning of "yet" stands for a consonant sound; at the end of "jetty" it stands for a vowel sound. In “thick” and "thin" the sequence *th* represents a single sound, not a *t* sound followed by an h sound. In "kite" the *e* represents no sound directly but distinguishes the vowel between k and *t* from the vowel in “kit.” These disharmonies arise from a number of causes. Economy in the use of letters is one factor. In addition, spoken forms are always changing over the centuries, whereas writing, particularly since the invention of printing, is very conservative. At one time the *e* at the end of words such as "kite" did stand for a vowel sound. This sound was lost between the 14th and 16th centuries, a time when other changes in the pronunciation of such words also occurred. The notorious *ough* spellings in English, standing for different sounds and sound sequences in "rough," "cough," "dough," "plough," "ought," and other such words, have arisen from historical changes that have driven spelling and pronunciation further apart.

This, of course, does not mean that spelling reforms are out of the question. Spelling reform has been talked of in relation to English for many centuries without much effect; but in some countries—for example, Norway and Holland—official action has prescribed certain reforms to be made, and these have then been taught in school and have gradually found their way into printed works. The sheer volume of printed matter preserved for use and consultation in the modern world adds much weight against the convenience otherwise accruing from reforms designed to correct the historically produced disharmonies between spelling and pronunciation.

Representing grammatical distinctions in spelling

Moreover, it is not always most useful for spellings to represent exactly the sound sequences in a word and nothing else; this is the task for which phoneticians have devised transcriptions. As far as the sounds themselves are concerned, the plural signs of "cats," "dogs," and "horses" are different: the final sound of "cats" is like the initial sound of "sink," that of "dogs" like the initial sound of “zinc,” and the plural of "horse" is indicated by a sound sequence rather like that in "is." But they are all indicated in writing by one and the same letter and always have been, because only one grammatical distinction, that of singular as against plural, is involved, and at this point in the language the actual differences in the sounds, important elsewhere, are irrelevant.

Letters, insofar as they stand for sounds, stand for consonants and vowels. But other sound features are involved in languages. In English words the location of the stress is important, and the words "import" as a noun and "import" as a verb are distinguished by this alone. All languages make use of sequences of rises and falls in pitch, called intonation, as part of spoken communication. These phenomena are unrepresented in orthography except for certain punctuation marks such as ? and ! and sometimes by italicization and underlining.

This is not a weakness in orthography. Writing is normally intended to be read and when necessary read aloud by people who already know the language and are therefore able to supply from their own competence the required detail. For specific purposes such as foreign-language teaching, as well as for the specific study of pronunciation and speech sounds in phonetics and phonology, various forms of transcription have been devised to indicate unambiguously by written signs the precise form of the spoken utterance, without regard to other considerations.

Written versus spoken languages. For these reasons one should distinguish the grammar of a written language (*e.g.*, written English) from the grammar of the corresponding spoken language (spoken English). The two grammars will be very similar, and they will overlap in most places; but the description of spoken English will have to take into account the grammatical uses of features such as intonation, largely unrepresented in writing, and the description of written English must deal adequately with the greater average length of sentences and some different syntactic constructions and word forms characterizing certain written styles but almost unknown in ordinary speech (*e.g.*, "whom" as the objective form of "who").

Ancient, or dead, languages

In studying ancient (dead) languages one is, of course, limited to studying the grammar of their written forms and styles, as their written records alone survive. Such is the case with Latin, Ancient Greek, and Sanskrit (Latin lives as a spoken language in very restricted situations, such as Roman Catholic services and as the official language of some closed religious communities, but this is not the same sort of Latin as that studied in classical Latin literature; Sanskrit survives also as a spoken language in similarly restricted situations in a few places in India). Scholars may be able to reconstruct something of the pronunciation of a dead language from historical inferences and from descriptions of its pronunciation by authors writing when the language was still spoken. They know a good deal about the pronunciation of Greek and Latin and a great deal about the pronunciation of Sanskrit, because ancient Indian scholars left a collection of extremely detailed and systematic literature on its pronunciation. But this does not alter the fact that when one teaches and learns dead languages today, largely for their literary value and because of the place of the communities formerly speaking them in our own cultural history, one is teaching and learning the grammar of their written forms. Indeed, despite what is known about the actual pronunciation of Greek and Latin, Europeans on the whole pronounce what they read in terms of the pronunciation patterns of their own languages.

Under present conditions, with universal literacy either an accepted fact or an accepted target, it is assumed that, wherever it is convenient or useful, writing may be employed for any purpose for which speech might have been used and by all sections of the community. This has not always been so. Literacy was until the 19th century the privilege of the few. In other periods and cultures, writing was the preserve of certain defined groups, such as the priesthood and the official class, and it was restricted to certain purposes, such as the annals of important events, genealogical tables, and records of inventories of things and persons. It is highly probable that writing first developed for particular types of use by particular groups of specialists within communities and subsequently, because of its obvious utility, spread outside these limits.

For further accounts of writing systems in greater detail, see WRITING, FORMS OF; ALPHABETS.

LINGUISTIC CHANGE

Every language has a history; and, as in the rest of human culture, changes are constantly taking place in the course of the learned transmission of a language from one generation to another. This is just part of the differences between human culture and animal behaviour. Languages change in all their aspects, in their pronunciation, word forms, syntax, and word meanings (semantic change). These changes are mostly very gradual in their operation, becoming noticeable only cumulatively over the course of several generations. But, in some areas of vocabulary, particular words closely related to rapid cultural change are subject to equally rapid and therefore noticeable changes within a generation or even within a decade. In the 20th century the vocabulary of science and technology is an outstanding example. The same is also true of those parts of vocabulary that are involved in fashionable slangs and jargons, whose raison d'être in promoting group, particularly age-group, solidarity depends on their being always fresh and distinctive. Old slangs date, as any reading of a novel or visit to a film more than ten years old is apt to show. The rapid obsolescence of young people's slangs is equally to be seen in the unsuccessful efforts of some well-intentioned older persons who vainly attempt to cultivate the speech styles of present-day youth groups in a misdirected attempt to bridge "the generation gap" (this last phrase is an example of mid-20th-century pseudoscientific slang).

Changes in vocabulary

Development of language families. In the structural aspects of language, their pronunciation and grammar, and in vocabulary less closely involved in rapid cultural movement, the processes of linguistic change are best observed by comparing written records of a language over extended periods. This is most readily seen by English speakers through setting side by side present-day English texts with 18th-century English, the English of the Authorized Version of the Bible, Shakespearean English, Chaucer's English, and the varieties of Old English (Anglo-Saxon) that survive in written form. Noticeably, as one goes back in time, the effort required in understanding increases, and, while people do not hesitate to speak of "Shakespearean English," they are more doubtful about Chaucer, and for the most part Old English texts are as unintelligible to a modern English speaker as, for example, texts in German. It is clear that the differences involved include word meanings, grammar, and, so far as this can be reconstructed, pronunciation.

Similar evidence, together with what is known of the cultural history of the peoples concerned, makes clear the continuous historical connections linking French, Spanish, Portuguese, Italian, and Romanian with the spoken ("vulgar") Latin of the western Roman Empire. This group constitutes the Romance subfamily of languages and is an example of how, as the result of linguistic change over a wide area, a group of distinct, though historically related, languages comes into being.

In the transmission of a language from parent to child, slight deviations in all aspects of language use occur all the time, and as the child's speech contacts widen he confronts a growing range of slight differences in personal speech forms, some of them correlating with social or regional differences within a community, these speech differences themselves being the results of the transmission process. As a consequence, the child's speech comes to differ slightly from that of his parents' generation. In urbanized communities an additional factor is involved: children have been shown to be effectively influenced by the speech habits of their peer groups once they have made contacts with them in and out of school.

Cumulative changes in languages

Such changes, though slight at the time, are progressively cumulative. Since ready intercommunication is a primary purpose of language, as long as a community remains unitary, with strong central direction and a central cultural focus, such changes will not go beyond the limits of intercomprehensibility. But in more scattered communities and in larger language areas, especially when cultural and administrative ties are weakened and broken, these cumulative deviations in the course of generations give rise to wider regional differences. Such differences take the form of dialectal differentiation as long as there is some degree of mutual comprehension but eventually result in the emergence of distinct languages. This is what happened in the history of the colloquial Latin of the western Roman Empire, and it can be assumed that a similar course of events gave rise to the separate Germanic languages (English, German, Dutch, Danish, Norwegian, Swedish, and some others), though in this family the original unitary language is not known historically but inferred as "Common Germanic" or "Proto-Germanic" and tentatively assigned to early in the 1st millennium BC as the period before separation began.

This is how language families have developed. Most but not all of the languages of Europe belong to the Indo-European family, so-called because in addition it includes the classical Indian language Sanskrit and most of the modern languages of northern India and Pakistan. It includes as subfamilies the two families just mentioned, Romance and Germanic, and several others. It is assumed that the subfamilies, and from them the individual languages of the Indo-European family, are ultimately derived from a unitary language spoken somewhere in eastern Europe or western Asia (its exact location is still under debate), perhaps 5,000 years ago. This unitary language has itself been referred to as "Indo-European," "Proto-Indo-European," the "common parent language," or the "original language" (*Ursprache*) of the family. But it must be emphasized that, whatever it may have been like, it was just one language among many and of no special status in itself. It was certainly in no way the original language of mankind or anything like it. It had its own earlier history, of which virtually nothing can be inferred, and it was, of course, very recent in relation to the time span of human language itself. What is really special about such "parent" or "proto-" languages is that they represent the farthest point to which our available techniques and resources enable us to reconstruct the prehistory of our attested and living languages. Similarly constituted families of languages derived from inferred common sources have been established for other parts of the world; for example, Altaic, covering Turkish and several languages of Central Asia, and Bantu, containing many of the languages of central and southern Africa. For further details of these and other language families see LANGUAGES OF THE WORLD and the specialist articles on the major language families, subfamilies, and languages.

"Parent" or "proto-" languages

If enough material in the form of written records from past ages were available, it would be possible to group all the world's languages into historically related families. In addition, an answer could perhaps be posited to the question of whether all languages are descended from a single original language or whether languages emerged independently among several groups of early peoples (the rival theories of monogenesis and polygenesis, a controversy more confidently disputed in the 19th century than today). In actual fact, written records, when they are available, go back only a fraction of the time in which human speech has been developed and used, and over much of the globe written records are nonexistent. In addition, there are no other linguistic fossils comparable to the fossils of geological prehistory. This means that the history and prehistory of languages will not be able to go back more than a few thousand years BC and will be much more restricted in language areas in which few or no written records are available, as in much of Africa and in South America. Many languages will remain not related with certainty to any family. Nevertheless, the methods of historical linguistics, involving the precise and systematic comparison of word forms and word meanings (see further LINGUISTICS), have produced remarkable results in establishing language families on the same basis as Indo-European was established, in far less favourable fields. But any attempt by these means to get back to "the origin of language" or to reconstruct man's original language, if indeed there was one, is quite beyond the reach of science and will remain so.

Causes of language change. The fundamental cause of linguistic change and hence of linguistic diversification is the minute deviations occurring in the transmission of speech from one generation to another. But other factors contribute to the historical development of languages and determine the spread of a language family over the world's surface. Population movements naturally play a large part, and movements of peoples in prehistoric times carried the Indo-European languages from a relatively restricted area into most of Europe and into northern India, Persia, and Armenia. But language and race are by no means the same thing, and the spread of the Indo-European languages resulted, in the main, from the imposition of one of them on the earlier population of the territories occupied. In the historical period, within Indo-European, the same process can be seen at work in the western Roman Empire. Latin superseded the earlier, largely Celtic languages of the Iberian Peninsula and of Gaul (France) not through population replacement (the number of Roman soldiers and settlers in the empire was never large) but through the abandonment of these languages by the inhabitants over the generations as they found in Latin the language of commerce, civilization, law, literature, and, above all, social advancement and prestige.

Population movements

Conquest does not always lead to the supersession of a language. Greek survived centuries of Turkish rule and indeed remained a focus of national feeling, as has happened elsewhere in history. Much depends on the various circumstances and on the mutual attitudes of those involved; what must be kept quite clear is the difference between movements of peoples and the spread of languages. When linguistically homogeneous people enter and occupy a virtually empty area, as with most of Australia, the two movements coincide.

Languages do not just spread and compete with each other for territorial use. They are in constant contact, and every language bears evidence of this throughout its history. Modern Greek is full of words of Turkish origin, despite efforts made at various times since independence to purify the language by official action. The Norman Conquest and a period in which French was the language of the ruling class in England effected great changes on English and contributed a very substantial number of French words to English vocabulary; hence the quantity of near synonymous pairs available today: "begin, commence"; "end, finish"; "kingly, royal"; "fight, combat"; and so on.

These historical processes take place without any direct volition on the part of speakers as regards the language itself. Latin was learned as part of personal advancement, not for its own sake. Loans were incorporated almost without their being noticed, along with the concomitant cultural changes and innovations. Deliberate action directly related to a language does occur. The creation of pidgins involves some degree of linguistic consciousness on the part of their first users. More deliberate, however, have been various attempts at preserving the purity of a language, at least for some uses, or at arresting the processes of change. The care bestowed on the preservation of the Sanskrit used in religious ritual in ancient India and recent attempts to free Modern Greek from much of its Turkish vocabulary have already been noticed. For a period, under Nazi rule, efforts were made to replace some foreign words in the German language by words of native origin, and there have been movements to replace later accretions in English by words derived from Old English forms. In the long run, such attempts never succeed in preventing or reversing change; at best they preserve collaterally supposedly purer forms and styles for certain purposes and in certain contexts.

Attempts to prevent or reverse linguistic change

With the picture painted above of the tendency for languages to fragment first into dialects and then into separate languages, it might be thought that dialects are relatively late in appearance in the history of a language family. This impression is reinforced by the fact that most nonstandard dialects are unrepresented as such in writing, and so comparatively little is known about dialectal differences within most languages as one goes back in time. In this respect the very detailed knowledge of the Ancient Greek dialect situation is quite untypical.

In fact, dialect divisions must have been a feature of linguistic communities as early as there is any knowledge of them. Dialect splitting is fostered by isolation and loss of contact between groups within a speech community, and the sparse populations of earlier days, often nomadic and spread over large areas relative to their numbers, will have encouraged this process. It is simply the case that all but literate dialects have been lost in the past, and an artificial homogeneity is attributed to most ancient languages and to the so-called reconstructed parent languages of families.

Present-day conditions tend toward the amalgamation of dialects and the disappearance of those spoken by relatively few people. Urbanization, mass travel, universal education, broadcasting, ease of communication, and social mobility all foster rather large regional and social dialects, with special occupational types of language within them, in place of the small, strictly localized dialects of earlier times. This is one reason for the urgency with which dialect studies are being pursued in many Western industrialized countries, such as England and parts of the United States. If work is not done soon, many dialects may perish unrecorded (see also DIALECTS).

For the same reasons, dialect divisions that earlier would have widened into distinct languages are now unlikely to do so. One may compare the emergence of the separate Romance languages from once unitary Latin with the splitting of South American Spanish and Portuguese into different dialects of these two languages. These dialectal divisions are not now expected to widen beyond the range of intercomprehensibility. These same conditions, together with the spread of literacy, are leading to the extinction of languages spoken by relatively small communities. Such is the fate of most of the North American Indian languages, and Irish, Welsh, and Scots Gaelic may ultimately survive only as learned second languages, preserved as cultural focuses for their communities. But in situations like this, both past and present, the intervening period of extensive bilingualism and the concomitant use of two languages has its effect on the changes taking place in the dominant language, which is influenced by the phonetic and grammatical composition of the speakers' former language.

Extinction of languages

Language typology. Language families, as conceived in the historical study of languages, should not be confused with the quite separate classifications of languages by reference to their sharing certain predominant features of grammatical structure. These classifications of languages give rise to what are known as typological classes.

In fulfilling the requirements of open-ended creativity that are imposed on language by human beings, grammatical structure has things in common in all known languages, particularly at the deeper levels of grammar. All known languages have words or wordlike elements combined in accordance with certain rules into sentences; all known languages in some way distinguish nounlike and verblike sentence components; and all known languages have the means of embedding or subordinating one sentence within another as an included clause (*e.g.*, "the sun set" and "we returned home": "When the sun set we returned home"; "Joan was playing tennis" and "Joan twisted her ankle": "Joan, who was playing tennis, twisted her ankle," or "while she was playing tennis, Joan twisted her ankle").

Descriptive analyses of all the languages of the world have not been prepared, and, of course, there is information about only a minute number of those that are no longer spoken—namely, those few that were written. However, there is enough known to make the assertion of such universal features as have been given with fair confidence. These are often referred to as language universals; their nature and extent is the subject of discussion and research.

Broad types of language structure

Within these very general guidelines, however, languages exhibit various types of structure. This can most readily be seen by comparing the relations between the forms of words and their syntactic functions in different languages. Such a comparison is the basis of three broad types of language that have been distinguished since the beginning of the 19th century. They are, in fact, more like characteristics than they are types, in that most languages contain traces of all three of them, but in different proportions.

Classical Chinese made little or no use of word-form variation, such as is found, for example, in Latin, for grammatical purposes. Sentence structure was expressed by word order, word grouping, and the use of specific grammatical words, or particles. Such languages have been called isolating or analytic. Modern Chinese languages are much less analytic; Vietnamese is probably the most fully representative of this type. Some languages string together, or agglutinate, successive bits, each with a specific grammatical function, into the body of single words. Turkish is a typical agglutinative language: compare Turkish *evleri*, "houses" (accusative case), in which *ev* is the root meaning "house," *-ler* marks plurality, and *-i* is the sign for accusative, with Latin *domūs*, in which *-ūs* combines the representation of accusative and plural without the possibility of assigning either category separately to one part of the word ending. Latin is in this respect an inflectional, or fusional, language. In a more extreme example, Latin *i*, "go!" cumulatively represents in one fused form the verb meaning "go," active voice, imperative mood, second person, and singular number, each a grammatically distinct category.

English, like many other languages, is representative of all three types. In its use of word order alone to distinguish grammatical differences ("the dog chased the cat"; "the cat chased the dog") it resembles Classical Chinese rather than Latin. In a word form such as "manliness," in which each bit can be assigned a grammatical function ("man" the basic noun, *-li-* the adjective formative, and *-ness* the abstract noun formative), it makes use of agglutination, whereas plurals such as "men" and "geese" and past tenses such as "came" and "ran" fuse distinct grammatical categories into a word form in which only arbitrarily can one allot some sound segments, or letters, to one and some to the other.

Assigning languages to different types in this way involves a delicate procedure of balancing one part of the grammar against another and deciding which type of structure predominates and how well the other types are represented. Languages predominantly of each of the types are found in communities at all levels of civilization and in all types of culture.

Change in grammatical structures

In the course of transmission, grammatical structures change, just as do pronunciation and meanings, and in time the cumulative effect may be the transference of a language from one overall type to another, although it remains descended from the earlier language and therefore is just as much part of the same historical family. Latin is very different typologically from French in its grammatical structure, but French is nevertheless the form that Latin took in France in the course of time. In the grammatical relevance of word order, the absence of case inflections in nouns, and the use of verbal auxiliaries instead of single word tense forms, French is more like English, a distant cousin within the Indo-European family, than it is like Latin, its immediate progenitor (compare French *j'ai donné*, English "I have given," Latin *dedi)*. The two sorts of language classification, historical and typological, serve different purposes and are differently based. Language families group languages together on the basis of descent; *i.e.*, unbroken transmission from an earlier common parent language. The evidence is mainly systematic correspondences among the shapes of words of similar meanings (*e.g.*, Greek *patēr*, Latin *pater*, French *père*, German *Vater*, English "father"). Languages are put into typological classes, with the reservation already mentioned, on the basis of certain overall similarities of structure iriespective of historical relations. Although these two classifications may coincide with some languages, as is the case to a great extent in the Bantu family, they do so only contingently; being based on different data and oriented differently, they do not logically or necessarily imply each other.

In a way these two systems of classification involve the two most important aspects in which languages must be seen to be properly understood: as products of a continuous historical process and also as self-sufficient systems of communication in any one period. Both as a component of cultural history and as a central part of culture itself, language is able to reveal, more than any other human activity and achievement, what is involved in mankind's distinctive humanity.

BIBLIOGRAPHY. A bibliography for LANGUAGE is likely to overlap at least partially the bibliography for LINGUISTICS, as well as those given in the related specialist articles. This bibliography draws attention to some books that may be useful without prior specialist knowledge, and that develop in further detail the major topics introduced in the article.

EDWARD SAPIR, ***Language*** (1921, reissued 1970), one of the most appealing books on language ever written; LEONARD BLOOMFIELD, ***Language*** (1933, reprinted 1963), longer and more technical than Sapir's book, but a classic work, though Bloomfield's treatment of semantics is somewhat dated; JEAN F. WALLWORK, ***Language and Linguistics*** (1969), a short introduction to the study of language, and ***Language and People*** (1978), a brief discussion of the social dimensions of language; HERBERT H. and EVE V. CLARK, ***Psychology and Language*** (1977), an introduction to theories of language acquisition; DWIGHT L. BOLINGER, ***Aspects of Language,*** 3rd ed. (1981), a very useful wide-ranging survey of contemporary approaches to the subject; DAVID ABERCROMBIE, ***Elements of General Phonetics*** (1967), and PETER N. LADEFOGED, ***A Course in Phonetics*** (1975), introductions to the linguistic study of human speech; JOHN LYONS, ***Language and Linguistics: An Introduction*** (1981), introduces the reader to some of the most recent significant theoretical developments in linguistics; N. MINNIS (ed.), ***Linguistics at Large*** (1971), a collection of papers on different aspects of language; ROBERT H. ROBINS, ***A Short History of Linguistics,*** 2nd ed. (1979), a brief historical account of the study of language from antiquity to the present. DELL HYMES (ed.), ***Language in Culture and Society: A Reader in Linguistics and Anthropology*** (1964), contains a number of useful articles on the relations between language and life in society.

(Ro.H.R.)

Languages of the World

Languages may be classified either genetically or typologically. A genetic classification assumes that certain languages are related in that they have evolved from a common ancestral language. This form of classification employs ancient records (such as those for Latin) as well as hypothetical reconstructions of the earlier forms of languages, called protolanguages. Because information on the genetic affiliations of languages is sufficiently extensive, world surveys of languages are necessarily oriented in that way—sometimes exclusively so and sometimes in conjunction with typological classifications. Typological classification is based on similarities in language structure. There is not enough known concerning individual frames of reference in language typology to permit a worldwide typological classification.

Typological classifications

Before the conclusive demonstration that unwritten languages could be classified genetically, they were often relegated to a typological classification, which at one time was denigrated by scholars. Since 1917, however, the prestige of some kinds of typology has risen—in particular, that of grammatical typology. The best known typological frame of reference represents the grammar of a language, either as a whole or as a subsystem. Once a genetic classification has been established, typological classification may be superimposed on it in order to show change of language type, as from a predominantly inflectional language (*e.g.*, Proto-Germanic) to a predominantly isolating language (modern English), or to show features that are shared by languages in different but neighbouring branches in the same family (*e.g.*, Celtic and Germanic in Indo-European). The ultimate grammatical typology is that which treats subsystems that are, in one sense or another, universal to all human languages.

Lexical typologies, based on similarities in vocabulary structure, have been used in cognitive anthropology and psycholinguistics (*e.g.*, perception of colours and use of colour terms). The sociolinguistic frame of reference in typology provides classifications for varieties of language in terms of their functions and their ways of identifying social groups and cultural spaces; in addition, it brings order and integration to problems concerning national standards that are faced by new nations that have many nonstandard and nonwritten languages as well as languages that make use of writing.

Families, isolates, and phyla

A few points of terminology should be explained before surveying and classifying the world's languages. Language family is the label often used for a conservative genetic classification, one that can be attested only when an abundance of cognates (related words) is available. Phylum is the label for a liberal genetic classification that is attested with fewer cognates; it encompasses language families. Although a given phylum will have greater extension than any of the families included in it, only fragments of phonology will be reconstructable in the protolanguage. In actual linguistic usage, however, the term family is often employed to refer to a phylum; *e.g.*, the Hamito-Semitic family, the Sino-Tibetan family.

The label language isolate is used for a language that is the only representative of a language family, as Basque or the extinct Sumerian language; the presumptive but unknown sister languages of isolates are dead and unrecorded. A language isolate may be classified, along with normal language families, under the rubric of an extensive phylum (*e.g.*, Korean is sometimes classified as a member of the Ural-Altaic phylum) or left wholly unclassified (*e.g.*, Ainu in Japan). The label pidgin-creole is used for a language that has had so much vocabulary change that cognates for reconstructing the protolanguage from which it descended cannot be found. A pidgin is a contact language used for communication between groups having different native languages. When a pidgin becomes the native language of a community it is customarily called a creole.

This survey treats geographical regions of unequal size: huge and sprawling areas for the peripheral regions of Africa, Oceania, and the Americas but relatively compact areas for the focal regions within the Euroasiatic world. Nine regions—six in Eurasia, in all of which writing and standard languages are widespread—constitute a convenient basis for comparison and contrast. The regions of the Euroasiatic world are: Europe, South Asia, North Asia, Southwest Asia, East Asia, and Southeast Asia.

LANGUAGES OF EUROPE

The great majority of the languages spoken in Europe are of Indo-European and Uralic (especially Finno-Ugric) affiliation. In terms of numbers of speakers, the people in Europe who speak the languages of these families ale now fewer than those in non-European countries who also speak such languages. If a language is to be localized primarily in the region in which most people speak it, than Europe is no longer the chief locale of Spanish, for example, but rather Latin America.

An unusually small degree of genetic diversity is found among European languages: there are fewer language families in Europe than in any other continental-sized region of the world. In addition, literary traditions that have resulted in the preservation of earlier forms of present-day languages are found to a high degree among these languages. Every European language with a writing tradition has developed at least one standard that is recognized nationally, and the national standard often coexists with recognized regional standards.

A few European languages are used internationally, as lingua francas, but there is a low degree of pidgin-creole usage in Europe today.

Typological classifications have been superimposed on genetic classificationr of European languages in particular. For example, the italic branch of Indo-European languages may be grouped with the Greek, Celtic, and Germanic branches on the basis of certain structural features, as can Armenian with Greek and Indo-Iranian, and so forth.

Indo-European languages. The languages of seven of the nine extant branches of the Indo-European language family are spoken in Europe. Variability in determining the number of particular languages reflects variation in the criteria used (*e.g.*, mutual intelligibility between neighbouring dialects, known common or separate history versus sociocultural factors such as separate literary traditions or status as national languages of politically independent units), as well as the time period for which the criteria are applied. Thus, for example, it is possible to say on linguistic grounds that there are nine extant languages in the Romance subgroup of the Italic branch: Portuguese, Spanish, Catalan, French, Romansh, Ladin, Friulian, Italian, and Romanian. In applying the criterion of separate literary tradition, the list would be expanded by the addition of Provençal and Sardinian. To apply the criterion of status as a national language would reduce the list because Ladin, Friulian, Provençal, and Sardinian are not national languages; but the picture is complicated by the fact that Sardinia was once politically independent, and Andorra, in which Catalan is spoken, has not always been independent.

Variables in determining the number of separate languages

Similar difficulties in counting separate languages exist for all the branches in which several languages are spoken. In terms of areas of high mutual intelligibility (which do not entirely reflect historical development), there are only five modern Germanic languages: English, Frisian, Netherlandic-German (including Afrikaans and Yiddish), Insular Scandinavian, and Continental Scandinavian. If literary tradition and national criteria are considered, the number is increased by the division of Netherlandic-German into Standard High German, Low German, Dutch-Flemish, Afrikaans, Luxemburgian, and Yiddish; the division of Insular Scandinavian into Icelandic and Faeroese; and the division of Continental Scandinavian into Norwegian (New Norwegian, or Nynorsk, and Dano-Norwegian, or Bokmål), Danish, and Swedish.

For the Slavic languages there are 13 literary standards, but between the nuclei formed by these literary norms there are scarcely any linguistic boundaries because transitional dialects—both in terms of shared features and intelligibility—connect adjacent areas. In terms of more immediate intelligibility and to some extent in terms of shared features, the Slavic literary norms can be grouped into three zones: East Slavic (Russian, Belorussian, Ukrainian), West Slavic (Polish, Kashubian, Low Sorbian or Lower Lusatian, High Sorbian or Upper Lusatian, Czech, Slovak), and South Slavic (Slovene, Serbo-Croatian, Macedonian, Bulgarian).

Language boundaries are more clear-cut for the modern living languages in the remaining European branches of Indo-European: Celtic (the physical separation of the speakers of the languages contributes to the separate identification of Welsh, Breton, Irish Gaelic, and Scottish Gaelic), Baltic (the literary and political separation coincide with the separation of Lithuanian and Latvian), Greek (the separate historical development and lack of mutual intelligibility separate Modern Greek from Tsakonian), and Albanian (the political unity of Albania contributes to the single-language identification of its two divergent dialects).

Dialects of two languages in the Indo-Iranian branch of Indo-European are or were also spoken in Europe: the Jassic dialect of Ossetic, an Iranian language, formerly spoken in Hungary; and the European dialect of Romany, which spread all across Europe and into America. It may be, however, that only in Wales, Finland, and the Balkans does Romany still serve as the native language of some Gypsies.

Extinct languages

A number of earlier Indo-European languages that died out without descendants are known from written records and comments by contemporaries; these include several in the Italic branch (as Oscan, Umbrian, Faliscan, Venetic), at least two in the East Germanic group (Gothic and that spoken by the Burgundians, Vandals, and others), and three in the Celtic branch (Gaulish, Cornish, and the only recently extinct Manx). The classification as Indo-European or non-Indo-European for many other extinct languages remains uncertain because of the scarcity of data; *e.g.*, Pictish spoken in Great Britain. One form of Minoan, that represented by the Mycenaean Linear B syllabary, was shown to be an archaic Greek dialect when deciphered.

For more information on the Indo-European languages of Europe, see INDO-EUROPEAN LANGUAGES; CELTIC LANGUAGES; ITALIC LANGUAGES; ROMANCE LANGUAGES: BALTIC LANGUAGES; SLAVIC LANGUAGES; GERMANIC LANGUAGES; ENGLISH LANGUAGE; ALBANIAN LANGUAGE.

Finno-Ugric languages. In addition to the Indo-European languages, all but two of the languages of the Finno-Ugric branch of the Uralic family are also spoken in Europe. As in the case of Indo-European languages, variation exists in the enumeration of separate languages. For example, several varieties of Lapp are mutually unintelligible but are often classified as dialects of a single Lapp language. The various types have also been classified according to geographic areas or national boundaries (Norwegian, Swedish, Finnish, and Russian Lapp). In Baltic-Finnic, the Finno-Ugric subgroup most closely related to Lapp, the Finnish, Karelian, Veps, Ingrian, Estonian, Livonian, and Votic languages are often linked by transitional dialects between the central areas of a given pair. The other Finnic languages, Mari (Cheremis) and Mordvin, and both of the languages of the Permic subgroup (Udmurt, or Votyak, and Komi, or Zyryan) are spoken much further to the east, in the central area of eastern European Russia.

Geographic classifications

To the north of these, from the mouth of the Northern Dvina River eastward into North Asia there are speakers of Nenets (Yurak), a language belonging to the Samoyedic branch of Uralic. The remaining Finno-Ugric language of Europe, which belongs to the Ugric subgroup, is Hungarian. See also URALIC LANGUAGES.

Other languages. Maltese. Maltese, spoken on Malta, is an Arabic dialect, so long isolated from other dialects of Arabic and so heavily influenced by Italian that the resultant loss of mutual intelligibility with other Arabic speakers might justify classifying it as a separate Semitic language.

Basque. Basque, spoken in the Pyrenees in Spain and France, is the only other living language of western Europe that does not belong to the Indo-European family. Numerous inconclusive attempts have been made to link Basque genetically with other languages. See also BASQUE LANGUAGE.

Turkic languages. In addition to Turkish, spoken by a number of people in Bulgaria and elsewhere in the Balkans, several languages of the Turkic language group (considered as a branch of the Altaic language grouping) are spoken entirely in eastern Europe. Chuvash, the most divergent Turkic language, is found mainly in the Chuvash Autonomous Soviet Socialist Republic, Tatar in the Tatar A.S.S.R. and adjacent areas and in Romania and Bulgaria, Bashkir in the Bashkir A.S.S.R., Gagauz in the Ukrainian S.S.R. and Moldavian S.S.R. and in the Balkans, and Karaim in the southern Ukraine and Lithuania (spoken by fewer than 1,000 people). Most or all of the speakers of Crimean Turkish were removed to the Uzbek S.S.R. after World War II. See also ALTAIC LANGUAGES.

Extinct languages. The existence of a number of long-extinct non-Indo-European languages of Europe is known through the records of the Greeks and Romans and also through the preservation of varying amounts of written records of them. The most extensive records are those in the still undeciphered Etruscan, which is known to have been spoken in Italy from the 8th century BC to the 4th century AD (see also ETRUSCAN LANGUAGE). Several languages were spoken in the Iberian peninsula, of which Iberian (preserved in a few inscriptions and many coins) was spoken along the Ebro River and at one time as far east as the Rhône River. Also probably non-Indo-European was the language of the undeciphered Minoan Mycenaean Linear A inscriptions found on Crete (see also ANCIENT EPIGRAPHIC REMAINS).

LANGUAGES OF SOUTH ASIA

The genetic classification of the languages of India, Bangladesh, Pakistan, and the border states (*e.g.*, Nepal, Sikkim, and Bhutan) includes two subgroups—Indo-Aryan (Indic) and Iranian—of a single branch of Indo-European (called Indo-Iranian), some indigenous language families (as Dravidian), a few language isolates (as Burushaski), and some Sino-Tibetan languages.

Indo-Aryan (Indic) languages

Indo-Iranian languages. Except for Romany and the few Dardic languages spoken in Afghanistan, all of the languages of the Indo-Aryan (Indic) subgroup of the Indo-Iranian branch of Indo-European are spoken in South Asia. It is difficult to identify language boundaries in the Indo-Aryan group because, between any pair of literary standards, "transitional" dialects grade into one another, with no clear-cut language barriers. The problem is further complicated by the enormous dialect differentiation in most of the Indo-Aryan languages. In terms of lack of mutual intelligibility between literary standards, there are more than 20 Indo-Aryan languages. Although Sanskrit is a classical Indo-Aryan language, preserved in writing, it also enters so deeply into the vocabulary of present-day languages as to become, in some cases, the salient mark differentiating two dialects of one language. Thus, Hindi of India differs linguistically from Urdu of Pakistan chiefly in that the former may be heavily Sanskritized in vocabulary and the latter not.

Ethnolinguistic loyalties may also increase the number of languages distinguished; *e.g.*, the separate recognition of Bengali and Assamese. Most of the Indo-Aryan languages are spoken by many millions of speakers; *e.g.*, Bengali-Assamese, West Hindi, Bihari, East Hindi, Marathi, Lahnda, Maithili, Gujarati, Oriya, Sinhalese (in Sri Lanka [formerly Ceylon]), Sindhi, and Nepali. There are also large numbers of speakers of Indo-Aryan languages (especially West Hindi) in South Africa, the South Pacific, and the American Guianas.

The languages of the Dardic subgroup differ sufficiently from the other Indo-Aryan languages as to be sometimes classified as Iranian rather than Indo-Aryan or as a separate sub-branch coordinate with the Indo-Aryan and Iranian sub-branches. Kashmiri, spoken in Jammu and Kashmir, is the only Dardic language with a literary tradition. Shina is also spoken in Jammu and Kashmir; other Dardic languages, which are spoken mostly in Pakistan, have relatively few speakers.

Speakers of at least four Iranian languages are also found in South Asia, including Pashto speakers in Pakistan and Baluchi speakers in Pakistan and India (see also INDO-IRANIAN LANGUAGES).

Dravidian languages. Although the greatest concentrations of Dravidian speakers are in southern India, the more than 20 languages of this family are widespread in India, and one language, Brahui, is isolated in Pakistan, separated from its nearest sister language by 800 miles.

Four Dravidian languages have long literary traditions and are spoken by many millions: Telugu, Tamil, Malayalam, and Kannada. Tamil speakers are also found in Sri Lanka, Malaysia, Indonesia, Burma, Vietnam, South Africa, and in scattered island and coastal areas around the world. Among other, less widespread Dravidian tongues of India are Gondi, Tulu, Kurukh, and Kui. No convincing remote relationships between the Dravidian family and other families have been proposed (see also DRAVIDIAN LANGUAGES).

Munda languages. The 16 or so Munda languages are all spoken in India. Some scholars classify them as a separate language family; others point to similarities with certain languages of Southeast Asia and include them in an Austro-Asiatic grouping with Mon-Khmer, Vietnamese, and Nicobarese. Santali is the Munda language with the greatest number of speakers (a few million); Mundari, Ho, Sora, Kharia, and Korku have significantly fewer speakers (see also AUSTRO-ASIATIC LANGUAGES),

Nehali, Khasi, Burushaski

Other languages. Nahali, spoken by a few hundred people in the Nimār District of Madhya Pradesh, and Khasi, spoken by fewer than 200,000 people in the Khāsi and Jaintia Hills District of central Assam, may both be language isolates—*i.e.*, the sole known members of their families. Both seem to be remotely related to the Austro-Asiatic languages, however, and some scholars have tentatively classified Nahali as a Munda language. The other language isolate of South Asia, Burushaski, spoken by some 30,000 people in Pakistan, is without even remote known relatives.

Sino-Tibetan languages. Speakers of languages of most of the branches of the Sino-Tibetan language family are to be found in South Asia. All the languages of the Bodo (Bodo-Garo) branch of the Bodo-Naga-Kachin language group are spoken in Assam. Naga (Tangsa) languages are spoken in scattered locations from eastern Nepal into Burma. The Kachin languages are centred in north Burma, but some dialects are spoken in Assam, where there are also some speakers of Kukish (Kuki-Chin) languages and of Burrnish (Burmese-Lolo) languages. Dialects of the various divisions of the Tibetan language are distributed from Kashmir to Bhutan and southward into India (*e.g.*, Balti, Sharpa, Lhoke, Spiti). Speakers of close to 50 Gyarung-Mishmi (or Himalayan) languages are found from northeastern India to northern Assam, with their greatest concentration in Nepal. There is some scholary disagreement as to how the numerous Sino-Tibetan languages should be classified into branches and groups (see also SINO-TIBETAN LANGUAGES).

Tai languages. Some speakers of Khamti, a Tai language, live in Assam, where Ahom, another Tai language, is still used as a ceremonial language in religious rituals but is no longer spoken.

Use of English. In all parts of postcolonial South Asia, including Sri Lanka (formerly Ceylon), some people know English; these speakers, although relatively few in number, are the people most likely to travel to a state in which a South Asian language unknown to them is spoken. Hence, English is de facto the current interstate and international language of South Asia, although many Indians would prefer to adopt another language, such as Hindi or a Dravidian language, as the national language.

LANGUAGES OF NORTH ASIA

The languages of North Asia are those spoken 'from the Arctic Ocean to South Asia and China and from the Caspian Sea and Ural Mountains in the west to the Pacific Ocean in the east. In genetic classification, most languages of North Asia belong either to the Uralic family, to one of the three families of the Altaic language grouping (Turkic, Mongolian, and Manchu-Tungus) or to Indo-European. The genetic affiliations of the Paleosiberian languages, spoken exclusively in this region, are uncertain at present. Scholars have hypothesized that some of the languages may have once been American Indian languages whose prehistoric speakers backtracked from the New World into North Asia. That it is possible for pre-industrial man to go back and forth over Arctic waters is shown by the residence of those Eskimos who are now found on both the Russian and Alaskan shores of the Bering Strait. It has been claimed that all languages indigenous to North Asia, except the Paleosiberian ones and the recently intrusive Russian language, are genetically related in a Ural-Altaic phylum. This liberal classification is questioned by many scholars. Whether or not the three Altaic language groups are related to the Uralic languages, there is no doubt that all the so-called Ural-Altaic languages share many typological features, such as vowel harmony, agglutination (a type of word formation in which word elements are added together but still retain a separate, definite meaning), and a restriction against combining a plural noun with a quantifier (as though, in English, the plural noun "girls" had to appear as a singular noun in the phrase "five girl," rather than "five girls").

Ural-Altaic phylum

Uralic languages. Languages from two branches of the Uralic family are spoken in North Asia; their speakers are few in number. Two of the Ugric languages (the subgroup of the Finno-Ugric branch that includes Hungarian)—Mansi (Vogul) and Khanti (Ostyak)—are spoken on the Ob River and its southwestern tributaries. All of the languages of the Samoyedic branch are spoken in North Asia: Nenets (Yurak), speakers of which are scattered from the mouth of the Yenisey westward to the mouth of the Northern Dvina; Enets (Yenisey), also spoken around the Yenisey; Nganasan (Tavgi), spoken on the Taymyr Peninsula in Siberia; and Selkup (Ostyak), spoken south of Enets between the Taz and Tym rivers. Another southern Samoyedic language, Kamas (Sayan), had only one speaker in 1963 (see also URALIC LANGUAGES).

Altaic languages. Turkic languages. The Turkic languages are remarkable for their lack of diversity in spite of their wide occurrence in all the Euroasiatic regions except South Asia and Southeast Asia. Several of the numerous Turkic languages might be considered as a single language if it were not for the fact that mutual intelligibility between groups is impaired by differential borrowing from the various unrelated languages encountered in different regions. Thus, a Turkish speaker from Turkey might understand or largely understand an Uzbek speaker with more ease than the Uzbek speaker would understand Turkish, which has many loanwords from Persian and Arabic. Educated speakers of Turkic languages are able to read books in other Turkic languages after some adjustment to their varying spelling conventions and sound correspondences. Because such differences are identified with different Turkic ethnic groups, it is customary to identify the larger of these ethnic groups as speaking different Turkic languages, although some degree of intelligibility exists between them, as it does between Uzbek, Bashkir, and Tatar. In addition, some languages have dialects that are transitional between two recognized language groups; *e.g.*, some dialects of Kara-Kalpak are said to be transitional to Turkmen, and others are said to be transitional to Uzbek.

Distribution of Turkic in North Asia

In North Asia, Turkic languages are distributed from the southern extension of North Asia northeastward through central Siberia and include Turkmen in the Turkmen S.S.R., Iran, and Afghanistan; Uzbek in the USSR, mostly in the Uzbek S.S.R., and in Afghanistan; Kirgiz in the Kirgiz S.S.R. and in neighbouring areas from Afghanistan to China; Kara-Kalpak in the Kara-Kalpak A.S.S.R.; and Kazakh in the Kazakh S.S.R. Six or seven Turkic groups immediately north of western Mongolia are much smaller both in terms of numbers of speakers and the area over which they spread; in northern Siberia the Yakut extend from the Yakut A.S.S.R. to include the dialect spoken to the west by the Dolgan. Other Turkic languages are spoken in Southwest Asia, Europe, and East Asia.

Mongolian languages. The Mongolian, or Mongol, languages are dispersed throughout Central Asia from Afghanistan to Manchuria, occupying large parts of North Asia and East Asia. The problem of recognizing language boundaries (*i.e.*, of distinguishing separate languages) in the Mongolian family is complicated by the fact that differences between dialects are exaggerated in areas where Mongolian speakers have borrowed features of different unrelated languages but are minimized in areas where one dialect is spoken as a lingua franca over a wide area. Among the Mongolian languages are Mogul, spoken in Afghanistan, where it has been influenced by Iranian and Turkic languages; Mongour, spoken in Kansu Province of China and in Tibet, with noticeable effects of both Tibetan and Chinese in the language (Paoan, spoken in Kansu Province, is linguistically close to Mongour, but may be a different language); and Daghur, spoken mainly in Inner Mongolia and heavily influenced by Tungus languages. Additional languages include Ordos in Inner Mongolia, Kharchin in China, Oirat in the Sino-Russian border area from the Kirgiz S.S.R. to the Altai Mountains, and Buryat from the Buryat A.S.S.R. into Inner Mongolia. Some degree of mutual intelligibility exists between some of these, but this may in part be the result of the lingua franca use of Khalkha, the official language of the Mongolian People's Republic.

Manchu-Tungus languages. Speakers of the Manchu-Tungus languages are scattered from central interior Siberia to the shores of the seas of Japan and Okhotsk, including the Kamchatka Peninsula and Sakhalin Island. Those not near the coast live generally along the banks of the major rivers—the Yenisey, Tunguska, Khatanga, Lena, Amur, and Sungari. Detailed information on most of the Manchu-Tungus languages is scanty, and language names usually coincide with politico-cultural groups, rather than being based on a comparison of linguistic features or knowledge of mutual intelligibility. Borrowing that resulted from contact with speakers of Samoyedic (Uralic) languages to the west and northwest, Mongolian languages and Chinese to the south, and the various Paleosiberian languages to the north and east has further complicated the subclassification of the Manchu-Tungus languages by increasing the superficial differences among them. Most speakers of Manchu-Tungus languages are bilingual in the official language of their country, and many are replacing their native languages by Russian or Chinese. After the Manchu in China and Manchuria, the next most well known and numerous of the Manchu-Tungus peoples are the Evenki (whose name is sometimes applied as a generic term for all the Tungus tribes). Other groups include the Evens (or Lamuts), Nanay (or Gold), and other tribes with only 2,000 or fewer members. For more information on the Turkic, Mongolian, and Manchu-Tungus languages, see ALTAIC LANGUAGES.

Difficulties in Manchu-Tungus classification

Paleosiberian languages. Most of the so-called Paleosiberian people now live in northeasternmost Siberia in the area between the East Siberian Sea and the Sea of Okhotsk, including the Kamchatka Peninsula, and along the coast of the Sea of Okhotsk as far south as the Amur River, and on Sakhalin Island; peoples of another Paleosiberian group live far to the west along the middle and upper Yenisey River. The languages of the Paleosiberian people form four groups that are not only not related to each other but also have not been demonstrated to be related to any other genetic groups. The northernmost and most widespread of these linguistic groups and the only one that includes more than one living language is the Luorawetlan family, which consists of Chukchi, Kamchadal, and Koryak. Some scholars now classify Kerek and Aliutor as separate languages; these have otherwise been classified as dialects of Koryak.

The Yukaghir family includes one living language, Yukaghir (spoken by a few hundred people south of the Arctic Circle on tributaries of the Kolyma River and in the tundra between the Indigirka and Alazeya Rivers), and one language—Chuvantsy—that was spoken until the 20th century on the Anadyr River. Gilyak, spoken on Sakhalin Island and in the coastal and inland Amur River country of the mainland, has no known linguistic relatives. Ket (or Yenisey-Ostyak) is the only language of the Yeniseian or Yenisey-Ostyak family that is still spoken. The speakers of Ket live along the upper and middle Yenisey River, as did the speakers of its sister languages, Kott (Cottian-Manu), which became extinct in the 19th century, and Assan (Asan) and Arin, both of which became extinct in the 18th century (see also PALEOSIBERIAN LANGUAGES).

Indo-European languages in northern Asia, in addition to Russian, introduced only relatively recently, include the Iranian languages in the southwestern extension of North Asia (Tadzhik Persian in the Tadzhik S.S.R. and Baluchi in the Turkmen S.S.R.) and the long-extinct Tocharian, which penetrated into Central Asia as far as Chinese Turkistan (see also TOCHARIAN LANGUAGE).

Writing and literacy **in this** region. The earliest stimulus toward writing in North Asia was from China. The latest stimulus, from Soviet Russia, has brought literacy to those Altaic peoples whose languages were unwritten in tsarist times. In contrast to the written Altaic languages, the Paleosiberian languages in general remain preliterate. The Soviet educational policy is to encourage the use of native languages for education and for teaching preliterates to write. The standard form of writing Tadzhik, for example, is in the Cyrillic alphabet, and knowledge of this alphabet facilitates later learning of Russian, which is used for supranational communication in modern North Asia. Hence, Russian is the modern lingua franca for North Asia today. In the Mongol Empire of the 13th century, Turkic languages were used as languages of administration across North Asia from the Caspian Sea to Manchuria and, initially, in adjacent Euroasiatic regions conquered by Genghis Khan.

Role of Russian and Soviet policy

LANGUAGES OF SOUTHWEST ASIA

Languages spoken today in the area from Iran westward to the Mediterranean (in Iran, Iraq, Saudi Arabia, Jordan, Syria, Lebanon, and Israel) are Semitic, Indo-European, or Turkic. The languages in the two marginal subareas of Southwest Asia (in Afghanistan and in Turkey and the Caucasus between the Black and Caspian seas) far exceed the languages of Europe in genetic diversity.

At one time, Sumerian, now preserved in writing, was spoken as the first language of civilization in the ancient Near East; this language was neither Semitic nor Indo-European (see also SUMERIAN LANGUAGE). Early literary traditions and literacy for the elite began in this central area of Southwest Asia and extended from the Sumerian, Old Persian, and Akkadian literatures to Asia Minor (Hittite) in the north and to the Nile (Egyptian) in Africa. Akkadian and Persian seem to have been the first two languages put to wide international use.

Indo-Iranian languages. Almost all of the score of living languages of the Iranian subgroup of the Indo-Iranian branch of Indo-European are spoken in Southwest Asia and occasionally extend beyond into neighbouring regions. Persian has three separate literary standards that are not confined to the countries in which they centre (Iran, Afghanistan, and Tadzhik S.S.R.). More than half of the speakers of Pashto live in Afghanistan and the rest in South Asia. Kurdish is spoken in an area extending southward from southern Armenian S.S.R. into Turkey, Syria, Iran, and Iraq. Perhaps two-fifths of the speakers of Baluchi live in Iran and southern Afghanistan. Several other Iranian languages (or dialects) have many fewer speakers; these include Luri and Bakhtyārī, spoken only in Iran, and Munji and Shughni, spoken largely in Afghanistan, with only a few of their speakers in Pakistan or the Tadzhik S.S.R. One Iranian language, Yaghnobi, is spoken only in the Tadzhik S.S.R. Three Iranian languages are spoken almost entirely in the Caucasus: Tat, Talysh (with some speakers in Iran), and Ossetic.

The Dardic subgroups

The half a dozen Nuristani (Kafiri) languages spoken in Afghanistan, sometimes classified as members of the Dardic subgroup of Indo-Aryan, have more recently been classified by some scholars as constituting a separate branch of Indo-Iranian. In addition, some Lahnda (Indo-Aryan) speakers also live in Afghanistan. Two very divergent dialects of another Indo-Aryan language, Romany, are spoken in Southwest Asia—Armenian Romany and Asiatic Romany (the dialect of the Palestinian Gypsies). For more information on the Iranian and Indo-Aryan languages, see INDO-IRANIAN LANGUAGES.

Other languages. The sole language of another branch of Indo-European, Armenian, is spoken predominantly in the Armenian S.S.R., Georgian S.S.R., and Azerbaijan S.S.R., but also in Syria, Lebanon, Iran, Turkey, and in communities of Armenians in the United States, Egypt, and France (see also ARMENIAN LANGUAGE).

The long-extinct languages of the Anatolian branch of Indo-European, including Hittite, were once spoken in Southwest Asia (see also ANATOLIAN LANGUAGES).

Five Turkic languages are spoken primarily in Southwest Asia: Turkish, spoken in Turkey and surrounding countries largely to the north; Azerbaijani, spoken in the Azerbaijan S.S.R. and Iran; Kumyk, Karachay, and Nogay, spoken in the Caucasus. Three Turkic languages spoken predominantly in North Asia are also spoken in Southwest Asia: Uzbek, Turkmen, and Kirgiz.

One language of the Mongolian family is spoken in Southwest Asia—Mogul in Afghanistan; and Brahui, a Dravidian language, has a small fraction of its speakers in Afghanistan and Iran.

Caucasian languages. In addition to the Indo-European and Turkic languages spoken in the Caucasus, there are the over 30 languages belonging to three Caucasian language families. These may he remotely related to each other in a Caucasian phylum, in which the Northeast Caucasian family is more clearly related to the Northwest Caucasian family than the South Caucasian family is to either. Georgian, a South Caucasian language, is the most widely known Caucasian language, with speakers in the Georgian S.S.R., the Azerbaijan S.S.R., and adjacent parts of Turkey and Iran; it is the only Caucasian language with a long literary tradition. Other South Caucasian (Kartvelian) languages are Laz (Chan) and Svan. The Northwest Caucasian (Abkhazo-Adyghian) languages include Kabardian (Circassian), Abkhaz, Abaza, Adyghian, and Ubykh (almost extinct, now spoken by only a few people in Turkey). The approximately 25 languages of the Northeast Caucasian (Nakho-Dagestanian) family are spoken by people living mostly in the Dagestan A.S.S.R. These languages include Chechen, Ingush, Avar, Dargwa, Lakk, Lezgian, and Tabasaran, all of which (except Chechen) have fewer than 500,000 speakers. There is some scholarly disagreement concerning the classification of the Caucasian languages (see also CAUCASIAN LANGUAGES).

Semitic languages. Five Semitic languages are still spoken in Southwest Asia: Arabic; Hebrew, primarily in Israel; dialects of East Aramaic, still spoken in Israel, Syria, Iran, Iraq, and the Armenian S.S.R.; West Aramaic dialects, still spoken in Lebanon and Syria; and Modern South Arabic, spoken in southern Saudi Arabia and on nearby islands. Of the extinct Semitic languages, the best known are Phoenician, Akkadian (Babylonian and Assyrian), Moabite, and Ugaritic (see also HAMITO-SEMITIC LANGUAGES).

LANGUAGES OF EAST ASIA

Conservative versus liberal classification of the languages

Languages in East Asia are those traditionally spoken in China, Japan, and Korea; *i.e.*, those that occupy the region between North Asia and Southeast Asia. A conservative genetic classification reflects immense genetic diversity for East Asia by claiming that Ainu, Japanese, and Korean are neither related to each other nor to any other language in East Asia and that the Chinese languages (or dialects) belong in one family, Miao-Yao languages in another, and Tai languages in still another. A liberal genetic classification leaves Ainu isolated, includes Korean and Japanese in the Altaic family, and classifies some or all of the other groups as Sino-Tibetan.

Altaic languages. Languages from three of the major families of North Asia are spoken in China. Uighur, a Turkic language, is spoken in Sinkiang and Kansu Provinces of China as well as in the U.S.S.R. and southwestern Mongolia. Another Turkic language, Kirgiz, has some speakers in China. Manchu is the best known of the Manchu-Tungus languages and that with the longest literary tradition (dating from as early as 1599). After the Manchus established the last Chinese dynasty in 1643, their language was gradually replaced in most parts of China by Mandarin--except for formal and ceremonial occasions—but it is still spoken in scattered localities in Manchuria and in Chinese Turkistan.

Striking similarities in syntax have led some linguists to postulate a remote relationship between the Altaic languages and Korean and, less frequently, Japanese.

Korean, Japanese, and Ainu. Korean is spoken in Korea as well as by sizable populations in China and Japan (see also KOREAN LANGUAGE).

The Japanese language family includes, besides Japanese, several mutually unintelligible dialects spoken on the Ryukyu Islands by people who are bilingual in Japanese. Japanese is spoken by 100,000,000 people in Japan and by small groups in Taiwan, Brazil, and the U.S., especially in Hawaii (see also JAPANESE LANGUAGE).

Ainu, the remaining language in insular East Asia for which not even a remote relationship with other languages seems likely, is spoken by approximately 16,000 people on Hokkaido Island of Japan, on Sakhalin Island, and on the Kuril Islands.

Chinese language (dialects). Most important in terms of numbers of speakers and their influence on the other languages in East and Southeast Asia are the Chinese languages (often called dialects). In terms of mutual intelligibility among adjacent dialects, there are several Chinese languages: Mandarin, Wu, Cantonese, Hsiang, Kan-Hakka, and Min (or North Min and South Min). Mandarin is the native language of over 70 percent of the Chinese and is spoken as a second language by many of the native speakers of the other languages, both Chinese and non-Chinese, in China. It has traditionally been the language of administration. Although speakers of two different Chinese languages may not be able to understand one another when they talk, communication between them is possible in writing; conversely, the same written message is read aloud differently by speakers of different Chinese languages. The functional advantages of Chinese writing explains its perseverance for four millennia, but these advantages are partly offset by the difficulties each generation must experience in learning the thousands of character signs that are needed for literacy. Traditionally most Chinese were supposed to be illiterate, but with simplified characters and romanization, the majority of the people in the People's Republic are now literate. The Chinese languages are notable for their enormous numbers of speakers, and Mandarin has the largest number of speakers of any of the world's languages (about 525,000,000 native speakers).

A remote relationship in one family (Sino-Tibetan) has been postulated for the Chinese languages and all the other non-Altaic families that have languages spoken in China. There is no doubt that all these languages bear many similarities to Chinese, but it is unknown to what extent such similarities might be the result of borrowing rather than common origin. A remote relationship in an Austro-Tai phylum has been proposed for two of these families (Tai and Miao-Yao) and Austronesian.

Remote relationships in East Asia

Tai and Miao-Yao languages. All of the languages of the Kam-Sui language group, which is related to the Tai family, are spoken in China (in Kweichow, Hunan, and Kwangsi Provinces), with some dialects extending into Southeast Asia. Speakers of Miao-Yao languages are scattered over south central China and extend into Vietnam, Laos, and Thailand. Dialects of the Miao language include Red Miao, White Miao, Green or Blue Miao, and the more divergent Black Miao. The Yao languages are Yao (also called Man or Mien), Laka, and Punu.

Tibeto-Burman languages. The Tibetan, or Tibetic, language group includes at least two Tibetan proper languages spoken in Tibet, Nepal, Sikkim, and India: Central Tibetan, including Lhasa, the standard dialect of Tibet, and Western Tibetan. In addition there are many other languages in Sikkim, Nepal, Assam, India, and Bangladesh that are closely related to Tibetan proper.

Languages that are more distantly related to Tibetan in a Tibeto-Burman branch of the Sino-Tibetan family are spoken in East Asia over the borders of Burma; these languages, often called Burmic, includes dialects of the Burmese-Lolo (Burmish) subgroup (including Burmese) and the Kachin subgroup. For more information on the Chinese, Tibetan, and Burmic languages, see SINO-TIBETAN LANGUAGES.

Three general types of syntax, which partly overlap the liberal genetic classification, can be distinguished among languages in East Asia. First, Ainu is isolated syntactically as well as genetically. The second type is shared by Korean and Japanese. All Chinese languages are strikingly alike in syntax, and this third type is approximated among some non-Chinese languages of the Sino-Tibetan family and among some languages of Southeast Asia whose genetic classification is tentatively indeterminate.

LANGUAGES OF SOUTHEAST ASIA (INCLUDING AUSTRONESIAN)

Genetic diversity in mainland Southeast Asia

Southeast Asia is generally taken to be a region that includes both a mainland subregion, south of China and east of India, and an insular subregion, which includes the insular half of Malaysia, all of Indonesia, and the islands of the Philippines. Virtually all the languages of insular Southeast Asia belong to a single language family—Austronesian (Malayo-Polynesian). Mainland Southeast Asia, on the other hand, has various representatives from the Austro-Asiatic, Tai, and Sino-Tibetan language groups. Hence, genetic diversity is greater in mainland than in insular Southeast Asia. Austronesian languages extend out of Southeast Asia to the most distant culture areas in Oceania (Polynesia and Micronesia), where they are the only languages known aboriginally. One modern Austronesian language (Malagasy) is even spoken on the African side of the Indian Ocean in Madagascar.

Curiously enough, it is in Melanesia, between the Bismarck Archipelago and New Hebrides, that the most diverse Austronesian languages are spoken today; this provides grounds for the conjecture that the Proto-Austronesian language was spoken there millennia ago and that the daughter languages diversified as their speakers migrated over half the world, with Malay and Cham backtracking eventually to the mainland of Southeast Asia, out of which the ancestors of Proto-Austronesian speakers must have come.

In general, the name of the country and the name of the national language are the same in both insular and mainland regions of Southeast Asia. Thus, Pilipino (based on Tagalog) is the name of the national language of the Republic of the Philippines, even though Pilipino is learned as a second language by most Filipinos. The fear in all of Southeast Asia of indirect neocolonial domination motivates continued distrust of the old languages of colonialism—English, French, Dutch, Spanish—and now also of Japanese and Russian. A pidgin-creole—Neo-Melanesian, or Melanesian Pidgin English—is used as a lingua franca by speakers of Austronesian and other languages from southern Papua through Melanesia into Micronesia. This use is now being encouraged in the Australian territories as a bridge to learning English.

Though the languages in the mainland subregion of Southeast Asia are genetically diverse, they show widespread ranges of the same typological features, such as the use of distinctive tones and classifiers, among unrelated or remotely related languages.

Mon-Khmer languages

Austro-Asiatic languages. The Mon-Khmer group includes more than 50 languages—more than any other family that is centred primarily in or entirely in Southeast Asia. Mon-Khmer languages are spoken from Burma to Vietnam. In Cambodia, Khmer (Cambodian) is the official language; its speakers are also found in Thailand. Mon is also spoken in Thailand as well as in Burma.

The language of mainland Southeast Asia with the greatest number of speakers is Vietnamese, spoken in Vietnam and by smaller numbers of speakers in Cambodia, Thailand, and Laos. Muong, spoken in the central highlands of North Vietnam, is recognized as a separate, but related, language.

Classified by some scholars as a northern group of Mon-Khmer languages are several languages spoken in Burma (east of Mandalay), northwest Thailand, northern Laos, and to a lesser extent in northern Vietnam and in China. These are sometimes classified by others as a separate Palaung-Wa, or Salween, family, including Khmu, spoken in Laos and extending into Thailand, and Palaung, spoken in Burma.

Three small groups of related languages in Malaya (sometimes called Malaccan) are considered to be related to the Austro-Asiatic languages. They are the Jahaic, or Semang, languages, spoken in the inland area of northern Malaya and across the border in Thailand; the Senoic, or Sakai, languages, with speakers south of Kuala Lumpur on the coast and inland further south; and the Semelaic, or Jakun, languages, spoken south of the Senoic languages (see also AUSTRO-ASIATIC LANGUAGES).

Tai and Sino-Tibetan languages. At least a dozen languages of the Tai language family are spoken in Southeast Asia: Thai, or Siamese, in Thailand; Lao, in Thailand, Laos, and Cambodia; Yuan, in Thailand; Shan, in Burma; Black Tai (Tai Noir), in Laos and Vietnam; Khün and Khamti, in Burma; and White Tai (Tai Blanc), Tay, Nung, Tho, and Kelao (Ch'i-lao), all in Vietnam.

Over 8,000,000 Chinese are distributed throughout Southeast Asia; of these, almost 4,000,000 are in Thailand, 3,000,000 in Vietnam, and smaller numbers in Burma, Cambodia, and Laos.

Of the other language groups in the Sino-Tibetan family in Southeast Asia, the Burmese-Lolo (Burmish) group has the widest distribution and the greatest number of speakers. Burmese is spoken as a second language by perhaps 90 percent of those in Burma who have another first or native language. The Lolo languages are spoken in Burma, Thailand, Laos, and Vietnam; they include Lisu, Lahu, Akha, Mung, Punoi, Pyen, and others, a few of which extend into Assam. Karen languages are spoken in Burma and Thailand and include Sgaw, Pho, Pa-o (or Taungthu), and Palaychi. Most of the languages of the Kuki-Chin (Kukish) group are spoken in Burma. Kachin languages are also spoken in Burma (see also SINO-TIBETAN LANGUAGES).

Nicobarese and Andamanese

Insular language groups. In insular Southeast Asia one small language group, Nicobarese, consisting of the languages spoken on the Nicobar Islands, is, in a liberal classification, classified as Austro-Asiatic. The other insular family, Andamanese, consisting of the languages spoken on the Andaman Islands by perhaps fewer than 200 people, has only recently been supposed to be remotely related to the Papuan languages of Melanesia.

Austronesian languages. There are perhaps 500 languages in the Austronesian (Malayo-Polynesian) family, spoken in Malaysia and the Indonesian archipelago; the Philippines; parts of Vietnam, Cambodia, and Taiwan; on the main island groups of the South and Central Pacific; on New Guinea; and on Madagascar. According to one classification, these languages include, in addition to small subgroups, at least two large subgroups: Western Austronesian (or Indonesian) and Eastern Austronesian (often called Oceanic), which includes the Polynesian languages and some of the Melanesian and Micronesian languages. Those Austronesian languages spoken on the Southeast Asian mainland (Malay in Malaysia, Cham and eight other languages mostly in Vietnam, with speakers of some of them also in Cambodia) belong to a Western Indonesian subgroup, which includes Javanese, Sundanese, and Malay, including Bahasa Indonesia, the national language of Indonesia. Closely related to the Western Indonesian subgroup is the subgroup comprising around 100 languages of the Philippines and a few languages of northern Borneo and northern Celebes (including Tagalog, which includes Pilipino, the national language of the Philippines, and Cebuano, Ilongo, and Ilocano). Classed with the West Indonesian and Philippine languages are a small group of languages of Celebes (*e.g.*, Buginese and Makasarese), a few languages of Borneo, and Malagasy (used on Madagascar).

Languages of Polynesia

The languages of Polynesia, including Maori in New Zealand, Tongan, Tahitian, and Hawaiian, form a subgroup that is part of a larger Eastern Oceanic subgroup of over 100 languages, which includes besides the Polynesian languages such languages as Fijian and a number of languages of the Solomon Islands. At least seven of the languages of Micronesia (including Gilbertese, Trukese, and Ponapean) form another subgroup. There are more than 100 Austronesian languages in New Guinea and more than 100 Austronesian languages, not counted as Eastern Oceanic, that are spoken on smaller islands of Melanesia. Those few with as many as 10,000 speakers are all used as lingua francas in wider areas than those of their native speakers (Dobu in the D'Entrecasteaux Islands, Banoni in southwestern Bougainville, Panayati in the Louisiade Archipelago), Among the Austronesian languages still spoken on Formosa are Ami, Atayalic, Paiwan, and Bunan. There is some scholarly disagreement concerning the classification of the Austronesian languages (see also AUSTRONESIAN LANGUAGES).

NON-AUSTRONESIAN LANGUAGES OF OCEANIA

In effect, the non-Austronesian language areas of New Guinea and Australia together constitute a wedge in the midst of three Austronesian areas: Polynesia to the east, Micronesia to the north, and Indonesia to the west. A few non-Austronesian languages are found on the Indonesian islands nearest to New Guinea (on Halmahera as well as on Timor and Alor).

An exceptionally liberal genetic classification claims that the many non-Austronesian languages in Melanesia and the few in Indonesia all belong to one phylum. Conservative classifications recognize several or even many different language families and avoid the older name for them (Papuan), because it might suggest either that the unrelated families of non-Austronesian languages are branches of one Papuan family or else that non-Austronesian languages are found only in New Guinea (where Papua is the name of a political territory). On the other hand, no classification is challenged when it is said that all Australian languages are ultimately related and that they are related neither to Austronesian nor to non-Austronesian languages outside of Australia.

Multilingual societies of Melanesia

In Melanesia, in essence the non-Austronesian world beyond Indonesia, there is much contact between Austronesian and non-Austronesian languages. Many of the Melanesian societies are multilingual, especially those in New Guinea; in addition to their native language, speakers often learn a few secondary languages—those of their immediate neighbours or, most frequently, Neo-Melanesian (a pidgin-creole with an English-based lexicon), or both. In part of Papua, Police Motu, a pidgin based on an Austronesian language, is used as a lingua franca far beyond the territory of the few thousand native speakers of Motu. In Australia the same interest in mastering a multiplicity of languages is widespread, and Aborigines have developed another English-based pidgin-creole, quite different from Neo-Melanesian. Another parallel between Australian languages and the non-Austronesian languages north of Torres Strait is the disinclination of both to recognize or develop any one dialect of a language as a standard.

Papuan languages. Several hundred Papuan or non-Austronesian languages extend from the Santa Cruz Islands north and west into the Solomons and the Bismarck Archipelago, across New Guinea to Halmahera, Timor, and Alor. Until the late 1950s all discussions of the languages of New Guinea that treated more than small, closely related groups of languages stressed the fact that the hundreds of languages spoken in a comparatively small area seemed to be completely unrelated to each other except for a few groups of immediate neighbours. Until then, little was actually known about more than a few of the languages of New Guinea. This situation was changed in the 1960s, with the publication of further survey work in the Highlands Districts, which stated explicit relationships among a large group of languages.

Since the initial recognition of this fairly large group of related languages in the Highlands, called the East New Guinea Highlands phylum, more and more languages of New Guinea have been found to be at least remotely related to it (in a Central New Guinea phylum). It must be kept in mind, however, that it is difficult to enumerate languages and families among the Papuan languages because too little information has been obtained for most of the languages to identify language boundaries or make possible detailed comparisons. There remain a number of families and isolated languages that seem not to be related to other Papuan languages. A new liberal classification presented by the U.S. linguist Joseph Greenberg in 1971, however, treats all the Papuan languages as genetically related in an Indo-Pacific phylum, which also includes Andamanese. Most Papuan languages are spoken by only a few hundred to a few thousand speakers (see also PAPUAN LANGUAGES).

Indo-Pacific phylum

Australian aboriginal languages. All of the aboriginal languages of Australia are remotely related to each other. A few dozen of the 200 or so Australian languages still spoken account for 90 percent of the total number of speakers; these include the Aranda languages, Tiwi, Walbiri, and Western Desert, two languages spoken by over a score of separately named small groups scattered over a territory 900 miles long. Scores of languages are now spoken by fewer than six people each. The greatest diversity among the languages is found in extreme northern and northwestern Australia (Arnhemland and the Kimberley District); a single remaining family (Pama-Nyungan). with 177 languages, is distributed over the rest of Australia (see also AUSTRALIAN ABORIGINAL LANGUAGES).

In grammatical typology the non-Austronesian languages north of Torres Strait are heterogeneous, while the Australian languages are syntactically homogeneous and almost identical in patterns of sound combinations. Both Australian languages and non-Austronesian languages have dialects that are linked in a chain such that speakers at either end do not understand the vocabulary of speakers at the other end, although speakers of adjacent dialects can understand each other.

The two or more languages that were spoken on Tasmania until the later part of the 19th century are not related to Australian languages, but may belong to the Indo-Pacific phylum.

LANGUAGES OF AFRICA

European languages in Africa

Languages that came into Africa from another homeland include, among others, all the European languages associated with 19th-century colonialism. Although the majority of countries in Africa regained their freedom in the 1960s, they continue to use the European languages of the colonial period along with the numerous languages indigenous to Africa. Languages from Southwest Asia preceded the languages of European colonization: migrations of peoples to North Africa brought the Ethiopians almost three millennia ago and the Arabic speakers many centuries ago. The Phoenician circumnavigation of Africa in ancient times left traces—Phoenician coins—on the coasts but none in the interior. And long ago migrants from Indonesia reached Madagascar, 250 miles off the African coast. Before and during the colonial period, Arab and Indian traders reached East Africa, where today a few Indo-Aryan languages are spoken among Asian businessmen. The interior of Africa was not known to any non-Africans before the colonial period, but its prehistory can now be partially reconstructed. For example, there is evidence that the homeland of the protolanguage of the numerous Bantu languages was in Cameroon or an adjacent area in West Africa (or in both areas); that a prehistoric migration brought the Bantu speakers to Central and East Africa; and that these Bantu forced the speakers of Bushman and Hottentot languages to leave their homeland around Lake Victoria and move south to the Kalahari.

In all the postcolonial nations today, either English or Arabic or French serves both as an international language and as a functioning national language. The question still unresolved for many African nations concerns which of their indigenous languages to develop through writing and to standardize as the official language or languages of education and of the political state. The numerous pidgin-creoles, as Krio, are recent and colonial in inspiration; Sango in the Central African Republic is surely indigenous but not so surely a pidgin-creole. Most of the dozen or so languages used in trade, as Swahili in East Africa and Hausa in West Africa, tend to have great changes in vocabulary like pidgin-creoles, but they are not classified as pidgin-creoles; instead they are varieties of normal languages that function as lingua francas. Lingua francas of one sort or another are a prerequisite for the markets found throughout tribal and peasant Africa.

Despite the genetic diversity in South Africa and the even greater diversity in West Africa, a part of each of these subregions can be shown, on the basis of typology, to be a linguistic area. Thus, most linguists have found that most languages in West Africa distinguish vocabulary items and word elements by tone; in South Africa the clicks characteristic of Khoisan languages are also found among neighbouring Bantu languages like Xhosa and Zulu. The early use of typology to anticipate genetic classification, however, led to the claim that Africa was full of mixed languages—*e.g.*, Mbugu in Tanzania. But Mbugu, despite having borrowed Bantu prefixes and culture words from Bantu, can be shown to have a single line of origin—to have descended from a single protolanguage (Proto-Cushitic)—on the basis of its grammatical constituents (pronouns and verb forms) and basic vocabulary items that are cognate with other Cushitic languages.

Five branches of the Hamito-Semitic language family in Africa

Hamito-Semitic languages. The Hamito-Semitic (Afro-Asiatic) language family (considered a phylum by some) includes five branches spoken across North Africa from Mauritania to Somalia and beyond into Southwest Asia: Chad, Semitic, Cushitic, Berber, and the now extinct Egyptian-Coptic. The Chad branch consists of over 100 languages spoken in Nigeria, Niger, Cameroon, Ghana, Chad, and the Central African Republic. By far the most widespread is Hausa, estimated to be spoken by as many as 25,000,000 people, for at least one-third of whom it is a second language.

Five Semitic languages are spoken in Africa, if modern colloquial Arabic is counted as a single language throughout its range across North Africa and the Arabian Peninsula and if Gurage in Ethiopia is also counted as a single language. The Semitic languages in Ethiopia include Amharic, Tigrinya, Tigré, and Gurage (but the people grouped as Gurage may be speaking several separate languages).

Cushitic languages are spoken in Ethiopia, Somalia, The Sudan, Tanzania, and Kenya. The languages with the greatest number of speakers are Galla in Ethiopia, Somali in Somalia and Ethiopia, Sidamo, Hadya, and Afar-Saho. Some scholars consider a group of languages traditionally classified as Cushitic to be a separate branch of Hamito-Semitic, called Omotic. Spoken in Ethiopia, they include Welamo, with far more speakers than the other Omotic languages, Ari, Shako, Zaysse, and others with only a few thousand or a few hundred speakers.

The languages of the Berber branch are spoken from the western desert of Egypt west to the Atlantic and extend to Senegal on the coast and to northern Nigeria in the interior. One language that may have been Berber, Guanche, was formerly spoken on the Canary Islands. Berber languages include Shilha (Shluh), spoken in Morocco; Tuareg (Tamashek) in Algeria, Libya, Niger, and Mali; and Tamazight in Morocco and Algeria (see also HAMITO-SEMITIC LANGUAGES).

Nilo-Saharan languages. The Nilo-Saharan languages in central interior Africa include the Chari-Nile languages and others that are not closely related to each other or to the Chari-Nile group. (The validity of this grouping has been questioned.) The largest Chari-Nile division, Eastern Sudanic, includes over 60 languages, which are spoken from Chad to Kenya and Tanzania; it includes a group of languages often classified as a separate family or branch (Nilo-Hamitic), which appears in some classifications in the Hamito-Semitic family rather than the Nilo-Saharan one.

Among the major Eastern Sudanic languages are Teso in Uganda and Kenya, Dinka in The Sudan, Luo in Kenya and Tanzania, and Lango in Uganda. Only 4 of the 30 or so languages of the Central Sudanic subgroup of Chari-Nile are spoken by groups of over 100,000 people: Sara in Central African Republic and Chad, Lugbara and Lendu, in Uganda and Congo, and Mangbetu in Congo.

Among the Nilo-Saharan languages that are not classified as Chari-Nile is the Saharan group. Kanuri, its largest member, is spoken by 2,500,000 people in Nigeria, Niger, Cameroon, and Chad. In the Maba group, Masalit is spoken by 100,000 people in The Sudan. Songhai, often classified as a language isolate, is spoken by 400,000 people in Niger, Mali, Upper Volta, Nigeria, and Dahomey. Fur, also sometimes considered as an isolate, is spoken by over 170,000 people, mostly in The Sudan.

Bantu languages

Niger-Congo languages. Languages in the Niger-Congo (or Niger-Kordofanian) family are spoken all across Africa from Mauritania to Kenya and south into South Africa. There are almost 900 Niger-Congo languages, which have been classified into six genetic subgroups. The Bantu languages (of the Benue-Congo subgroup) far outnumber those of any other family in Africa, both in terms of number of languages and in terms of number of speakers per language (except that the total number of speakers of a single Semitic language, Arabic, exceeds the total of over 55,000,000 speakers of Bantu languages). Fifteen Bantu languages are spoken by more than 1,000,000 people; the following each have a few million speakers: Swahili, Rwanda, Rundi, South Sotho, Luba-Lulua, Xhosa, and Zulu. Other subgroups in the Niger-Congo family include only a few dozen languages, as those in the Mande subgroup in West Africa, which are spoken from Mauritania to Ghana (including Bambara, Mende, and Vai). The Gur (Voltaic) languages, spoken from Mali and the Ivory Coast to Nigeria, include Mossi, with 2,000,000 speakers, and numerous other languages with significantly fewer speakers. The West Atlantic languages, spoken from Senegal to Nigeria, include Fulani, spoken by 4,500,000–6,000,000 people (estimates vary), Wolof and Temne (each with over 500,000 speakers), and several other languages of less numerical import. Of the languages of the Adamawa-Eastern subgroup, spoken from The Sudan to Cameroon, only Sango, through its use as a lingua franca, may be known by as many as 1,000,000 people. The Kwa subgroup of Niger-Congo includes Twi (Akan), Yoruba (in Nigeria and extending into Dahomey), and Igbo (also known as Ibo; in Nigeria), each of which has over 1,000,000 speakers. Some scholars link the Kordofanian languages of Kurdufān Province in The Sudan with the Niger-Congo languages in a Niger-Kordofanian phylum.

Khoisan languages. The Khoisan family consists of about four dozen languages spoken in southern Africa and two click languages (Sandawe and Hadza) spoken in Tanzania that are not closely affiliated with any one group in the Khoisan family. Uncertainties in the number of languages and the number of language groups arise from the profusion of labels for various groups and the lack of detailed linguistic comparisons among large numbers of them. Most of the Khoisan languages have been considered to be on the verge of extinction, if not known to be already extinct, but recent estimates of the numbers of peoples grouped on the basis of their culture (Hottentot and Bushman) show many thousands of speakers. The Khoisan language estimated to have the most speakers is Nama, with about 40,000. For more information on the Nilo-Saharan (Chari-Nile), Niger-Congo, and Khoisan languages, see AFRICAN LANGUAGES.

LANGUAGES OF THE AMERICAS

Languages indigenous to the Americas were brought from Asia by the forebears of modern American Indians (including Eskimos), who left Asia after the dog was domesticated but before other animals were domesticated. Something is known about the culture of these Indians but nothing about their languages, which are known only after contact with European languages.

European languages in the Americas

Today there are six European languages in the Americas that serve as languages of both education and government administration. (Only one Indian language, however, functions in this dual role—Guarani of Paraguay.) These official languages and their number of primary political divisions are Spanish (18), Portuguese (1), Dutch (3)—all in Latin America and the Caribbean; English (2 in North America and 5 in the Caribbean);

French (1 in North America and 3 in the Caribbean); and Danish (1 in Greenland). Before the colonial period in Latin America and during the first century or two of that period, the following American Indian languages could also be classed as official or semi-official: Nahuatl (Nahua), the language of the Aztecs in Mexico and Central America; Chibcha-Muisca in Colombia; Quechua, the language of the Incas, in the Andean area; Tupí in Brazil; and Guarani in and around Paraguay. In addition to American Indian languages, two pidgin-creole languages are official in their own political divisions, Sranan (Taki-Taki) in Surinam and Papiamento in Curaçao. Other pidgin-creoles in the Caribbean, such as Haitian Creole, are being increasingly written.

Genetic diversity among languages of continental-sized areas can be expressed in terms of the number of minimum genetic classes taken as the usual basis for discussion by specialists of that area. Research may lead to a downward (or upward) revision, and a new number of minimum genetic classes is used as a basis for further discussion. For North America (north of Mexico) and for the 20th century, the basis for discussion has shifted three times so far: from about 50 families in the classification of the U.S. scholar J.W. Powell to six phyla in the classification of the U.S. anthropological linguist Edward Sapir, which was revised at the 1964 Conference on North American Indian Languages by splitting and reclassification (*e.g.*, of Sapir's Hokan-Siouan) and by merging (*e.g.*, the Muskogean family and a few isolates were added to Algonkian in the new Macro-Algonkian phylum). This third classification is summarized below. Proposals for a minimum number of genetic classes in South America range from more than 100 families to three phyla (in a recent liberal classification).

The Plains Indian sign language (hand talking) is still known, but Chinook Jargon and other pidgin-creoles in North America fell into disuse as soon as American Indians became bilingual in English, French, or Spanish.

Ten culture areas of North American Indians

North and Central American Indian languages. For North America north of Mexico, the summary of culture areas (before any American Indians were relocated by Europeans) by the U.S. anthropologist Harold E. Driver is a convenient basis on which to superimpose the various ways in which language classifications (genetic and typological) combine with cultures that are ecologically adapted to each of ten areas—the Arctic, Subarctic, Northwest Coast, Plateau, Plains, Prairies, East, California, Great Basin, and Southwest. The three variables (genetic, typological, and cultural) coincide approximately in the Arctic (the one language family, Eskimo-Aleut, does not include typologically diverse languages, but it does spread over a culture area that is not entirely homogeneous). In the Subarctic two language families are represented, Algonkian and Athabascan, which are distinct typologically as well as genetically. Northwest Coast and adjacent Plateau languages are genetically very diversified but surprisingly homogeneous in a diffusional kind of phonological typology. The languages in the treeless Plains and the midwestern Prairies are genetically and typologically diverse; all the language families represented, except Caddoan, are intrusive in the sense that their homelands lie outside the Plains and Prairie areas.

Language families in the East give an impression of a little typological similarity combined with considerable genetic diversity. On the opposite coast, California is surprisingly homogeneous in culture and in language typology but heterogeneous in genetic classification of languages. There are few languages and only two language families represented in the Great Basin, which is homogeneous in all respects. The adjacent Southwest is anomalous in all three variables considered here. Where it is culturally homogeneous, as between Pueblo societies, it is genetically and typologically diverse in language: four different language families are represented in Pueblo societies. Non-Pueblo societies of the Southwest are diverse culturally as well as linguistically.

Eskimo-Aleut. The three languages of the Eskimo-Aleut family are still spoken in their prediscovery areas from Greenland to Siberia and also on Komandor Island between the Aleutians and Kamchatka (see also ESKIMO-ALEUT LANGUAGES).

Athabascan. Over 20 languages of the Athabascan family are still spoken in four different culture areas: the Yukon and Mackenzie areas of the Subarctic (the centre of Athabascan diversity, with 17 living languages, including Chipewyan-Slave-Yellowknife and Carrier), the Northwest Coast (where only Hupa and Chasta Costa may still be spoken), the Southwest (where the Navajo dialect of what may be considered a single Southwestern Apachean language has more speakers than any other Indian language north of Mexico), and the Plains (where two Athabascan languages are more recently intrusive—Sarcee [Sarsi] from the Subarctic and Kiowa Apache from the Southwest). Three language isolates spoken in the Northwest Coast (Eyak, Tlingit, and Haida) are remotely related to Athabascan in the Na-Dené phylum, but Eyak is so much more closely related to the Athabascan family that it might be considered a divergent member of the family.

Macro-Algonkian phylum

Algonkian. The Algonkian family includes 13 languages still spoken, which belonged in the culture areas of the eastern Subarctic (*e.g.*, Cree, Ojibwa, Micmac, Malecite), the Prairies (*e.g.*, Fox, Potawatomi), the Plains (*e.g.*, Blackfoot, Cheyenne, Arapaho), and the East (where most Algonkian languages became or are now becoming extinct, with only the removed Shawnee and Delaware surviving in Oklahoma). Remotely related to the Algonkian languages in a Macro-Algonkian phylum are languages spoken further to the south in the East—the Muskogean family (including Choctaw-Chickasaw and Creek-Seminole) and several language isolates that are no longer spoken, as well as two almost extinct languages of the Northwest Coast that are more closely related to Algonkian (Wiyot and Yurok).

Macro-Siouan. The Macro-Siouan phylum is named for its most extensive component, the Siouan family, the extant languages of which belong in the Plains and Prairies, including Dakota, Crow, Winnebago, and Omaha-Osage. (The Siouan languages of the East, such as Ofo and Biloxi, are no longer spoken.) Less widely distributed than Siouan is the Iroquoian family (six languages, largely of the East, including Cherokee and Mohawk), the Caddoan family (Caddo in the East, Wichita and Pawnee in the Prairies), and two language isolates of the East (Catawba and Yuchi), more closely related to Siouan than to the other families in Macro-Siouan.

Hokan. The Hokan phylum includes several small families and a number of language isolates scattered from the Northwest Coast through California, with extensions into the Great Basin and the Southwest, and as far south as Meso-America. Hokan languages spoken by the greatest numbers of speakers include those in two families in Mexico, the Tlapanecan and Tequistlatecan, and in the Yuman family in Arizona and California.

Penutian. The Penutian phylum is the only group of languages in North America for which relationships with languages in South America have been traced convincingly. The Penutian languages are thus distributed from the Northwest Coast and Plateau areas through California (with a possible extension into the Southwest) and Meso-America into Bolivia, Chile, and Argentina. Many of the more than 24 Penutian languages north of Mexico are either no longer spoken or are spoken by fewer than ten people. In Meso-America, however, many native languages have a considerable number of speakers; *e.g.*, Mixe, in the Zoque family, has over 30,000 speakers, and the Mayan family includes some languages with several hundred thousand speakers, as Maya, Quiché, Kekchi, Cakchiquel, and Mam.

The Tanoan and Uto-Aztecan families

Aztec-Tanoan. The Aztec-Tanoan phylum consists of two families: the Tanoan (Kiowa-Tanoan) family with three languages in the Southwest, including those spoken by the Taos and the Santa Clara, and one language in the Plains (Kiowa); and the Uto-Aztecan family, with about a score of languages spoken from the Plateau and California into Meso-America, with relatively late extensions into the Plains. California Uto-Aztecan languages include Cahuilla and Luiseño; Great Basin languages include Paiute and Shoshoni, with the Ute and Comanche dialects in the Plains; Southwestern languages include Hopi and Pima-Papago; Meso-American languages include Nahuatl, the language of the descendants of the Aztecs. The million speakers of the several varieties of Nahuatl far outnumber the total number of speakers of all the other Uto-Aztecan languages.

Oto-Manguean. Languages of one North American phylum are located entirely in Meso-America—the Oto-Manguean phylum, consisting of five small families. The languages with the largest number of speakers are Otomí, Mixtec, and Zapotec.

Unaffiliated languages. In North America one large family (the Salish family in the Northwest Coast and Plateau) and several smaller families and language isolates (as the Wakashan family in the Northwest Coast and Tarascan in Meso-America) remain undetermined in phylum affiliation. Remote relationships that have been proposed for some of these are in conflict with other proposed relationships, with no overwhelming evidence presented for any one of the proposals.

South American Indian languages. Language names for South America are much more numerous than those for North America, but information on actual languages is generally sporadic and often lacking entirely. Even when the list of names is reduced to 350 for languages said to be still spoken, the data to which the names refer consist, for the most part, of brief word lists; or nothing more may be known than the fact that a tribe X is said to speak differently from a tribe Y. Though it is possible to know that certain languages are probably closely related, it is not always possible to say how closely; *i.e.*, whether they might be dialects of, or occasionally just different names for, the same language. At the opposite extreme of genetic relationship, it is clear that there are large groups of remotely related languages, but the paucity of data makes possible conflicting proposals. For at least one group of languages, those of the high cultures of South America—the Inca and the Aymara—and some of their neighbours, the problem of establishing genetic relationship is complicated by the problem of sorting out borrowings among them.

Andean-Equatorial phylum

The Andean-Equatorial phylum includes the greatest number of still-spoken languages (almost 200) and the three South American Indian languages with the greatest number of speakers (Quechua, Guarani, and Aymara). The living Andean-Equatorial languages comprise some 14 families and several language isolates. The Arawakan family includes the most languages—around 100—and has the widest distribution: across northern South America from French Guiana to Colombia and southward as far as Paraguay; formerly, Arawakan languages were also spoken in Central America and the islands of the Caribbean. Most Arawakan languages are spoken by not more than a few hundred people. More than two dozen languages of the Tupian family are still spoken over a large part of South America, mostly south of the Amazon River. Tupian languages include Guarani (Tupi-Guarani), spoken in a number of dialects by about 2,900,000 people in Paraguay, Brazil, Argentina, and Bolivia. Quechua, of the Quechumaran group, is spoken by around 12,000,000 people in Peru, Ecuador, Colombia, Bolivia, Argentina, and Chile. Some Quechua dialects are so divergent that they might be regarded as separate languages. The other Quechumaran language group, Aymaran, is spoken by about 1,900,000 people in Peru and Bolivia. Most other languages in the Andean-Equatorial phylum are spoken by only a few thousand persons.

The Ge-Pano-Carib phylum includes almost as many languages still spoken as the languages of the Andean-Equatorial phylum, but the former are all spoken by relatively small tribes, so that the total number of speakers of these languages is only a small fraction of the number of speakers of Andean-Equatorial languages. In terms of numbers of languages, the largest family in the Ge-Pano-Carib phylum is the Cariban (Carib) family, with some 60 languages still spoken in Venezuela, the Guianas, Brazil, and Colombia. Cariban languages were also formerly spoken in the Caribbean islands. Most Cariban languages have no more than 1,000 speakers. The other large family in the phylum, the Macro-Ge family, includes more than 25 languages in Brazil.

The languages of the Macro-Chibchan phylum, of which 39 may still be spoken, are distributed from Guatemala and Honduras southward into, and possibly beyond, Peru. The largest component of the phylum is the Chibchan family, of which 16 languages are still spoken from Nicaragua to northwest Colombia—these include Cuna, spoken on the Mulatas (San Blas) Islands of Panama as well as on the mainland of Panama and Colombia; Guaymi, spoken in Panama; and Paez in Colombia. For further information on the Indian languages of the Americas, see NORTH AMERICAN INDIAN LANGUAGES; MESO-AMERICAN INDIAN LANGUAGES; SOUTH AMERICAN INDIAN LANGUAGES.

BIBLIOGRAPHY. E.H. STURTEVANT, ***Linguistic Change*** (1917, reprinted 1961), was an important text in Indo-European comparative linguistics, until replaced by WINFRED P. LEHMANN, ***Historical Linguistics: An Introduction*** (1962). EDWARD SAPIR, ***Language: An Introduction to the Study of Speech*** (1921, reprinted 1957), presents a revision of the 19th-century whole-language typology. For an example of modern application of subsystem typology, see A.K. RAMANUJAN and COLIN MASICA, "Toward a Phonological Typology of the Indian Linguistic Area," in ***Current Trends in Linguistics***, vol. 5, pp. 543–577 (1969); and for an example of lexical domain typology, see BRENT BERLIN and PAUL KAY, ***Basic Color Terms: Their Universality and Evolution*** (1969). ANTOINE MEILLET and MARCEL COHEN (eds.), ***Les Langues du monde***, new ed. (1952), is a survey of languages of the world organized in terms of their genetic classification. Languages are also classified genetically in two more recent surveys: C.F. and F.M. VOEGELIN, "Languages of the World," in ***Anthropological Linguistics***, vol. 6–8 (1964–66); and THOMAS A. SEBEOK (ed.), ***Current Trends in Linguistics***, 13 vol. (1963–75)—individual volumes deal with geographic areas, and typological information on particular languages is presented. JOSEPH H. GREENBERG, "The Indo-Pacific Hypothesis" (pp. 807–871), and S.A. WURM, "The Papuan Linguistic Situation" (pp. 541–657), both in ***Current Trends in Linguistics***, vol. 8 (1971), exemplify variation in proposals for remote genetic relationships among the same set of languages. EDWARD SAPIR, 'Central and North American Languages," in ***Encyclopædia Britannica***, 14th ed. (1929), presented a classification of these languages in terms of six phyla; this classification was revised by a conference of specialists that resulted in the publication of c.~and F.M. VOEGELIN (comps.), ***Map of North American Indian Languages***, rev. ed. (1967). The culture areas of North American Indians are described in HAROLD E. DRIVER, ***Indians of North America***, 2nd ed. rev. (1969). See also the bibliographies for the various language and language family articles mentioned in the text.

(C.F.V./F.M.V.)

Laos

Laos is a small, landlocked, undeveloped country on the Indochinese Peninsula. It is bounded on the north by China, northeast and east by Vietnam, south by Cambodia, and west by Thailand and Burma. Following the contour of the Indochinese Peninsula, Laos extends more than 600 miles from northwest to southeast, with a total area of approximately 91,400 square miles (236,800 square kilometres).

In the 1950s Laos became a focus of struggle among the major powers of the world. Despite international agreements to neutralize and isolate the country from external threat and to support its unity and development, Laos continued until 1975 to be the scene of others' battles, and then became the object of political competition between Vietnam and the Soviet Union, on one hand, and the People's Republic of China, on the other, while its own people remained divided and in conflict.

Located in an area of strategic importance among contending powerful neighbours, Laos has neither the natural barriers nor a population large enough to withstand foreign encroachment. The majority of its estimated 3,546,000 people live in the lowlands and in the valley of the Mekong River in the western and central part of the country. The remainder live in small groups in the eastern hill areas. The people are divided by differences of religion, language, and custom. Another source of disunity is the fact that, until the 19th century, Thailand, Burma, and Vietnam either held territory that at present is a part of Laos or exercised suzerainty and indirect rule over parts of the land and its people. Although the peoples of Laos can look back to a golden age when the country was strong and united under its own kings, for the past two centuries it has been divided and weak, subject to contention among its neighbours.

Laos gained legal unity and the formal status of a nation on October 22, 1953, when France recognized the Kingdom of Laos as fully independent and sovereign; in December 1955 it was admitted to the United Nations. In reality, however, it was at best a divided state. In 1953 the Communist Pathet Lao, a dissident group supported by the Viet Minh military forces of Vietnam, established itself in the northeast and claimed to be the legal government. The Geneva Conference of 1954, while recognizing the unity of Laos and the legality of the kingdom, permitted the Pathet Lao forces to occupy the provinces of Phôngsali and Houaphan (Sam Neua), "pending a political settlement." No real settlement occurred, despite efforts by the leaders of Laos to integrate the dissidents, and internal warfare continued until 1975 as foreign forces supported rival Laotian factions and used Laos as a battlefield in the Vietnam war. Intra-Laotian strife became intertwined with the larger conflict. In 1975 the Pathet Lao gained control of the country, and the monarchy was ended and the Lao People's Democratic Republic inaugurated in December.

This article deals with contemporary Laos. For historical aspects, see LAOS, HISTORY OF; for an associated physical feature, see MEKONG RIVER.

Landscape. Dominating the landscape of Laos are its inhospitable, forest-covered mountains, which in the north rise to a maximum of 9,248 feet (2,819 metres) above sea level. The principal range lies along a northwest-southeast axis, but in the north there are lower, secondary ranges along a northeast-southwest axis. The ranges are broken by narrow, deep valleys through which rivers flow. A vital area strategically and politically is the Plain of Jars (a name derived from large prehistoric stone jars discovered there; Laotian Thông Haihin), a part of the Plateau de Xiangkhoang that often changed hands during the war. In the central and northern parts of the country, the mountains form part of the Chaîne Annamitique. The only lowlands lie alongside the eastern bank of the Mekong. The general slope of the land is upward to the east.

Drainage

The rivers on the eastern mountain slopes flow toward Vietnam and the South China Sea; the rivers on the western side drain into the Mekong, which flows in a southerly direction from the province of Yunnan in China to the South China Sea, forming Laos' boundaries with Burma and (for the most part) with Thailand. The floodplains of the lowlands are formed from the alluvial soils carried by the river and its tributaries. The only other important fertile area is the Plateau des Bolovens, at an elevation of about 3,500 feet (1,100 metres). The hills and mountainsides can be brought into temporary cultivation by the primitive method of slash and burn, but they quickly lose their fertility and the cultivators must move to other areas and repeat the process.

The climate in Laos reflects its location between the latitudes of 14° and 22° N. During the rainy season, May to October, the winds blow from the southwest and deposit an average rainfall of between 50 and 90 inches (1,300 and 2,300 millimetres). In the Plateau des Bolovens precipitation reaches 160 inches (4,100 millimetres) per year. The dry season, from November to April, is dominated by northeast winds. Temperatures average between 60° and 70° F (16° and 21° C) in the cool months of December through February, increasing to more than 90° F (32" C) in March and April, just before the rains. In the wet season, the average temperature is 80" F (27" C).

Laos has tropical rain forests of broadleaf evergreens in the north and monsoon forests of mixed evergreens and deciduous trees in the south. In the monsoon forest areas the ground is covered with tall, coarse grass called tranh; the trees are mostly second growth, with an abundance of bamboo, scrub, and wild banana.

Peoples. More than half the population is concentrated in the lowlands, where most are engaged in wet rice farming. The isolated valley communities preserve different traditions and different dialects. Village populations range between 50 and 2,000, usually located close to rivers and roads that give the people access to itinerant traders as well as to each other. Most villages are laid out around a main street or open area, farmlands being adjacent to the residential areas. Every village, if it can, has a pagoda and supports at least one Buddhist monk. The pagoda compound usually includes a public building that serves as a school and a meeting hall. Village leadership is usually divided, the headman having authority in secular matters and the monk in religious.

The hill peoples

The hill peoples are usually organized on tribal lines and live in smaller groupings. They are hunters and gatherers of forest products as well as farmers; their techniques of shifting cultivation prevent them from establishing permanent villages. Their political and social structures are varied. The tribal Tai have a stratified social structure and a political hierarchy, while the Mon-Khmer tend to have a simple political organization with a single headman. The Meo (Miao) tribesmen have a tradition of a king and chiefs; in theory their villages are tribal, but in fact they usually have a headman. Hill peoples living close to the Lao and Tai tend to acquire the languages and cultures of their neighbours and to engage in limited trade with them; those living at higher elevations remain unacculturated.

Urban life in Laos is limited to the capital, Vientiane, the former royal capital, Louangphrabang, and four or five large towns. With the exception of Louangphrabang, all are located near the Mekong in the floodplain area. Their populations are predominantly Lao, with many Chinese, Vietnamese, and Europeans. The Lao elite tends to be Westernized in its life-style and dress. By comparison with the cities of Thailand, Malaysia, or Vietnam, those of Laos are small and provincial.

Linguistic–political groups. The peoples of Laos are divided by language, culture, and location. Lao officials distinguish four basic linguistic–political groups: Lao-Lu, or valley Lao; Lao-Tai, or tribal Tai; Lao-Theng, or Mon-Khmer; and Lao-Soung, or Meo and Man. Mountain people sometimes are called Kha ("Slave"), a pejorative term.

The Lao-Lu live in the lowlands, on the banks of the Mekong and its tributaries, and in the cities. They speak Laotian Tai, which is closer to the language spoken by Thais in Thailand than it is to the language of the local Tai-speaking tribes. It is the language that the minorities living near Lao areas are gradually adopting.

The Lao-Tai include such local groups as the Black Tai and Red Tai, both names referring to the dress of the women; the Tai Neua, or Tai of the north; the Tai Phuan; and the Phon Tai. The Lao-Tai live in all parts of the country, chiefly in upland areas. The various dialects of the Lao-Tai are mutually intelligible.

Mon-Khmer

The Lao-Theng, better known as Mon and Khmer, or Mon-Khmer, include many groups of people scattered throughout Laos, northeastern Burma, northern Thailand, and southern China and are thought to be descendants of the earliest populations in the region. These people do not form a single coherent group but rather include between 25 and 30 distinct groups, some closely related, others only tenuously identified as part of this linguistic group.

The Lao-Soung, which include the Meo (Miao) and the Man (Yao), are believed to have come from southern China since the late 18th century. They are divided into subgroups, and neither constitutes a large proportion of the population of Laos. (For the languages, see further TAI LANGUAGES.)

The other distinct linguistic groups are few. Tibeto-Burman-speakers, who came from southern China, live in the north and northwest. Chinese and Vietnamese live primarily in the urban areas. Initially, French was the language of the Lao elite and of the cities, but by the 1970s English had begun to displace it. Under the leadership of the Lao People's Revolutionary Party, Vietnamese has become the third language of the elite.

Cultural and religious differences. Prior to the emergence of the Lao People's Democratic Republic in 1975,

Social structure

it was accurate to say that the Lao peoples had a distinct pattern of culture and dress. They also had a well-defined social structure. differentiating between royalty and commoners. The members of the elite were said to number about 2,000, which included only a few outsiders who were not descendants of nobility. Most of the elite lived in the cities. drawing their incomes from rural land rents or from urban occupations. After **1975** a new elite emerged representing the victorious leftist forces. Many of this group were of aristocratic origin.

Traditionally the **Lao-Tai,** or tribal Tai. were organized in groups larger than villages. called muong. Each was ruled by a hereditary ruler called the chao muong. Within this broad grouping. however. there were ethnic variations. Among the Black Tai. the nobility consisted of two descent groups. the Lo and the Cam. who provided the rulers of the *muong*. The religious leaders came from two other descent groups. the Luong and the Ka. The Black Tai tribal organization had three levels: the village; the commune. which was composed of a number of villages; and the overall *muong*. The latter two were ruled by nobles. while the village headman was selected from among the commoners by the heads of households. The Red Tai had a similar social structure. with the addition of a council of five to aid the chao muong. The nobility owned the land and had the right of service from the commoners. Not all of them were Buddhists; those living in the higher and more isolated areas retained their traditional culture.

The **Lao-Theng,** or **Mon-Khmer,** have no political or social structure beyond the village. They are led by a village headman. who is their link to the central government; but his role in the village is not clear. The people traditionally have been spirit worshippers; one subgroup. the Lamets. practices ancestor worship. Some have adopted Buddhism.

Among the **Lao-Soung,** the Meo have a tradition of a king and subchiefs and a large-scale organization. although in practice this is usually limited to the village. The village consists of several extended families. In some villages. all the heads of households are members of a single clan. and the head of the clan is the headman of the village. Where several clans reside together in a large village there are several headmen. one being the nominal head and the link to the government. The headman has real authority in the village and is aided by a council. The Meo have extended their organization beyond the village for military purposes. They are spirit and ancestor worshippers.

The religion of most of the people of Laos is Theravāda Buddhism. professed by most Lao and by a small number of other ethnic groups. The rest of the people are animists or spirit worshippers. Many see no contradiction in being both. since Buddhism shows the way to Nirvāṇa while spirit worship helps a person to cope with daily and local problems. Among the hill peoples. especially those who have migrated from southern China. there are groups that mix Confucian ideas with Buddhism and animism. There were Catholic and Protestant missionaries in the country before **1975.** but it is estimated that no more than **2** or **3** percent of the population is Christian. The Vietnamese. who live both in the cities and in the northeastern rural areas. practice a mixture of Mahāyāna Buddhism and Confucianism.

No complete census has ever been conducted in Laos. but estimates by the United Nations placed the population at 3.546.000 in **1978.** The birth rate was estimated to be 44.1 per 1.000. the death rate **20.3,** and the annual rate of increase **2.3** percent. To build the population. the government in **1976** prohibited the sale of birth-control devices. Considerable migration takes place in and out of China. Vietnam. and Thailand.

Economy. Mining. The natural resources of Laos include coal. iron. copper. lead. gold. tin. and precious stones. Only tin has been extracted on a commercial scale. though the others have been mined in primitive and unsystematic ways. About 1.000 tons of tin ore per year are produced from the mines at Phôntiou and Nong Sun. Agriculture. The chief occupation of the people is

MAP INDEX

Political subdivisions

Attapu	14·45n	107·00e
Borikhan	18·30n	103·40e
Charnpasak	14·45n	105·45e
Champhon	16·25n	105·25e
Hôngsa	19·40n	101·10e
Houakhong	20·45n	101·00e
Houaphan	20·20n	104·00e
Khamrnouan	8·00n	105·00e
Louang-phrabang	20·20n	102·30e
Pak-Lay	18·20n	101·15e
Phôngsali	22·00n	102·00e
Saravan	15·40n	106·45e
Savannakhkt	16·40n	105·00e
Sithandon	14·15n	105·50e
Vangviang	19·00n	102·35e
Vapikhamthong	15·40n	105·50e
Vientiane	18·10n	102·35e
Xaignabouri	19·20n	101·30e
Xédôn	15·00n	106·10e
Xiannkhoana	19·15n	103·30e

The name of a political subdivision if not shown on the map is the same as that of its capital city.

Cities and towns

Attapu	14·48n	106·50e

B = Ban

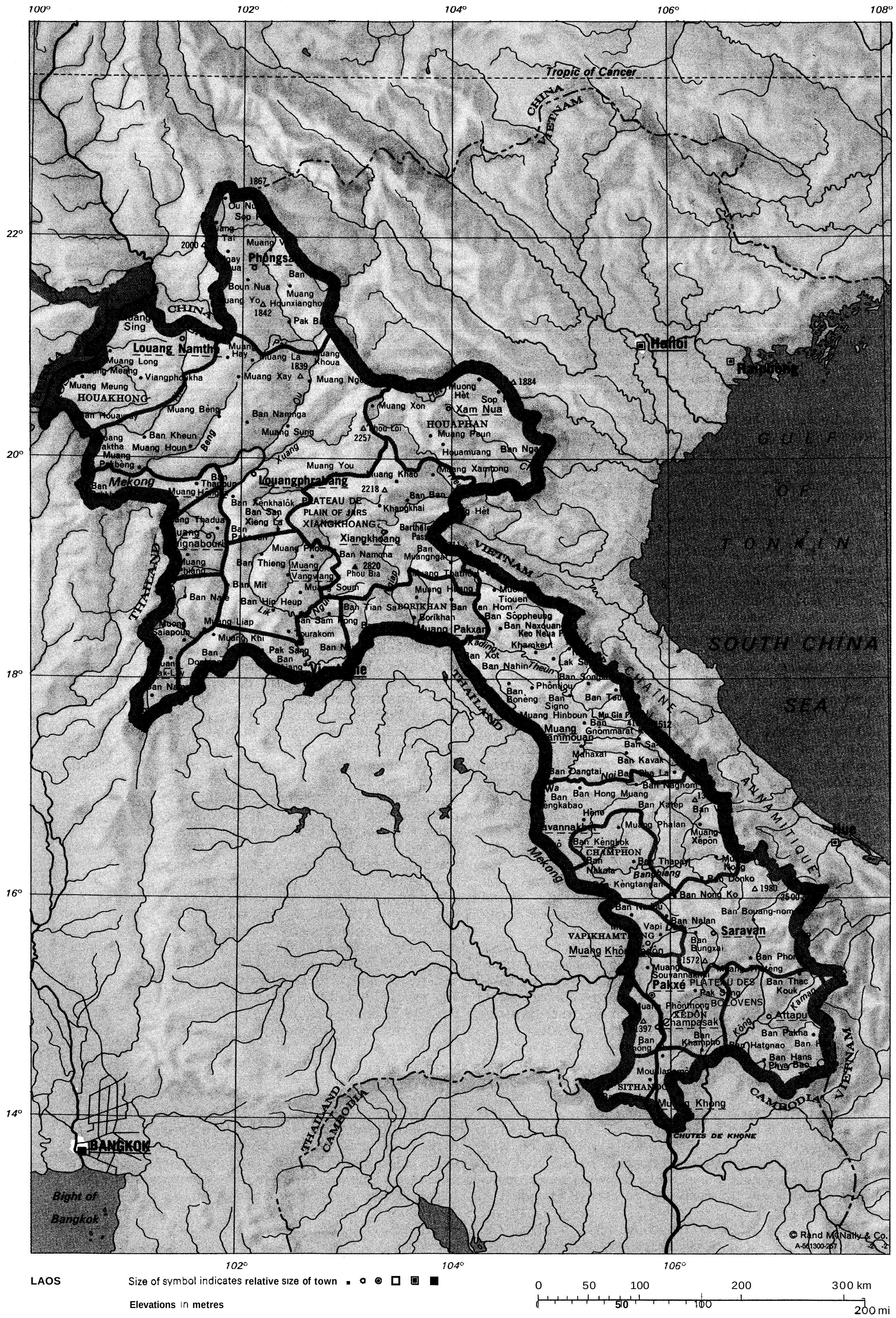
B. Ban	19·38n	103·34e
B. Bonèng	17·58n	104·35e
B. Bouangnom	15·47n	106·47e
B. Bungxai	15·42n	106·14e
B. Chak	14·17n	105·25e
B. Cha La	17·11n	106·05e
B. Dang Tai	17·06n	104·57e
B. Dong Ko	16·12n	106·17e
B. Donhiang	18·05n	101·48e
B. Gnômmarat Keo	17·35n	105·10e
B. Hans Phya Bao	14·31n	106·51e
B. Hatgnao	14·40n	106·35e
B. Hatkiang	18·11n	102·40e
B. Hèt	14·44n	107·29e
B. Hin Heup	18·38n	102·20e
B. Hong Muang	17·04n	105·12e
B. Houayxay	20·18n	100·26e
B. Katèp	16·48n	105·52e
B. Kavak	17·18n	105·37e
B. Kèngkabao	16·48n	104·45e
B. Kèngkok	16·26n	105·12e
B. Kèngtangan	16·05n	105·22e
B. Khampho	14·37n	106·17e
B. Kheun	20·13n	101·07e
B. Mit	18·51n	101·55e
B. Muangngat	19·05n	104·04e
B. Nadou	15·51n	105·38e
B. Nagnom	17·02n	105·44e
B. Nahin	18·14n	104·13e
B. Nakala	16·17n	105·11e
B. Nalan	15·50n	106·04e
B. Nalè	18·42n	101·34e
B. Namcha	19·09n	102·53e
B. Namnga	20·22n	102·19e
B. Nam Tao	17·50n	101·15e
B. Naxon	18·12n	103·05e
B. Ngam	20·11n	104·53e
B. Nong Ko	16·01n	106·04e
B. Pakha	14·38n	107·24e
B. Pakkhop	19·49n	100·36e
B. Pakneun	19·14n	101·50e
B. Pak Sang	18·13n	102·08e
B. Phone	15·25n	106·42e
B. Phông Pho	14·36n	105·52e
B. Phya	21·35n	102·55e
B. Sam Pong	18·32n	102·47e
B. San Xieng La	19·27n	102·26e
B. Signo	17·51n	105·04e
B. Songkhon	17·58n	105·10e
B. Sôppheung	18·33n	104·17e
B. Teung	17·54n	105·29e
B. Thaboc	18·22n	103·12e
B. Thanoun	19·50n	101·29e
B. Thapayi	16·19n	105·41e
B. Tian Sa	18·43n	103·14e
B. Van Hom	18·44n	104·01e
B. Vat	16·54n	106·25e
B. Xènkhalôk	19·42n	101·54e
B. Xot	18·11n	104·05e
Borikhan	18·33n	103·43e
Botene	17·35n	101·12e
Boun Nua	21·38n	101·54e
Champasak	14·53n	105·52e
Hène	16·42n	105·17e
Houamuang	20·09n	103·38e
Khamkeut	18·15n	104·43e
Khangkhai	19·28n	103·15e
Lak Sao	18·11n	104·58e
Louang Namtha	20·57n	101·25e
Louang-phrabang	19·52n	102·08e
Mahaxai	17·25n	105·12e
Mounlapamôk	14·20n	105·52e

M. = Muang

M. Bèng	20·22n	101·44e
M. Hay	21·03n	101·49e
M. Hinboun	17·35n	104·36e
M. Hôngsa	19·43n	101·20e
M. Houn	20·09n	101·27e
M. Hounxiang-houng	21·37n	102·18e
M. Huang	18·54n	103·42e
M. Khammouan	17·24n	104·48e
M. Khao	19·47n	103·29e
M. Khi	18·27n	101·46e
M. Không	14·07n	105·51e
M. Khôngxédôn	15·34n	105·49e
M. Khoua	21·05n	102·31e
M. La	20·52n	102·07e
M. Liap	18·29n	101·40e
M. Long	20·57n	100·48e
M. Meung	20·43n	100·28e
M. Ngoy	20·43n	102·41e
M. Nong	16·22n	106·30e
M. Ou Nua	22·18n	101·48e
M. Ou Tai	22·07n	101·48e
M. Pakbèng	19·54n	101·08e
M. Pak-Lay	18·12n	101·25e
M. Paktha	20·06n	100·36e
M. Pakxan	18·22n	103·39e
M. Peun	20·13n	103·52e
M. Phalan	16·39n	105·34e
M. Phiang	19·06n	101·32e
M. Phônthong	15·05n	105·39e
M. Phoun	19·07n	102·43e
M. Sing	21·11n	101·09e
M. Soum	18·46n	102·36e
M. Souvan-nakhili	15·23n	105·49e
M. Sung	20·19n	102·27e
M. Thadua	19·26n	101·50e
M. Thatèng	15·26n	106·23e
M. Thathôm	19·00n	103·36e
M. Va	21·53n	102·19e
M. Vangviang	18·56n	102·27e
M. Vapi	15·40n	105·55e
M. Xaignabouri	19·15n	101·45e
M. Xarntong	19·51n	103·51e
M. Xay	20·42n	101·59e
M. Xépôn	16·41n	106·14e
M. Xon	20·27n	103·19e
M. Yo	21.31n	101·51e
M. You	19·49n	102·50e
Muong Het	20·48n	104·03e
Muong Tiouen	18·51n	104·23e
Napè	18·18n	105·06e
Ngay Neua	21·50n	101·54e
Nong Hèt	19·29n	103·59e
Pak Ban	21·14n	102·28e
Pak Song	15·11n	106·13e
Pakxé	15·07n	105·47e
Phôngsali	21·41n	102·06e
Phôntiou	17·53n	104·37e
Saravan	15·43n	106·25e
Savannakhét	16·33n	104·45e
Sop Hao	20·33n	104·27e
Tourakorn	18·26n	102·32e
Viangphoukha	20·41n	101·04e
Vientiane	17·58n	102·36e
Xam Nua	20·25n	104·02e
Xénô	16·35n	104·50e
Xiangkhoang	19·20n	103·22e

Physical features and points of interest

Annarnitique. Chaine (Annamese Cordillera). mountains	17·00n	106·00e
Banghiang, *river*	16·03n	105·15e
Barthklemy Pass	19·26n	104·06e
Beng, river	19·53n	101·08e
Bia. Phou. mountain	18·59n	103·09e
Boiovens. Plateau des	15·20n	106·20e
Chu, river	19·53n	105·45e
Don. river	15·07n	105·48e
Het. river	20·49n	104·01e
Jars. Plain of	19·27n	103·10e
Kading. river	18·19n	104·00e
Kaman. river	14·48n	106·51e
Keo Neua Pass	18·23n	105·09e
Khone. Chutes de (Khone Falls). rapids	13·56n	105·56e
Kông, river	14·28n	106·20e
Lik. river	18.31n	102.31e
Loi. Phou. mountain	20·16n	103·12e
Mekong. river	13·55n	105·58e
Mu Gia Pass	17·40n	105·47e
Neun. river	19·42n	104·03e
Ngiap. river	18·24n	103·36e
Ngum, river	18·09n	103·06e
Noi. river	17·05n	105·02e
Ou, river	20·04n	102·13e
Pak, river	21·05n	102·31e
Tha. river	20·07n	100·36e
Theun, river	18·14n	104·38e
Wa, river	16·55n	104·48e
Xiangkhoang, Plateau de	19·30n	103·10e
Xuang, river	19·58n	102·15e



LAOS

Size of symbol indicates relative size of town ▪ ○ ◎ □ ▣ ■

Elevations in metres

Laos, Area and Population

	area		population*
	sq mi	sq km	1973 estimate
Provinces (*khoueng*)†			
Attapu	4,600	11,900	99,000
Borikhan	2,400	6,200	51,000
Champasak	1,000	2,600	103,000
Houakhong	5,000	12,900	142,000
Houaphan	6,300	16,300	211,000
Khammouan	10,200	26,400	259,000
Louangphrabang	14,400	37,200	450,000
Phôngsali	6,100	15,800	127,000
Saravan	7,100	18,400	166,000
Savannakhét	8,400	21,700	456,000
Sithandon	2,000	5,300	67,000
Vapikhamthong	1,300	3,200	122,000
Vientiane	5,600	14,400	351,000
Xaienabouri	7,100	18,400	213,000
Xédôn	2,500	6,600	146,000
Xiangkhoang	7,500	19,500	218,000
Total Laos	91,400‡	236,800	3,181,000

*No census has ever been taken. †Provinces newly established in 1973: Champhone, Hôngsa, Paklay, and Vangvieng, all parts of older provinces. ‡Converted area figures do not add to total given because of rounding.
Source: United Nations figures.

Rice

agriculture, with an estimated 90 percent engaged in rice farming. Since the creation of the Lao People's Democratic Republic in 1975, natural calamities, including both drought and excessive rainfall, have prevented the people from achieving self-sufficiency in rice production. In the late 1970s the nation had to look to imports and gifts to make up for inadequate production. Modest amounts of corn (maize), sugarcane, tobacco, and cotton are produced. The forests provide teak and other woods, gum benzoin, cardamom, and stick lac. A major agricultural commodity that does not find its way into the statistics is opium, grown mainly by the Meo and traded illegally.

Nearly every household raises livestock—including cattle and buffalo—and poultry. Leather and hides are traded.

Under the revolutionary government, a serious effort is being made to relocate hill peoples onto the plains and to get them to adopt wet rice farming practices in place of their traditional slash and bum. It has been reported that some of the Meo have been forced to resettle in agricultural cooperatives in the region of the Plain of Jars. The majority of hill dwellers are resisting change, however, despite pressure and persuasion to alter their agricultural practices, social customs, and traditions. If the government persists in this policy it may provoke outright opposition to, and revolt against, the most serious challenge to Meo traditional life.

Hydroelectricity. A major hydroelectric project is under way on the Nam Ngum (nam, "river"), an effluent of the Mekong, about 40 miles (70 kilometres) north of Vientiane. Financed by Japan, West Germany, and several other countries, the project produces surplus electricity that is sold to Thailand.

Industry. The tiny industrial sector manufactures bricks and ceramics, matches, cigarettes, soft drinks and beer, rubber sandals, cloth, pottery, and plastic bags. By mid-1977 the government reported that nearly 80 percent of the small and medium-sized factories in the areas formerly under the royal government were back in production.

Trade and finance. The chief exports are timber, tin, coffee, leather and hides, cardamom, gum benzoin, and stick lac. The major buyers formerly were Singapore, Malaysia, Thailand, the United States, and Hong Kong. The major suppliers of imports were Thailand, Japan, the United States, France, and the United Kingdom. After 1975, however, this pattern of trade was disrupted, and new trading relationships are beginning to develop with Thailand and Vietnam and with Communist European states. Imports always vastly exceed exports in value.

In 1975 financial support from Western countries all but stopped; to halt inflation and reduce black markets, the old kip (Laotian unit of currency) was replaced in June 1976 with a new kip, at a greatly reduced value. This, together with greater state control of the economy and gifts from the Soviet Union, the People's Republic of China, Vietnam, and the United Nations, helped the government during its first difficult years.

Transportation. A major obstacle to the economic and social development of Laos is its transportation system. Rivers and roads are the major avenues of communication, supplemented by air transport. The Mekong River is the major north–south commercial artery; all but two sections of it—Chutes de Khone (Khone Falls) and the rapids of Khemmarat (Khemarat)—are navigable either all or part of the year. Large barges operated by Chinese ply the deeper sections of the rivers between towns, but most of the water traffic is carried in Lao-operated sampans and pirogues. The latter average 25 to 35 feet in length and can carry up to seven passengers and half a ton of cargo. Some of the mountain people build bamboo rafts and float their goods to market, selling the raft along with the goods.

River traffic

During French rule, a primitive network of roads was created. The main artery joined Saigon with Louangphrabang, and several lesser roads led eastward through the four mountain passes to Vietnam and to the main towns and population centres. During the 1960s, with U.S. assistance, an all-weather road was added between Vientiane and Pakse; another was to be built between Vientiane and Louangphrabang. In addition, the North Vietnamese developed a complex of roads and trails across eastern Laos for their own use during the war. Laos itself has no railways, but a railroad from Bangkok, Thailand, to the Lao border serves as a major artery.

The newest means of transport is the airplane. In the mid-1970s there were nine airports with permanent surface runways. The Wattay international airport at Vientiane connects with Bangkok, Phnom Penh and Hong Kong. This traffic was halted briefly in 1976, but by the next year regular air traffic with Thailand, Cambodia, and Vietnam resumed. There were plans to extend air service to Burma and China.

Administration and social conditions. Government. Before December 2, 1975, Laos was a constitutional monarchy. Real power rested with the prime minister and the legislature. There had been nominal political parties since 1947, when the constitution was promulgated, but they were loose coalitions of leaders with little popular following. In 1953 there were two main parties, the National Progressive and the Independent, and two splinter parties, the National Union and the Democratic. In 1958 the Lao Patriotic Front, backed by the Communist Pathet Lao, participated in the elections, as did the Peace Party; together they won several seats. In the late 1950s two new parties emerged, the Rally of the Lao People and the Committee for the Defense of the National Interest.

Beginning in 1973, when a cease-fire agreement was reached, a series of coalition governments sought to exercise power. The traditional elite continued to rule in urban areas, while the bulk of the people in the countryside remained outside the national political process. The coalition government gave way in December 1975 to a new body, the National Congress of People's Representatives, which, under the guidance of the Pathet Lao, met and passed several resolutions that ended the monarchy and replaced it with the Lao People's Democratic Republic. The King abdicated.

Supreme People's Council

The government is headed by a president, who is also chairman of the Supreme People's Council, an interim (pending elections) legislative body. The secretary general of the Lao People's Revolutionary Party (the only political party) serves as prime minister. The party, which replaced the Pathet Lao as leader of the Socialist revolution, was formerly called the People's Party of Laos and is the Communist core of the Lao Patriotic Front. It is organized much like other Communist parties, with a Central Committee headed by a Politburo. Its membership is estimated to be about 15,000.

The country is divided into 20 provinces, each subdivided into districts, which, in turn, are subdivided into towns and villages. At each level of local government there are "people's revolutionary committees," which receive directions from the Central Committee of the Lao People's Revolutionary Party.

Military forces. Between independence, in 1953, and the cease-fire, in 1973, the military forces grew to become a major institution. In addition to the Royal Laotian Army, Royal Lao Air Force, Laotian River Flotilla, and Lao National Police Corps, there were the separate armies of the Pathet Lao and the neutralists. The government forces numbered about 65,000 and received aid from the United States. The Pathet Lao army of 20,000, with 15,000 irregulars, was supported by North Vietnam and China. For a time during the 1960s the neutral force of 10,000 stood between the two, but in the late 1960s this group threw its support to the Pathet Lao.

Army

After the victory of the Pathet Lao (officially Lao People's Liberation Army from 1965), the Royal Laotian Army was in part integrated with it and the rest was disbanded. In the late 1970s the army had about 46,000 men, and there was a very small navy and an air force with about 2,000 personnel.

Medical care. The country has around 40 hospitals, supplemented by infirmaries and rural dispensaries. After 1975 and with the departure of the majority of medical doctors, the government began building village infirmaries in most of the provinces and training medical workers. Using medicinal herbs, these village medical workers provide most of the primary health care. Malaria and gastroenteritis are the major health problems.

Education. Education is being reorganized. After 1975 the government set up a number of agricultural schools, sent teachers to give literacy classes in provincial villages, and opened new primary, secondary, and teacher-training schools. It claimed that by mid-1978 some 2,000,000 of its people were literate.

Cultural life. The basis of Laotian culture is religion and tradition. Art, literature, music, and drama draw mainly from these sources, and there appears to be little Western influence outside Vientiane.

Religious basis of cultural life

Theravāda Buddhism entered the country in the 14th century and along with Hinduism has been a major influence on cultural and intellectual life. The story of the Buddha and Hindu myths are the subjects of the carvings and sculptures found in all religious places. In the south, Khmer influences on the peoples of Laos are strong; in the north, Burmese and Thai influences are readily apparent. As elsewhere in Southeast Asia, religious symbols, stories, and themes have been modified and localized. The snake, for example, representations of which adorn religious and royal buildings, represents the benevolent spirit of the water and the protector of the king.

The Laotians have a variety of folk arts, including weaving, basketmaking, wood and ivory carving, and silverwork and goldwork. There are a number of Laotian musical instruments, of which the *khene,* a bamboo wind instrument, is most widely known. Music is not written down but is played from memory. The musicians and their music are closely linked to religion.

Dancing is a profession rather than a form of recreation; the professional dance troupes travel throughout the country performing for religious celebrations or on important holidays. Their main themes are drawn from the Indian epics. All professional dancers are male, the female roles being performed by young men and boys.

Laotian literature is predominantly religious and linked to the Buddhist tradition. There is also a secular literary stream based on themes of the Hindu epic poems, which have been transmuted into popular language, as in the Laotian epic the *Sin Xay.* The popular poems and songs are often satirical.

Laos had five newspapers in 1966, all published by the government or by prominent leaders and of very limited circulation. Since 1975 there has been only one paper, *Siang Pasason* ("Voice of the People"), the news organ of the governing party. The National Radio of Laos broadcasts on several stations, and the Foreign Service provides programs in Khmer, English, French, Lao, Thai, Vietnamese, and many dialects; both are sponsored by the government. There is no television service.

Prospects for the future. Under the revolutionary government, the warfare that divided the country was brought to an end. The social and economic changes introduced were not uniformly popular, and many of the hill peoples, especially the Meo, opposed the reforms and provided for future internal war. The end of the war did not end Laos' involvement in t[illegible] military and political [illegible] of it neighbours. The [illegible]ic of Kampuchea by Vietnam, followed by the assault of the People's Republic of China upon Vietnam, had an effect on Laos. China openly criticized the strong influence of Vietnam, which shows no sign of loosening its hold upon Laos. Since Vietnam is now tied politically and economically to the Soviet Union, there is little likelihood that Western nations will offer Laos technical or financial aid.

BIBLIOGRAPHY

General: RENE DE BERVAL *et al., Présence du royaume Laos, pays du million d'éléphants et du parasol blanc* (1956; Eng. trans., *Kingdom of Laos: The Land of the Million Elephants and of the White Parasol,* 1959), a general symposium by French and Laotian writers on the country's history, economy, society, and culture; FRANK M. LE BAR and ADRIENNE SUDDARD (eds.), *Laos: Its People, Its Society, Its Culture* (1960), a general survey by American scholars; T.D. ROBERTS (ed.), *Area Handbook for Laos* (1972), a highly readable survey.

Landscape: E.H.G. DOBBY. *Southeast Asia.* 10th ed. (1967), a standard work; C. ROBEQUAIN, *L'Évolution économique de l'Indochine française* (1939; Eng. trans., *Economic Development of French Indo-China,* 1944), an older but useful study.

Peoples: J.M. HALPERN, *Government, Politics* and *Social Structure in Laos: A Study of Tradition and Innovation* (1964), a detailed treatment by a leading authority; *Economy and Society of Laos: A Brief Survey* (1964), a good detailed overview; P. KUNSTADTER (ed.), *Southeast Asian Tribes, Minorities, and Nations, 2* vol. (1967), with four excellent chapters on the peoples of Laos; FRANK M. LE BAR, G.C. HICKEY, and J.K. MUSGRAVE, *Ethnic Groups of Mainland Southeast Asia* (1964), a good general treatment of the peoples of Laos.

National economy: A. SCHAAF, C. HART, and R.H. FIFIELD, *The Lower Mekong: Challenge to Cooperation in Southeast Asia* (1963); UNITED NATIONS, *Economic and Social Survey of Asia and the Pacific,* annual (1948-), and *Statistical Yearbook for Asia and the Pacific,* annual (1970-), are reliable sources of economic data and analysis.

Government, administration, and social conditions: A.J. DOMMEN, *Conflict in Laos: The Politics of Neutralization* (1964); P.F. LANGER and J.J. ZASLOFF, *North Vietnam and the Pathet Laos* (1970); SISOUK NA CHAMPASSAK, *Storm over Laos: A Contemporary History* (1961).

(J.Si.)

Laos, History of

The Lao people, the predominant ethnic group in present-day Laos, are a branch of the Tai peoples who by the 8th century AD established a powerful kingdom, Nanchao, in southwestern China. From Nanchao the Tai gradually penetrated southward into the Indochinese Peninsula; their migration was accelerated in the 13th century by the Mongol invasions of southern China by Kublai Khan. The Lao, together with other Tai peoples, gradually supplanted various primitive tribes (collectively known as Kha, or "Slaves") that from the 5th century on had lived in what is now Laos under the suzerainty of the Indianized Khmer Empire of Cambodia. During the 12th and 13th centuries they established the principality of Muong Swa (later Luang Prabang, now Louangphrabang), which was ruled by various Tai leaders and the history of which survives in Laotian legend and myth.

The first Laotian state

Lan Xang. Recorded Laotian history begins with Fa Ngum, who founded the first Laotian state, Lan Xang (Kingdom of the Million Elephants), with the help of the Khmer sovereign at Angkor. He was a great warrior and, in 1353–71, conquered territories including all of present-day Laos and much of what is today northern and eastern Thailand. He extended the Indo-Khmer civilization to the upper Mekong River and introduced Theravada Buddhism, preached by Khmer missionaries from Angkor.

In 1373 Fa Ngum was succeeded by his son Phya Sam Sene Thai, who did much to organize the pattern of administration and defense for the kingdom. After his death in 1416, a long period of calm, broken only by a Vietnamese invasion in 1478, allowed his successors to complete the work of organizing Lan Xang. This period of peace and tranquillity ended with Photisarath (reigned

1520–47), who involved Lan Xang in a struggle against Burma and Siam that lasted two centuries. Photisarath waged three wars against Siam and succeeded in placing his son Sethathirath on the throne of the Tai state of Chiang Mai (Chiengmai), marking Lan Xang's maximum territorial expansion. On Photisarath's death, Sethathirath returned to occupy his father's throne. His reign was marked by the loss of Chiang Mai to the Burmese, by the transference of the capital from Luang Prabang to Vientiane, and by the repulsion of two Burmese invasions c. 1565 and 1570.

When he died (1571) the Burmese seized Vientiane (1574) and ravaged the country, which lapsed into anarchy until Souligna-Vongsa ascended the throne in 1637 and restored order. He fixed the frontiers with Vietnam and Siam by means of treaties and led two victorious expeditions against the principality of Chieng Khouang in the south. A defender of Buddhism and a patron of the arts, he embellished Vientiane and made it a centre of intellectual brilliance. His reign is considered by Laotians to be a Golden Age.

Vietnamese rule

When Souligna-Vonasa died in 1694, one of his nephews seized the throne with the help of a Vietnamese army, thus placing Lan Xang under Vietnamese rule and initiating a period of chaos ending in the partition of the kingdom of Lan Xang. Other members of the royal family refused to accept Vietnamese vassalage. With the northern provinces under their control, they declared themselves independent (1707) and established the separate kingdoms of Luang Prabang and Vientiane. The south seceded in turn and set itself up as the kingdom of Champassak (1713). Split into three rival kingdoms, Lan Xang ceased to exist.

Under foreign rule. During the 18th century the three Laotian states. continually at loggerheads. tried to maintain their independence from the Burmese and the Siamese who were contending for the control of western Indochina. Their weakness, resulting from their disunity, inevitably caused them to fall prey to the Siamese.

Vientiane, which had sided with the Burmese, was invaded (1778), annexed, and made a state subject to Siam (1782). Luang Prabang, which had supported the Siamese, was invaded by the Burmese (1752), who imposed their rule upon it until the Siamese supplanted them (1778). In the south, Champassak, which had supported Burmese revolt against the Siamese, was also invaded (1778) and transformed into a dependency of Siam. Each of these kingdoms was placed under the control of a Siamese commissioner. The kings of Champassak, Vientiane, and Luang Prabang were allowed to rule in their respective kingdoms but had to pay tribute to Bangkok. Their appointments to the throne were made in Bangkok.

Chao Anou (king of Vientiane 1805–28) attempted to shake off this yoke. First, he strengthened the bonds of allegiance uniting Vientiane to Vietnam (1806), whose influence in Indochina had grown to rival that of Siam. Next, he persuaded Bangkok to give his son the governorship of Champassak, thus extending his frontiers as far as the old southern boundaries of Lan Xang. Thinking that the British, who had just conquered Burma, were going to attack Siam, he led three armies against Bangkok (c. 1826). But the Siamese regrouped their forces, marched on Vientiane, and defeated Anou, who fled to Vietnam. Vientiane was pillaged and destroyed. In 1819 Anou attempted another attack but was again defeated. Vientiane was made a Siamese province.

For the Siamese the annexation of Vientiane was the first step toward the creation of a great empire. They next extended and consolidated their colonization of the left bank of the Mekong to protect themselves from an eventual Vietnamese expansion westward. They therefore garrisoned Champassak (1846) and Luang Prabang (1885) and stationed troops as far as the Chaîne Annamitique. Siamese expansion toward the northeast—where the mountain states were placed under the cosuzerainty of Vietnam and Luang Prabang—provoked the protests of the French, who had established a protectorate over Vietnam. France entered into negotiations with Bangkok (1886) to define the Siamese-Vietnamese frontier and won the right to install a vice consul in Luang Prabang. The office was entrusted to Auguste Pavie, who, owing partly to his popularity with the Laotians, succeeded in winning Luang Prabang over to France. After a number of Franco-Siamese incidents in the Mekong Valley, French ships made a show of strength off Bangkok. On the advice of the British, Siam withdrew from the left bank of the Mekong and gave official recognition to the French protectorate in the evacuated territory (1893). French annexation was completed by treaties with Siam (Thailand) in 1904 and 1907.

French protectorate

The French organized this territory as the Protectorate of Laos and allowed it autonomy in local matters. The kingdom of Luang Prabang survived, but the other provinces were placed under the direct authority of a French official. France paid little attention to Laos until 1941, when, under Japanese pressure, the Vichy government restored to Thailand the territories acquired in 1904. In March 1945 the Japanese drove the French from Indochina and proclaimed the independence of Laos.

Two movements sprang up at this time. The first was anti-Japanese and was represented by the court of Luang Prabang and Prince Boun Oum of Champassak; the second was anti-French (the Free Laos movement, or Lao Issara), was located in Vientiane, the former French colonial administrative centre, and was led by Prince Petsarath. These two movements remained in conflict until the return of French troops, which compelled the supporters of the Lao Issara to flee to Thailand. In 1946 France, in a temporary agreement, recognized the unity and internal autonomy of Laos under the king of Luang Prabang, Sisavang Vong. Finally, after the promulgation of a constitution and general elections, a Franco-Laotian convention was signed on July 19, 1949, by which Laos was granted a limited degree of self-government within the French Union. All important power, however, remained in French hands.

Although many of the Lao Issara leaders were prepared to work with the French under this new arrangement, their decision was opposed by a more radical group led by Prince Souphanouvong. Under Souphanouvong's presidency a new political movement, the Pathet Lao (Lao Country), was created (1950) that joined forces with the Viet Minh of Vietnam in opposing the French. The Pathet Lao remained unreconciled when the French took further steps toward granting independence to Laos in October 1953, while still retaining control of all military matters in the kingdom. Between 1950 and early 1954 the Pathet Lao gained strength in northeastern Laos and had a firm grip on two of the country's provinces when the Geneva Conference brought the First Indochina War to an end.

Geneva Conference

Independent Laos. At the Geneva Conference the 14 participating nations (including France, Great Britain, the U.S., China, and the Soviet Union) agreed on the establishment of Laos as a unified, independent buffer state between Thailand and North Vietnam, which were allied, respectively, to the West and to the Communist bloc. But this agreement, a compromise among the great powers, concerned only the international aspects of Laos' neutral status; the Laotians were left with the problem of making it work on the local level. The deep political divisions within Laos made it impossible for the country to function effectively as a neutral buffer state.

Relative calm prevailed during 1955–58 as a government of national union, including Pathet Lao representatives, sought to implement the Geneva accords. When elections held in 1958 showed there was significant support for the Pathet Lao, the right reacted by forcing the neutralist prime minister, Prince Souvanna Phouma, out of office and installing their own candidate, Phoui Sananikone. Then in July 1959 the right-wing forces in control of Vientiane imprisoned Prince Souphanouvong (he escaped a year later). In further violation of the Geneva agreement, the rightists accepted military aid from the U.S. and the Philippines. The Pathet Lao retaliated by seizing control of Phbngsali and Houaphan (Sam Neua) provinces, and a period of inconclusive hostilities followed. When Phoui showed some readiness to move toward a neutralist position, he was deposed by rightist army officers and his place was taken by Gen. Phoumi Nosavan. A period of great confusion followed as Phoumi Nosavan and his supporters were themselves deposed in a coup led by a parachute battalion commander, Kong Le, in August 1960, only to succeed in recapturing Vientiane, with considerable U.S. and Thai support, in the following December. By this stage the situation in Laos was a cause for growing tension between the U.S. and the Soviet Union.

Souvanna Phouma coalition

A change in U.S. policy under the new administration of Pres. John F. Kennedy brought acceptance of the concept of a neutral Laos, and in May 1961 the Geneva Conference was reconvened to seek a formula that would allow Laos to occupy a neutral position as a buffer zone. In June 1962 Prince Souvanna Phouma formed a new coalition government that included Pathet Lao, neutralist, and rightist representatives. Despite international support for such a solution, Laos could not disentangle its affairs from the war being fought in neighbouring Vietnam. Nor were the major political groupings within the country ready to give up control over those areas they dominated by means of their armed forces.

Increasingly from 1963, developments in Laos were linked with those in Vietnam. Sections of eastern Laos were a vital part of the Vietnamese Communist supply lines to the south (the Ho Chi Minh Trail), and, as the Vietnamese Communists gave their aid to the Pathet Lao, the U.S. joined in military support of the forces of the government in Vientiane. Despite the external assistance received by both sides, neither the government in Vientiane nor the Pathet Lao was able to break the pattern by which the territory of Laos was divided between them, with little apparent possibility of either gaining a decisive victory. In these circumstances, and with the U.S. and North Vietnam negotiating for a cease-fire in Paris, the Vientiane government and the Pathet Lao began negotiations in 1972 and signed a cease-fire in February 1973.

Beginning in that year, yet another effort was made to find a way to rule Laos through a coalition of rightists, neutralists, and the Pathet Lao. After further protracted negotiations, a Provisional Government of National Unity was inaugurated in April 1974 that had an equal number of members drawn from those who had been associated with the Vientiane government and from the Pathet Lao. Once associated with this provisional government, the Pathet Lao demonstrated a unity and an organizational ability that contrasted with the disunity of their rightist opponents. This enabled the Pathet Lao to take advantage of the opportunities offered by the declining fortunes of the anti-Communist regimes in Saigon and Phnom Penh. When those two cities fell in April 1975, morale dropped sharply among representatives of the right in Laos, and from May onward, as many of the most important rightist politicians and soldiers fled to Thailand, the Pathet Lao moved swiftly and firmly to establish control over the whole of the country.

Lao People's Revolutionary Party

After the Pathet Lao gained control over Laos, the existence of a Laotian Communist Party (the Lao People's Revolutionary Party, founded in 1955 as the People's Party of Laos) was revealed, and it is this party that controls the Lao People's Democratic Republic that was inaugurated in December 1975. Under Prince Souphanouvong as first head of state and Kaysone Phomvihane as first prime minister, the Lao People's Democratic Republic maintains close links with Vietnam, with which in July 1977 it concluded a series of wide-ranging military and economic agreements. Contradictory reports have emerged concerning the policies being followed by the new rulers of Laos, but there seems little doubt that the political, social, and economic programs they are pursuing represent a great break with the past.

BIBLIOGRAPHY. P.B. LAFONT, ***Bibliographie du Laos*** (1964), a listing of articles and books published from 1663 to 1960; ***Annales du Laos*** (1926), a French translation of Lao documents; G. COEDES, ***Les Peuples de la péninsule indochinoise*** (1962); D.G.E. HALL, ***A History of South-East Asia,*** 3rd ed. (1968); P. LE BOULANGER, ***Histoire du Laos français: Essai dune étude chronologique des principautds laotiennes*** (1931), a history of Laos from 1353 to 1930; W.G. BURCHETT, ***Mekong Upstream*** (1957), on the origins of the Pathet Lao movement, and J.J. ZASLOFF, ***The Pathet Lao: Leadership and Organization*** (1973); A.J. DOMMEN, ***Conflict in Laos: The Politics of Neutralization*** (1964); B.F. FALL, ***Anatomy of a Crisis: The Laotian Crisis of 1960–61*** (1969); N.S. ADAMS and A.W. MCCOY, ***Laos: War and Revolution*** (1970); D. LANCASTER, ***The Emancipation of French Indo-China*** (1961); K.P. LANDON, "Thailand's Quarrel with France in Perspective," ***Far Eastern Quarterly,*** 1:25–42 (1941); A. PAVIE *et al.*, ***Mission Pavie: Gdographie et voyages,*** 7 vol. (1900–19); S. PENSRI, ***Les Relations entre la France et la Thailande (Siam) au XIX^e siècle, d'après les archives des Affaires Étrangères*** (1962), on the French penetration; K.D. SASORITH, ***Contribution à l'histoire du mouvement dinddpendance nationale Lao*** (1948), on the Lao Issara movement, by its chief spokesman; H. TOYE, ***Laos: Buffer State or Battleground*** (1968); D.K. WYATT, "Siam and Laos 1767–1827," ***Journal of South East Asian History,*** 4:13–32 (1963).

(P.-B.L./M.E.O)

Lao-tzu

Lao-tzu (in Pinyin romanization Lao-zi), "Master Lao," is the name by which the author of the Lao-tzu, or Tao-te Ching, one of the most famous books of ancient China, is usually known. He is considered the first philosopher of the Taoist school (Tao-chin), which, among the numerous movements of thought in the period of the Waning States (5th–3rd century BC), was the only one, along with the Confucianist school, to survive. The Tao-te Ching is not only important from the philosophical point of view; it is also one of the most sacred scriptures of the Taoist religion, which considered Lao-tzu a great saint and even worshipped him. Furthermore, this little book is by far the most translated work of all the literature of China, and Taoism was one of the major influences in Chinese thought until the Communist revolution (see TAOISM; TAOISM, HISTORY OF; TAOIST LITERATURE).

But if Lao-tzu is an important figure he is also a very obscure one, and the information provided by the historical sources is disappointing.

Ssu-ma Ch'ien's history

The principal document is the biography of Lao-tzu in the Shih-chi ("Historical Records"), by Ssu-ma Ch'ien. This historian, who wrote in about 100 BC, had little solid information concerning the philosopher. He says that Lao-tzu was a native of Ch'ii-jen, a village in the district of Hu in the state of Ch'u, which corresponds to the modern Lu-yi in the east part of Honan Province. His family name was Li, his proper name Erh, his appellation Tan. He was appointed to the office of shih at the royal court of the Chou dynasty (*c.* 1122–221 BC). Shih today means "historian," but in ancient China the shih were scholars specializing in matters such as astrology and divination and were in charge of sacred books. After noting the civil status of Lao-tzu, the historian proceeds to relate a celebrated but questionable meeting of the old Taoist with the younger Confucius (551–479 BC). The story has been much discussed by the scholars; it is men-

By courtesy of the Royal Ontario Museum, Toronto

Lao-tzu (centre), detail from a Taoist temple fresco, southern Shansi, China. Yüan dynasty (AD 1279–1368). In the Royal Ontario Museum, Toronto.

tioned elsewhere, but the sources are so inconsistent and contradictory that the meeting seems a mere legend. During the supposed interview, Lao-tzu blamed Confucius for his pride and ambition, and Confucius was so impressed with Lao-tzu that he compared him to a dragon that rises to the sky, riding on the winds and clouds.

No less legendary is a voyage of Lao-tzu to the west. Realizing that the Chou dynasty was on the decline, the philosopher departed and came to the Hsien-ku pass, which was the entrance to the state of Ch'in. Yin Hsi, the legendary guardian of the pass (*kuan-ling*), begged him to write a book for him. Thereupon, Lao-tzu wrote a book in two sections of 5,000 characters, in which he set down his ideas about the Tao (literally "Way," the Supreme Principle) and the *te* (its "virtue"): the *Tao-te Ching*. Then he left and "nobody knows what has become of him," says Ssu-ma Ch'ien.

After the account of the voyage of Lao-tzu and of the redaction of the book, Ssu-ma Ch'ien alludes to other men with whom Lao-tzu was sometimes identified. One was Lao-Lai-tzu, a Taoist contemporary of Confucius; another was a great astrologer named Tan, who had an interview with Duke Hsien of Ch'in (384–362). Ssu-ma Ch'ien adds, "Maybe Lao-tzu has lived one hundred and fifty years, some say more than two hundred years." Since the ancient Chinese believed that superior men could live very long, it is natural that the Taoists credited their master with an uncommon longevity, but this is perhaps a rather late tradition because Chuang-tzu, the Taoist sage of the 4th century BC, still speaks of the death of Lao-tzu without emphasizing an unusual longevity.

To explain why the life of Lao-tzu is so shrouded in obscurity, Ssu-ma Ch'ien says that he was a gentleman recluse whose doctrine consisted in nonaction, the cultivation of a state of inner calm, and purity of mind. Indeed, throughout the whole history of China, there have always been recluses who shunned worldly life. The author (or authors) of the *Tao-te Ching* was probably a person of this kind who left no trace of his life.

The question of whether there was a historical Lao-tzu has been raised by many scholars, but it is rather an idle one. The *Tao-te Ching*, as we have it, cannot be the work of a single man; some of its sayings may date from the time of Confucius; others are certainly later; and the book as a whole dates from about 300 BC. Owing to these facts, some scholars have assigned the authorship of the *Lao-tzu* to the *t'ai-shih* (astrologer) Tan; while others, giving credit to a genealogy of the descendants of the philosopher, which is related in the biography by Ssu-ma Ch'ien, try to place the life of Lao Tan at the end of the 4th century BC. But this genealogy can hardly be considered as historical. It proves only that at the time of Ssu-ma Ch'ien a certain Li family (see above) pretended to be descended from the Taoist sage; it does not give a basis for ascertaining the existence of the latter. The name Lao-tzu seems to represent a certain type of sage rather than an individual.

Hagio-graphical legends

Beyond the biography in the *Shih-chi* and sporadic mentions in other old books, several hagiographies were written from the 2nd century AD onward. These are interesting for the history of the formation of religious Taoism (Tao-chiao). During the Later Han dynasty (AD 23–220), Lao-tzu had already become a mythical figure who was worshipped by the people and occasionally by an emperor. Later, in religious circles, he became the Lord Lao (Lao-chiin), revealer of sacred texts and saviour of mankind. There were several stories about his birth, one of which was influenced by the legend of the miraculous birth of Buddha. Lao-tzu's mother is said to have borne him 72 years in her womb, and he to have entered the world through her left flank. One legend gives an explanation of his family name, Li: the baby came to light at the foot of a plum tree (*li*) and decided that *li* (plum) should be his surname. Two legends were particularly important in the creed of the Taoists. According to the first, the Lao-chiin was believed to have adopted different personalities throughout history and to have come down to the earth several times to instruct the rulers in the Taoist doctrine. The second developed from the story of Lao-tzu's voyage to the west. In this account the Buddha was thought to be none other than Lao-tzu himself. In the 3rd century AD, an apocryphal book was fabricated on this theme with a view to combatting Buddhist propaganda. This book, the *Lao-tzu Hua-hu ching* ("Lao-tzu's Conversion of the Barbarians"), in which Buddhism was presented as an inferior kind of Taoism, was often condemned by the Imperial authorities.

The figure of Lao-tzu has never ceased to be generally respected in all circles. To the Confucianists he was a venerated philosopher; to the people, he was a saint or a god; and to the Taoists he was an emanation of the Tao and one of their greatest divinities. Under the T'ang dynasty (618–907), whose rulers belonged to the Li clan, he was worshipped as an Imperial ancestor.

BIBLIOGRAPHY. The texts of the *Shih-chi* concerning Lao-tzu have been translated and thoroughly discussed in a debate between H.H. DUBS and D. BODDE in the *Journal of the American Oriental Society*, vol. 61, 62, and 64 (1941, 1942, and 1944). For less technical accounts of the question, see ARTHUR WALEY, *Lao-tzu: The Way and the Power* (1934, reprinted 1956); H. WELCH, *Taoism: The Parting of the Way*, rev. ed. (1966); and M. KALTENMARK, *Lao tseu et le taoïsme* (1965; Eng. trans., *Lao Tzu and Taoism*, 1969). On the legends, see H. DORE, "Lao-tsé et le taoïsme," *Recherches sur le superstitions en Chine*, vol. 18 (1938), not very scientific, but gives much information; and ANNA K. SEIDEL, "The Image of the Perfect Ruler in Early Taoist Messianism: Lao-tzu and Li Hung," *History of Religion*, 9:216–247 (1969–70); and *La Divinisation de Lao tseu dans le taoïsme des Han* (1969), two studies that bring together much new material. A most complete bibliography by M. SOYMIE and F. LITSCH has been published in *Etudes taoïstes*, 3:18–40 (1968).

(Ma.K.)

Laplace, Pierre-Simon, Marquis de

His dynamical investigations into the stability of the solar system earned for mathematician Pierre-Simon Laplace the title the Newton of France. He successfully applied the Newtonian theory of gravitation to the solar system by accounting for all of the observed deviations of the planets from their theoretical orbits and developed a conceptual view of evolutionary change in the physical universe. Moreover, he demonstrated the usefulness of the probabilistic interpretation of scientific data.

Laplace was born on March 23, 1749, at Beaumont-en-Auge, Normandy, where his father was a peasant farmer. Little is known of his early life except that he quickly showed his mathematical ability at the military academy at Beaumont. At 18 he left his humble surroundings for Paris, determined to make his way in mathematics. He then composed a letter on principles of mechanics for the mathematician Jean d'Alembert, who recommended him to a professorship at the École Militaire.

Major work

In 1773 he began his major lifework—applying Newtonian gravitation to the entire solar system—by taking up a particularly troublesome problem: why Jupiter's orbit appeared to be continuously shrinking while Saturn's continually expanded. The mutual gravitational interactions within the solar system were so complex that mathematical solution seemed impossible; indeed, Newton had concluded that divine intervention was periodically required to preserve the system in equilibrium. Laplace announced the invariability of planetary mean motions, carrying his proof to the cubes of the eccentricities and inclinations. This discovery in 1773, the first and most important step in establishing the stability of the solar system, was the most important advance in physical astronomy since Newton. It won him associate membership in the Academy of Sciences the same year.

Applying quantitative methods to a comparison of living and nonliving systems, Laplace and the chemist Antoine Lavoisier in 1780, with the aid of an ice calorimeter that they had invented, showed respiration to be a form of combustion. Returning to his astronomical investigations with an examination of the entire subject of planetary perturbations—mutual gravitational effects—Laplace in 1786 proved that the eccentricities and inclinations of planetary orbits to each other will always remain small, constant, and self-correcting. The effects of perturbations were therefore conservative and periodic, not cumulative and disruptive. The opposite and secular inequalities of Jupiter and Saturn (accelerations and deceleration, respectively), for example, were due to a changing effect with a period of *929* years. Their inequalities were therefore not cumulative, but periodic.

Turning to the subject of the attraction between spheroids, Laplace in 1784–85 proved that the theorem concerning spheroids of revolution is true for any spheroids with common focuses and explored the problem of the attraction of any spheroid upon a particle situated outside or upon its surface. Through his discovery that the attractive force of a mass upon a particle, regardless of direction, could be obtained directly by differentiating a single function, Laplace laid the mathematical foundation for the scientific study of heat, magnetism, and electricity.

Laplace removed the last apparent anomaly from the theoretical description of the solar system in *1787* with the announcement that lunar acceleration depends on the eccentricity of the Earth's orbit. Although the mean motion (average angular velocity) of the Moon around the Earth depends mainly on the gravitational attraction between them. it is slightly diminished by the pull of the Sun on the Moon. This solar action depends, however, on changes in the eccentricity of the Earth's orbit resulting from perturbations by the other planets. As a result, the Moon's mean motion is accelerated so long as the Earth's orbit tends to become more circular; but, when the reverse occurs, this motion is retarded. The inequality is therefore not truly cumulative, Laplace concluded, but is of a period running into millions of years. The last threat of instability thus disappeared from the theoretical description of the solar system.

Traité de mécanique céleste and other publications

Laplace's *Exposition du système du monde,* a semipopular treatment of his work in celestial mechanics and a model of French prose published in *1796,* included his "nebular hypothesis"-cooling and contracting of a gaseous nebula—which strongly influenced future thought on planetary origin. His *Traitk de Mkcanique ckleste (Celestial Mechanics,* 1966), appearing in five volumes between *1798* and *1827,* summarized the results obtained by his mathematical development and application of the law of gravitation. He offered a complete mechanical interpretation of the solar system by devising methods for calculating the motions of the planets and their satellites and their perturbations, including the resolution of tidal problems. It made him a celebrity.

In 1814 Laplace published a popular work for the general reader, *Essai philosophique sur les probabilités.* This work was the introduction to the second edition of his comprehensive and important *Thkorie analytique des probabilités* ("Analytic Theory of Probability"), first published in *1812,* in which he described many of the tools he invented for mathematically predicting the probabilities that particular events will occur in nature. He applied his theory not only to the ordinary problems of chance but also to the inquiry into the causes of phenomena, vital statistics, and future events, while emphasizing its importance for physics and astronomy.

Probably because he did not hold strong political views he escaped imprisonment and execution during the Revolution. Laplace was president of the Bureau des Longitudes (Board of Longitude), aided in the organization of the decimal system, helped found the Société d'Arcueil, a scientific society, and was created a marquis. He served for six weeks as Minister of the Interior under Napoleon, who thought his record as an administrator was undistinguished. He died in Paris on March *5, 1827.*

BIBLIOGRAPHY. There is no definitive biography of Laplace, but the best account of his work is by SIR EDMUND T. WHITTAKER in *The Mathematical Gazette,* 33:1–12 *(1949);* ERIC T. BELL, *Men of Mathematics (1937,* reprinted 1961), also contains a lively biographical essay; see also H. ANDOYER, *L'Oeuvre scientifique de Laplace (1922).* The original documents relating to his life have almost all perished, most of them in a fire in *1925* that destroyed the home of his great-great-grandson, the Comte de Colbert-Laplace, and some during World War II in the bombardment of Caen. An important source of information is M.P. CROSLAND, *The Society of Arcueil: A View of French Science at the Time of Napoleon I (1967).* Laplace's collected works were published in Paris by Gauthier-Villars in 15 vol. *(1878–1912).* Translations of Laplace's works include *A Philosophical Essay on Probabilities (1902)* and *The System of the World, 2* vol. *(1830);* both are classics. The latter is a semipopular treatment of his *Celestial Mechanics,* 4 vol., trans. by NATHANIEL BOWDITCH, JERZY NEYMAN, and LUCIEN M. LE CAM (1828–*39,* reprinted *1966).*

(G.J.W.)

Laptev Sea

A marginal sea of the Arctic Ocean, the Laptev (Laptevykh More in Russian) takes the form of a wide bay situated between the shore of Siberia, the Taymyr Peninsula (Poluostrov), the islands of Severnaya Zemlya, and the New Siberian and Kotelny islands. It is connected in the west with the Kara Sea by the straits of Vilkits, Shokal, and Red Army and in the east with the East Siberian Sea by the Dmitry Laptev (Dmitriya Lapteva), Eterikan, and Sannikova straits. The Laptev Sea (formerly called the Siberian Sea) was named in *1935* in honour of Khariton and Dmitry Laptev, the brothers who were the first to map its shores (1735–40).

Physical characteristics. The sea's area is about 276,*000* square miles *(714,000* square kilometres), the volume of water *97,000* cubic miles, the average depth *1,896* feet, and the greatest depth *9,774* feet. The greatest length from southeast to northwest is *800* miles. The largest bays—Khatangsky, Olenyoksky, Buorkhaya, and others—cut into the low-lying, gently sloping shore. Numerous rivers, the largest being the Lena, Khatanga, Yana, Olenyok, and Anabar, flow into it, carrying about *150* cubic miles of freshwater, of which the Lena accounts for *77* percent. Several rivers form extensive deltas. There are several dozen islands with a total area of *1,461* square miles, primarily in the western part of the sea. The islands vary in landscape and origin. The steep, hilly shores of Severnaya Zemlya descend in glaciers to the sea, producing icebergs. The eastern shore of the Taymyr Peninsula is also hilly. The shores of the Peschany, Dunay, and other islands are low-lying and gently sloping. Low but steep shores are characteristic of the New Siberian Islands.

Occurrence of glacial remnants

Ancient rivers and glaciers played an important role in the formation of the relief of the bottom and of the shores of the sea, and even today, in the precipices of the shores of the New Siberian Islands, layers of relic ice of considerable thickness are found. Digging through the layers of ice has revealed numerous remains of animals of the preglacial period, including mammoths. The floor of the sea is a gently sloping plain, breaking off abruptly toward the Arctic Ocean. Depths of less than *160* feet occupy about *53* percent of its area, and depths greater than *3,300* feet, *22* percent. The bottom of the deepwater part is covered with silt, the shallower areas with sand and silt. In the eastern part of the sea, under a thin layer of deposits, occurs a layer of very old "relic" ice. As regards salinity, the thawing of ice and the inflow of fresh river water might have resulted in a layer of freshwater *53* inches thick. In the winter the salinity in the southeastern part of the sea is *20–25* parts per thousand (‰), in the northern part up to 34‰, and in the summer the salinity of the water falls to 5‰–10‰ in the southeast and 30‰–32‰ in the north.

Surface flows form a cyclonic water gyral, or rotation. Tides are primarily semidiurnal; the amplitude on the average is about *1.5* feet, excluding the Khatangsky Bay, where it approaches *6.5* feet. Tidal influence is exerted from the Arctic Basin, the tide range exceeding *8.2* feet.

Climate. The Laptev Sea has one of the most severe climates of the Arctic seas. The polar night (days on which the sun does not rise above the horizon) lasts about three months in the south and five months in the north. Air temperature below *32"* F (0° *C)* is observed in the north of the sea about *11* months and in the south nine months. The average temperature in January is *—24"* to —29° F *(—31"* to *—34" C),* the minimum about *—58"* F *(—50" C).* In July in the northern part the average temperature is just above the freezing point, in the southern part about 43° F (6° C), with a maximum 50° F (10° C). On the shores the maximum temperature can reach 75° F (24° C). The number of days with frost in July in the south is six, in the north 25. In winter there are frequent gales, blizzards, and snowstorms; in the summer, snow squalls and fogs. For most of the year the sea is covered with ice.

Water temperatures

The Laptev Sea is distinguished by its low water temperature. In the winter the temperature of the sub-ice level is 30.6° F (—0.8° C) in the southeastern part of the sea and 28.8° F (—1.8° C) in the northern part; in the deep regions it is from 29.1° to 28.9° F (—1.6° to —1.7° C). In the deepwater part of the sea, warmer Atlantic waters (up to 34.7° F, or 1.5° C) penetrate in a layer 800–1,000 feet (250–300 metres) deep. Beneath this layer the temperature is about 30.6° F (—0.8° C). In the summer, in icefree regions, a thin layer of water warms to 46°–50° F (8°–10° C) in the bays, 36°–37° F (2°–3° C) in the central part, and 32°–34° F (0°–1° C) in the north.

Economic activity. At the mouths of the rivers people earn their living catching salmon and other fish. Mammals found in the Laptev Sea include seal, sea hare, walrus, and polar bear.

The Laptev Sea is on the Northern Sea Route connecting Vladivostok and Arkhangelsk. Timber, building materials, and furs are the main cargoes. The main port is Tiksi. (Y.G.N./A.O.S.)

La Rochefoucauld

La Rochefoucauld became known in the 17th century as the leading exponent of the *maxime*, a peculiarly French type of epigram. The *maxime* utilizes utmost brevity to express, with the effect of shock and surprise, a truth or attitude from which palliation and explanation are absent.

Role in the Fronde

Heritage and political activities. François de La Rochefoucauld was born in Paris on September 15, 1613, the son of François, comte de La Rochefoucauld, and his wife, Gabrielle du Plessis-Liancourt. The son was known until his father's death in 1650 as the Prince de Marcillac. In 1628 he was married to Andrée de Vivonne, by whom he had four sons and three daughters. He served in the army against the Spaniards in Italy in 1629, in the Netherlands and Picardy in 1635–36, and again in Flanders in 1639. The public lives of both father and son were conditioned by the policies of Louis XIV's government, which by turns threatened and flattered the nobility. Though his father was created duke and made governor of Poitou, he was later deprived of that post when the loyalty of the family was called into question. His son was allowed by Cardinal Mazarin, the infant king's chief minister, to resume the governorship in 1646. The fact that his château at Verteuil was demolished by the crown, apparently without notice, in 1650 throws light on a main cause of the series of revolts between 1648 and 1653 known as the Fronde: the distrust and fear felt by the monarchy for the local independence of the nobility.

La Rochefoucauld, anonymous 17th-century portrait in the palace of Versailles.

Lauros–Giraudon

La Rochefoucauld was more vulnerable than most of his contemporaries, because throughout his life he seems to have been susceptible to feminine charm. In 1635 the Duchesse de Chevreuse had lured him into intrigues against the Cardinal de Richelieu, the chief minister of Louis XIII, an adventure that only procured for La Rochefoucauld a humiliating interview with Richelieu, eight days of imprisonment in the Bastille, and two years of exile at Verteuil. Later, his hatred for Mazarin and his devotion to Anne de Bourbon, duchesse de Longueville, sister of the Great Condé, who was the leader of the Fronde, led to an even more disastrous outcome. His own account of the weary alternation of plots and campaigns of the mutinous nobles throughout the revolts (1648–53) may be read in his *Mémoires.* His loyalty to the House of Condé did not increase his popularity with the crown and prevented him from pursuing any single policy for reform of royal or ministerial government. How far toward treason he allowed himself to be led, when the intentions of the reforming princes and nobility were superseded by personal ambitions, is shown by the draft of the so-called Treaty of Madrid of 1651, which laid down conditions of Spanish help to the French nobility. La Rochefoucauld not only signed the treaty but is thought by one scholar to have drafted it.

Two other features of his public career deserve mention, since they explain much of his writing—courage and litigation. The man who was to pen the aphorisms on courage and cowardice had certainly been in the forefront of battle. Within six years he was wounded in no fewer than three engagements. The injuries to his face and throat were such that he retired from the struggle, his health ruined and his peace of mind lost.

Financial and legal difficulties

His financial difficulties were no doubt intensified by war, his lands were heavily mortgaged, and but for the astute help of his agent he might not have been able to keep his establishment in central Paris, as he did from 1660 onward. He was forced to pay not only for fine living but for endless litigation. There is evidence of no fewer than five lawsuits in the space of three years, chiefly against other noble families, over questions of precedence and court ceremonial.

Yet in 1655 his literary endeavours were still before him. Thanks to the lasting and intellectually stimulating friendships with Mme de Sablé, one of the most remarkable women of her age, and Mme de Lafayette, he seems to have avoided politics for a while and gradually won his way back into royal favour, a feat sealed by his promotion to the knightly order of the Saint-Esprit at the end of 1661. Reading and intellectual conversation occupied his time as well as that of other men and women of a circle who listened to private readings of Pierre Corneille's classical tragedies and Nicolas Boileau's didactic poem on the principles of poetic composition, *L'Art poétique.* The circle was enlivened by a new game that consisted of discussing epigrams on manners and behaviour, expressed in the briefest, most pungent manner possible. The care with which La Rochefoucauld kept notes and versions of his thoughts on the moral and intellectual subjects of the game is clear from the surviving manuscripts. When the clandestine publication of one of them in Holland forced him to publish under his own name, it was clear that he had satisfied public taste: five editions of the *Maximes,* each of them revised and enlarged, were to appear within his lifetime.

The Maximes. The first edition of the *Maximes,* published in 1665, was called *Réflexions ou sentences et maximes morales* and did not contain epigrams exclusively; the most eloquent single item, which appeared only in the first edition and was thereafter removed by the author, is a three-page poetic description of self-interest, a quality he found in all forms of life and in all actions. The manuscripts also contain epigrams embedded in longer reflections; in some cases the various versions show the steps by which a series of connected sentences was filed down to the point of ultimate brevity. Beneath the general single statement, however, can be found a personal reaction to the Fronde, or to politics, often violent in its expression. For example

> ***Les crimes deviennent innocents, même glorieux, par leur nombre et par leurs qualités; de là vient que les voleries publiques sont des habiletés, et que prendre des provinces injustement s'appelle faire des conquêtes. Le crime a ses héros, ainsi que la vertu.*** (Crimes are made innocent and even virtuous by their number and nature; hence public robbery becomes a skillful achievement and wrongful seizure of a province is called conquest. Crime has its heroes no less than virtue has.)

It may have been hostile reception or the fear of revealing a political attitude that made him abandon this kind of epigram except for the almost unrecognizable No. 185: *"Il y a des hkros en mal comme en bien"* ("Evil as well as good has its heroes"). Modern readers forget that La Rochefoucauld's contemporaries would read recent history into statements that appear cryptic and opaque to posterity.

The Fronde was to La Rochefoucauld one of those moments of history that seemed to reveal men's motives at their worst. His exposure of the self-seeking that lay beneath conventional homage to morality has earned for him the reputation of a cynic, but his keener contemporaries are no less severe. The pungency and absence of explanation make his epigrams seem more scornful than similar statements embedded in memoirs. But La Rochefoucauld was concerned to convey something more than scorn, and beneath his professions of idealism he pinpointed a restless and unquenchable thirst for self-preservation. Virtue in the pure state was something he did not find:

> ***Les vertus se perdent dans l'intérêt comme les fleuves se perdent dans la mer*** (Virtues are lost in self-interest as rivers are lost in the sea).

This image of the sea recurred:

> ***Voilà la peinture de l'amour-propre, dont toute la vie n'est qu'une grande et longue agitation; la mer en est une image sensible; et l'amour-propre trouve dans le flux et reflux de ses vagues continuelles une fidèle expression de la succession turbulente de ses pensées et de ses éternels mouvements*** (Such is the picture of self-love, of which all life is one continuous and immense ferment. The sea is its visible counterpart and self-love finds in the ebb and flow of the sea's endless waves a true likeness of the chaotic sequence of its thoughts and of its everlasting motion).

Expression of moral and political attitudes

La Rochefoucauld has been called an Epicurean but his imaginative insights attached him to no doctrine. Like Michel de Montaigne and Blaise Pascal, he was aware of the mystery around man that dwarfs his efforts and mocks his knowledge, of the many things about man of which he knows nothing, of the gap between thinking and being, between what man is and what man does: ***"La nature fait le mérite et la fortune le met en oeuvre"*** ("Nature gives us our good qualities and chance sets them to work"). Some epigrams show a respect for the power of indolence, and others reveal an almost Nietzschean respect for strength. All these insights seem common to the French classical school of which he is so brilliant a member—though as an aristocrat he disdained being called a writer. These insights also accounted for his fame and influence on his disciples: in England Lord Chesterfield, the orator and man of letters, and the novelist and poet Thomas Hardy; in Germany the philosophers Friedrich Nietzsche and Georg Christoph Lichtenberg; in France the writers and critics Stendhal, Charles-Augustin Sainte-Beuve, and Andre Gide.

Artistry of the *Maxirnes*

Yet his chief glory perhaps is not as thinker but as artist. In the variety and subtlety of his arrangement of words he made the *maxime* into a jewel. It is not always the truth of the maxim that is so striking, but its exaggeration which can surprise one into a new aspect of the truth. He describes and defines—he has no time for more—but of the single metallic image he makes amazing use. He handles paradox to such effect that a final word can reverse the rest:

> ***On ne donne rien si libéralement que ses conseils*** (We give nothing so generously as . . . advice). ***C'est une grande folie de vouloir être sage tout seul*** (It is great folly to seek to be wise . . . on one's own).

La Rochefoucauld authorized five editions of the ***Maximes*** from 1665 to 1678. Two years after the last publication, he died in Paris during the night of March 16–17, 1680.

Though he did a considerable amount of writing over the years La Rochefoucauld actually published only two works, the *Mémoires* and the ***Maximes.*** In addition, about 150 letters have been collected and 19 shorter pieces now known as ***Réflexions diverses.*** These, with the treaties and conventions that he may have drawn up personally, constitute his entire work and of these only the ***Maximes*** stand out as a work of genius. Like his younger contemporary, Jean de La Bruyère, La Rochefoucauld was a man of one book.

BIBLIOGRAPHY. Three recent editions of La Rochefoucauld's writings in the "Pléiade Series" (1947, 1957, and 1962), give an idea of the movement of scholarship. Still indispensable, in spite of their date, are the *Oeuvres . . . ,* ed. by D.L. GILBERT, 4 vol. (1868–83); and E. MAGNE, ***Le Vrai visage de La Rochefoucauld*** (1923). For a general treatment, see W.G. MOORE, ***La Rochefoucauld: His Mind and Art*** (1969). Two good working editions of the ***Maximes*** are those by J. TRUCHET (1969), and by D. SECRETAN (1967). A mass of recent evidence on the background has been published by R. MOUSNIER in ***Lettres et mémoires adressés au chancelier Se'guier, 2*** vol. (1964).

(W.G.Mo.)

La Salle, Robert Cavelier, sieur de

René-Robert Cavelier, sieur de La Salle, a French explorer in North America, attempted not only to revolutionize the fur trade by the use of sailing vessels on the Great Lakes and the Mississippi, but also to found an empire for France in the West. His plans were too vast for achievement with the resources available to him, but by leading the first expedition to travel from the lakes of Canada to the Gulf of Mexico, La Salle made a prominent place for himself in the history of North American exploration.

La Salle, engraving by an unknown artist.
BY courtesy of the Bibliotheque Municipale, Rouen, France; photograph, Ellebe

He was born at Rouen, Normandy, on November *22,* 1643, and was educated at a Jesuit college. La Salle first studied for the priesthood, but at the age of 22 he found himself more attracted to adventure and exploration and in 1666 set out for Canada to seek his fortune. With a grant of land at the western end of Île de Montreal, La Salle acquired at one stroke the status of a seigneur (*i.e.,* landholder), and the opportunities of a frontiersman.

Land grant in Montreal

The young landlord farmed his land near the Lachine Rapids, and, at the same time, set up a fur-trading outpost. Through contact with the Indians who came to sell their pelts, he learned various Indian dialects and heard stories of the lands beyond the settlements. He soon became obsessed with the idea of finding a way to the Orient through the rivers and lakes of the Western frontier.

If experience modified the visions of the dreamer, it enhanced the knowledge and skill of the pathfinder and trader. Having sold his land, La Salle set out in 1669 to explore the Ohio region. His discovery of the Ohio River, however, is not accepted by modern historians.

La Salle found a kindred spirit in Comte de Frontenac, the "Fighting Governor" of New France (the French possessions in Canada) from 1672 to 1682. Together, they pursued a policy of extending French military power by establishing a fort on Lake Ontario (Fort-Frontenac), holding the Iroquois in check, and intercepting the fur trade between the Upper Lakes and the Dutch and English coastal settlements.

Their plans were strongly opposed by the Montreal merchants, who feared the loss of their trade, and by the missionaries (especially the Jesuits), who were afraid of losing their influence over the Indians of the interior. Nevertheless, Fort-Frontenac was built where Kingston now stands, and La Salle was installed there as seigneur in 1675 after a visit to the French court, as Frontenac's representative. The governor had recommended him as "a man of intelligence and ability, more capable than anybody else I know here to accomplish every kind of enterprise and discovery. . . ." Louis XIV was sufficiently impressed by him to grant him a title of nobility.

At Fort-Frontenac, La Salle had control of a large share of the fur trade, and his affairs prospered. But his restless ambition drove him to seek greater ends. On another visit to France in 1677 he obtained from the King authority to explore "the western parts of New France" and permission to build as many forts as he wished, as well as to hold a valuable monopoly of the trade in buffalo hides.

Since the project had to be carried out at his own expense, however, he borrowed large sums in both Paris and Montreal, and he began to be enmeshed in a tangle of debts that was to blight all of his later enterprises. La Salle's proposals also roused still further the enmity of the Jesuits, who resolutely opposed all his schemes.

Preparations for the expedition to the Mississippi

When he returned to Canada in 1678, La Salle was accompanied by an Italian soldier of fortune, Henri de Tonty, who became his most loyal friend and ally. Early in the following year, he built the "Griffon," the first commercial sailing vessel on Lake Erie, which he hoped would pay for an expedition into the interior as far as the Mississippi. From the Seneca Indians above the Niagara Falls he learned how to make long journeys overland, on foot in any season, subsisting on game and a small bag of corn. His trek from Niagara to Fort-Frontenac in the dead of winter won the admiration of a normally critical member of his expeditions, the friar Louis Hennepin.

La Salle's great scheme of carrying cargo in sailing vessels like the "Griffon" on the lakes and down the Mississippi was frustrated by the wreck of that ship and by the destruction and desertion of Fort-Crevecoeur on the Illinois River, where a second ship was being built in 1680. Proud and unyielding by nature, La Salle tried to bend others to his will and often demanded too much of them, though he was no less hard on himself. After several disappointments, he at last reached the junction of the Illinois with the Mississippi and saw for the first time the river he had dreamed of for so long. But he had to deny himself the chance to explore it. Hearing that Tonty and his party were in danger, he turned back to aid them.

After many vicissitudes, La Salle and Tonty succeeded in canoeing down the Mississippi and reached the Gulf of Mexico. There, on April 9, 1682, the explorer proclaimed the whole Mississippi Basin for France and named it Louisiana. In name, at least, he acquired for France the most fertile half of the North American continent.

The following year La Salle built Fort-Saint-Louis at Starved Rock on the Illinois River (now a state park), and here he organized a colony of several thousand Indians. To maintain the new colony he sought help from Quebec; but Frontenac had been replaced by a governor hostile to La Salle's interests, and, instead of help, he received orders to surrender Fort-Saint-Louis. He refused to comply and left North America to appeal directly to the king. Welcomed in Paris, La Salle was given an audience with Louis XIV who favoured him by commanding the governor to make full restitution of La Salle's property.

Expedition to Mexico

The last phase of his extraordinary career centred on his proposal to fortify the mouth of the Mississippi and to invade and conquer part of the Spanish province of Mexico. He planned to accomplish all this with some 200 Frenchmen, aided by buccaneers and an army of 15,000 Indians—a venture that caused his detractors to question his sanity. But the king saw a chance to harass the Spaniards, with whom he was at war, and approved the project, giving La Salle men, ships, and money.

The expedition was doomed from the start. It had hardly left France when quarrels arose between La Salle and the naval commander. Vessels were lost by piracy and shipwreck, while sickness took a heavy toll of the colonists. Finally, a gross miscalculation brought the ships to Matagorda Bay in Texas, 500 miles west of their intended landfall. After several fruitless journeys in search of his lost Mississippi, La Salle met his death on March 19, 1687, at the hands of mutineers near the Brazos River. His vision of a French empire died with him.

La Salle provoked much controversy both in his own lifetime and later. Those who knew him best praised his ability unsparingly. He was considered "one of the greatest men of the age" by Tonty, who, like Frontenac, was among the very few who were able to understand the proud spirit of the dour Norman. Henri Joutel, who served under La Salle through the tragic days of the Texas colony until his death, wrote both of his fine qualities and of his insufferable arrogance toward his subordinates. In Joutel's view, this arrogance was the true cause of La Salle's death.

Undoubtedly, La Salle was hampered by faults of character and lacked the qualities of leadership. On the other hand, he possessed prodigious vision, tenacity, and courage. His claim of Louisiana for France, though but a vain boast at the time, pointed the way to the French colonial empire that was eventually built by other men.

BIBLIOGRAPHY. LOUISE P. KELLOGG, *Early Narratives of the North-West, 1634–1699* (1917), contains translations of original narratives of La Salle's companions; FRANCIS PARKMAN, *The Discovery of the Great West*, 5th ed. (1871, reissued 1956), embodies material from La Salle's own letters and other contemporary documents; LEON LEMONNIER, *Cavelier de la Salle et l'exploration du Mississippi* (1942), gives a balanced, modern French view of the subject; JOHN B. BREBNER, *The Explorers of North America, 1492–1806* (1964), summarizes impartially various estimates of La Salle's character and achievements.

(D.C.G.S.)

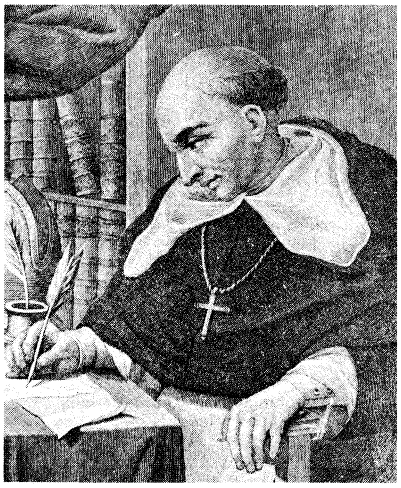
Las Casas, Bartolome de

Bartolomé de las Casas, an early-16th-century Spanish theologian and missionary in the Americas, was the first to expose the oppression of the Indian by the European and to call for the abolition of Indian oppression. A prolific writer and in his later years an influential figure of the Spanish court, he yet failed to stay the progressive enslavement of the indigenous races of Latin America. The son of a small merchant, Las Casas was probably born in Seville in August 1474. He is believed to have gone to Granada as a soldier in 1497 and to have enrolled to study Latin in the academy at the cathedral in Seville.

Early years in America

In 1502 he left for Hispaniola, in the West Indies, with the governor, Nicolás de Ovando. As a reward for his participation in various expeditions, he was given an *encomienda* (a royal land grant including Indian inhabitants), and he soon began to evangelize the Indians, serving as *doctrinero,* or lay teacher of catechism. Perhaps the first person in America to receive holy orders, he was ordained priest in either 1512 or 1513. In 1513 he took part in the bloody conquest of Cuba and, as priest-*encomendero* (land grantee), received an allotment of Indian serfs.

Although during his first 12 years in America Las Casas was a willing participant in the conquest of the Caribbean, he did not indefinitely remain indifferent to the fate of the natives. In a famous sermon on August 15, 1514, he announced that he was returning his Indian serfs to the Governor. Realizing that it was useless to attempt to defend the Indians at long distance in America, he returned to Spain in 1515 to plead for their better treatment. The most influential person to take up his cause was Francisco Jiménez de Cisneros, the archbishop of



Las Casas, engraving by an unknown artist.
By courtesy of the Organization cf American States

Toledo and future co-regent of Spain. With the help of the Archbishop, the *Plan para la Reformación de Ins Indias* was conceived, and Las Casas, named priest-procurator of the Indies, was appointed to a commission to investigate the status of the Indians. He sailed for America in November 1516.

Las Casas returned to Spain the next year. In addition to studying the juridical problems of the Indies, he began to work out a plan for their peaceful colonization by recruiting farmers as colonists. His stirring defense of the Indians before the Spanish Parliament in Barcelona in December 1519 persuaded King Charles V, who was in attendance, to accept Las Casas' project of founding "towns of free Indians"—*i.e.*, communities of both Spaniards and Indians who would jointly create a new civilization in America. The location selected for the new colony was on the Gulf of Paria in the northern part of present-day Venezuela. Las Casas and a group of farm labourers departed for America in December 1520. The failure to recruit a sufficient number of farmers, the opposition of the *encomenderos* of Santo Domingo, and, finally, an attack by the Indians themselves all were factors that brought disaster to the experiment in January 1522.

Historia de las Indias

Upon his return to Santo Domingo, the unsuccessful priest and political reformer abandoned his reforming activities to take refuge in religious life; he joined the Dominican order in 1523. Four years later, while serving as prior of the convent of Puerto de Plata, a town in northern Santo Domingo, he began to write the ***Historia Apologe'tica.*** One of his major works, the *Apologética* was to serve as the introduction to his masterpiece, the *Historia **de las Indias.*** The ***Historia,*** which by his request was not published until after his death, is an account of all that had happened in the Indies just as he had seen or heard of it. But, rather than a chronicle, it is a prophetic interpretation of events. The purpose of all the facts he sets forth is the exposure of the "sin" of domination, oppression, and injustice that the European was inflicting upon the newly discovered colonial peoples. It was Las Casas' intention to reveal to Spain the reason for the misfortune that would inevitably befall her when she became the object of God's punishment.

He interrupted work on the book only to send to the Council of the Indies in Madrid three long letters (in 1531, 1534, and 1535), in which he accused persons and institutions of the sin of oppressing the Indian, particularly through the ***encomienda*** system. After various adventures in Central America, where his ideas on the treatment of the natives invariably brought him into conflict with the Spanish authorities, Las Casas wrote ***De único modo*** (1537; "Concerning the Only Way of Drawing All Peoples to the True Religion"), in which he set forth the doctrine of peaceful evangelization of the Indian. Together with the Dominicans, he then employed this new type of evangelization in a "land of war" (a territory of still-unconquered Indians)—Tuzutlan, near the Golfo Dulce (Sweet Gulf) in present-day Costa Rica. Encouraged by the favourable outcome of this experiment, Las Casas set out for Spain late in 1539, arriving there in 1540.

While awaiting an audience with Charles V, Las Casas conceived the idea of still another work, the ***Brevísima relación de la Destruccidn de las Zndias*** ("A Brief Report on the Destruction of the Indians"), which he wrote in 1542 and in which the historical events described are in themselves of less importance than their theological interpretation: "The reason why the Christians have killed and destroyed such an infinite number of souls is that they have been moved by their wish for gold and their desire to enrich themselves in a very short time." (***Destruccidn,*** Page 36).

Las Casas and the New Laws

Las Casas' work finally seemed to be crowned with success when King Charles signed the so-called New Laws (Leyes Nuevas). According to these laws, the ***encomienda*** was not to be considered a hereditary grant; instead, the owners had to set free their Indians after the span of a single generation. To ensure enforcement of the laws, Las Casas was named bishop of Chiapas in Guatemala, and in July 1544 he set sail for America, together with 44 Dominicans. Upon his arrival in January 1545, he immediately issued *Avisos* y *reglas **para confesores de españoles*** ("Admonitions and Regulations for the Confessors of Spaniards"), the famous ***Confesionario,*** in which he forbade absolution to be given to those who held Indians in *encomienda.* The rigorous enforcement of his regulations led to vehement opposition on the part of the Spanish faithful during Lent of 1545 and forced Las Casas to establish a council of bishops to assist him in his task. But soon his uncompromisingly pro-Indian position alienated his colleagues, and in 1547 he returned to Spain.

Las Casas then entered upon the most fruitful period of his life. He became an influential figure at court and at the Council of the Indies. In addition to writing numerous *memoriales* (petitions), he came into direct confrontation with the learned Juan Ginés de Sepúlveda, an increasingly important figure at court by reason of his *Democrates II* ("Concerning the Just Cause of the War Against the Indians"), in which he maintained, theoretically in accordance with Aristotelian principles, that the Indians "are inferior to the Spaniards just as children are to adults, women to men, and, indeed, one might even say, as apes are to men." Las Casas finally confronted him in 1550 at the Council of Valladolid, which was presided over by famous theologians. The argument was continued in 1551, and its repercussions were enormous.

The servitude of the Indians nevertheless was already irreversibly established, and, despite the fact that Sepúlveda's teachings had not been officially approved, they were, in effect, those that were followed in the Indies. But Las Casas continued to write books, tracts, and petitions, testimony to his unwavering determination to leave in written form his principal arguments in defense of the American Indian.

During his final years Las Casas came to be the indispensable adviser to both the Council of the Indies and to the King on many of the problems relating to the Indies. In 1562 he had the final form of the *Prólogo* to the ***Historia de las Zndias*** published, although in 1559 he had left written instructions that the work itself should be published only "after forty years have passed, so that, if God determines to destroy Spain, it may be seen that it is because of the destruction that we have wrought in the Indies and His just reason for it may be clearly evident." At the age of 90 Las Casas completed two more works on the Spanish conquest in the Americas. Two years later, on July 17, 1566, he died in the Dominican convent of Nuestra Seiiora de Atocha de Madrid, having continued to the end his defense of his beloved Indians, oppressed by the colonial system that Europe was organizing.

Assessment

At the suggestion of Francisco de Toledo, the viceroy of Peru, the King ordered all the works, both published and unpublished, of Las Casas to be collected. Although his influence with Spain and the Indies declined sharply, his name became well-known in other parts of Europe, thanks to the translations of the ***Destruccidn*** that soon appeared in various countries. In the early 19th century, the Latin American revolutionary Simón Bolivar himself was inspired by some of the letters of Las Casas in his struggle against Spain, as were some of the heroes of Mexican independence. His name has come into prominence again in the latter half of the 20th century, in connection with the so-called *Indigenistas* movements in Peru and Mexico. The modern significance of La Casas lies in the fact that he was the first European to perceive the economic, political, and cultural injustice of the colonial or neocolonial system maintained by the North Atlantic powers since the 15th century for the control of Latin America, Africa, and Asia.

BIBLIOGRAPHY. The most complete bibliography on Las Casas is *Bartolomé de las Casas (1474–1566) . . .*, compiled by LEWIS HANKE and MANUEL GIMENEZ FERNANDEZ (1954). The most important work to date, despite the fact that the author could not complete it, is MANUEL GIMENEZ FERNANDEZ, *Bartolomé de las Casas:* vol. 1, *Delegado de Cisneros para la reformación de las Indias. 1516–1517* (1953), and vol. 2, *Capellán de S.M. Carlos I, poblador de Cumaná, 1517–1523* (1960). A critical polemic work is that of RAMON MENENDEZ PIDAL, El *Padre de las Casas: su doble personalidad* (1963). For English readers the four studies of LEWIS HANKE supersede earlier works: *Bartolomé de las Casas: An Interpretation of His Life and Writings* (1951) ; *. . . Bookman, Scholar and Propagandist* (1952); *. . . Historian* (1952); and *Aristotle and the American Indians* (1959). The most complete edition of Las Casas' works is JUAN ANTONIO LLORENTE (ed.), *Colección de las obras del venerable obispo de Chiapas don Bartolomé de las Casas, defensor de la libertad de los americanos* (1822).

(E.Du.)

Laser and Maser

Lasers and masers are devices that produce a unique kind of radiation. Lasers produce an intense beam of light of a very pure single colour. Masers produce similar radiation, but in the microwave (radio) part of the electromagnetic spectrum. The names are acronyms derived from Microwave (or Molecular) Amplification by Stimulated Emission of Radiation (MASER) and Light Amplification by Stimulated Emission of Radiation (LASER). The difference between a maser and a laser is only that each operates in a different part of the spectrum, the maser in the radio spectrum and the laser in the light spectrum. Though the maser was invented first, the laser has proven much more useful.

Lasers represent the fulfillment of one of mankind's oldest dreams of technology, that of providing a light beam intense enough to vaporize the hardest and most heat-resistant materials. Lasers have been used to drill holes in diamonds for wire-drawing dies, to weld the retina of an eye to its supports to prevent detachment, and to perform microsurgery on parts of single cells.

Laser and maser principles. Atoms and molecules exist at low and high energy levels. Those at low levels can be excited to higher levels, usually by heat, and on reaching the higher levels they give off light when they return to a lower level. In ordinary light sources the many excited atoms or molecules emit light independently and in many different colours (wavelengths). If, however, during the brief instant that an atom is excited, light of a certain wavelength impinges on it, the atom can be stimulated to emit radiation that is in phase (that is, in step) with the wave that stimulated it. The new emission thus augments or amplifies the passing wave; if the phenomenon can be multiplied sufficiently, the resulting beam, made up of wholly coherent light (that is, light of a single frequency or colour in which all the components are in step with each other) will be tremendously powerful.

Einstein recognized the existence of stimulated emission in 1917, but not until the 1950s were ways found to use it in devices. U.S. physicist Charles H. Townes and colleagues built the first maser; shortly after, Townes and A.L. Schawlow showed that it was possible to construct a similar device using optical light; that is, a laser. Two Soviet physicists proposed related ideas independently. The first laser, constructed by the U.S. scientist T.H. Maiman in 1960, used a rod of ruby; since then many types of lasers have been built.

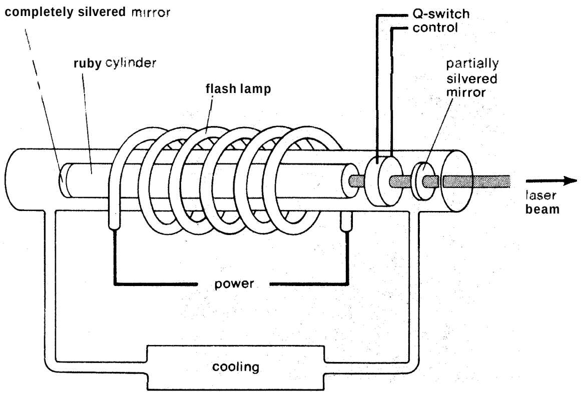


Figure 1: Q-switch, a special switching device that produces giant output pulse. In this example, a ruby laser is being used (see text).

TYPES OF LASERS

Of the several different types of lasers, produced by different means and useful for different purposes, the following are the most important.

Optically pumped solid lasers. One way to achieve the excitation of atoms to the higher energy level for laser action to take place is by illuminating the laser material with light of a frequency higher than that which the laser is to emit. This process is called optical pumping; the light pump must be of high intensity, as the process is usually rather inefficient.

An optically pumped laser consists of a rod of the material chosen, with the ends polished flat and parallel and coated with mirrors to reflect the laser light. The sides are left clear to admit the light from the pumping lamp, which may be a pulsed gas discharge, flashing on and off like a photographer's electronic flash bulb. It may be wound around the laser rod, positioned alongside, or focussed on it by a mirror (see Figure 1). The first operating laser utilized a rod of pink ruby, an artificial crystal of sapphire (aluminum oxide). Many other rare-earth elements have since been used, the most widely used being neodymium. Power outputs in the form of brilliant flashes of light of thousands of watts can be obtained.

Liquid lasers. Solid lasers have the disadvantage of occasional breakdown and damage at higher power levels because of the intense heat generated within the material and by the pumping lamp. The liquid laser is not susceptible to such damage; the crystalline or glassy rod is replaced by a transparent cell containing a suitable liquid, such as a solution of neodymium oxide or chloride in selenium oxychloride. Such cells can be made as large as desired to increase power output. Only a small number of inorganic liquids, however, will function as lasers.

Advantage of liquid lasers

Dye lasers. Certain organic dyes are capable of fluorescing; that is, re-radiating light of a different colour. Though the excited state of their atoms lasts only a small fraction of a second and the light emitted is not concentrated in a narrow band, many such dyes have been made to exhibit laser action, with the advantage that they can be tuned to a wide range of frequencies.

Dyes such as rhodamine *6G,* which emits orange-yellow light, can be made to lase (provide laser action) by excitation by another laser. Rhodamine 6G was the first dye for which continuous, rather than pulsed, operation was achieved, making possible the production of a continuous beam of tunable laser light. Another dye, methylumbelliferone, with the addition of hydrochloric acid, can be made to lase at wavelengths varying across the light spectrum from ultraviolet to yellow, producing laser light of almost any desired frequency within this range.

Gas lasers. Atoms in a gas discharge can be excited to radiate and produce light, as in a neon sign. Occasionally, a particular energy level will cause an exceptionally high number of atoms to accumulate within it; if mirrors are positioned at the ends of the discharge tube, laser action results. Though the conditions are unusual and occur for only a few of the many wavelengths at which the discharge emits, most gases can be made to exhibit laser

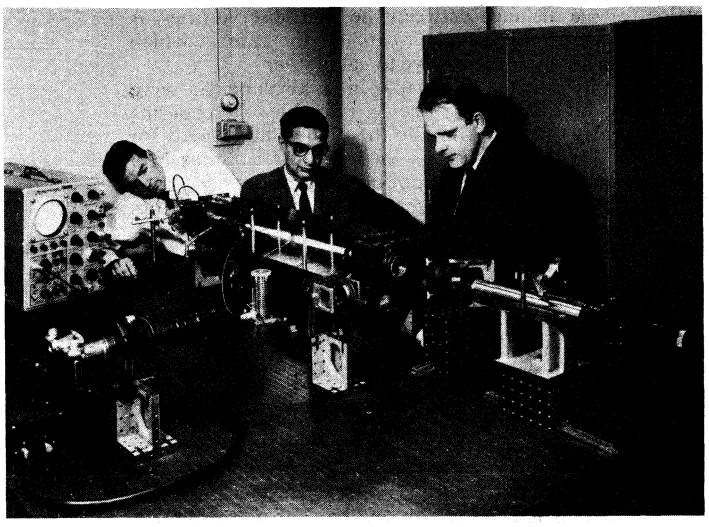


Figure 2: The first gas discharge laser, constructed by (left to right) Donald R. Herriott, Ali Javan, and William R. Bennett, Jr., and put in operation in late 1960.
By courtesy of Bell Laboratories

action at some wavelength under certain discharge conditions. The beam produced provides a nearly ideal straight line, valuable for alignment purposes in construction work (see Figure 2).

Gas dynamic lasers. If a hot gas is allowed to cool rapidly, the number of molecules in a low-energy state may decrease more rapidly and fall below the number in a higher energy state, thus permitting laser action. This condition can-be achieved by expanding burning carbon monoxide mixed with nitrogen through jet nozzles. High power outputs of over 30,000 watts can be obtained.-

Chemical lasers. Certain chemical reactions produce enough high-energy atoms to permit laser action to take place. Laser action can occur in carbon dioxide, for example, if it is present when the elements hydrogen and fluorine are reacting to produce hydrogen fluoride. Large amounts of energy can be released when only moderate amounts of the appropriate materials react.

Semiconductor lasers. A semiconductor laser consists of a flat junction of two pieces of semiconductor material, each of which has been treated with a different type of impurity. When a large electrical current is passed through such a device, laser light emerges from the junction region. Power output is limited, but the low cost, small size, and relatively high efficiency make these devices suited for short-distance communication (telephone, television, etc.) and for distance-measuring equipment.

Lasers as amplifiers and oscillators. Most lasers use the basic structure of a long, narrow column of active material terminated by small mirrors facing each other. Without the mirrors, the structure can be used to amplify a powerful laser beam, producing a still more intense output. With the mirrors, the structure becomes an oscillator, generating a beam whose wavelength is determined by two main factors: the spacing between the mirrors and the characteristics of the laser medium.

Lasers producing short, intense pulses. A shutter placed between the amplifying column and the end mirrors of a laser can prevent laser action as long as it is closed. If conditions are otherwise correct for laser action and the shutter is suddenly opened, the stored energy is released as a giant pulse of light lasting only a tiny fraction of a second, and having a peak power capacity that may be as high as several hundred thousand kilowatts. This is known as Q-switching. The Q-switch may be a mechanical shutter or, more usually, a liquid or solid optical shutter that is normally opaque but can be made transparent by the application of an electrical pulse. The shutter may also be an opaque dye which becomes transparent when exposed to laser light.

Normally a laser oscillates in several modes; that is, at several different frequencies. By synchronizing these modes, a process called mode-locking, even shorter, more powerful pulses can be obtained. Such pulses are useful in scientific investigations and in puncturing holes so rapidly that the surrounding material is not affected.

Tunable laser devices. For scientific purposes, it is highly desirable to produce lasers that can be tuned to a variety of frequencies, and this is possible with certain types. Dye lasers are especially good for this purpose, because they typically operate over a fairly wide frequency band; the desired wavelength can be selected by replacing one of the two mirrors with a mirror that reflects only the desired frequency, such as a diffraction grating. Certain solid lasers can be tuned over a smaller range by adjusting the temperature and orientation of the crystal. Others are capable of generating harmonics; that is, integral multiples of the frequency of an incoming laser beam. This characteristic of the crystal lithium iodate has been used in conjunction with an infrared beam to generate green laser light, which has good transmission characteristics through water.

MASERS

A maser oscillator requires a source of excited atoms or molecules and a resonator to store their radiation. The excitation must force more atoms or molecules into the upper energy level than in the lower, in order for amplification by stimulated emission to predominate over absorption. For wavelengths of a few millimetres or longer, the resonator can be a metal box whose dimensions are chosen so that only one of its modes of oscillation coincides with the frequency emitted by the atoms; that is, the box is resonant at the particular frequency, much as a kettle drum is resonant at some particular audio frequency. The losses of such a resonator can be made quite small, so that radiation can be stored long enough to stimulate emission from successive atoms as they are excited. Thus, all the atoms are forced to emit in such a way as to augment this stored wave. Output is obtained by allowing some radiation to escape through a small hole in the resonator.

The first maser used a beam of ammonia molecules that passed along the axis of a cylindrical cage of metal rods, with alternate rods having positive and negative electric charge. The nonuniform electric field from the rods sorted out the excited from the unexcited molecules, focussing the excited molecules through a small hole into the resonator. The output was less than one microwatt (10^{-6} watt) of power but the wavelength, being determined primarily by the ammonia molecules, was so constant and reproducible that it could be used to control a clock that would gain or lose no more than a second in several hundred years. This maser can also be used as a microwave amplifier. Maser amplifiers have the advantage that they are much quieter than those that use vacuum tubes or transistors; that is, they add very little noise to the signal being amplified. Very weak signals can thus be utilized. The ammonia maser amplifies only a very narrow band of frequencies and is not tunable, however, so that it has largely been superseded by other kinds, such as solid-state ruby masers.

Solid-state and travelling-wave masers. Amplification of radio waves over a wide band of frequencies can be obtained in several kinds of solid-state masers, most commonly crystals such as ruby at low temperatures. Suitable materials contain ions (atoms with an electrical charge) whose energy levels can be shifted by a magnetic field so as to tune the substance to amplify the desired frequency. If the ions have three or more energy levels suitably spaced, they can be raised to one of the higher levels by absorbing radio waves of the proper frequency.

The amplifying crystal may be operated in a resonator that, as in the ammonia maser, stores the wave and so gives it more time to interact with the amplifying medium. A large amplifying bandwidth and easier tunability are obtained with travelling-wave masers. In these, a rod of a suitable crystal, such as ruby, is positioned inside a wave-guide structure that is designed to cause the wave to travel relatively slowly through the crystal.

Solid masers have been used to amplify the faint signals returned from such distant targets as satellites in radar

Masers for satellite communications

and communications. Their sensitivity is especially important for such applications because signals coming from space are usually very weak. Moreover, there is little interfering background noise when a directional antenna is pointed at the sky, and the highest sensitivity can be used (see also SATELLITE COMMUNICATION). In radio astronomy, masers made possible the measurement of the faint radio waves emitted by the planet Venus, giving the first indication of its temperature.

Gas masers. Generation of radio waves by stimulated emission of radiation has been achieved in several gases in addition to ammonia. Hydrogen cyanide molecules have been used to produce a wavelength of 3.34 millimetres. Like the ammonia maser, this one uses electric fields to select the excited molecules.

One of the best fundamental standards of frequency or time is the atomic hydrogen maser introduced by U.S. scientists N.F. Ramsey, H.M. Goldenberg, and D. Kleppner in 1960. Its output is a radio wave whose frequency of 1,420,405,751.786 hertz (cycles per second) is reproducible with an accuracy of one part in thirty million million (30×10^{12}). A clock controlled by such a maser would not get out of step more than one second in 100,000 years.

In the hydrogen maser, hydrogen atoms are produced in a discharge and, like the molecules of the ammonia maser, are formed into a beam from which those in excited states are selected and admitted to a resonator. To improve the accuracy, the resonance of each atom is examined over a relatively long time. This is done by using a very large resonator containing a storage bulb. The walls of the bulb are coated so that the atoms can bounce repeatedly against the walls with little disturbance of their frequency.

Another maser standard of frequency or time uses vapour of the element rubidium at a low pressure, contained in a transparent cell. When the rubidium is illuminated by suitably filtered light from a rubidium lamp, the atoms are excited to emit a frequency of 6.835 gigahertz (6.835×10^9 hertz). As the cell is enclosed in a cavity resonator with openings for the pumping light, emission of radio waves from these excited atoms is stimulated.

LASER APPLICATIONS

The light produced by lasers is in general far more monochromatic, directional, powerful, and coherent than that from any other light sources. Nevertheless, the individual kinds of lasers differ very greatly in these properties as well as in wavelength, size, and efficiency. There is no single laser suitable for all purposes, but some of the combinations of properties can do things that were difficult or impossible before lasers were developed.

Lasers for alignment purposes

A continuous visible beam from a laser using a gas, such as the helium–neon combination, provides a nearly ideal straight line for all kinds of alignment applications. The beam from such a laser typically diverges by less than one part in a thousand, approaching the theoretical limit. The beam's divergence can be reduced by passing it backward through a telescope, although fluctuations in the atmosphere then limit the sharpness of a beam over a long path. Lasers have become widely used for alignment in large construction; *e.g.*, to guide machines for drilling tunnels and for laying pipelines. Laser beams are used to align jigs employed in producing large jet aircraft, permitting accuracies of a hundredth of an inch (0.25 millimetres) over distances of 200 feet (60 metres) or more.

A pulsed laser can be used in a light radar, sometimes called LIDAR, and the narrowness of its beam permits sharp definition of targets. As with radar (*q.v.*), the distance to an object is measured by the time taken for the light to reach and return from it, since the speed of light is known. LIDAR echoes have been returned from the Moon, facilitated by a multiprism reflector that was placed there by the first astronauts to land there. Now distances can be measured from an observatory on Earth to the lunar mirror with an accuracy of about one foot (30 centimetres). Simultaneous measurements of the mirror's distance and direction from two observatories, on different parts of the Earth, could give an accurate value for the distance between the two observatories. A series of such measurements can tell the rate at which continents are drifting relative to each other.

A vertically directed laser radar in an airplane can serve as a fast, high-resolution device (capable of rendering detail clearly) for determining the plane's altitude or for mapping fine details such as the contours of steps in a stadium or the shape of the roof of a house.

With a pulsed laser radar, returns can be obtained from dust particles and even from air molecules at higher altitudes. Thus air densities can be measured and air currents can sometimes be traced. A simpler LIDAR can serve to measure the height of cloud cover over airports.

Precise measurement of distances

The high coherence of a laser's output is very helpful in measurement and other applications involving interference of light beams. If a light beam is divided into two parts that travel different paths, when the beams come together again they may be either in step so that they reinforce, or out of step so that they cancel. Thus the brightness of the recombined wave changes from light to dark, producing interference fringes, when the difference in path lengths is changed by one-half of a wavelength (about ten millionths of an inch or 0.00025 millimetres for visible light). Such devices are called laser interferometers. Very small displacements can be detected, and larger distances can be measured with precision. With lasers, these measurements can be carried out over very long distances. Laser interferometers are being used to monitor small displacements in the Earth's crust across geological faults. In manufacturing, laser interferometers are being used to gauge fine wires, to monitor the products of automated machine tools, and to test optical components.

Lasers can be so monochromatic that a small shift in the light frequency can be detected. Light reflected from an object that is moving toward the laser is raised in frequency by an amount depending on the velocity of the object (Doppler effect). For a receding object the frequency is lowered. In either case, if some of the original and the shifted light are recombined at a photodetector, a signal at the difference frequency (the difference in frequency between the original and the shifted light) is observed, and even small velocities can be measured.

A laser can be constructed in which the light travels around a square or triangular ring. Waves are generated to travel around the ring in both the clockwise and counterclockwise directions. If the ring is stationary, these waves have the same frequency, but if it is rotated the frequency difference is proportional to the rate of rotation. Thus the ring laser can function as a gyroscope to sense rotations, even though it has no moving parts.

The brightness and coherence of laser light make it especially suitable for visual effects and photography that simulate third dimensional depth; *e.g.*, holography (*q.v.*).

The light from many lasers is relatively powerful and can be focussed by a conventional lens system to a small spot of great intensity. Thus even a moderately small pulsed laser can vaporize a small amount of any substance and drill narrow holes in the hardest materials. Ruby lasers, for example, are used to drill holes in diamonds for wire drawing dies and in sapphires for watch bearings. For biological research, a finely focussed laser can vaporize parts of a single cell, thus permitting microsurgery of chromosomes. Small spots, closely spaced, can be vaporized in an opaque film to record information for large computer memories.

The heating effects produced by a pulsed laser beam are both highly selective and extremely rapid. Thus ink can be removed from paper, making possible a laser eraser. The ink absorbs the laser light, is vaporized, and burns, while the paper remains unaffected.

Lasers for surgery

Strong heating can be produced by a laser at a place where no mechanical contact is possible. Thus one of the earliest applications of lasers was for surgery on the retina of the eye. Other surgical applications, using larger lasers, are being investigated.

Lasers are also used for small-scale cutting and welding. They can trim resistors to exact values by removing material, and can alter connections within integrated arrays of microcircuit elements.

A pulse of light from a laser can vaporize a sample of a substance for analysis by suitable instruments. By this method an extremely small sample can be analyzed without introducing contaminants.

The high brightness, pure colour, and directionality of laser light make it ideally suitable for experiments on light scattering. Even a small amount of light that is scattered with a change of wavelength or direction can be readily identified. In particular, a type of scattering known as the Raman effect produces characteristic wavelength shifts by which molecular species can be identified. With laser sources and sensitive spectrography, small samples of transparent liquids, gases, or solids can be analyzed. It is even possible to measure contaminants in the atmosphere at a considerable distance by Raman scattering of light from a laser beam.

Laser beams can be used for communications. Because the light frequency is so high (around 5 × 10'' hertz for visible light), the intensity can be rapidly altered to encode very complex signals. In principle, one laser beam could carry as much information as all existing radio channels; experimentally, seven television programs have been transmitted at one time. Laser light can, however, be blocked by rain, fog, or snow so that for reliable communications on Earth, the laser beam would need to be enclosed in protecting pipes. Meanwhile, lasers are useful for special purpose communications. For long distance communication through outer space, the great directionality of laser beams makes effective use of small amounts of power.

SAFETY CONSIDERATIONS

Some lasers, especially the higher power carbon dioxide and solid-state, produce beams so powerful that they could cause serious injury even at a great distance. These must be operated only when no inflammable materials are in its path and when there is no possibility of anyone being struck by the beam.

The largest numbers of lasers in use, however, are low power types, such as the helium–neon lasers widely used for alignment and measurement. The intensity of the laser light is not high enough to cause injury directly, but, since it is a highly directional beam, its light can be concentrated into a small spot by focussing. In particular, the lens of the eye can focus laser light to a small, intense spot on the retina and can damage this sensitive tissue. The light reflected from a mirror or even a shiny surface is nearly as dangerous as the direct beam. For such lasers it is essential to avoid looking directly into the beam or its reflection. Pulsed lasers are more hazardous because of the high peak intensity and because there is no chance to respond to the sudden flash.

There is little hazard, however, when the light of a low power continuous-wave laser is used to illuminate a diffuse surface. The light scattered by the object is spread in all directions and an eye at any substantial distance can intercept only a small fraction of the light.

Most lasers are far from being the death rays of science fiction. With reasonable precautions they can be used safely for many purposes. Safety standards and operating procedures are under study by various government public health and safety organizations, and high-power lasers are subject to regulations in many areas.

FUTURE PROSPECTS

Research on properties and applications of masers and lasers is active throughout the world. The search for ways of obtaining coherent light at additional wavelengths continues, even as gaps are being filled by tunable lasers. One persistent frontier of research is the problem of generating still shorter wavelengths for scientific purposes and possible other applications; constructing maser or laser-like devices, will become increasingly difficult as the wavelength is decreased. The rate of spontaneous emission increases rapidly as the wavelength is decreased, and it becomes correspondingly more difficult to maintain enough atoms in excited states to provide amplification. Moreover, for wavelengths approaching or in the extremely short X-ray region, no substance is as transparent as many materials are for visible light. Extra excitation will be needed to overcome losses due to this lack of transparency. It does seem possible that some kind of device to produce coherent X-rays by stimulated emission eventually will be constructed. Its uses can hardly be foreseen now, although holography to reveal the positions of atoms in molecules is a possibility.

X-ray lasers possible

Continuous-wave gas lasers in the visible or near-visible region provide very good standards of length. Their wavelength is chiefly determined by the particular gas atoms. There are small uncertainties introduced by the spacing of the end mirrors and by collisions in the gas discharge tube, however. The wavelength of an unstabilized helium–neon laser may vary by about ten parts in 1,000,000.

Much better accuracy is obtained by carefully adjusting the laser to oscillate at the precise centre of the band of wavelengths that the gas discharge can generate. Still better accuracy, about one part in 10^{12}, can be achieved by new methods that locate the exact centre frequency of the absorption resonance of molecules such as iodine. These molecules can be located outside the laser so as not to be disturbed by the gas discharge, and the wavelength that they absorb is very precisely reproducible. A laser of this sort might ultimately serve as a primary standard of length. More studies are needed to find the best of the several possible atomic or molecular wavelength standards and to ensure that it can be accurately reproduced in different laboratories.

It has become possible to measure the frequency of a light wave by using lasers. A microwave oscillator is first measured in terms of an atomic frequency standard such as the hydrogen maser or an atomic beam of the element cesium. High harmonics of the microwave oscillator are then generated by a silicon crystal and used to measure and synchronize the frequency of a laser in the infrared. In turn, harmonics of this laser are generated and used to control the frequency of a shorter wavelength infrared laser. After several such steps, the overtones of a near infrared laser produce visible light whose frequency is an exact, known multiple of a measured microwave frequency

In the near future, it should be possible to measure precisely the frequency and wavelength of the same visible laser beam, synchronized to some atom or molecule. It may then be best to define the basic standard of length in terms of the wavelength of the light of the chosen atom, and to define the standard of time in terms of the frequency of the same atomic oscillation.

BIBLIOGRAPHY. *Lasers and Light: Readings from Scientijic American,* with introductions by A.L. SCHAWLOW (1969); and DAVID FISHLOCK (ed.), *A Guide to the Laser* (1967), contain authoritative articles at an introductory level on all aspects of this subject. M. BROTHERTON, *Masers and Lasers* (1964); JOHN M. CARROLL, *The Story of the Laser,* new ed. (1970); and H. ARTHUR KLEIN, *Masers and Lasers* (1963), are readable introductions to the background and basic physical principles. RONALD BROWN, *Lasers: Tools of Modern Technology* (1968); and STANLEY LEINWOLL, *Understanding Lasers and Masers* (1965), are introductory surveys of the principles and uses of lasers. Treatments of the basic theory of lasers for those having some acquaintance with physics and mathematics include: G. BIRNBAUM, *Optical Masers* (1964); B.A. LENGYEL, *Introduction to Laser Physics* (1966) and *Lasers,* 2nd ed. (1971); W.V. SMITH and P.P. SOROKIN, *The Laser* (1966); D. ROSS, *Laser: Lichtverstarker und -oszillatoren* (1966; Eng. trans., *Lasers: Light Amplifiers and Oscillators,* 1969); ARTHUR MAITLAND and M.H. DUNN, *Laser Physics* (1970); SAMUEL L. MARSHALL (ed.), *Laser Technology and Applications* (1968); and A.E. SIEGMAN, *An Introduction to Lasers and Masers* (1971). LEON GOLDMAN, *Biomedical Aspects of the Laser* (1967), is a summary of directions and results in the applications of lasers to these fields. See also MICHAEL J. BEESLEY, *Lasers and Their Applications* (1971).

(Ar.L.S.)

Lassalle, Ferdinand

Ferdinand Lassalle was the leading spokesman for German socialism, after an unsuccessful revolution there in 1848, and one of the founders of the German labour movement.

Lassalle was born at Breslau on April 11, 1825 of Jew-

Lassalle, c. 1860.
Archiv *fur* Kunst und *Geschichie*

ish parents; his father, Heymann Lasal, or Loslauer, was a wholesale silk merchant and town councillor.

Early years

Ferdinand Lassalle—the spelling of the name dates from a stay in Paris in 1846—attended the Breslau classical high school but was expelled when he forged a signature on a school report. He attended a trade school in Leipzig in 1840, returned to Breslau in 1841, and passed his school-leaving examination in 1843. Diaries of this period show him as a "tumultuous personality," full of contradictions. On the one hand he says, "I have voluntarily renounced a life devoted to beauty in order to become a shop clerk," and on the other, "Now is the time to fight for the most sacred principles of humanity."

In 1843–44 he began to study philosophy, history, philology, and archaeology at the University of Breslau. In 1844–45 he continued his studies in Berlin, interrupting them only briefly in 1845 for an interlude in Breslau. In Berlin he first encountered the ideas of the German philosophers G.W.F. Hegel and Ludwig Feuerbach and of the French Utopian thinkers. Intending to take his degree and to qualify as a university lecturer with a thesis on the philosophy of Heracleitus, he made repeated studies of the subject in Paris between 1845 and 1847. Here he met the French social theorist Pierre-Joseph Proudhon and Heinrich Heine, the German poet.

Champion of Countess Hatzfeld

In 1846, in Düsseldorf, he met the unhappily married Countess Sophie Hatzfeld, who was trying to divorce her husband. Although not a lawyer, Lassalle conducted 35 lawsuits in her behalf and in 1854 finally obtained a divorce for her. Henceforth, he received an annual pension of 4,000 thalers from the countess, thus becoming financially independent. His lifelong relationship with the countess, though it was nothing more than that of son and mother, "stimulated gossip about Lassalle and immensely impeded his political career." Lassalle lived in Düsseldorf from 1848 to 1857 and took part in the revolution of 1848–49, by which the liberal middle class tried to attain a constitutional monarchy that would grant such civil rights as freedom of assembly and freedom of the press. During those days he established contact with Karl Marx and Friedrich Engels, the Socialist leaders. When Lassalle urged the militia to open revolt in November 1848 he was arrested and held in prison until his trial in July 1849. One of the few radical leaders who did not emigrate during the revolution, he remained in Germany after its collapse. Although he was repeatedly arrested, indicted, and sentenced to prison, Lassalle counted his years in Düsseldorf, where he was able to be active both as a writer and labour organizer, among the happiest of his life. In the period of reaction following the abortive revolution, he travelled to Switzerland, to the World's Fair in Paris in 1855, and to the Orient in 1856. He completed the Heracleitus manuscript and the tragedy *Franz von Sickingen* (1859), which assigns to personality a role in determining the course of history.

Lassalle and Bismarck

In 1857 Lassalle went back to Berlin, and in 1859 he settled permanently in the capital, where he became active as a political journalist. He met Marx in 1861, but, although they continued to correspond, they gradually became estranged. In contrast to Marx, Lassalle believed that the revolutionary phase had come to an end and that only a legal and evolutionary approach could hold hopes of success. With this goal in mind he held discussions with the Prussian prime minister Otto von Bismarck in 1863–64. Fourteen years after Lassalle's death Bismarck said of him, "He was one of the most intelligent and amiable men I have ever associated with, a man of great ambition and by no means a republican." Finding himself in a difficult political situation, Bismarck was, in the early 1860s, seeking allies in his struggle against the majority liberal opposition, while Lassalle was considering the concept of a monarchical welfare state. This was to be based on a universal suffrage for the three classes rather than on the existing suffrage that favoured the upper classes. He thus hoped, by integrating the working class into political and social life, to achieve a transition from a bourgeois state based on private property to a democratic constitutional state. Lassalle and Bismarck were attracted to each other by their many common characteristics. Lassalle in particular was distinguished by his charismatic personality and his paternalist notions of democracy, understandable in the context of a largely politically apathetic population.

The year 1862 produced a crisis in Lassalle's thinking when the rising in Italy led by Garibaldi did not, contrary to Lasalle's expectations, spread to other countries. Berlin remained utterly unreceptive to his ideas. Realizing that lecturing and distributing pamphlets to artisans' clubs and citizens' associations were not producing sufficient results, Lassalle began agitating in workingmen's associations in order to make his political aims known to the masses. Bismarck's conflict with the Prussian parliament led Lassalle to re-examine his liberal beliefs. He now began to study political economy. Lecturing on constitutional problems, he stated: "Initially, constitutional questions are not questions of law, but rather questions of power."

Organization of the "Suffrage Army"

In December 1862, Lassalle was asked by the executive committee of the "Central Committee to Convoke a General Congress of German Workers" to write a program for the congress. Lassalle at once recognized in the congress an opportunity to organize a "Suffrage Army." "Organize yourselves as a general German workingmen's association to agitate legally and peacefully, but untiringly and ceaselessly, for the introduction of universal and direct suffrage in all German provinces! This is the banner you must raise! This is the sign under which you will be victorious!"

In 1863–64 Lassalle hurled himself into the struggle for workers' rights, especially in the Rhineland. "Only the working class matters to me," he declared. When the ADAV (Allgemeiner Deutscher Arbeiterverein, or General German Workers' Association) was founded on May 23, 1863, in Leipzig, Lassalle was elected president for a five-year term. In Cologne he collaborated with a Socialist writer, Moses Hess, but other associates rebelled against Lassalle's authoritarian leadership and the cult of his personality he did nothing to discourage. His generally incendiary speeches were often followed by lawsuits.

Exhausted and disappointed over the insignificant results of his propaganda activity, Lassalle announced, "I am sick and tired of politics!" In Switzerland, where he went for a rest in July 1864, he met Helene von Dönniges. He courted her passionately, but encountering opposition from the young girl's family, he challenged her father and her former fiancé, Yanko von Racowitza, to a duel. Racowitza accepted, and on August 28, in a little forest near Geneva, the senseless duel was fought. Lassalle was struck in the abdomen and died August 31. He was buried in the Jewish cemetery in Breslau.

Assessment

Lassalle was for many decades considered a reformist heretic by the worker's movement, which then adhered to the deterministic notions of popular Marxism according to which the dictatorship of the proletariat was foreordained by history. By others Lassalle continued to be romantically glorified as a pioneer of Socialism.

Only since the time of Eduard Bernstein and the era of revisionism, when the German Social Democratic Party, aiming at becoming a party of the masses, set down the aims of parliamentary democracy and participation in government, has the modern significance of Lassalle been acknowledged. It is not the theorist or the organizer of a workers' party who is remembered, but, in the words of the German Social Democratic leader Carlo Schmid, a Lassalle "who in place of scientific analysis constantly fixed his sights on the true aim on history's horizon: the liberation of man from the position of object and the elimination of man's alienation from himself through the power of his own will."

BIBLIOGRAPHY. The best biographies are GUSTAV MAYER, ***Bismarck und Lassalle*** (1928); EDUARD BERNSTEIN, ***Ferdinand Lassalle und seine Bedeutung für die Arbeiterklasse,*** 2nd ed. (1919) and ***Ferdinand Lassalle seine Bedeutung in der Geschichte der Sozialdemokratie,*** in vol. 1 of Lassalle's ***Reden und Schriften*** (1892; Eng. trans., ***Ferdinand Lassalle As a Social Reformer,*** 1893, reprinted 1969); THILO RAMM, ***Ferdinand Lassalle als Rechts- und Sozialphilosoph,*** 2nd ed. (1956); WILHELM MATULL, ***Ferdinand Lassalle*** (1964); and SHLOMO NA'AMAN, *Lassalle* (1970), both in German.

English-language studies include GEORG M.C. BRANDES, ***Ferdinand Lassalle*** (1911, reprinted 1970); and DAVID FOOTMAN, ***The Primrose Path: A Life of Ferdinand Lassalle*** (1946; *U.S.* title, ***Ferdinand Lassalle, Romarztic Revolutionary,*** 1947, reprinted 1969). Lassalle's love affair is the subject of GEORGE MEREDITH'S novel ***The Tragic Comedians*** (1881).

(Wi.M.)

Latin America and the Caribbean, Colonial

The colonial era in Latin America and the Caribbean began in the late 15th century with the first voyages of discovery in the New World (then thought to be part of the Indies) and ended in the 19th century with the successful movements for independence. The histories of these regions before and after the colonial era are covered in other major articles.

This article is divided into the following sections:

- I. The Spanish conquest of America
  - The great voyages
    - The voyages of Christopher Columbus
    - Other early explorations
  - The conquest of Mexico
  - The conquest of Peru
- II. Spain's colonial empire
  - Colonial government
    - The Council of the Indies
    - Viceroys, captains general, and audiencias
    - Provincial government
    - Restrictions on public officials
    - Municipal government
    - Other government offices
  - Spain's Indian policy
  - The colonial economy
  - Commerce, smuggling, and piracy
  - The church in the Indies
    - The missions
    - The Jesuits in Paraguay
    - The Inquisition
- III. Colonial Brazil
  - The formation of colonial Brazil
    - Colonization
    - Slave hunters
    - Rivals to the Portuguese
    - Shift to the south
    - Trade
  - Government and church
  - Masters and slaves
- IV. The Bourbon reforms and Spanish America
  - Reforms
  - Colonial economic growth
  - Decentralization
  - Defense of the empire
- V. The wars of independence
  - Background for the wars
    - Enlightenment and revolutionary influences
    - Beginnings of the independence movement
  - The independence of South America
    - Spanish South America
    - Brazil
  - The independence of Mexico

I. The Spanish Conquest of America

THE GREAT VOYAGES

Search for a sea route to the East

The discovery of America was a by-product of European efforts in the 15th century to find a sea route to the East and thereby to end the monopoly of Italian and Levantine middlemen over the lucrative trade in spices and other Oriental products. Portugal took a decisive lead; it had the advantages of a long Atlantic seaboard with excellent harbours, a large number of fishermen and sailors, and an aristocracy that had learned to supplement its meagre revenue from the land with income from trade and shipbuilding. In 1497 a fleet of four Portuguese ships commanded by Vasco da Gama sailed from Lisbon on a voyage that inaugurated the age of European imperialism in Asia. After rounding the Cape of Good Hope, Vasco sailed into the Indian Ocean and up the coast of East Africa and then to Calicut, the great spice-trade centre on the west coast of India.

The voyages of Christopher Columbus. The search for a sea road to the Indies inspired more than one solution. An obscure Italian seafarer, Christopher Columbus, became convinced that it was possible to reach the East from Europe by sailing westward across the Atlantic and that his proposed route was shorter than the route around Africa, a conception that underestimated the size of the earth and overestimated the size and eastward extension of Asia. About 1484 Columbus, who then resided in Lisbon, offered to make a voyage of discovery for John II of Portugal, but the king turned it down. Columbus next turned to Castile, where, after eight years of discouraging delays and negotiations, Queen Isabella agreed to support the "Enterprise of the Indies." The contract made by the Queen with Columbus named him admiral, viceroy, and governor of the lands he should discover and promised him a generous share in the venture's profits.

The "Enterprise of the Indies"

On August 3, 1492, Columbus sailed from Palos with three small ships—the "Pinta," the "Santa Maria," and the "Niña"—manned not by the jailbirds of legend but by experienced crews under competent officers. The voyage was remarkably prosperous, with fair winds the whole way out. On October 12 they made landfall at an island in the Bahamas, which Columbus named San Salvador. Cruising southward through the Bahamas, Columbus came to the northeastern coast of Cuba, which he mistook for part of Cathay (China). Next he sailed westward to explore the northern coast of an island that he named Española (Hispaniola), where he lost his flagship, the "Santa Maria." He then returned to Spain to report his supposed discovery of the Indies.

In response to Portuguese charges of encroachment on an area in the Atlantic reserved to Portugal by a previous treaty with Castile, King Ferdinand and Queen Isabella appealed for help to Pope Alexander VI, himself a Spaniard. The pontiff issued a series of bulls (1493) that assigned to Castile all lands discovered or to be discovered by Columbus and drew a line from north to south 100 leagues (345 miles) west of the Azores and Cape Verdes. To the west of this line was to be a Spanish sphere of exploration; to the east, Portuguese. To John II this line seemed to threaten Portuguese interests in the south Atlantic and the promising route around Africa to the East. Yielding to Portuguese pressure, Ferdinand and Isabella signed, in 1494, the Treaty of Tordesillas, establishing a boundary 270 leagues (930 miles) farther west.

Columbus returned to Espaiiola at the end of 1493 with a fleet of 17 ships carrying 1,200 colonists. The settlers soon gave themselves up to gold hunting and preying on the Indians; and Columbus, a foreigner, lacked the power and the personal qualities to control them. In 1496 he returned to Spain to report his new discoveries and to answer charges sent by disgruntled settlers.

The first two voyages had not paid their way, but the Spanish sovereigns still had faith in Columbus and outfitted a third fleet in 1498. On this voyage he discovered the island of Trinidad and the mouths of the Orinoco. He arrived in Española to find chaos. The Spaniards, disappointed in their hopes of quick wealth, blamed Columbus for their misfortunes and rose in revolt. To appease the rebels Columbus had to issue pardons and grant land and Indian slaves. Meanwhile, a stream of complaints against Columbus had caused the sovereigns to send out an agent, Francisco de Bobadilla, to supersede Columbus and investigate the charges against him. Bobadilla seized Columbus and sent him to Spain in chains. Although Isabella ordered Columbus' release, he never again exercised the functions of viceroy and governor in the New World. Columbus made one more voyage—in 1502–04—an unsuccessful search for a strait that would lead into the Indian Ocean. From Española, where he was not permitted to land, he crossed the Caribbean to the coast of Central America and followed it southward to the Isthmus of Panama. He finally departed for Española but was forced to land on Jamaica, where he and his men were marooned for a year.

Complaints against Columbus

Other early explorations. Other explorers followed and gradually made known the immense extent of the mainland coast of South America. In 1499 Alonso de Ojeda, accompanied by the Florentine Amerigo Vespucci, sailed to the mouths of the Orinoco and explored the coast of Venezuela. Vespucci personally directed another voyage, in 1501–02, under the flag of Portugal; this expedition, sent to follow up the discovery of Brazil by Pedro Álvares Cabral in 1500, explored the Brazilian coast and discovered the Rio de la Plata before turning back. To Vespucci it was obvious that the landmass thought by Columbus to be a part of Asia was really a new continent —a "fourth part of the world." Vespucci's letters circulated widely in the early 1500s and gave him the fame of being the first European to set foot on the South American continent. A German geographer, Martin Waldseemuller, honoured Vespucci by assigning on a map the name America to the area of Brazil. The name caught on and presently was applied to the whole of the New World.

The naming of America

A growing shortage of Indian labour and a general lack of economic opportunities for new settlers on Española incited Spanish slave hunters and adventurers to explore and conquer Puerto Rico, Jamaica, and Cuba between 1509 and 1511. In the same period efforts to found colonies on the coast of northern Colombia and Panama failed disastrously, and the remnants of two expeditions were united under the conquistador Vasco Núñez de Balboa to form the settlement of Darién on the Isthmus of Panama. Moved by Indian tales of a great sea, south of which lay a land overflowing with gold, Balboa led an expedition across Panama to the shores of the South Sea (Pacific Ocean). This aroused the jealousy of his father-in-law, Pedrarias Dávila, sent out by Charles V in 1514 as governor of the Isthmus, who contrived charges of treason and desertion; Balboa was tried, condemned, and beheaded in 1519.

The discovery of the South Sea helped confirm Vespucci's view that the so-called Indies formed no part of Eastern Asia. After 1513 the work of discovery centred on the search for a waterway to the East through or around the American continent. Ferdinand Magellan's circumnavigation of the globe in 1519–22 was too long to have commercial value. The net result was to enhance the value of America in Spanish eyes. Disillusioned with the dream of easy access to the East, Spain turned with concentrated energy to the task of extending its American conquests and to the exploitation of the human and natural resources of the New World.

THE CONQUEST *OF MEXICO*

In 1517 a slave-hunting expedition outfitted by Gov. Diego de Velázquez de Cuéllar of Cuba discovered the Yucatan peninsula of Mexico, inhabited by Maya Indians whose pyramids, temples, and gold ornaments revealed a native culture far more advanced than any the Spaniards had previously encountered. Encouraged by the gold and other signs of Indian wealth brought back by the expedition, Velázquez outfitted a new venture, which he entrusted to his kinsman Juan de Grijalba. Grijalba coasted down the Yucatan Peninsula and, in June 1518, reached the limits of the Aztec Empire. From near what is now the port of Veracruz, Grijalba sent one ship back to Cuba with the gold that had been gained by barter with the coastal Indians and sailed on westward with three other ships, perhaps as far as the river Pánuco, marking the northern limits of the Aztec Empire. He returned to Cuba in November.

Velázquez sent a third expedition, with some 600 men under Hernán Cortés, to conquer the Mexican mainland in February 1519. Because Velázquez had not yet secured from the emperor Charles an agreement authorizing conquest and settlement of the mainland, Cortés' instructions permitted him only to trade and explore.

Cortés' expedition to Mexico

Cortés' fleet first touched land at the island of Cozumel, off the coast of Yucatan. In March 1519, Cortés landed on the coast of Tabasco and defeated local Indians. In April he dropped anchor near the site of modern Veracruz, founding the town of Villa Rica de la Vera Cruz and appointing its first officials, into whose hands he surrendered the authority he had received from Velázquez. These officials then conferred on Cortés the title of captain general with authority to conquer and colonize the newly discovered lands (Cortés thus drew on Spanish medieval traditions of municipal autonomy to vest his disobedience with a cloak of legality).

Some days later the Aztec king Montezuma's (Moctezuma) ambassadors appeared in the Spanish camp. Apparently convinced that Cortés was the god Quetzalcóatl, who was returning to reclaim his lost realm, the envoys brought precious gifts—the finery of the great gods Quetzalcóatl, Tláloc, and Tezcatlipoca; a gold disk in the shape of the sun, as big as a cartwheel; an even larger disk of silver, in the shape of the moon; and a helmet full of small grains of gold—while pleading with Cortés not to seek a meeting with their king. By plying Cortés-Quetzalcóatl with gifts Montezuma hoped to dissuade him from advancing into the interior and reclaiming the god's lost throne. Suavely Cortés informed the ambassadors that he had come a long way to see and speak with Montezuma, and he could not return without doing so.

Cortés, becoming aware of the bitter discontent of tributary towns with Aztec rule, began to play a double game. He encouraged the Totonac Indians of the coast to seize and imprison Montezuma's tax collectors, and he then promptly obtained their release and sent them to the king with expressions of regard and friendship. He took two other steps before beginning the march on the Aztec capital at Tenochtitlán (now Mexico City). First, he sent dispatches to the emperor Charles seeking approval for his actions by describing the extent and value of discoveries; and second, to stiffen the resolution of his followers by cutting off all avenues of escape, he scuttled and sank all his remaining ships on the pretext that they were not seaworthy. Then, with his small army, he began a march on Tenochtitlán.

Advancing into the sierra, Cortés met and defeated in battle the Tlaxcalan Indians, traditional enemies of the Aztecs; the Tlaxcalans then formed an alliance with the white invaders. Next Cortés marched on Cholula, centre of the cult of Quetzalcóatl, where he claimed that the Cholulans were conspiring to attack him and staged a mass slaughter of the Cholulan nobility and warriors after they had assembled at his bidding in a great courtyard. Montezuma sent new envoys who brought rich gifts to Cortés but urged him to abandon his plan of visiting the Aztec capital; all of Montezuma's stratagems failed, and he welcomed Cortés at the entrance to the capital as a rightful ruler returning to his throne. The Aztec ruler even allowed himself to be kidnapped from his palace and taken to live as a hostage in the Spanish quarters.

Greeting of Cortés by Montezuma

A blunder on the part of his lieutenant Pedro de Alvarado thwarted Cortés' plan to use Montezuma as a puppet ruler. In Cortés' absence Alvarado ordered a massacre of the leading Aztec chiefs and warriors during a religious festival. This caused an uprising that forced the Spaniards to retreat to their own quarters. The tribal council deposed the captive Montezuma and elected a new chief, who launched vigourous attacks on the white invaders. In the midst of the struggle Montezuma died—killed by his own people as he appealed for peace, according to Spanish accounts; strangled by the Spaniards themselves, according to Indian sources. Threatened by a long siege and famine, Cortés evacuated Tenochtitlán at a heavy cost in lives. The surviving Spaniards and their Indian auxiliaries at last reached Tlaxcala.

In December 1520 Cortés, strengthened by Spanish reinforcements from Cuba, again marched on Tenochtitlan. A desperate struggle began in late April 1521. On August 23, after a siege of four months, the last Aztec king Cuauhtémoc surrendered; and Cortés took possession of the ruins that had been the city of Tenochtitlán.

From the Valley of Mexico the tide of conquest flowed in all directions. Guatemala was reduced by Pedro de Aivarado; Honduras, by Cortés himself. In 1527 Francisco de Montejo began the conquest of Yucatan (in 1542 the Maya Indians rose in a revolt that was crushed with great slaughter). Meanwhile, expeditions from Darién subjugated the Indians of Nicaragua; and thus the two streams of Spanish conquest, both originally starting from Espafiola, came together again.

For a brief time Cortés was undisputed master of the old Aztec Empire, renamed the "Kingdom of New Spain." The crown granted him the title of "Marquis of the Valley of Oaxaca" and the tribute and labour services of 23,000 Indian vassals. But the characteristic royal distrust of the great conquerors soon asserted itself. He was removed from his office of governor, and in 1539 he returned to Spain.

THE CONQUEST OF PERU

The golden kingdom beyond the South Sea

The conquest of Mexico challenged other Spaniards to match the exploits of Cortés and his companions and to discover the golden kingdom rumoured to lie beyond the South Sea. In 1519 Pedrarias founded the town of Panama on the western side of the isthmus, which became a base for exploration along the Pacific coast. Three years later Pascual de Andagoya crossed the Gulf of San Miguel and returned with more information about a land of gold called Birú (Peru). Pedrarias then entrusted command of a voyage of discovery southward to Francisco Pizarro, who, in turn, recruited two partners—Diego de Almagro, an adventurer of obscure origin, and Hernando de Luque, a priest who acted as financial agent for the trio. Two preliminary expeditions (1524 and 1526) yielded enough finds of gold and silver to confirm the existence of the elusive kingdom. Pizarro then left for Spain to obtain royal sanction for the exploration of Peru. He returned to Panamá with the titles of captain general and *adelantado* (provincial governor) accompanied by his four brothers and other followers.

In December 1530 Pizarro again sailed southward from Panamá with a force of some 200 men and landed in the spring on the Peruvian coast. Civil war was raging in the Inca Empire. Atahualpa, son of the late emperor Huayna Capac by a secondary wife, had defeated and imprisoned the lawful heir to the throne, Huáscar, and was moving toward the imperial capital of Cuzco when he received news of the arrival of white strangers. After an exchange of messages and gifts between Pizarro and Atahualpa, the two armies advanced toward a meeting at the town of Cajamarca, high in the mountains.

Pizarro planned to win a quick, relatively bloodless victory by seizing the emperor as Cortés had done with Montezuma. When Atahualpa and his escort appeared in the square of Cajamarca, they found it deserted, for Pizarro had concealed his men in some large buildings opening on the square. At a signal from Pizarro his soldiers, supported by cavalry and artillery, rushed forward to kill hundreds of terrified Indians and take the Inca prisoner.

Atahualpa attempted to gain his freedom by offering to fill his spacious cell with gold higher than a man could reach. Pizarro accepted the offer; but when the room had been filled to the stipulated height, he informed the Inca that he was to remain in "protective custody." Pizarro proposed to use Atahualpa as a puppet ruler to ensure popular acceptance of the new order, but he became convinced that Atahualpa was organizing a resistance movement against the Spaniards. After a farcical trial, a Spanish court found the Inca guilty of polygamy, idolatry, and the murder of his brother Huáscar, and it condemned him to burn at the stake—a sentence commuted to strangling when he accepted baptism. Atahualpa's enormous ransom of gold was divided among the Spaniards, and Hernando Pizarro was sent to Spain with Emperor Charles' share of the plunder. Hernando's arrival in Spain with his load of gold caused feverish excitement, and a new wave of Spanish fortune hunters flowed to the New World.

Atahualpa's ransom of gold

Francisco Pizarro, posing as the defender of the legitimate Inca line, now proclaimed Huáscar's brother, Manco, as the new Inca. But a formidable revolt, organized and led by Manco himself, broke out in many parts of the empire. A large Indian army besieged Cuzco for ten months but failed to take the city. Defeated by superior Spanish weapons and tactics and by food shortages in his army, Manco retreated to a fastness in the Andean Mountains; there he and his successors maintained a kind of Inca government-in-exile until 1572, when a Spanish military expedition entered the mountains, broke up the imperial court, and captured the last Inca, Tupac Amaru, who was beheaded in a solemn ceremony at Cuzco.

The siege of Cuzco had barely ended when fighting broke out between one group of the conquerors, headed by the Pizarro brothers, and another, led by Diego de Almagro, over possession of the city. Before these struggles ended Francisco Pizarro had been murdered and two Almagros, father and son, had suffered death on the block. A new round of fighting began in 1544 when a new viceroy arrived in Peru to proclaim the protective Indian legislation known as the "New Laws of the Indies." Led by Gonzalo Pizarro, the desperate conquistadors rose in revolt. The rebellion collapsed after the arrival of a new crown envoy, who suspended the New Laws and offered pardons and rewards to all repentant rebels. Gonzalo Pizarro, however, resisted to the last and was captured and executed. Peace and order were not solidly established in Peru until the administration of Viceroy Francisco de Toledo, who arrived in 1569.

II. Spain's colonial empire

COLONIAL GOVERNMENT

Political organization

The political organization of the Spanish Empire in America reflected the centralized, absolutist regime by which Spain itself was governed. In the Indies, as in Spain, there was a frequent contrast between the formal concentration of authority in the hands of royal officials and the actual exercise of supreme power on the local level by the great landowners.

The pattern of Spain's administration of its colonies was formed in the period between 1492 and 1550. The final result reflected the steady growth of centralized rule in Spain and the application of a trial and error method to the problems of colonial government. To Columbus, Cortés, Pizarro and other great expeditionary leaders, the Spanish kings granted sweeping powers that made these men practically sovereign in the territories they had won or proposed to conquer. Once the importance of these conquests had been revealed, however, royal jealousy of the great conquistadors quickly appeared; their authority was soon revoked or strictly limited, and the institutions that had been used in Spain to achieve centralized political control were transferred to America for the same end. By the mid-16th century the political organization of the Indies had assumed the definitive form that it was to retain, with slight variations, until late in the 18th century.

The Council of the Indies. The Council of the Indies, chartered in 1524, stood at the head of the Spanish imperial administration almost to the end of the colonial period. Although great nobles and court favourites .were appointed to the Council, especially in the 17th century, its membership consisted predominantly of lawyers. Under the king, whose active participation in its work varied from monarch to monarch, it was the supreme legislative, judicial, and executive organ of colonial government. One of its most important functions was the nomination of all high colonial officials to the king. It also framed a vast body of legislation for the Indies—the famous Laws of the Indies, first codified in 1681—and sought to obtain detailed information on the history, geography, resources, and population of the colonies (the relaciones, which incorporated this information, represent a rich mine of materials for students of colonial Spanish America). Often staffed by conscientious and capable officials in the early Habsburg period, the quality of the Council's personnel tended to decline under the inept princes of the 17th century.

Principal royal agents

Viceroys, captains general, and audiencias. The principal royal agents in the colonies were the viceroys, the captains general, and the *audiencias* (high courts). Viceroys and captains general had essentially the same functions, differing only in the greater importance and extent of the territory assigned to the jurisdiction of the former; each was the supreme civil and military officer in his territory. At the end of the Habsburg era, in 1700, there were two great American viceroyalties—the viceroyalty of New Spain, with its capital at Mexico City, included all Spanish possessions north of the Isthmus of Panama; that of Peru, with its capital at Lima, embraced all of Spanish South America except the coast of Venezuela. Captains general, theoretically subordinate to the viceroys but in practice virtually independent of them, governed large subdivisions of these vast jurisdictions. Smaller subdivisions, called presidencias, were governed by audiencias, with the judge-president acting as governor but with military authority usually reserved to the viceroy.

A colonial viceroy enjoyed an immense delegated authority, which was augmented by the distance that separated him from Spain. By background he might be a lawyer or even a priest, but more often he came from one of the great noble and wealthy houses of Spain. A court modelled on that of Castile, a numerous retinue, and the constant display of pomp and circumstance testified to his exalted status. In theory his freedom of action was limited by the laws and instructions issued by the Council of the Indies, but recognition of the need to adapt the laws to existing circumstances gave him a vast discretionary power. The 16th century saw some able and even distinguished viceroys in the New World; in the 17th century, however, the quality of the viceroys declined (in 1695, for example, the viceroyships of Peru and Mexico were, in effect, sold to the highest bidders).

The audiencia

A viceroy or captain general was assisted in the performance of his duties by an audiencia, which served as his council of state. The joint decisions of viceroy and audiencia had the force of law, giving the audiencia a legislative character, roughly comparable to that of the Council of the Jndies, in relation to the king. Although the viceroy was not obliged to heed the advice of the audiencia, its immense prestige and its right to correspond directly with the Council of the Indies made it a potential check on the viceregal authority. The crown thus developed a system of checks and balances that assured ample deliberation and consultation on all important questions but that also encouraged indecision and delay.

Provincial government. Provincial administration in the Indies was entrusted to royal officials, who governed districts of varying size and importance from their chief towns and who usually held the title of corregidor. Some were appointed by the viceroy; others, by the crown. They possessed supreme judicial and political authority in their districts and represented the royal interest in the town councils (cabildos). If not trained as a lawyer, a corregidor was assisted by a legal counsel (asesor) in the trial of judicial cases. Certain civil and criminal cases could be appealed from the municipal magistrates (alcaldes) to the corregidor, and from him to the audiencia.

Corregidores were of two kinds. Some presided over Spanish towns; others, *corregidores* de indios, administered Indian towns (pueblos), which paid tribute to the crown. A principal duty of a corregidor de indios was to protect the natives against fraud or extortion on the part of whites, but the corregidor was himself the worst offender in this respect. Perhaps the worst abuse of this authority arose in connection with the practice of *repartimiento*—the mandatory purchase of goods from the corregidor by the Indians of his district. Originally designed to protect the Indians from the frauds of private Spanish traders, the cortegidor's exclusive right to trade with the Indians became a means for his speedy enrichment at the expense of the natives.

Restrictions on public officials. A series of regulations was designed to ensure good and honest performance on the part of public officials. Viceroys and oidores (members of an audiencia) were forbidden to engage in trade or own land within their jurisdictions or to accept gifts or fees; even their social life was restricted. All royal officials faced a judicial review (residencia) of their conduct at the end of their term of office. This took the form of a public hearing at which all who chose could appear before the "judge of residence" to present charges or testify for or against the official in question. At the end of the process the judge found the official guilty or innocent of all or part of the charges and handed down a sentence, which could be appealed to the Council of the Indies. Another device, the *visita*, was an investigation of official conduct, usually made unannounced, by a visitador specially appointed for this purpose by the crown; or, in the case of lesser officials, by the viceroy in consulation with the audiencia. As a rule, the visita was no more effective than the residencin in preventing or punishing official misdeeds.

Royal control of local officials

Municipal government. The only colonial institution that satisfied to some degree local aspirations for self-rule was the town council, known as the cabildo or *ayuntamiento*. At an early date, however, the crown assumed the right to appoint the councilmen (*regidores*) and municipal judges; under Philip II and his successors it became the established practice for the king to sell these posts to the highest bidder, with a right of resale or bequest, on condition that a certain portion of the price be paid to the crown as a tax at each transfer. Throughout the colonial period the municipal governments were self-perpetuating oligarchies of rich landowners, mineowners, and merchants who frequently received no salaries and who used their positions to distribute municipal lands to themselves, to assign themselves Indian labour, and to serve the narrow interests of their class. Vigilantly supervised by the corregidor, who frequently intervened in its affairs, the cabildo soon lost such autonomy as it may have possessed in the early days.

Other government offices. The officials and agencies described above represented a small part of the apparatus of colonial government. There were large numbers of such officers as secretaries (escribanos), police officers, tax collectors of "the royal fifth," and alcaldes with special jurisdiction. Under Charles V such offices were often in the gift of high Spanish officials, who sold them to persons who proposed to go to the Indies to exploit their fee-earning possibilities; beginning with Philip II, many of these offices were withdrawn from private patronage and sold directly by the crown, usually to the highest bidder. In the second half of the 17th century, the sale of offices spread from fee-earning positions to higher salaried posts; as a result, in this period corruption became structural in the government of the Indies.

Power of the landowners

Although royal authority—represented by viceroys, *oidores*, corregidores, and other officials—was more or less supreme in the capitals and the surrounding countryside, the same was not true of the more distant regions. In such areas the royal authority was very remote, and the power of the great landowners was virtually absolute. On their large, self-sufficient estates they dispensed justice in the manner of feudal lords, holding courts and imprisoning peons in their own jails, and raising and maintaining their own private armies. Sometimes these powerful individuals combined their de facto military and judicial power with an official title that made them representatives of the crown in their vicinities. The contrast between the nominal concentration of power in the central government and the effective supremacy of great landowners on the local level was a legacy of the colonial period to independent Latin America, and it still remains a characteristic of the political life of many Latin American republics.

SPAIN'S INDIAN POLICY

From the first days of the conquest the Spanish government faced a problem of defining its attitude toward the American natives (wrongly called Indians) and of determining what relations should exist between the conquerors and the conquered. The Indian question had several facets; the first and most urgent was to harmonize the demand of the conquistadors for cheap Indian labour—frequently employed in a wasteful and destructive manner—with the crown's interest in preserving a large tribute-paying Indian population. There was a political issue, too; the Spanish kings were determined to prevent a concentration of land and Indians in the hands of colonists that might lead to the rise of feudal lords independent of royal authority.

The church also had a major interest in the Indian problem. If the Indians died out, there would be no pagan souls to save, and the good name of the church would suffer. Moreover, the church largely relied on Indian labour for the construction and service of its churches and monasteries in the Indies. For these reasons the church, in general, sympathized with and aided crown efforts to protect the Indians against excessive exploitation.

The Spanish dispute over Indian policy quickly assumed the form of a struggle of ideas. Spanish thought of the 16th century was strongly scholastic in character, and jurists and theologians argued over such questions as the nature and cultural level of the Indians, whether they were a subhuman race who might properly be conquered and made to serve the Spaniards, and the rights and obligations the papal donation of America to the Spanish monarchs conferred upon them. Behind these disputations, however, was a struggle between the crown, the church, the colonists, and the Indians themselves over who should control Indian labour and tribute—the foundations of the Spanish Empire in America.

The *encomienda*

Española was the first testing ground of Spain's Indian policy. Eager to prove to the crown the value of his discoveries, Columbus compelled the natives to bring in a daily tribute of gold dust. Later, yielding to the demands of rebellious settlers, Columbus distributed the Indians among them, along with the right to use the forced labour of the natives. This temporary arrangement, formalized in the administration of Gov. Nicolás Ovando and sanctioned by the crown became the ***encomienda,*** which consisted of the assignment to a colonist of a group of Indians who were to serve him with tribute and labour, while he assumed the obligation of protecting his Indians, paying for the support of a parish priest, and helping to defend the colony. In practice the *encomienda* in the West Indies proved to be a hideous slavery that decimated the Indian population of Española.

The first protests against this state of affairs were made by a group of Dominican friars who arrived in Española in 1510. King Ferdinand responded to their agitation by approving a code of Spanish–Indian relations—the Laws of Burgos (1512–13). These laws contained detailed regulations prescribing good treatment of Indian labourers; but these provisions were not enforced, and they did little more than sanction and regularize the existing situation.

A former ***encomendero,*** Bartolomé de Las Casas (*q.v.*), joined the struggle against enslavement and mistreatment of the Indians. Las Casas argued that the papal grant of America to the crown of Castile had been made solely for the purpose of conversion, and it gave the Spanish king no temporal power or possession in the Indies; the Indians had rightful possession of their lands by natural law and the law of nations; all Spanish wars and conquests in the New World were illegal. Las Casas' mature program called for the suppression of all ***encomiendas,*** liberation of the Indians from all forms of servitude except a small tribute to the crown in return for its gift of Christianity, and even the restoration of the ancient Indian states and rulers. Over these states the Spanish king would preside as "emperor over many kings" in order to fulfill his sacred mission of bringing the Indians to the Catholic faith and the Christian way of life. Las Casas' proposals seemed radical, but objectively they served the royal aim of preventing the rise of a powerful colonial feudalism in the New World. Not humanitarianism but self-interest explains the partial official support that Las Casas' reform efforts received during the reign of Charles V (ruled 1516–56).

The New Laws of the Indies

The climax of royal intervention came with the proclamation of the New Laws of the Indies (1542), which appeared to doom the *encomienda.* They prohibited the enslavement of Indians, ordered the release of Indian slaves to whom legal title could not be proved, barred compulsory personal service by the Indians, regulated tribute, and declared that existing ***encomiendas*** were to lapse on the death of the holder.

The New Laws provoked a great revolt in Peru; in New Spain they caused a storm of protest by the *encomenderos* and a large part of the clergy. Under this pressure the crown again retreated; it reaffirmed the laws forbidding Indian slavery and forced labour, but the right of inheritance by the heir of an *encomendero* was recognized and extended by stages to a third, fourth. and, sometimes, even a fifth life. Thereafter, or earlier in the absence of an heir, the ***encomienda*** reverted to the crown. In the natural course of events, the number of *encomiendas* steadily diminished and that of crown towns increased.

Meanwhile, the economic value of the ***encomienda*** to the colonists was declining. They had lost the right to demand labour from their tributaries (1549); they had also lost their fight to make the *encomienda* perpetual. The heaviest blow of all, however, was the catastrophic decline of the Indian population in the second half of the 16th century. In New Spain, according to recent calculations, the Indian population dropped from about 25,000,000 in 1519 to slightly over 1,000,000 in 1605. Disease, especially disease of European origin, against which the Indians had no acquired immunity, was the major immediate cause; but overwork, social disorganization, and loss of will to live were largely responsible for the terrible mortality associated with the great epidemics and even with epidemic-free years.

As the number of their tributaries fell and their income declined proportionately, many *encomenderos* and other Spaniards began to engage in the more lucrative pursuits of agriculture, stock raising, and mining. The decline of the Indian population, sharply reducing the flow of foodstuffs to Spanish cities and mining centres, stimulated a rapid growth of Spanish estates (haciendas), producing grain and meat.

The *repartimiento*

A new system, the ***repartimiento—under*** which all adult male Indians had to give a certain amount of their time in rotation throughout the year to work in Spanish mines and factories, on farms and ranches, and on public works—replaced the ***encomienda.*** The crown thus sought to regulate the use of an ever-diminishing pool of Indian labour. The Indians received a token wage for their work, but the ***repartimiento*** was also essentially a disguised slavery. In Peru, where great numbers of Indians were conscripted for labour in the silver mines of Potosi and in the Huancavelica mercury mine, the ***repartimiento*** (there known as the ***mita)*** was particularly disastrous.

The ***repartimiento*** did not provide a dependable and continuing supply of labour, and Spanish ***hacendados*** turned increasingly to the use of so-called free labour. From the first, such labour was closely associated with the system of debt peonage, which helped to bind and hold workers in a time of rapid population decline. The heavy weight of tribute and *repartimiento* burdens on the ever-diminishing native population, and the contraction of Indian communal lands as a result of Spanish encroachments, caused many Indians to become farm labourers working for wages, mostly paid in kind. An advance of money or goods bound the peon to work for his employer until the debt was paid, a miracle that rarely occurred. Despite its later evil reputation, peonage had some advantages for many Indians; it usually freed an

Indian from the recurrent tribute and repartimiento burdens, and it often gave him a piece of land that he could work for himself and his family.

If the hacienda offered many Indians a means of escape from intolerable burdens, it aggravated the difficulties of those who remained on their ancestral lands. The hacienda expanded at the expense of the Indian pueblo, absorbing whole towns and leaving others without enough land for its people when the population decline ended in the first half of the 17th century and a slow recovery began. The hacienda lured labourers from the pueblo, making it difficult for the Indian town to meet its tribute and *repartimiento* obligations.

Debt servitude assumed its harshest form in the numerous workshops (*obrajes*), producing cloth and other goods, that developed in many areas in the 16th and 17th centuries. Convict labour, assigned to employers by Spanish judges, was early supplemented by the "free" labour of Indians, who were often lured into these workshops by an offer of liquor or a small sum of money. Once inside the gates, they were never let out again.

Negro slavery

Side by side with the disguised slavery of repartimiento and debt servitude existed Negro slavery. For a variety of reasons—including the fact that Spaniards and Portuguese were accustomed to holding Muslim Negro slaves and the belief that Negroes were better able to support the hardships of plantation labour—Spanish defenders of the Indian did not display the same zeal on behalf of the enslaved Africans. The rapid rise of sugar cane agriculture in the West Indies in the early 1500s brought an insistent demand for Negro slave labour to replace the vanishing Indian. There arose a lucrative slave trade, chiefly carried on by foreigners under a system of contract (asiento) between an individual or a company and the Spanish crown. The high cost of slaves limited their use to the more profitable plantation cultures or to domestic servitude in the homes of the wealthy.

A small class of genuinely free, paid workers also came into existence at an early date; it included resident and migrant farm workers, miners, unskilled urban labourers, and skilled workers who practiced their trades in the Spanish towns, sometimes as journeymen, in Spanish-controlled guilds from which Indians, mestizos (persons of mixed Indian and European ancestry) and mulattoes were excluded as masters. Except for skilled workers, wages tended to remain at the subsistence level throughout the colonial period.

THE COLONIAL ECONOMY

The Conquest disrupted the traditional economy of the Indians and transformed the character and tempo of Indian economic activity. When the frenzied scramble for treasure had exhausted the available gold and silver objects, the encomienda became the principal instrument for extracting wealth from the vanquished. The Aztec and Inca peoples were accustomed to paying tribute to their rulers and nobility, but the Spaniards' demands were unlimited. Driven by visions of infinite wealth, the Spaniards exploited the Indians mercilessly.

As noted above, the Indian population decline, causing acute food shortages in the Spanish towns, created new economic opportunities for Spanish farmers and ranchers; moreover, it left vacant large expanses of land, which Spanish colonists occupied for wheat raising or as sheep and cattle ranges. By the end of the 16th century the Spanish-owned haciendas produced the bulk of agricultural commercial production and pressed ever more aggressively on the shrinking Indian sector of the colonial economy. The establishment of an entail (mayorazgo) assured perpetuation of the consolidated property in the hands of the owner's descendants.

Spanish colonial agriculture

Spanish colonial agriculture early produced wheat on a large scale for sale in such urban centres as Mexico City, Lima, Veracruz, and Cartagena, and maize, for the large Indian consumers' markets in Mexico City, Lima, and other cities. Sugar was brought from the Canary Islands to Española and became the foundation of the island's prosperity. From the West Indies, sugar quickly spread to Mexico and Peru. Sugar refining, with its large capital outlays for equipment and Negro slaves, was the largest scale enterprise in the Indies.

Wine and olives as well as sugar were produced in quantity in the irrigated coastal valleys of Peru. The silk industry flourished briefly in Mexico, but it soon declined because of labour shortages and competition from Chinese silk brought from the Philippines to the port of Acapulco. Other products cultivated by colonists on a capitalist plantation basis were tobacco, cacao, and indigo. Cochineal, a blood-red dye, was a unique Mexican and Central American export highly valued by the European cloth industry.

Spain enriched American economic life by introducing various domestic animals—chickens, mules, horses, cattle, pigs, and sheep. The mules and horses revolutionized transport, and the cattle and smaller domesticated animals greatly enlarged the continent's food resources, while providing hides for export to Spain and other European centres of leather manufacture, and hides and tallow for the domestic market, especially in mining areas. Sheep raisers found a large wool market in the textile factories that arose in many parts of the colonies.

In populous central Mexico, cattle trampled the Indian crops, causing untold damage, and the close grazing of sheep caused massive erosion by torrential rain on the slopes of valleys. By the end of the 16th century, however, the Mexican cattle industry had become stabilized; the exhaustion of virgin pasture lands, the mass slaughter of cattle for their hides and tallow, and the official efforts to halt grazing on Indian harvest lands, greatly reduced the herds. Gradually, the cattle ranches and sheep herds moved from the densely settled south and central areas to new permanent grazing grounds in the semiarid north.

A rapid increase of horses, mules, and cattle also took place in the empty grasslands (pampas) of the Río de la Plata, modem Argentina. The inhabitants of this area, fcrbidden to trade directly with the outside world and lacking precious metals or abundant Indian labour, traded illegally with Dutch and other foreign traders, who carried their hides and tallow to Europe; they also sent mules and horses, hides and tallow, to the mining regions of Upper Peru (Bolivia). Another centre of the cattle industry was the West Indies.

Mining

Mining, as the principal source of royal revenue in the form of the royal fifth (quinto) of all gold, silver, or other precious metals obtained in the Indies, received the crown's special attention and protection. Silver was the principal mining product. The great silver mine of Potosí in Upper Peru was discovered in 1545 and produced enormous quantities of the metal between 1579 and 1635; the rich Mexican silver mines of Zacatecas and Guanajuato were opened up in 1548 and 1558, respectively. The introduction of the patio process for separating the silver from the ore with mercury (1556) gave a great stimulus to silver mining. In the same period, important gold placers were found in central Chile and in the interior of New Granada (Colombia). The mining industry brought prosperity to a few and failure to the great majority.

Inefficient production methods, lack of capital to finance technological improvements, flooding, and similar problems, led to a sharp decline of silver production from about 1630 to the end of the century. This mining crisis depressed trade with Europe, since silver was the most important export item in Spanish America's balance of trade, and also had an adverse effect on internal commerce. Colonial agriculture and stock raising, which had expanded to satisfy the demand of the mining centres for grain, meat, hides, tallow, and work animals, also suffered.

Throughout the colonial period the majority of the natives continued, as before the Conquest, to supply their own needs for pottery, clothing, and other household goods. In the Spanish towns craft guilds, modelled on those of Spain, arose in response to the high prices for all Spanish imported goods. These guilds maintained control over the quantity and quality of production in luxury industries serving the needs of the colonial upper class. The 17th century saw a rapid growth of obrajes, many of which produced cheap cotton and wool goods for popular consumption. Most were privately owned, but some were operated by Indian communities to meet their tribute payments. Other primitive factories produced such items as soap, chinaware, and leather. A 17th-century depression, restricting colonial capacity to purchase imports, promoted this growth of colonial industry.

COMMERCE, SMUGGLING, AND PIRACY

Colonial trade

Control over all colonial trade, under the Royal Council of the Indies, was vested in the Casa de Contratación (House of Trade), established in 1503 in Seville. Commerce with the colonies was restricted until the 18th century to the wealthier merchants of Seville and Cádiz, who were organized in a guild that exercised great influence in all matters relating to colonial trade. Trade was concentrated in the three American ports—Veracruz in New Spain, Cartagena in New Granada, and Nombre de Dios (later Portobelo) on the Isthmus of Panama. The Seville merchant oligarchy and related merchant groups in the Indies (particularly the merchant guilds at Mexico City and Lima) deliberately kept the colonial markets understocked and in general played into each other's hands at the expense of the colonists, who were forced to pay exorbitant prices for all European goods acquired legally. Inevitably, the system generated colonial discontent and stimulated the growth of contraband trade.

To enforce a closed-port policy and protect merchant vessels against foreign attack, an elaborate fleet system was developed in the 16th century to convoy ships between American ports and to and from Spain. In the 17th century, as a result of Spain's economic decline and growing contraband trade, the fleet sailings became increasingly irregular.

Spanish industry, handicapped by its guild organization and technical backwardness, could not supply the colonies with cheap and abundant manufactures in return for colonial foodstuffs and raw materials. Prices to the colonial consumer also were raised by a multitude of taxes. As a result, the more advanced industrial nations of northern Europe sought to break into the large and unsatisfied Spanish-American markets, rejecting Spain's claim of domination over all the Western Hemisphere except the portion that belonged to Portugal. The foreign challenge to Spain's monopoly assumed the forms of smuggling, piracy, and colonization, as well as efforts to seize and occupy portions of the Indies.

Foreign challenge to Spain

England, under Queen Elizabeth, emerged as the principal threat to Spain's empire in America. Sir John Hawkins' slave-trading voyage to the West Indies in 1562 opened the English drive to penetrate the Spanish-American market and culminated in the near destruction of Hawkins' trading fleet by a Spanish naval force at Veracruz in 1568. English voyages of reprisal followed; they included the expedition of Francis Drake (1577), undertaken with the secret sponsorship and support of Queen Elizabeth, whose objects were to seize Spanish treasure ships, ravage Spanish colonial towns, and display English maritime prowess through a second circumnavigation of the globe.

In the 17th century foreign piracy and smuggling were supplemented by efforts to found colonies not only on the American mainland but in the Caribbean. The Dutch, at war with Spain with brief intervals since 1576, launched a military and commercial offensive against the Spanish West Indies; their principal instrument was the Dutch West India Company, organized in 1621. The Dutch captured the whole homebound Veracruz treasure fleet off the coast of Cuba in 1628. Capture of Curaçao, hard off the coast of Venezuela (1634), gave the Dutch an invaluable smuggling base and emboldened the French and English to seize both unoccupied and occupied Spanish islands—Barbados and St. Christopher, Martinique and Guadeloupe. In 1655 an English fleet, defeated in an effort to capture Santo Domingo, easily captured Jamaica. In the same period French corsairs began to settle the northwest comer of Española, virtually abandoned by Spaniards since 1605; by 1655 this region had become the French colony of Saint-Domingue, with a governor appointed by the trading Compagnie des Indes.

In this period piracy in the West Indies became a highly-organized and large-scale activity often enjoying the open or covert protection of the English governors of Jamaica and the French governors of Saint-Domingue. Piracy entered a decline following the Treaty of Madrid (1670) between England and Spain, by which the British government agreed to aid in suppression of the corsairs in return for Spanish recognition of its sovereignty over the British West Indian islands. French buccaneers, however, continued active until the Treaty of Rijswijk (1697), by which Spain formally recognized French possession of Saint-Domingue.

Smuggling

Pirates and privateers, however, inflicted fewer losses on Spain than those caused by foreign smugglers. Contraband trade steadily increased in the 16th and 17th centuries; and the European establishments in Jamaica, Saint-Domingue, and the lesser Antilles became bases for contraband trade with the Spanish colonies. Buenos Aires was another funnel through which Dutch and other foreign traders poured immense quantities of goods that penetrated as far as Peru. Smuggling flourished even at Seville and Cádiz where, by the end of the 17th century, French companies operating behind the facades of Spanish merchant houses dominated the legal trade with the Indies.

Spanish economists of the 17th century understood the principal cause of Spain's plight: its economic weakness. They offered sound criticisms and constructive proposals for reform. But they were powerless to change the course of Spanish policy, dictated by small mercantile and aristocratic cliques whose special interests and privileges were incompatible with the cause of reform.

THE CHURCH IN THE INDIES

Royal control over church affairs, in both Spain and the Indies, was founded on the institution of royal patronage (the ***patronato real***). Under diplomatic pressure from King Ferdinand, in 1508 Pope Julius II accorded to Spain's rulers the exclusive right to nominate all Church officials, collect tithes, and found churches and convents in America, ostensibly to assist them in the work of converting New World pagans.

The missions. Beginning with Columbus' second voyage, one or more clergymen accompanied every expedition that sailed for the Indies; and they came in growing numbers to the conquered territories. The friars who came to America in the first decades after the Conquest were, in general, an elite group. The products of a revival of asceticism and discipline in the medieval church, and especially of a reform movement, this vanguard group frequently combined with missionary zeal a sensitive social conscience and love of learning.

Conversion of Indians

The friars converted prodigious numbers of natives. In Mexico, the Franciscans claimed to have converted more than 1,000,000 by 1531; when persuasion failed, pressures of various kinds, including force, were used to obtain conversions. (To facilitate the missionary effort the friars studied the native languages and wrote grammars and vocabularies. Scholars devoted themselves to preserving and recording the history, religion, and customs of the ancient Indian peoples.) The work of conversion was less than wholly successful. The result of the missionary effort was generally a fusion of pagan and Christian religious ideas. To this day Indians in such lands as Guatemala and Peru continue to perform rites dating from the time of the Maya and the Inca.

The clergy had to battle not only the Indian tendency toward backsliding but also divisions within their own ranks. A serious conflict was waged during the 16th century between the secular and the regular clergy. Finally, a royal decree of 1583 stated the principle that secular clergy were to be preferred over friars in all appointments to parishes. From first to last, however, the colonies were a scene of strife between groups of clergy over their fields of jurisdiction.

A gradual loss of a sense of mission and of morale among the regular clergy also contributed to the decline of their intellectual and moral influence. By the late 16th century there were frequent complaints against the church's excessive number of convents and growing wealth. The principal sources of this wealth were legacies and other gifts from rich donors; invested in land and mortgages, it brought in more wealth. The last important order to arrive in Spanish America, the Society of Jesus (1572), had the largest number of rich benefactors and the most efficient administration.

Material wealth

Inevitably, this concern with material wealth weakened the ties between the clergy and the Indian and mixed-blood masses. Hand in hand with a growing materialism went an increasing laxity of morals (except among the Jesuits); concubinage became so common among the clergy of the later colonial period that it seems to have attracted little official notice or rebuke.

The missionary impulse of the first friars survived longest on the frontier, "the rim of Christendom." Franciscans first penetrated the great northern interior of New Spain; they also accompanied the Juan de Oñate expedition of 1598 into what is now New Mexico and dominated the mission field there until the end of the colonial period; and they were also found in such distant outposts of Spanish power as what are now Florida and Georgia. The early Jesuits worked among the Indians of California, were active in converting and pacifying the tough Chichimeca Indians of the north central plateau of Mexico and had exclusive charge of the conversion of the Indian tribes of the northwest coast of Mexico. Following the expulsion of the Jesuits from the Indies in 1767, the Franciscans replaced them in directing missionary work in California.

The mission was one of three closely linked institutions —the other two being the presidio, or garrison, and the civil settlement--designed to serve the ends of Spanish imperial expansion and defense on the northern frontier. This three-pronged attack on the frontier was not very successful. Certain tribes on the northern frontier, such as the Apache of Arizona, New Mexico, and Texas and the Comanche of Texas, never were reduced to mission life. The missionaries had greater success among such sedentary tribes as the Pueblo Indians of New Mexico; but even among these peaceful tribes, Indian revolts and desertions were frequent. In 1680 the supposedly Christianized Pueblo Indians revolted, slaughtered the friars, and maintained a decade of tenacious resistance to Spanish efforts at reconquest. The civil settlements were no more successful. By the end of the colonial period there were only a few scattered towns on the northern frontier, and continuous Indian raids made life and property insecure. Ultimately, the whole task of defending and civilizing the frontier fell on a chain of presidios stretching approximately along the present border between the United States and Mexico. In the end Spain was forced to adopt a policy of neutralizing the Apache and Comanche by the periodic distribution of gifts to these warlike tribes. When the outbreak of the Wars of Independence stopped the flow of gifts, the hostile Indians drove through the useless line of presidios into the interior of Mexico.

Expulsion of the Jesuits

The Jesuits in Paraguay. The most successful missionary effort, at least from an economic point of view, was that of the Jesuit establishment in Paraguay, which included more than 30 missions or reductions; these formed the principal field of Jesuit activity in America. Their strict discipline, centralized organization, and absolute control over the labour of thousands of docile Indians producing cotton, tobacco, hides, and other products enabled the Jesuits to make their missions a highly profitable business enterprise. Every effort was made to limit contact with the outside world. The life of the Indians was rigidly regimented in dress, housing, and the routines of work, play, and rest. Jesuit mission activity in the colonies ended as a result of a royal decree (1767) ordering the expulsion of the order from the colonies. Motives for this action included the conflict between the nationalistic church policy of the Bourbons and Jesuit emphasis on papal supremacy, suspicion of Jesuit meddling in state affairs, and belief that the Jesuit mission system constituted a state within a state.

The Inquisition. The Inquisition formally entered the Indies with the establishment by Philip II of tribunals of the Holy Office at Mexico and Lima in 1569. Before that time its functions were performed by clergy who were vested with or assumed inquisitorial powers. Its great privileges, its independence of other courts, and the dread with which Spaniards generally regarded the charge of heresy made the Inquisition an effective check on "dangerous thoughts," whether religious, political, or philosophical. Most of the cases tried by its tribunals, however, dealt with offenses against morality or with such minor deviations from orthodox religious conduct as blasphemy. Like the Spanish Inquisition, the Inquisition in the Indies relied largely on denunciations by informers and employed torture to secure confessions. Indians were originally subject to the jurisdiction of Inquisitors but were later exempted because as recent converts of supposedly limited mental capacity they were not fully responsible for their deviations from the faith.

III. Colonial Brazil

THE FORMATION OF COLONIAL BRAZIL

Brazil's existence was unknown when the Treaty of Tordesillas of 1494, between Spain and Portugal, assigned to Portugal a large stretch of the South American coastline. In 1500 Pedro Álvares Cabral, who had been sent with a large fleet to India, was driven far off his course and touched on the Brazilian coast. Cabral claimed the land for Portugal and sent a report of his discovery to the king. Portugal's limited resources, already committed to exploitation of the wealth of Africa and the Far East, made impossible full-scale colonization of Brazil. But the presence of a valuable dyewood (brazilwood) attracted merchant capitalists, who obtained concessions to engage in the brazilwood trade with the Indians. A few settlers —some castaways, others degredados (criminals exiled from Portugal to distant parts of the empire)—arrived and were often well received by the local Indians. In 1532 the first Portuguese town in Brazil, named São Vicente, was founded near the present town of Santos.

Hereditary captaincies

Colonization. Portugal's heavy commitments in the spice-rich East forced its kings to assign to private individuals major responsibility for colonization of Brazil. The Brazilian coastline was divided into 15 parallel strips extending inland to the uncertain line of Tordesillas. These strips were granted as hereditary captaincies to a dozen individuals, each of whom agreed to colonize, develop, and defend his captaincy or captaincies at his own expense. The captaincy system represented a fusion of feudal and capitalist elements. The grantee was both a vassal owing allegiance to his suzerain, the king, and an entrepreneur who hoped to derive large profits from his estates and from taxes obtained from colonists to whom he gave land. Few of the captaincies proved successful because few grantees possessed the capital and administrative ability required to attract settlers and defend against Indian attacks and foreign intruders.

Slave hunters. By the mid-16th century sugar had replaced brazilwood as the foundation of the Brazilian economy. Favoured by its soil and climate, the Northeast (now the states of Pernambuco and Bahia) became the seat of a sugar cane civilization featured by the large estates (fazendas), monoculture, and slave labour. The problem of labour was first met by raids on Indian villages, but Indian labour was unsatisfactory because the natives worked poorly and offered resistance ranging from attempts at flight to suicide. After 1550, planters turned increasingly to the use of Negro slave labour imported from Africa; but the supply of Negro slaves was often cut off or sharply reduced by the activity of Dutch pirates and other foreign foes, and the services of Brazilian slave hunters were in large demand throughout the colonial period. The most celebrated slave hunters were the bandeirantes (from a word meaning "banner" or "military company") from the upland settlement of São Paulo; often part Indian, they made slave raiding in the interior their principal occupation. As the Indians near the coast dwindled in numbers or fled before the invaders, the bandeirantes pushed ever deeper southward and westward, expanding Brazil's frontiers in the process.

Jesuit missionaries were the first to protest the enslavement and mistreatment of the Indians. The Jesuit program for the settlement of their Indian converts in *aldeas* ("villages"), where they would live under the tutelage of the priests, provoked the slave hunters and the planters. The Portuguese crown pursued for two centuries a policy of compromise that satisfied neither Jesuits nor planters. A decisive turn came in the mid-18th century under the Marques de Pombal, the foreign minister and later prime minister of José I, who, in 1759, expelled the Jesuits from Portugal and Brazil and secularized their missions. Pombal forbade Indian raids and enslavement and thus accepted the Jesuit thesis of Indian freedom; but, unlike the Jesuits, his policy did not segregate the Indians from the Portuguese community; it made them available as paid workers by the colonists, and it encouraged mingling between the two races. Meanwhile, the growth of the African slave trade, also encouraged by Pombal, reduced the demand for Indian labour and this brought a greater measure of peace to the Indians.

French and Dutch incursions

Rivals to the Portuguese. The dyewood, sugar, and tobacco of Brazil early attracted foreign powers. The French made sporadic efforts to entrench themselves on the coast, and, in 1555, they founded Rio de Janeiro as the capital of what they called Antarctic France. But French colonization in Brazil was weakened by Catholic-Huguenot strife at home, and, in 1567, the Portuguese ousted the French and occupied Rio de Janeiro.

The Dutch posed a more serious threat to Portuguese sovereignty over Brazil. The Dutch West India Company seized and occupied (1630–54) the richest sugar-growing portions of the Brazilian coast. Ultimately, weakened by tenacious Brazilian resistance and a simultaneous struggle with England, the Dutch withdrew, taking their capital and the lessons they had learned in the production of sugar and tobacco and transferring both to the West Indies. Thus the Caribbean islands were soon competing with Brazilian sugar in the world market, with a resulting fall of prices, and by the end of the 16th century the Brazilian sugar industry was on the point of collapse.

Shift to the south. The discovery of gold (1690) in the southwestern region opened a new economic cycle and began a major shift of Brazil's economic and political centre from north to south. Large numbers of colonists from the Northeast, accompanied by their slaves and servants, swarmed into the mining area, causing an acute shortage of field hands in the older regions, which continued until the gold boom had run its course by the mid-18th century. In 1709 the mining region was elevated to the status of a captaincy, with the name Minas Gerais. A few years later came the discovery that certain stones in the area, hitherto thought to be crystals, were really diamonds; and many adventurers turned from gold to diamond washing.

By 1750 the river gold washings of Minas Gerais were nearly exhausted; the diamond district also suffered a progressive exhaustion of deposits. But the mineral cycle left a permanent mark in the form of new settlements in the southwest, not only in Minas Gerais but in what later became the provinces of Goiás and Mato Grosso—Brazil's far west. The mining decline also spurred efforts to promote the agricultural and pastoral wealth of the region. The southward movement of population and economic activity was formally recognized, in 1763 when Rio de Janeiro became the seat of the viceregal capital.

Meanwhile, the Northeast experienced a partial revival based on increasing European demand for sugar, cotton, and other semitropical products. Brazilian cotton production made significant advances between 1750 and 1800, but then declined rapidly because of competition from the more efficient cotton growers of the United States. The beginnings of the coffee industry also date from the late colonial period.

The cattle industry

Cattle raising contributed to the advance of the Brazilian frontier and to the growing importance of the south. The intensive commercial agriculture of the coast and the concentration of population in such coastal cities as Baia and Pernambuco created a demand for meat that gave an initial impulse to cattle raising. Because the expansion of plantation agriculture in the coastal zone did not leave enough land for grazing, the cattle industry had to move inland; and by the second half of the 17th century the penetration of the San Francisco Valley was well under way. Powerful cattlemen, with their herds, their *vaqueiros* ("cowboys"), and their slaves, entered the backcountry; drove out the Indians; and established fortified ranches and villages for their retainers. The cattle industry later expanded to the extreme southern region of Rio Grande do Sul, colonized by the government to defend against Spanish expansionist designs. The cattle industry provided meat for the coastal cities and mining camps, draft animals for the plantations, and hides for export to Europe.

Trade. During the Portuguese union with Spain (1580–1640), Brazil's commerce was restricted to Portuguese nationals and ships. The Dutch, who had been the principal carriers of Brazilian sugar and tobacco to European markets, responded with extensive smuggling and with a direct attack on the sugar-growing northeast. Following its successful revolt against Spain, Portugal made a trade treaty with England whereby British merchants were permitted to trade between Portuguese and Brazilian ports. English ships, however, frequently neglected the formality of touching at Lisbon and plied a direct trade with the colony. Because Portuguese industry was incapable of supplying the colonists with the requisite quantity and quality of manufactured goods, England provided cargoes of textiles and other manufactures. The decree of free trade of 1808, issued by Prince John following the flight of the Portuguese royal family from Portugal to Brazil after French invasion of his country, only confirmed the virtual English monopoly of trade with Brazil.

GOVERNMENT AND CHURCH

Colonial administration

The donatory system of government soon proved unsatisfactory, because few donatories could cope with the tasks of defense and colonization they had assumed. A governmental reform followed. In 1549 Tomé de Sousa was sent out as governor general to head a central colonial administration for Brazil; and Baia, situated about midway between the flourishing settlements of Pernambuco and São Vicente, became his capital. Gradually the hereditary rights and privileges of the donees were absorbed by governors appointed by the king. As the colony expanded, new captaincies were created. In 1763, as noted above, the governor of Rio de Janeiro replaced his colleague at Baia as head of the colonial administration with the title of viceroy. In practice, however, his authority over the other governors was negligible.

During the period of Spanish–Portuguese union their colonial policies were aligned by the creation, in 1604, of the Conselho da India, whose functions resembled those of the Spanish Council of the Indies. In 1736 a newly created ministry of Marinha e Ultramar (marine and overseas) assumed the functions of the Conselho; under the king, this body framed laws for Brazil, appointed governors, and supervised their conduct. The governor or viceroy combined in himself military, administrative, and even some judicial duties. His power tended to be absolute but was tempered by the constant intervention of the home government, which bound him with precise, strict, and detailed instructions; by the counterweights of other authorities, especially the high courts (*relações*), which were both administrative and judicial bodies; and the existence of special administrative organs, such as the intendencies created in the gold and diamond districts, which were completely independent of the governor. His authority was also diminished by the vastness of the country, the scattered population, the lack of social stability, and the existence of enormous landholdings in which the feudal power of great planters and cattle barons was virtually unchallenged.

The Senado de Camara, or municipal council, was the most important institution of local government. Elected either by a restricted property-owning electorate or chosen by the crown, its membership represented the ruling class of planters, merchants, and professional men; its authority was limited by frequent intervention of the royal judge (ouvidor), who usually combined his judicial functions with the administrative duties of *corregidor.* Generally speaking, the greater the size and wealth of the city, and the farther it was from the viceregal capital, the greater were its powers.

The evils of Spanish colonial administration—inefficiency, bureaucratic attitudes, slowness, and corruption—were also prominent in the Portuguese colonial system. In vast areas of the colony, however, administration and courts were virtually nonexistent. Outside the few large towns, local government often meant the rule of great landowners, for it was from their ranks that royal governors invariably appointed the *capitães môres,* or district militia officers. With unlimited power to enlist, command, arrest, and punish, the *capitão môr* became a symbol of despotism and oppression. (The feudalism that still dominates the Brazilian backcountry can be traced to these colonial origins.)

Some improvement, at least on the higher levels of administration, took place in the 18th century under Pombal, who abolished the remaining hereditary captaincies, reduced the special privileges of the municipalities, and increased the power of the viceroy. He sought to promote the economic advance of Brazil in order to rehabilitate Portugal, whose state was truly forlorn.

Church affairs

In Brazil, as in the Spanish colonies, church and state were intimately united. By comparison with the Spanish monarchs, however, the Portuguese kings were almost niggardly in their dealings with the church; but their control over its affairs, exercised through the *padroado* —the ecclesiastical patronage granted by the pope to the Portuguese king in his realms—was as absolute. Rome, however, maintained a strong indirect influence through the Jesuits, who were very influential in the Portuguese court until expelled from Portugal and Brazil in 1759.

With some honourable exceptions, notably that of the Jesuits, the tone of clerical morality and conduct in Brazil was low. The clergy provided such educational and humanitarian establishments as existed in the colony; and from its ranks—which were open to talents and even admitted individuals of mixed blood, despite the formal requirement of a special dispensation—came most of the distinguished names in Brazilian colonial science, learning, and literature.

MASTERS AND SLAVES

Race mixture played a decisive role in the formation of the Brazilian people. The scarcity of white women, the freedom of the Portuguese from puritanical attitudes, and the despotic power of great planters over Indian and Negro slave women, all gave impetus to miscegenation. Of the three possible combinations—white-Negro, white-Indian, Negro-Indian—the first was most common.

In principle, and to a considerable degree in practice, colour lines were strictly drawn. But the enormous number of mixed unions and the resulting large progeny, some of whom were regarded with affection by white fathers and provided with some education and property, inevitably led to some blurring of colour lines and to a fairly frequent phenomenon of "passing," with a tendency to classify individuals racially, if their colour was not too dark, on the basis of social and economic position rather than by physical appearance.

Slavery in Brazil

Slavery played as important a role in Brazil's organization as did race mixture in its ethnic make-up. Slavery corrupted both master and slave, festered harmful attitudes toward the dignity of labour, and retarded the nation's economic development. The virtual monopoly of labour by slaves sharply limited the number of socially acceptable occupations in which whites or free mixed bloods could engage. This gave rise to a large class of vagrants, beggars, "poor whites," and other degraded elements who would not or could not compete with slaves in agriculture and industry. Given the lack of incentive to work on the part of the slave, the level of efficiency and productivity of his labour was very low.

The 20th-century Brazilian sociologist Gilberto Freyre has emphasized the patriarchal relations between masters and slaves in the sugar plantation society of the Northeast. But the slaves described by Freyre were usually house slaves, who occupied a privileged position and whose situation was different from that of the great majority of slaves, who worked on the sugar and tobacco plantations of Bahia and Pernambuco. A royal dispatch of 1700 denounced the barbarity with which owners of both sexes treated their slaves. The very low rate of reproduction among slaves and frequent suicides speak eloquently about their condition; many slaves ran away and formed *quilombos* (settlements of fugitive slaves) in the bush. The most famous of these was the so-called republic of Palmares, whose destruction required a major military campaign in the 17th century.

The nucleus of Brazilian social and economic organization was the *fazenda,* based on Negro slavery and centred about the *casa grande* (the "big house"); it constituted a patriarchal community that included the owner and his family, his chaplain and overseers, his slaves, his sharecroppers *(obrigados)*, and his *agregados,* or retainers. The system implied relations of mutual aid and a paternalistic interest in the welfare of the landowner's people; but it did not exclude intense exploitation of those people or the display of the most ferocious cruelty if they should dispute his absolute power.

Planter aristocracy

In the sugar-growing northeast, the great planters became a distinct aristocratic class. Most colonial towns were mere appendages of the countryside, dominated politically and socially by the rural magnates. These men often left supervision of their estates to majordomos and overseers, preferring to live in the cities. In the cities lived other social groups that disputed or shared power with the great landowners: high officials of the colonial administration; dignitaries of the church; wealthy professional men, especially lawyers; and merchants, usually European-born, who monopolized the export–import trade and financed the industry of the planters. The conflict between native-born landowners and European-born merchants, aggravated by nationalist resentment of upstart immigrants, sometimes led to armed struggle. An illustration is the War of the Mascates (1710–11) between Olinda, provincial capital of Pernambuco, dominated by the sugar planters, and its neighbouring seaport of Recife, controlled by the merchants.

IV. The Bourbon reforms and Spanish America

REFORMS

After the War of the Spanish Succession (1702–13), Spain was left with a more manageable, more truly Spanish, empire consisting of the kingdom of Castile and Aragon, and the Indies, and then turned its attention to implementing a program of reform. The ensuing revival of Spain is associated with three princes of the House of Bourbon—Philip V (ruled 1700–46) and his two sons, Ferdinand VI (ruled 1746–59) and Charles III (*q.v.*; ruled 1759–88). The work of national reconstruction reached its climax under Charles III, who attempted to revive Spanish industry and agriculture.

The outbreak of the French Revolution, which followed by a few months the death of Charles III in December 1788, brought the reform era effectively to a halt. Frightened by the overthrow of the French monarchy, Charles IV and his ministers turned sharply to the right; the leading reformers were banished or imprisoned, and the importation of French rationalist and revolutionary literature was forbidden. Yet the clock could not be entirely turned back, either in Spain or its colonies. Under Charles IV, for example, an expedition sailed from Spain (1803) to carry the procedure of vaccination to the Spanish dominions in America and Asia.

With colonial reform the Bourbons moved slowly and cautiously, because of the powerful vested interests identified with the status quo. There was no intention of giving more self-government to the colonists or of permitting them to trade more freely with the non-Spanish world. On the contrary, the Bourbons centralized colonial administration still further, with a view to making it more efficient, and their commercial reforms were designed to diminish smuggling and strengthen the exclusive commercial ties between Spain and its colonies.

Commercial and administrative reforms

Under the first Bourbons, efforts were made to check smuggling in the Caribbean by the use of ***guardacostas,*** which prowled the main lanes of trade in search of ships loaded with contraband. The depredations of these ***guardacostas*** led to English demands for compensation and finally to war between England and Spain in 1739.

The first Bourbons made few changes in the administrative structure of colonial government, contenting themselves with efforts to improve the quality of administration by more careful selection of officeholders. One major reorganization was the separation of the northern Andean region (present-day Ecuador, Colombia, and Venezuela) from the viceroyalty of Peru and its elevation to the status of a new viceroyalty, named New Granada, with its capital at Santa Fé (modern Bogota). This change reflected a desire to provide better protection for the Caribbean coast and especially the fortress of Cartagena; it also reflected the rapid growth of population in the central highlands of Colombia. Within the new viceroyalty, Venezuela was named a captaincy general, with its capital at Caracas and virtually independent of Santa Fé.

Colonial reform, like domestic reform, reached its peak under Charles III. In 1765 commerce with the West Indies was thrown open to seven ports besides Cádiz and Seville; this, coming at a time when Cuban sugar production was beginning to expand, stimulated the island's economy. This privilege was extended to other regions until, by a decree of free trade of 1778, commerce was permitted between all qualified Spanish ports and all the American provinces except Mexico and Venezuela, which were opened to trade on the same terms in 1789. Restrictions on intercolonial trade were also progressively lifted, but this trade was largely limited to non-European products. A major beneficiary of this was the Rio de la Plata area, which, in 1776, was opened to trade with the rest of the Indies. Meanwhile, the Casa de Contratación steadily declined in importance until it closed its doors in 1790. A similar fate overtook the venerable Council of the Indies; most of its duties were entrusted to a powerful colonial minister appointed by the king.

Failure of colonial reforms

The entrance of new trading centres and merchant groups into the Indies trade, the reduction of duties, and the removal of irksome restrictions, had the effect of increasing commerce, reducing prices, and perhaps diminishing contraband. But the Bourbon commercial reform ultimately failed in its aim of reconquering colonial markets for Spain for several reasons—first, the opposition of the still dominant merchant oligarchs of Cádiz, who resisted intercolonial trade and efforts to replace French and English manufactures in the export trade to the colonies with noncompetitive Spanish products; second, Spain's continuing industrial weakness, which the best efforts of the Bourbons could not overcome; and, third, Spain's closely related inability to keep its sea-lanes to America open in time of war with England, when foreign traders again swarmed in Spanish American ports.

Colonial economic growth. Perhaps the most significant result of the Bourbon commercial reforms was the stimulus given to economic activity in Spanish America. To what extent this economic growth should be ascribed to the Bourbon reforms and to what degree it resulted from the economic upsurge in western Europe in the 18th century cannot be stated with certainty. Stimulated by the Bourbon reforms and a growing European demand for sugar, coffee, tobacco, hides, and other staples, production of these products rose sharply. But this increase in agricultural production resulted from more extensive use of land and labour rather than from the use of improved implements or techniques. The inefficient ***latifundio*** (a large, poorly cultivated estate), using poorly paid peon labour, and the slave plantation accounted for the bulk of agricultural production. Such natural disasters as drought or excessive rains easily upset the precarious balance between food supplies and population, producing frightful famines (as in 1785–87 in central Mexico, when thousands died of hunger or diseases).

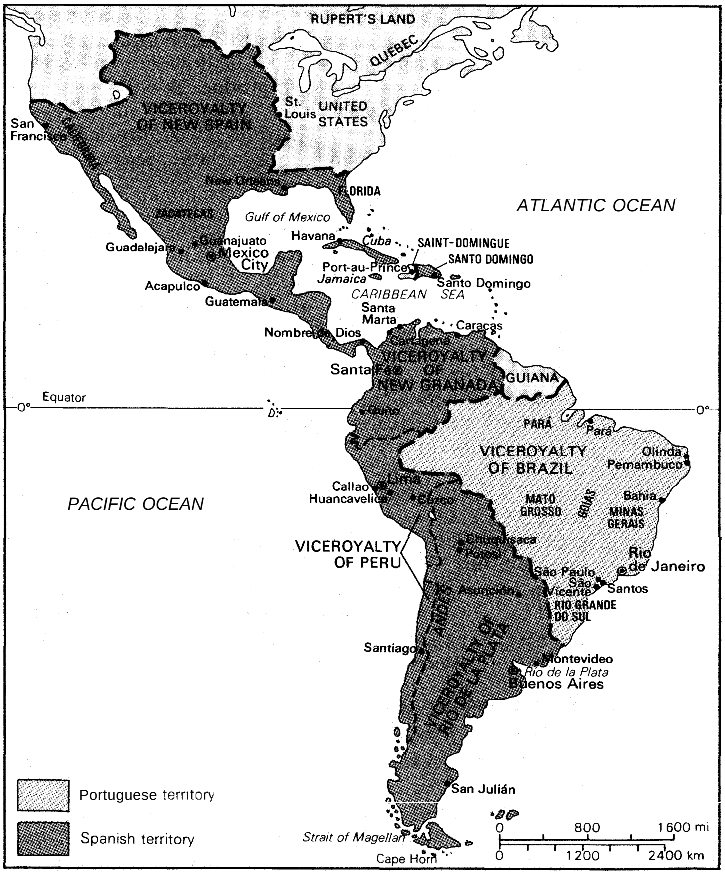
What sugar, cacao, and coffee were for the Caribbean area, hides were for the Río de la Plata. Rising European demand for leather and the opening of direct trade with Spanish ships in 1735 sparked an economic upsurge in the Plata area. By the end of the 18th century the ***estancias*** (cattle ranches) were often huge, with as many as 80,000 or 100,000 head of cattle. Meat gained in value as a result of the demand for salt beef. Markets for salt beef were found above all in the Caribbean area, especially Cuba, chiefly for feeding the slave population. The growth of cattle raising in La Plata was attended by concentration of land in ever fewer hands and took place at the expense of agriculture, which remained in a depressed state.

Silver mining revival

The 18th century also saw a strong revival of silver mining in the Spanish colonies. Peru and Mexico shared in this advance, but the Mexican mines forged far ahead of their Peruvian rivals. The crown, especially under Charles III, made serious efforts to encourage the industry by such means as tax reductions and the dispatch of foreign and Spanish experts to Mexico and Peru. These efforts were largely frustrated by the traditionalism of the mine owners, lack of capital to finance changes, and mismanagement. Yet the production of silver steadily increased; supplemented by the gold of Brazil, it helped to spark the industrial revolution in northern Europe and stimulated commercial activity on a worldwide scale. More especially, American silver helped the Bourbons meet the enormous expenses of their chronic wars.

Colonial manufacturing, on the other hand, began to decline late in the 18th century, principally because of competition from cheap foreign wares. The textile and wine industries of western Argentina and the textile production of Quito in Ecuador decayed as they lost their markets to lower-priced foreign wines and cloth. In the Mexican city of Puebla, production of chinaware slumped catastrophically between 1793 and 1802. Although Spain adopted mercantilist legislation designed to restrict colonial manufacturing, this legislation seems to have been a small deterrent to the growth of large-scale manufacturing. More important were lack of investment

Reprinted by permission of the publisher, from *People and Politics of Latin America*, copyright © 1955 by Ginn, 1945 by Ginn & Co., 1938 by Mary W. Williams, 1930 by Mary W. Williams; all rights reserved



Spanish and Portuguese America in 1784.

capital, preference for land and mining as fields of investment, and a semiservile system of labour that was harmful equally to the workers and to productivity.

DECENTRALIZATION

Under Charles III the work of territorial reorganization continued. The viceroyalty of Peru was further curtailed by the creation (1776) of the viceroyalty of the Rio de la Plata, with its capital at Buenos Aires. In 1783 the establishment of a royal *audiencia* at Buenos Aires completed the liberation of the Plata provinces from the distant rule of Lima. The inclusion of Upper Peru in the new viceroyalty, with the resulting redirection of the flow of Potosi silver from Lima to Buenos Aires, signified a major victory for the landlords and merchants of Buenos Aires over their mercantile rivals of Lima.

The trend toward decentralization reflected not only a struggle against foreign military and commercial penetration but also an awareness of the problems of communication and government posed by the great distances between the various provinces. One indication of this was the greater autonomy and the increased number of the captaincies general in the 18th century. Venezuela was raised to a captaincy general in 1777, Chile, in 1778.

The intendant system

Between 1782 and 1790 the intendant system was introduced to the colonies. The intendants—provincial governors who ruled from the capitals of their provinces—were expected to relieve the overburdened viceroys of many of their duties, especially in financial matters. The offices of *corregidor* and *alcalde mayor,* holders of which were notorious as oppressors of the Indians, were replaced in Indian districts by that of *subdelegado,* nominated by the intendants and confirmed by the viceroys. Many intendants at the height of the reform era were capable, cultivated men who not only worked to increase economic activity and revenue collection but also promoted education and other cultural projects. Most *subdelegados,* however, soon became as notorious as their predecessors for their oppressive practices; a common complaint was that they compelled the Indians to trade with them, although forbidden by the Ordinance of Intendants. The failure of the Indian and mixed-blood masses to profit by the 18th-century economic advance—whose principal beneficiaries were Creole (American-born Spaniards) landowners, mine owners, and merchants—helped to produce popular revolts of 1780–81 (of Tupac Amaru in Peru and of the *comuneros* in New Granada), which, although suppressed, shook Spanish power to its foundations in those areas.

DEFENSE OF THE EMPIRE

Increased revenue was a major objective of the Bourbon commercial and political reforms, needed for strengthening the sea and land defenses of the empire. Spanish losses in the Seven Years' War (1756–63), and especially the loss of Havana and Manila to the English (1762), resulted in efforts to improve the defense system of the colonies. Fortifications of important American ports were strengthened and colonial armies were created. Regular units were stationed permanently in the colonies or rotated between peninsular and overseas service, and a colonial militia was filled by volunteers or conscripts. To make military service attractive to the Creole upper class, which provided the officer corps of the new force, the crown granted extensive privileges and exemptions to Creole youth who accepted commissions, thus adding to the lure of prestige and honours protection from civil legal jurisdiction and liability, except for certain specified offenses. This led to the development of a special officer class, which survives in many Latin American countries.

V. The wars of independence

BACKGROUND FOR THE WARS

The Bourbon reforms combined with the growing European demand for colonial products in the 18th century to bring material prosperity and other benefits to the upper class Creoles. The improvements and refinements introduced by enlightened viceroys and intendants made life in the colonial cities more attractive; educational reforms and opportunities widened the intellectual horizons of Creole youth. These advances, however, only enlarged Creole aspirations and discontents. Increased production strained against the trade barriers maintained by Spanish mercantilism. Awareness of their economic importance made intolerable to the Creoles their virtual exclusion from commanding posts in government and the church.

Creole discontent

The conflict of interests between Spain and its colonies found expression in the cleavage between Creoles and peninsular Spaniards. This cleavage was a major cause of the Spanish American wars of independence.

Enlightenment and revolutionary influences. Enlightenment ideas contributed to Creole restlessness and discontent. The forbidden writings of G.T. Raynal, Montesquieu, Voltaire, and Rousseau undoubtedly were read by educated Creoles; and scientific works based on the premises of Descartes, Leibnitz, and Newton circulated freely in the colonies and helped to spread Enlightenment ideas. By 1800 such influences had partly renovated the intellectual climate of Spanish America and had given a rationalist, pragmatic stamp to Creole thought.

The United States War of Independence contributed to the growth of "dangerous ideas" in the Spanish colonies. Spain was aware of the ideological as well as the political threat to its empire posed by the United States. After 1783 a growing number of U.S. ships touched legally or illegally at Spanish American ports, sometimes introducing such subversive documents as the writings of the revolutionists Thomas Paine and Thomas Jefferson.

The French Revolution exerted greater influence on the Creole mind. The Colombian Antonio Nariiio translated and printed the French Declaration of the Rights of Man of 1789; he was sentenced to prison in Africa for ten years but later led the successful independence movement in Colombia. The French Revolution soon took a radical turn, and the Creole elite became disenchanted with it as a model. The most important direct result of the French Revolution was a slave revolt in the French part of Haiti under Toussaint-Louverture (*q.v.*) and other black leaders; by January 1, 1804, black revolutionaries in Haiti had established the first liberated territory in Latin America, but their achievement dampened rather than aroused support for independence among the Creole elite of the colonies (see HAITI, HISTORY OF).

Beginnings of the independence movement. In 1806 the revolutionary Francisco de Miranda landed on the shores of his native Venezuela with a force of some 200 foreign volunteers, but his call for an uprising met with no response, and he hastily withdrew. The colonial independence movement might have remained ineffectual if not for decisions and actions by European powers with very different ends in view.

Spanish–French alliance

A major cause of the revolutionary crisis was Spain's involvement in the European wars unleashed by the French Revolution. Spain became France's ally in 1796, and English sea power promptly drove Spanish shipping from the seas, virtually cutting off communication between Spain and its colonies. The alliance with France had other results. An English fleet sailed with a regiment of soldiers against Buenos Aires (1806). The English soldiers entered Buenos Aires meeting only token resistance, but Creoles and peninsular Spaniards soon rallied to expel their unwanted liberators; a volunteer army attacked and routed the occupation force, capturing the English general and 1,200 of his men. A second British invasion was beaten back with heavy losses. Impressed by the tenacious defense, the British commander agreed to evacuate both Buenos Aires and the town of Montevideo. With this victory, the Creoles of Buenos Aires had tasted power and would not willingly relinquish it again.

In Europe, Napoleon gradually reduced Spain to a helpless satellite. In 1807 Napoleon obtained from Charles IV permission to invade Portugal through Spain; and the Portuguese royal family and court escaped to Brazil in a fleet under British convoy. Popular resentment at the French presence in Spain forced Charles IV and his son Ferdinand to abdicate; Napoleon tried to place his brother Joseph on the Spanish throne.

An insurrection against the French occupation forces began in Madrid and spread across Spain beginning in 1808; it led to the formation of a national Cortes, or parliament, which met in Cadiz from 1810 to 1814 under the protection of English naval guns. The constitution, approved by the Cortes in 1812, provided a limited monarchy and freedom of speech and assembly, and it abolished the Inquisition. But the Cortes made few concessions to the American colonies; it invited American delegates to join its deliberations but made clear that the system of peninsular domination and commercial monopoly would remain essentially intact.

Creole plans for gaining power

In Spanish America events had transformed the remote prospect of independence into a realistic goal. Confident that the French armies would crush all opposition, Creole leaders prepared to take power with the pretext of loyalty to the "beloved Ferdinand." The confusion caused among Spanish officials by the arrival of rival emissaries who proclaimed Ferdinand or Joseph Bonaparte the legitimate king of Spain aided the Creole plans. In the spring of 1810, with the fall of Cádiz seemingly imminent, the Creole revolutionaries moved into action; charging viceroys and other royal officials with doubtful loyalty to Ferdinand, they organized demonstrations in Caracas, Buenos Aires, Santiago, and Bogotá that forced these authorities to surrender control to Creole-dominated local juntas. Their hopes of a peaceful transition to independence, however, were doomed to failure; their claims of loyalty did not deceive the groups truly loyal to Spain, and fighting soon broke out between patriots and loyalists.

THE INDEPENDENCE OF SOUTH AMERICA

The Latin American struggle for independence lacked a unified direction or strategy that was caused not only by the vast distances and other geographical obstacles to unity, but also by the economic and cultural isolation of the various Latin American regions. Moreover, the Latin American movement for independence lacked a strong popular base—the Creole elite, itself part of an exploitative white minority, feared the Indians, Negroes, and oppressed castes and usually sought to keep their intervention in the struggle to a minimum. This lack of regional and class unity helps explain why it took Latin America so long to achieve independence.

The struggle for independence had four main centres. In Spanish South America there were two principal theatres of military operations--one flowed southward from Venezuela, the other ran northward from Argentina; these two currents joined at Peru, the last Spanish bastion on the continent. The third centre, Brazil, achieved its own swift and relatively peaceful separation from Portugal. Finally, Mexico had to travel a difficult, devious road toward independence.

Bolivar's involvement in the movement

Spanish South America. Simón Bolivar of Venezuela is the symbol and hero of the struggle for independence in northern South America. Soon after his return to Venezuela from a visit to Europe (1804–07), he became involved in conspiratorial activity directed at overthrow of the Spanish regime. In April 1810, the Creole party in Caracas forced the captain general to abdicate, and a Creole-dominated junta took power. In 1811 a Venezuelan congress proclaimed the country's independence and framed a republican constitution that abolished special privileges and Indian tribute, but retained Negro slavery and made Catholicism the state religion. Meanwhile, the veteran revolutionary Francisco de Miranda had returned from England and assumed command of the patriot army.

Differences soon broke out between Miranda and his young officers, especially Bolivar. Amid these disputes came the earthquake of March 26, 1812, which caused great loss of life and property in Caracas and other patriot territory, but spared the regions under Spanish control. The royalist clergy proclaimed this disaster a divine retribution against the rebels. A series of military reverses completed the discomfiture of the revolutionary cause. With his forces disintegrating, Miranda negotiated a treaty with the royalist commander and then tried to flee the country, taking with him part of the republic's treasury. Bolivar and some of his comrades, regarding Miranda's act a form of treachery, seized him before he could embark and turned him over to the Spaniards. Miranda died in a Spanish prison four years later.

Bolivar was saved from a Spanish reaction by the influence of a family friend. Under a safe conduct, he departed for Colombia, which was still partially under patriot control. Given command of a small force to clear the Magdalena River of enemy troops, he employed a strategy featured by swift movement and aggressive tactics; he also judged his soldiers on merit without regard to social background or colour. His success gained Bolivar the rank of general in the Colombian army and won him the approval of a plan of his for liberation of Venezuela.

In a forced march of three months, Bolivar led a force of 500 men across jungles and swamps toward Caracas; as he approached the capital, the Spanish forces withdrew. He entered Caracas in triumph and received from the city council the title of "liberator"; soon afterward the congress of the restored republic voted to grant him dictatorial powers.

Bolivar's success was short-lived. The fall of Napoleon, in 1814, brought Ferdinand VII to the Spanish throne, released Spanish troops for use in America, and heartened the royalists. Meanwhile, the ***llaneros,*** or cowboys, of the Venezuelan plains joined the royalist cause. A mass of ***llaneros*** invaded Caracas, crushing all resistance; and in July 1814 Bolivar hastily abandoned the city and retreated toward Colombia with the remains of his army.

Bolivar found Colombia on the verge of chaos; despite the imminent threat of a Spanish invasion, the provinces quarrelled with each other and defied the authority of the weak central government. Concluding that the situation was hopeless, Bolivar left in May 1815 for the British island of Jamaica. Meanwhile, a Spanish army landed in Venezuela, completed the reconquest of the colony, and then sailed to lay seige to Cartagena; the city surrendered in December, and the rest of Colombia was pacified within a few months. Of all the Spanish American provinces, only Argentina remained in revolt.

Bolivar's "Letter from Jamaica"

Bolivar had an unshakeable faith in the triumph of independence. From Jamaica he wrote a famous document --"The Letter from Jamaica"—in which he affirmed his faith and argued that monarchy was foreign to the genius of Latin America; only a republican regime would be accepted by its peoples. Bolivar boldly forecast the destiny of the different regions, taking account of their economic and social structures (Chile, for example, seemed to him to have a democratic future; Peru, on the other hand, was fated to suffer dictatorship because it contained gold and slaves).

From Jamaica, Bolivar went to Haiti, where he received a sympathetic hearing and the offer of some material support. After two efforts to gain a foothold on the Venezuelan coast failed, Bolivar decided to establish a base in the Orinoco River Valley, distant from the centres of Spanish power. Roving patriot bands still operated in this region, and Bolivar hoped to win the allegiance of the ***llaneros,*** who were becoming disillusioned with their Spanish allies. In September 1816 Bolivar sailed for the Orinoco River Delta and made his headquarters at the town of Angostura (modern Ciudad Bolivar).

The patriot guerrilla bands accepted Bolivar's leadership, and he also gained the support of the ***llaneros.*** The end of the Napoleonic Wars had idled a large number of British soldiers; and many of these veterans came to Venezuela, forming a British legion that distinguished itself in battle on the patriot side. Bolivar was also helped by English merchants, who made loans that enabled him to secure men and arms, and by the mulish attitude of Ferdinand VII, whose refusal to make any concessions to the colonists caused the British government to lose patience and regard favourably the prospect of Spanish American independence.

Liberation of Colombia and Venezuela

On the eve of the decisive campaign of 1819, Bolivar summoned to Angostura a congress that vested him with dictatorial powers. His strategy for the liberation of Venezuela and Colombia was to strike a blow at Spanish forces from a completely unexpected direction; he advanced with an army of some 2,500 men along the Orinoco and Arauco rivers across the plains and then ascended the Colombian Andes until he reached the plateau where lay Bogota. On the field of Boyacá the patriot army surprised and defeated the royalists in a short, sharp battle and entered Bogotá. He then prepared for the liberation of Venezuela; in June 1821 patriot troops crushed the last major Spanish force in Venezuela at Carabobo. Save for some coastal towns and forts held by beleaguered royalists, Venezuela was free.

The independence of Spanish America remained precarious as long as the Spaniards held the immense mountain bastion of the central Andes. While Bolivar prepared an offensive from Bogota against Quito, he sent his lieutenant José Antonio de Sucre by sea from Colombia's Pacific coast to seize the port of Guayaquil. Even before Sucre arrived, the Creoles of Guayaquil revolted, proclaimed independence, and placed the port under Bolivar's protection. Advancing into the Ecuadorian highlands, Sucre defeated a Spanish army on the slopes of Mount Pichincha, near Quito. Royalist resistance crumbled on news of Sucre's victory; and the provinces composing the viceroyalty of New Granada — the future republics of Venezuela, Colombia, Ecuador, and Panama — were free. They were temporarily united into a large state named Great Colombia (Gran Colombia), established at the initiative of Bolivar by the union of New Granada and Venezuela in 1821.

Liberation of Argentina

Ever since the defeat of the British invasions of 1806–07, the Creole party, nominally loyal to Spain, had effectively controlled Buenos Aires. In May 1810, when word came that French troops had entered Seville and threatened Cádiz, an open town meeting convened to decide the future government of the colony. This first Argentine congress voted to depose the viceroy and establish a junta to govern in the name of Ferdinand. In 1813 a national assembly gave the country the name United Provinces of the Rio de la Plata and enacted such reforms as the abolition of *mita, encomienda*, titles of nobility, and the Inquisition. A declaration of independence, however, was delayed until 1816.

The junta promptly attempted to consolidate its control over the vast viceroyalty. The western interior provinces were subdued after sharp fighting. Montevideo, across the Rio de la Plata, remained in Spanish hands until 1814, when it fell to an Argentine siege. The junta met even more tenacious resistance from the gauchos of the Uruguayan pampa, led by Jose Gervasio Artigas, who demanded Uruguayan autonomy in a loose federal connection with Buenos Aires; but Artigas became caught between pressure from Buenos Aires and Portuguese forces who claimed Uruguay for Brazil, and he had to flee to Paraguay (Uruguay did not achieve complete independence until 1828). The Creoles of Paraguay defeated a force sent from Buenos Aires to liberate Asunción and proceeded to depose Spanish officials and proclaim the independence of Paraguay.

The junta's efforts to liberate the mountainous province of Upper Peru failed because of steep terrain, long lines of communication, and the apathy of the Indian population.

Upper Peru in Spanish hands represented a standing threat to the security of the La Plata provinces. The military genius of José de San Martin of Paraguay offered a solution to the problem. San Martin was a colonel in the Spanish army when revolution broke out in Buenos Aires. He promptly sailed for La Plata to join the patriot junta and was soon raised to command the army of Upper Peru. San Martin proposed a march across the Andes to liberate Chile, where a Spanish reaction had toppled the revolutionary regime established by Bernardo O'Higgins and other patriot leaders in 1810. Having liberated Chile, the united forces of La Plata and Chile would descend upon Peru from the sea.

San Martin's liberation of Chile and Peru

San Martin obtained an appointment as governor of the province of Cuyo, whose capital, Mendoza, lay at the eastern end of a strategic pass leading across the Andes to Chile. He spent two years recruiting, training, and equipping his Army of the Andes. The army began the crossing of the cordillera in January 1817, and in 21 days it issued on Chilean soil. A decisive victory at Chacabuco in February opened the gates of Santiago to San Martin. Another victory at Maipú (1818) ended the threat to Chile's independence. Rejecting Chilean invitations to become supreme ruler of the republic, a post assumed by O'Higgins, San Martin secured a number of ships in England and the United States and, in August 1820, sailed for Peru in a fleet of seven ships of war and 18 transports. He landed his army about 100 miles (160 kilometres) south of Lima but delayed moving on the Peruvian capital; he hoped to obtain its surrender by economic blockade, propaganda, and direct negotiation with Spanish officials. His strategy was successful; in June 1821 the Spanish army evacuated Lima and retreated toward the Andes. San Martin entered the capital and proclaimed the independence of Peru. But he then had to deal with counterrevolutionary plots and the resistance of Lima's elite to his program of social reform. Meanwhile. a large Spanish army manoeuvred in front of Lima, challenging San Martin to a battle which he dared not join with his much smaller force. San Martin became convinced that only monarchy could bring stability to Spanish America, and he sent a secret mission to Europe to find a prince for the throne of Peru.

San Martin met with Bolivar in Guayaquil (July 26–27, 1822); the proceedings are surrounded with an atmosphere of mystery. Argentine historians hold that San Martin came to Guayaquil in search of military aid but was rebuffed by Bolivar, who was unwilling to share with a rival the glory of bringing the struggle for independence to an end; San Martin then magnanimously decided to leave Peru and allow Bolivar to complete the work he had begun. Venezuelan historians argue that San Martin came to Guayaquil primarily to recover that city for Peru; they deny that he asked Bolivar for more troops and insist that he left Peru for reasons having nothing to do with the conference. Both interpretations tend to diminish the stature and sense of realism of the two liberators—San Martin must have understood that Bolivar alone combined the military, political, and psychological assets needed to solve the factional problems in Peru and to gain final victory over the powerful Spanish army in the sierra; given the situation in Lima, San Martin's presence there could only hinder the performance of those tasks. Viewed in this light, the decision of Bolivar to assume sole direction of the war and of San Martin to withdraw reflected a realistic appraisal of the Peruvian situation and the solution it required.

San Martin returned to Lima to find that in his absence his enemies had usurped his power. In September 1822, before the first Peruvian congress, he announced his resignation as protector and his impending departure. San Martin's departure left Lima and the territory under its control in serious danger of reconquest by the strong Spanish army in the sierra. Bolivar allowed the situation to deteriorate until May 1823, when the Peruvian Congress called on him for help. The scare produced by a brief reoccupation of the capital by the Spanish army prepared the Creole leaders to accept Bolivar's absolute rule.

Bolivar arrived in Peru in September 1823 and needed almost a year to achieve political stability and to weld his army and the different national units under his command into a united force. After a difficult ascent of the sierra, patriot forces won a victory near the lake of Junin (August 6, 1824). To Sucre fell the glory of defeating the Spanish army in the last major engagement of the war, at Ayacucho (December 9, 1824). Only scattered resistance at some points in the highlands and on the coast remained to mop up. The work of continental liberation was achieved (see further ARGENTINA, HISTORY OF; BOLIVIA, HISTORY OF; CHILE, HISTORY OF; COLOMBIA, HISTORY OF; ECUADOR, HISTORY OF; PARAGUAY, HISTORY OF; PERU, HISTORY OF; URUGUAY, HISTORY OF; VENEZUELA, HISTORY OF).

Brazil. Brazil made a swift, almost bloodless transition to independence. The idea of Brazilian independence first arose in the late 18th century as a Creole reaction to the Portuguese policy of tightening political and economic

The Portuguese court in Brazil

control over the colony. The first significant conspiracy against Portuguese rule was organized in Minas Gerais, but it was easily discovered and crushed. The French invasion of Portugal (1808), followed by the flight of the Portuguese court to Rio de Janeiro, brought large changes and benefits to Brazil. The Portuguese prince regent John opened Brazil's ports to the trade of friendly nations, permitted the rise of local industries, and founded a "Bank of Brazil." Brazilian Creoles took satisfaction in Brazil's new role and the growth of educational, cultural, and economic opportunities for their class; but this feeling was mixed with resentment at the thousands of Portuguese who came with the court and competed with Brazilians for jobs and favours.

The revolution of 1820 in Portugal precipitated Brazil's break with the mother country. The Portuguese revolutionaries framed a liberal constitution for the kingdom, but assumed a conservative posture toward Brazil; they demanded the immediate return of Prince John to Lisbon, an end to the system of dual monarchy that he had devised, and the restoration of the Portuguese commercial monopoly. John approved the new constitution and sailed for Portugal, leaving behind his son and heir Dom Pedro as regent and advising him that, in the event the Brazilians demanded independence, he should assume leadership of the movement and set the Brazilian crown on his head.

Soon it became clear that the Portuguese Cortes intended to abrogate all the liberties and concessions won by Brazil since 1808. One of its decrees insisted on the immediate return of Dom Pedro from Brazil in order that he might complete his political education. But Dom Pedro, urged on by Creole advisers who saw a golden opportunity for an orderly transition to independence, rejected the Portuguese demand, issued his famous Fico ("I remain"), and in December 1822, having overcome slight resistance from Portuguese elements, was formally proclaimed constitutional emperor of Brazil (see also BRAZIL, HISTORY OF).

THE INDEPENDENCE OF MEXICO

A social revolution

In Mexico the movement for independence took an unexpected turn; the Indian and mixed-blood masses joined the struggle and for a time converted it from a quarrel between the Creole and Spanish-born peninsular elites into a social revolution.

In 1810 a Creole plot for independence was being hatched in the important industrial and mining centre of Querétaro; its leaders included Miguel Hidalgo, a priest in the town of Dolores known for his sympathy with the Indians. Informed that their plot had been denounced to Spanish officials, the conspirators decided to launch their revolt before preparations were complete. On September 16, 1810, Hidalgo called on the Indians of his parish to rise against their Spanish rulers; he appealed to the religious fanaticism of the natives by proclaiming the Virgin of Guadalupe the patron of his movement. After his first victories Hidalgo issued decrees abolishing slavery and Indian tribute and ordering the restoration of lands to the Indian communities. These measures gave the Mexican revolution a popular character largely absent from the movement for independence in South America, but they alienated the many Creoles who desired independence without social revolution.

Hidalgo could not weld his Indian horde into a disciplined army or capitalize on his early victories. Within a year he was captured, condemned as a heretic by an Inquisitorial court, and shot. A mestizo priest, José Maria Morelos, assumed leadership of the revolutionary struggle; he liberated most of southern Mexico, then summoned a congress that proclaimed independence and framed a republican constitution. Morelos extended Hidalgo's social reforms by prohibiting all forced labour and forbidding the use of racial terms. Like Hidalgo, Morelos also had differences with Creole associates that hampered his conduct of the war, and his premature efforts to install a constitutional regime also hindered his military efforts. In 1815 he, too, was captured by the Spaniards and shot.

Conservative–rebel alliance

The revolution then declined into a guerrilla war waged by many rival chiefs. Royalist armies gradually extinguished the remaining centres of resistance. The Spanish revolution of 1820 abruptly changed this state of affairs. Fearing the loss of their privileges, conservative clergy, army officers, and merchants in Mexico schemed to separate from the mother country and establish independence under conservative auspices. Their instrument was the Creole officer Agustin Iturbide, who had waged implacable war against the patriots. Iturbide offered peace and reconciliation to the principal rebel leader, Vicente Guerrero; his plan combined independence, monarchy, the religious supremacy of the Roman Catholic Church, and the civil equality of Creoles and Spaniards. For the moment Iturbide's program offered advantages to both sides.

The united forces of Iturbide and Guerrero swiftly overcame scattered loyalist resistance. On Septembr 28, 1821, Iturbide proclaimed Mexican independence; and eight months later a congress selected by Iturbide confirmed him as Agustin I, emperor of Mexico. Iturbide's empire had no popular base, however, and within a few months he had to abdicate and depart for Europe, with a warning never to return. In 1824 Iturbide landed on the Mexican coast with a small party, was promptly captured by troops of the new republican regime, and was shot (see also MEXICO, HISTORY OF).

BIBLIOGRAPHY

General works: CHARLES GIBSON, *Spain in America* (1966), the best short survey, incorporates the latest scholarship and has a valuable bibliographical essay; JOHN H. PARRY covers the same ground somewhat more amply in his well-written *The Spanish Seaborne Empire* (1966). SILVIO A. ZAVALA, *El mundo americano en la época colonial,* 2 vol. (1967), is a work of synthesis by a leading Mexican historian. STANLEY J. and BARBARA H. STEIN, *The Colonial Heritage of Latin America: Essays on Economic Dependence in Perspective* (1970), brief but insightful, shows the continuity of Latin American economic patterns from colonial times to the present. Among older works, CLARENCE H. HARING, *The Spanish Empire in America* (1947, reprinted 1963), remains indispensable for colonial institutions. BAILEY W. DIFFIE, *Latin American Civilization: Colonial Period* (1945), is a work of formidable scholarship and sharply defined viewpoints on various controversial topics. For a comprehensive collection of source materials on the colonial period, see BENJAMIN KEEN (comp.), *Readings in Latin America Civilization,* 1492 *to the Present,* 2nd ed. (1967).

Spanish Conquest of America: CARL O. SAUER, *The Early Spanish Main* (1966), critically examines the impact of the discovery on the Indian population of the Caribbean. WILLIAM H. PRESCOTT, *History of the Conquest* of *Mexico,* 3 vol. (1843, reprinted 1966), and *History of the Conquest of Peru,* 2 vol. (1908, reprinted 1963), remain classics, unsurpassed for breadth of conception and literary charm, but their romantic attitudes clearly reveal the books' age. JOHN HEMMING, *The Conquest of the Incas* (1970), supersedes all previous accounts.

Spain's colonial emprre: PHILIP W. POWELL, *Soldiers, Indians, and Silver: The Northward Advance of New Spain,* 1550–1600 (1952); and EUGENE H. KORTH, *Spanish Policy in Colonial Chile* (1968), examine Spanish–Indian relations. LEWIS U. HANKE, *The First Social Experiments in America* (1935), and *The Spanish Struggle for Justice in the Conquest of America* (1949), discuss the intellectual history of the Conquest; while his *Aristotle and the American Indians* (1959), studies the controversy between Las Casas and Juan Ginés de Sepdlveda over the nature and capacity of the Indians. For Messianic and Utopian influences in the Conquest, see JOHN L. PHELAN, *The Millenial Kingdom of the Franciscans in the New World,* rev. ed. (1969). For a reappraisal of Spain's Indian policy that questions some of Lewis Hanke's premises, see BENJAMIN KEEN, "The Black Legend Revisited: Assumptions and Realities," *Hispanic American Historical Review,* 49:703–719 (1969). On race relations in the colonial period, see the thoughtful survey of MAGNUS MORNER, *Race Mixture in the History of Latin America* (1967). Controversy surrounds the question of the size of pre- and post-Conquest Indian populations. S F. COOK and WOODROW BORAH made an illuminating series of studies that indicate a very large Indian population in 1519, followed by a demographic catastrophe in the sixteenth and seventeenth centuries, consolidating their findings in *The Indian Population of Central Mexico, 1531–*1610 (1960). For comparison, see the sharply divergent findings of ANGEL ROSENBLAT, *La población de América en* 1492:

Viejos y nuevos calculos (1967). FRANCOIS CHEVALIER, ***La Formation des grands domaines au Mexique*** (1952; Eng. trans., ***Land and Society in Colonial Mexico: The Great Hacienda,*** 1963), a landmark in the study of colonial land systems, traces the interplay of demographic and economic trends in the development of the hacienda. JAMES LOCKHART deftly analyzes the structure of Spanish society in colonial Peru in ***Spanish Peru, 1532–1560*** (1968).

The ***encomienda*** has been the subject of careful study. Two pioneer works are L.B. SIMPSON, ***The Encomienda in New Spain,*** rev. ed. (1950); and SILVIO A. ZAVALA, ***La encomienda indiana*** (1935). CLARENCE H. HARING, ***Trade and Navigation Between Spain and the Indies in the Time of the Hapsburgs*** (1918), remains a standard work. On land systems, see J.M. OTS CAPDEQUI, ***España en América: El régimen de tierras en la época colonial*** (1959). On mining, see MODESTO BARGALLO, ***La minería y la metalurgia en la America Española durante la época colonial*** (1955), for a general survey.

For the relations between church and state, see JOHN L. MECHAM, ***Church and State in Latin America,*** rev. ed. (1966); and WILLIAM E. SHIELS, ***King and Church: The Rise and Fall of the Patronato Real*** (1961).

Colonial Brazil: CAIO PRADO, *Formacão do Brasil Contemporañeo* (1942; Eng. trans., ***The Colonial Background of Modern Brazil,*** 1967), is an excellent introductory survey. See also CHARLES R BOXER, ***The Dutch in Brazil, 1624–1654*** (1957), and ***The Golden Age of Brazil, 1695–1750*** (1962). DAURIL ALDEN, ***Royal Government in Colonial Brazil*** (1968), is a model monograph. GILBERTO FREYRE, *Casa Grande e Senzala,* 2 vol. (1943; Eng. trans., ***The Masters and the Slaves,*** 1946), a work of great literary charm, portrays Portuguese colonizers as almost free from colour prejudice and Brazilian slavery as relatively mild. For opposing points of view, see CHARLES R. BOXER, ***Race Relations in the Portuguese Colonial Empire, 1415–1825*** (1963); and MARVIN HARRIS, ***Patterns of Race in the Americas*** (1964). On the economics of colonial Brazil, see ALEXANDER MARCHANT, ***From Barter to Slavery*** (1942), which stresses the mixed feudal-capitalist character of early Brazilian economy; and CAIO PRADO, ***op. cit.***

The Bourbon reforms and Spanish America: JOHN LYNCH, ***Spanish Colonial Administration, 1782–1810*** (1958), discusses colonial political reform. For the transmittal of Enlightenment ideas to Latin America, see the essays in ARTHUR P. WHITAKER (ed.), ***Latin America and the Enlightenment,*** 2nd ed. (1961).

Wars of independence: ROBIN A. HUMPHREYS, "The Fall of the Spanish American Empire," ***History,*** 37:213–277 (1952), is a perceptive interpretive essay, while he and JOHN LYNCH (eds.) bring together varied viewpoints in their anthology, ***The Origins of the Latin American Revolutions, 1808–1826*** (1965).

(B.K.)

Latvian Soviet Socialist Republic

Latvia, which had been an independent republic since 1920, was constituted as one of the 15 republics of the Soviet Union on July 21, 1940, and was proclaimed a Soviet Socialist republic on August 5. Situated in the Soviet west on the shores of the Baltic Sea and the Gulf of Riga, it covers an area of 24,600 square miles (63,700 square kilometres). In the north it borders on the Estonian Soviet Socialist Republic, in the east on the Russian S.F.S.R., and in the south on the Lithuanian S.S.R. The overall length of Latvia's borders is 1,120 miles (1,800 kilometres), of which 307 miles (494 kilometres) are coastal. The capital is Riga (Riga in Latvian). For the history of Latvia, see BALTIC STATES, HISTORY OF THE.

The natural landscape. ***Relief.*** Latvia is essentially an undulating plain, with fairly flat lowlands alternating with hills. The eastern part is more elevated, the most prominent feature being the central Vidzeme (Livonia) elevation, which reaches a maximum height of 1,020 feet (311 metres). In the southeast the highest point is 948 feet (289 metres). The Kurzeme (Courland) elevation in the west is cut by the Venta River into western and eastern parts. Between the central Vidzeme and Latgalē (Latgallian) elevations lies the East Latvian Lowland, partly crossed by moraine ridges that impede drainage; there are many peat bogs in this area.

The shores of the Baltic and the Gulf of Riga are only slightly indented, and there are considerable stretches of excellent sandy beaches.

Drainage and soils. Latvia contains a multitude of rivers belonging to the Baltic drainage area. The largest are the Western Dvina, locally called the Daugava (with a total length of 224 miles in Latvia), the Gauja (Russian Gauya), the Venta, and the Lielupe. Amid the hills, many of which are forested, are numerous lakes, ranging from a few acres up to 12 square miles (30 square kilometres) in area. Soils are predominantly podzolic, though calcareous soils characterize the Zemgale Plain. Swampy soils are found in some areas, particularly the East Latvian Lowland. Erosion is a problem in the more intensely cultivated hilly areas.

The Daugava **River**

Mineral deposits. Mineral resources are confined to gravel, sand, dolomite, limestone, clay, and peat. Oil has been discovered in the Courland Peninsula, but reserves have not yet been explored.

Climate. The climate is influenced by the prevailing air masses coming from the Atlantic. Humidity is high, and the skies are usually somewhat cloudy; there are only 30 to 40 days of sunshine per year and 150 to 180 completely overcast days. Average precipitation is 22–24 inches (about 550–600 millimetres) on lowlands and 28–31 inches (about 700–800 millimetres) on elevations. Southwesterly and southerly winds prevail. The frost-free season lasts from 125 to 155 days. Summers are cool and rainy more often than not. The mean air temperature in June is 63° F (17" C) with occasional jumps to about 93" F (34" C). Winter sets in slowly and lasts from the middle of December to the middle of March. The mean January temperature ranges from 28° F (—2" C) on the coast to 19° F (—7° C) in the east. There are occasional drops to —40" F (—40" C).

Vegetation. Some 67 percent of Latvia is covered with forests, meadows, pastures, swamps, and wasteland. Forests, which account for 38 percent of the total area, are a dominating feature of the republic's natural scene; about 10 percent of the forests are cultivated. The larger forest tracts are to be found in the northern part of the Courland Peninsula, along the left bank of the Daugava, and in the northeast. Conifers (pine and spruce) predominate. Of the deciduous species, birch, aspen, and alder occur more commonly. Meadows are found everywhere, both in the river valleys and among the hills.

Forest animals

Fauna. Latvia's fauna is typical for a region with mixed forests; there are squirrel, fox, hare, lynx, and badger. Somewhat less common are ermine and weasel. Conservation measures have resulted in an increase in the number of deer and elk and reintroduced beaver.

The numerous birds include the nightingale, oriole, blackbird, woodpecker, owl, grouse, partridge, finch, tomtit, quail, and lark. Storks and herons are found in the marshes and meadows.

Population. According to estimates of 1976, the population of Latvia was 2,499,000. The bulk of the people are Latvians (56.8 percent at the census of 1970, down from 62 percent in 1959 and from more than 75 percent in 1935) and Russians (29.8 percent, up from 26.6 percent in 1959). The Latvians, or Letts, speak one of the two surviving Baltic languages, the other being Lithuanian (see BALTIC LANGUAGES). In 1935 about 68 percent of them were Lutheran, 26 percent Roman Catholic. The ancestors of the present-day Latvians were the Latgals (Latgallians), who in the 9th century lived in the northeastern part of the present republic and who absorbed the Kurs (Cours, Couronians) and Livs. Two other early Latvian tribes were the Selonians and the Zemgals (Semigallians).

The total number of towns and urban settlements in the republic is 92. A considerable proportion of these have populations of 50,000 or more. In 1975 the urban population was 1,623,000 (65 percent), a proportion that is tending to increase. Immigrant Russians tend to settle in the cities; about 80 percent of the Russian population was urban in 1970, as against only 52 percent of the Latvians. The density of population was 100.8 per square mile (38.9 per square kilometre), and the growth rate in 1970 was 34 per 10,000 (though the natural growth rate of the Latvians themselves was only 29).

The economy. Industrialization in Latvia began in the latter part of the 19th century, and by the 1970s it was the most heavily industrialized republic of the Soviet Union.

More than 56 percent of the annual national income is derived from industry, 21.7 percent from agriculture, 6.6 percent from building, 4.4 percent from transport and communications, and 11.3 percent from trade and other activities.

Energy. Latvia produces approximately half of its own energy requirements, the remainder being derived from imported fuel and from the unified power system of the northwest regions of the Soviet Union. On the Daugava River stand the major hydroelectric stations—Pļaviņas (825,000 kilowatts), Kegums (68,000 kilowatts), and Riga (384,000 kilowatts). There are thermoelectric stations in Riga and other cities. Total production of electric energy in 1975 was 2,885,000,000 kilowatt-hours. All the stations are integrated in a single power grid, which in turn is incorporated in the power grid of the Soviet northwest.

Manufactures. Machine building and metal engineering are the leading manufacturing activities. Labour-intensive goods—that is, items utilizing small quantities of raw materials and much labour—such as radios and scientific instruments, are produced in quantity. Durable consumer goods, such as refrigerators, washing machines, motorcycles, and motor scooters, are also produced. The heavy engineering sector turns out ships, rolling stock, streetcars, power generators, diesel motors, and agricultural implements. The light consumer goods industry, concentrating on textiles, shoes, and hosiery, is sufficiently well developed that its products can be exported to other Soviet republics. There are many food-processing enterprises.

Agriculture and fisheries. Agricultural specialization is in dairy farming and meat production. Of the agricultural land, some 65 percent is used for crops, the remainder mainly for pasture. Of the crops, grain is the most important, industrial crops (sugar beets and flax) occupying only about 3.2 percent of the total crop area.

Collective and state farms

Collectivization of agriculture was accomplished, against resistance, in 1947–50. In the mid-1970s there were some 403 collective farms and 230 state farms. Agriculture is mechanized, permitting intensive farming. At the end of 1975 there were 1,389,000 head of cattle, including 586,000 milk cows; 1,195,000 pigs, and 7,684,000 poultry. The republic produced in that year 1,799,000 tons of milk, 250,000 tons of meat, and 666,000,000 eggs. The total grain yield reached 1,243,000 tons.

Latvia accounts for 5.8 percent of the Soviet Union's fish catch.

Transport. All types of transport are found in Latvia. The overall length of railway lines is 1,600 miles (about 2,600 kilometres), Latvia ranking first among the union republics in the density of its rail network. Railways account for 97 percent of freight traffic (22,000,000,000 ton-miles, or 32,000,000,000 ton-kilometres, in 1970). The republic's highways total 15,100 miles, of which 8,100 miles are hard-surfaced. The length of navigable internal waterways is 224 miles (360 kilometres). Much Soviet foreign trade is conducted via the seaports of Riga and Ventspils, which are open the year round. Riga has air links with Moscow and other large cities of the Soviet Union, and there is some internal air service.

Administration and social conditions. ***Government.*** The highest formal organs of government in Latvia are the Supreme Soviet, members of which are elected for a period of four years, and the Presidium. The Supreme Soviet appoints the Council of Ministers and elects the Supreme Court and the Presidium. The latter exercises power between sessions of the Supreme Soviet. Administration is the responsibility of the Council of Ministers, which drafts economic development plans and oversees their implementation. For administrative purposes the republic is divided into 26 *rayony* (districts) and 56 cities. At the local level, government is exercised by district, city, and village soviets, elected for terms of two years.

Justice is administered by people's courts, the judges of which are elected for terms of two years.

The Communist Party of Latvia

The Communist Party of Latvia, like its counterparts in the other union republics, is in fact the only source of political power, under the Communist Party of the Soviet Union. When the Latvian S.S.R. was constituted, the party was minuscule, but membership had reached about 144,000 by 1976. Western students of the Soviet Union conclude, however, that the party is dominated by non-Latvians (mainly Russians and other Slavs) and by russified Latvians who have lived in Russia for large parts of their lives. The Komsomol (Young Communist League) has about 276,000 members.

Education. General literacy was achieved in Latvia in the 1890s. The educational system is composed of eight-year and 10-year general schools, of which there were some 1,100 with 400,000 pupils, in 1974–75; specialized secondary schools (41,000 students); and 10 higher educational establishments (44,000 students, among whom, however, the proportion of Latvians is declining). About 82,000 children attend preschool establishments. There is a variety of institutions for extracurricular studies.

Teaching in the general schools is in Latvian or Russian or both; the number of bilingual schools (240 in 1967, out of a total of 1.200) is increasing. In Latvian-language schools the study of Russian is compulsory.

Scientific work is carried on at the institutes of the Latvian Academy of Sciences (founded in 1946), in higher educational establishments, and in research institutes. The number of scientific workers totalled 11,200 in 1974.

Health and welfare. In 1974 Latvia had 9,500 doctors, or 38 per 10,000 persons (one of the highest rates in the world); there were 31,000 hospital beds, or 125 per 10,000 persons. The republic's sanatoriums accommodate around 10,000 patients. Social insurance embraces factory and office workers and collective farmers. Its budget is administered by trade-union organizations, which pay out benefits in cases of disability, pregnancy, childbirth, or death and which administer pensions.

Cultural life. Amateur art thrives in Latvia. Clubs and individual enterprises have drama groups, choirs, ensembles, orchestras, and dance companies. The song festivals that have been held in Latvia since 1873 are still very popular; every five years the local districts and towns hold their own festivals and then send their best choirs, orchestras, and dance companies to the national festival, held at a park in Riga. The republic has a conservatory of music, an academy of arts, and a number of specialized secondary educational establishments for students of music, painting, and the applied arts.

Noted Latvian composers include Jāzeps Medins, Jānis Medins, and Emēlis Melngailis.

Modern Latvian literature dates from the late 19th century; the national epic, *Lāčplēsis* ("Bear Slayer"), by Andrejs Pumpurs, was published in 1888. Jānis Rainis, who died in 1929, is generally considered to be the most important Latvian writer. Three houses now publish literature in Latvian, Russian, and other languages. In 1974 they published 2,430 titles (of which, however, only a little more than 51 percent were in Latvian, down from 81 percent in 1945). Newspapers and magazines are published in Latvian and Russian. Like the other Baltic republics, Latvia is better supplied with radio receivers than the rest of the Soviet Union and receives foreign broadcasts. Television broadcasts are part local, part from Moscow; a large part of television transmission is in Russian. The Riga Film Studio produces full-length feature films as well as documentaries, short subjects, cartoons, and newsreels.

An important national tradition is the festival of Midsummer Eve (St. John's Eve, or Janu Naktis) and Day, which, though officially abolished by the government in 1960, continues to be observed.

Prospects. National five-year plans in the 1970s called for further growth of the electrotechnical, radiotechnical, medicinal, electronic, consumer goods, meat, milk, and fish industries. Further agricultural specialization, notably in milk, meat, and bacon, was also emphasized. It was hoped that living standards also would be improved.

Of great concern to Latvians, both in the republic and abroad, is the possibility of loss of ethnic identity, as a consequence of immigration of non-Latvians, together with a very low rate of natural increase among the Latvians themselves (the lowest rate in the Soviet Union), combined with the widespread use of the Russian language.

(P.V.G./Ed.)

Laud, William

Archbishop of Canterbury from 1633 to 1645 and a leading figure on Charles I's privy council, William Laud played a large part in shaping the policies of religious uniformity and authoritarian royal government, resistance to which was one of the major causes of the English Civil War.

By courtesy of the Archbishop of Canterbury, copyright reserved by the Courtauld Institute of Art and the Church Commissioners

Laud, painting by Sir Anthony Van Dyck, 1833. In Lambeth Palace, London.

Academic and churchman

Laud was born in Reading on October 7, 1573, the son of a prominent clothier. From Reading Grammar School he went on to St. John's College, Oxford, and until he was nearly 50 combined the successful but unspectacular careers of academic and churchman. He was soon associated with the small clerical group, followers of the patristic scholar Lancelot Andrewes, who, in opposition to Puritanism, stressed the continuity of the visible church and the necessity, for true inward worship, of outward uniformity, order, and ceremony. In 1608 Laud entered the service of Richard Neile, bishop of Rochester, with whose help he secured a succession of ecclesiastical appointments. From 1611 he was a royal chaplain and came gradually to the notice of King James I. His lifelong conflict with John Williams, later bishop of Lincoln and archbishop of York, began when both sought advancement through the patronage of Charles's favourite, the Duke of Buckingham. During Buckingham's years of power, Laud was his chaplain and confidant; and he established a dominant voice in church policies and appointments. He became a privy councillor in 1627 and, a year later, bishop of London.

Persecution of Puritans

In his London diocese, Laud devoted himself to combatting the Puritans and to enforcing a form of service in strict accordance with the Book of Common Prayer. The wearing of surplices, the placing of the communion table—railed off from the congregation—at the east end of the chancel, and such ceremonies as bowing at the mention of the name of Jesus were imposed, though cautiously enough to avoid unmanageable opposition. Churches, from St. Paul's Cathedral down to neglected village chapels, were repaired, beautified, consecrated. To religious radicals, all such reforms seemed moves toward popery, At Oxford, where Laud was chosen president of St. John's in 1611 and chancellor in 1629, new statutes, new endowments, and new buildings improved the university, both as a centre of learning and as a training ground for Laudian religion. On the death of George Abbott in 1633, Laud became archbishop of Canterbury; but he had already, by instructions issued in the King's name and by his ruthless energy in the royal prerogative courts of Star Chamber and High Commission, extended his authority—with varying success--over the whole country. From 1634 to 1637 visitations of every diocese (including, after strong resistance by Williams, that of Lincoln) showed the extent of deficiencies within the Anglican Church and the strength of Puritan practices. A succession of detailed orders from the Archbishop laid down the remedies. Preaching, to Puritans the essential task of the ministry, was to Laud a most dangerous source of "differences" in religion to be curtailed and controlled. In London his attack on Puritan "lectureships" culminated in the overthrow of the "feoffees for impropriations," the City organization for buying up tithes and church patronage for the benefit of Puritan clergy. The printed word was dangerous, too: celebrated Puritan propagandists such as Alexander Leighton and William Prynne were mutilated and imprisoned. Occasionally, Laud was less harsh than his enemies admitted, especially to the clergy. But he rejected all conciliation of the Puritan movement, whose strength and qualities he never understood. He had, in fact, much in common with some forms of it: the unrelenting quest for the godly life, the intolerant certainty of his own rectitude, the hatred of corruption and extravagance. He could do much to diminish inefficiency, pluralism, absenteeism, and sheer idleness. But his wider efforts to overcome the poverty of clergy and parishes and restore something of the church's position as a great and powerful landowner had extremely limited success.

To Laud, the strength of the church was inseparable from that of the state. Conflict between royal and ecclesiastical power was a possibility he never faced: under Charles I both could be exalted simultaneously. Holding no state office, he used his position on the privy council and his influence over the King to attack "the Lady Mora" (delay) in what he considered her first personification, the treasurer Richard Weston, and afterward in other ministers. His most effective direct impact on government was in the social policy he applied through the council and the courts. Exacting landlords and unscrupulous officials were attacked, and the poor were protected against everyone except the state itself. In all this his one constant ally was Thomas Wentworth (later the earl of Strafford), from 1633 lord deputy in Ireland. Laud and Wentworth corresponded regularly and frankly on their joint struggle to establish "thorough," as their rigorous policy came to be called. But by 1637 both began to see, dimly, the storm that was about to break upon them. The further trial of Prynne, together with other radical Puritans such as Bastwick and Burton, demonstrated not success for Laudian suppression but rather huge popular support for the opposition. The resistance of the gentry was consolidated by the extended demand for "ship-money," the most hated of Charles's non-parliamentary levies. Attempts by Charles and Laud to impose Anglican forms of worship in Scotland provoked fierce resistance there. English forces were sent northward, and in 1639 the "Bishops' Wars" began.

Trial and execution

In the spring of 1640 Parliament met for the first time in 11 years and with it the clerical assembly, the Convocation, which laid down in a new set of canons the principles of the Laudian church. They explained the prescribed ceremonies as "fit and convenient" rather than essential. But they added to the popular hatred of Laud shown in mass demonstrations, petitions, and leaflets. In December, formally accused of high treason, he was taken to the Tower. His trial, managed enthusiastically by Prynne, began only in 1644, in the midst of the Civil War. As with Strafford, the Commons had to abandon legal proof and resort to an ordinance of attainder, accepted hesitantly by the lords. On January 10, 1645, the Archbishop was beheaded.

Laud was never much liked, even by his allies. A humourless, dwarflike figure, uninterested in court pleasures, unmarried, tactlessly impartial in his condemnations, he could never establish a party of influential supporters. During the war and interregnum, royalists and peacemakers generally preferred to forget him. At the Restoration, in 1660, outward Laudian forms were accepted but by a church less significant than ever to the community and the individual. Few in the 18th century saw Laud as a martyr. In the 19th century the historian Thomas Babington Macaulay's fierce contempt for the "ridiculous old bigot" inspired the schoolbooks of many generations. The Oxford Movement, a movement of High Anglican reform in the 1840s, tried unconvincingly to re-establish him as a religious leader, and High Anglican clergy have remained his principal supporters. But at the turn of the 19th century, the Civil War historian Samuel Rawson Gardiner stressed Laud's abilities and integrity and regarded the links with authoritarian politics as his "misfortune."

In the 20th century, the eminent English historian H.R. Trevor-Roper has set against his narrow-minded methods the comprehensive idealism of his social policy, "coloured over by the accepted varnish of an appropriate religious doctrine." Laud, as he himself was well aware, failed; but his devotion to a coherent purpose and his repudiation of hypocrisy, compromise, and corruption in allies and enemies of whatever class were rare and admirable qualities.

BIBLIOGRAPHY. H.R. TREVOR-ROPER, ***Archbishop Laud, 1573–1645,*** 2nd ed. (1962), the best modern biography; W.M. LAMONT, ***Godly Rule,*** ch. *3* (1969), stressing Laud's idea of the "Elect Church"; CHRISTOPHER HILL, ***Economic Problems of the Church, from Archbishop Whitgift to the Long Parliament***, pt. *3* (1956), on Laud's large but unsuccessful schemes for restoring the church's economy; J.W. ALLEN, ***English Political Thought, 1603–1644,*** vol. *1*, pt. 2 (1938, reprinted 1967), showing the vagueness of his political ideas; E.C.E. BOURNE, ***The Anglicanism of William Laud*** (1947), an uncompromisingly favourable clerical view.

(D.H.P.)

Laurales

Laurales is a fairly large order of flowering plants, consisting of 11 families with a total of about 70 genera and more than 2,500 species. The order is of interest not merely because of its abundance but also for its utilitarian, morphological (structural), and phylogenetic (evolutionary) importance.

Economic importance of the laurel family

General features. ***Utility.*** Of the 11 families, the laurel (Lauraceae) is the most useful to man. The leaves and bark of several species such as those of ***Lindera*** (benzoin), *Litsea*, and ***Laurus*** (laurel) are of medicinal value. Some, like the South African ***Ocotea bullata*** (stinkwood), yield useful timber. Most of the world's supply of natural camphor, used in the manufacture of colloidin and explosives, comes from *Cinnamomum camphora*, a native of Japan and Taiwan. Essential oil containing up to 80 percent of safrole, used in the perfume industry, is obtained from the roots of ***Sassafras** officinale* and the wood of two Brazilian trees, ***Ocotea preteosa*** and ***O.** cymbarum.* Cinnamon spice is derived from the bark of ***Cinnamomum zeylanicum,*** a native of Ceylon and southern India. Cinnamon bark and leaf oil are also used in medicine and perfume. ***Persea americana,*** the avocado pear, an evergreen tree native to the rain forests of Central America, is economically important in the U.S. South and Hawaii, and in Brazil, South Africa, and tropical Australia. Its edible fruits contain a higher percentage of protein than any other fruit.

Other families of Laurales are economically less important. *Peumus boldus* (family Monimiaceae), of Chile, yields boldo wood used in cabinet-making. ***Doryphora sassafras,*** of the same family, is a source of essential oil. Some species of the families Calycanthaceae and Chloranthaceae are grown as ornamentals.

Size range and diversity of structure. Members of the order produce their own food by photosynthesis (***i.e.***, are autotrophic) with the exception of *Cassytha*, a rootless, vinelike stem parasite without proper leaves; it grows on many species of flowering plants. Four of the families—Austrobaileyaceae, Amborellaceae, Lactoridaceae, and Calycanthaceae—are exclusively shrubs; the other seven are predominantly trees. Some members of the laurel family (Lauraceae) attain great size; *Cinnamomum* ***camphor~***, for example, reaches a height of about 100 feet ***(30*** metres) and ***Persea americana*** grows up to 60 feet (18 metres) or more.

Primitive and advanced structures

The Laurales order shows great diversity in vegetative and floral structure. Unlike most other flowering plants, the wood in ***Amborella*** and ***Sarcandra*** (family Chloranthaceae) is devoid of vessels (water-conducting tissues made of special hollow cells). In ***Austrobailrya*** (family Austrobaileyaceae), a woody vine, the wood has vessels, but exhibits other primitive characteristics. On the other hand, the families Lactoridaceae, Calycanthaceae, and Lauraceae show some features considered to be advanced.

The plants of the genus ***Austrobaileya*** are more simple. Their phloem (the part of the vascular system that transports food), unlike that of other flowering plants, is devoid of companion cells—a characteristic of coniferous plants. Typical sieve tubes (the normal arrangement of phloem tissue) are also absent in this genus; instead, its long and overlapping phloem-tissue cells lack the characteristic terminal sieve plates, with small holes that lead to the next cell. Further structural simplicity is seen in the phloem sf ***Austrobaileya,*** and also the family Calycanthaceae, which lack sclerenchyma (thick-walled cells). On the other hand, the presence of sieve cells with single, transverse end plates is an advanced development in the Calycanthaceae family.

The order demonstrates considerable diversity in floral structure even though many families, particularly the Amborellaceae, Trimeniaceae, and Lactoridaceae are not completely described as yet. The genus ***Austrobaileya*** has probably the most primitive floral structure of the order. Its flower is bisexual and hypogynous (the flower parts are joined at the base of the female structure. or ovary, a condition sometimes called "superior ovary"), and its numerous floral parts are arranged spirally. The stamens (male pollen-producing structures) are laminar (flat, or petallike) and the carpels (simple ovaries) each have numerous ovules, or developing seeds. In the families Gomortegaceae and calycanthaceae, also, the flowers are bisexual (with both male and female organs). In ***Amborella*** (family Amborellaceae) the flower is unisexual (one plant with either male or female flowers); all other families are polygamous (with both unisexual and bisexual flowers on the same plant).

The flower form is usually hypogynous ("superior ovary") in the families Trimeniaceae, Lactoridaceae, and Lauraceae; rarely, as in ***Hypodaphnis,*** it is epigynous, or inferior ovary, with the petals, sepals, and stamens arising at the upper end of the ovary, which they enclose, forming what is called the floral tube. In all the other eight families the flower form is either perigynous ("half inferior") or epigynous.

Stamens, carpels, and ovules

In the family Lactoridaceae the stamens are laminar with protruding connectives and the pollen is held in permanent tetrads (groups of four), as in the Winteraceae family and some members of the Annonaceae family (both of the order Magnoliales). In the families Hernandiaceae, Chloranthaceae, Lauraceae, and Gyrocarpaceae the gynoecium (female flower parts) consists of a single carpel with a single ovule. Other families have more than one carpel; in the families Trimeniaceae and Gomortegaceae they are united; in the others, free. In the Lactoridaceae family the ovule number per carpel is six, in the Calycanthaceae family it is one or two; all others have a single ovule.

The fruit is winged in the families Hernandiaceae and Gyrocarpaceae. In the families Trimeniaceae, Monimiaceae, Gomortegaceae, Chloranthaceae, and Lactoridaceae the seed is endospermous (*i.e.*, it contains endosperm, a nutritive starchy tissue for the developing embryo), and the embryo is small. In contrast, exalbuminous (without endosperm) seeds with large embryos occur in the families Hernandiaceae, Calycanthaceae, Lauraceae, and Gyrocarpaceae.

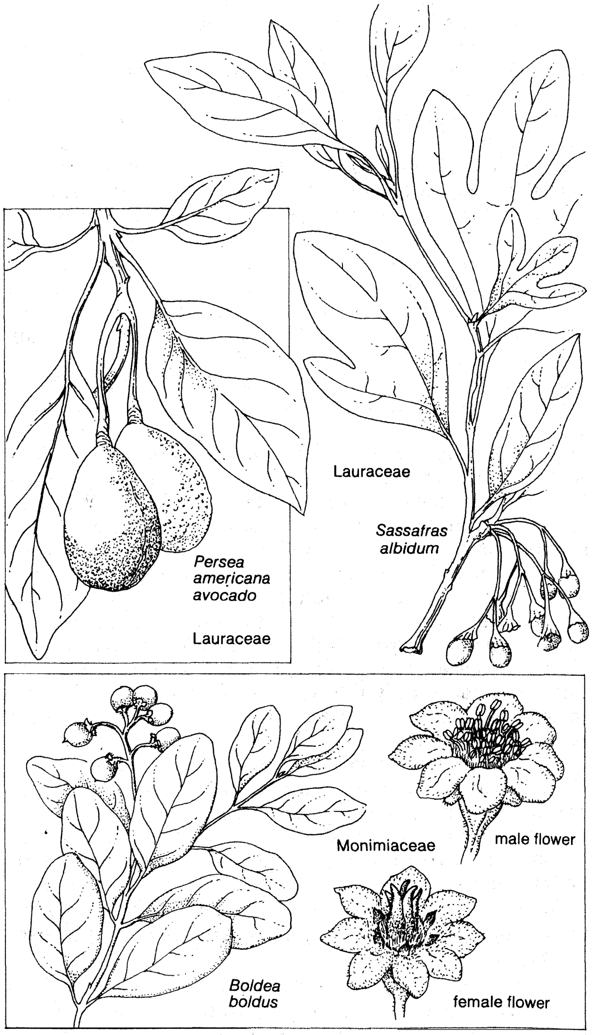
Distribution and abundance. Plants of the Laurales order occur generally in the tropics or subtropics. Some families are widely distributed; others in the order are restricted. Lactoridaceae, for example, is confined to the

Islas Juan Fernández, 400 miles (645 kilometres) west of Chile, others to places like Chile, New Caledonia, Australia, or islands of the Pacific Ocean. A few species occur in warm temperate regions: in North America, extending to Canada; in Africa; and in southern Europe, where *Laurus nobilis* (family Lauraceae) extends to the Mediterranean.

Phylogenetic importance. The Laurales order occupies an important place in the phylogeny, or evolutionary history, of the flowering plants. Of the ten genera of dicotyledons (plants with two "seed leaves") with primitive, vesselless xylem (water-conducting tissue) two are included in the Laurales order; the rest in the Magnoliales order. The only flowering plant species without companion cells and true sieve tubes (Austrobaileya) also belongs to the Laurales order. On the other hand, in several families, floral evolution has advanced to the level of epigyny (inferior ovary) and unisexuality. Thus, many of the evolutionary trends observed among the flowering plants as a whole are retained in this and in the related and probably more primitive order, the Magnoliales.

Futhermore, distribution of several members of this order is restricted to areas that are relatively little explored botanically. The living and fossil plants of such regions, when more completely studied, will yield valuable data for understanding the origin and evolution of the flowering plants. Phylogenetic study of the Laurales order might well be further intensified and extended.

Natural history. Life cycle. Very little is known about the life cycle in six of the families of the order (Austrobaileyaceae, Amborellaceae, Trimeniaceae, Gomortegaceae, Hernandiaceae, and Lactoridaceae, and such knowledge is far from complete in families like Monimiaceae.



Representative plants from the two largest families of the laurel order.

The flowers of the Calycanthaceae family are remarkably adapted for pollination by beetles. The food provided by the tips of some of its tepals (similar-appearing flower petals and sepals) and the prolonged connectives of the stamens attract the beetles, which also eat the nectar secreted by the inner staminodes (sterile stamens, usually flat and petallike). In doing both, the beetles serve in pollination. It has been suggested that the position of the carpels on the concave receptacle might offer protection to the ovules from the foraging beetles.

Unusual mode of nectar secretion

The mode of secretion of nectar from the glandular appendages of the stamens in the Lauraceae and Monimiaceae families has been studied. In the Lauraceae family, the epidermis ("skin") and cuticle (a waxy coating) of the bud break down just before the opening of the flowers; the nectar then is secreted through the intercellular spaces of the flower tissues. In the Monimiaceae family, there is no secretion of nectar before the beginning of flower opening, and the epidermis is not as extensively damaged as in the Lauraceae family. Numerous bees, blowflies (Calliphora), and mites visit the flowers for the nectar.

There is little uniformity in pollination and other processes of the life cycle within the order. In some members of Laurales, such as species of *Beilschmiedia* (Lauraceae) and Hortonia (Monimiaceae), the staminal appendages do not function as nectaries. In *Persea americana* (avocado) the flowers open twice and pollination is brought about by bees. In Sri Lanka (formerly Ceylon), *Cinnamomum* zeylanicum (Ceylon or true cinnamon) flowers in May and fruits mature within two or three months; this is probably true of other tropical species.

Ecology. The Laurales order is mostly tropical and subtropical and its species are mostly evergreen. Some members of the Lauraceae and Calycanthaceae families also grow in warm temperate regions, where they are deciduous. Calycanthus fertilis, Sassafras albidum, *Per*sea borbonia (sweet bay), *Litsea* aestivalis, and *Lindera* benzoin (spice-bush), for example, are among the characteristic species in the deciduous forests of eastern North America, where climatic conditions are similar to northwestern Europe and parts of eastern Asia.

Some species of the Eauraceae family, like *Cinnamo*mum *camphora* (camphor tree), are not particularly demanding in their environmental and soil requirements. The camphor tree grows well at elevations of 3,000—4,000 feet (900—1,200 metres) and even up to 45° (latitude) north and south of the Equator. In contrast, *Cinnamomum* zeylanicum is sensitive and thrives best only on the white sandy soils of Sri Lanka.

Laurel order compared with Magnolia order

Form and function. Despite the great diversity of structure among families of the order, some structural features common to all serve to distinguish the Laurales from other orders. Like the related Magnoliales, the Laurales order is composed of species that are all woody and that have ethereal (aromatic) oil cells. Those species are further characterized by pollen grains with no more than two grooves, and by a primitive nodal anatomy (arrangement of vascular bundles at the juncture of leaf and stem) of the type called unilacunar. The stamens have nectar-bearing appendages, and, in most species, the anthers split open when ripe by means of valves. The female structures usually consist of a single carpel. Unlike the Magnoliales order, which has generally primitive leaflike carpels and stamens, most Laurales species have more specialized floral organs.

Evolution and paleontology. Well-preserved fossils resembling living genera of the Lauraceae family, such as Sassafras, *Cinnamomum,* and *Persea,* have been found in Lower and Upper Cretaceous rocks (about 65,000,000 to 136,000,000 years old) in several localities in Europe and America. Since those genera at that time were already as highly specialized as their modem counterparts, it may be inferred that they must have had a fairly long history of evolution before the Cretaceous Period. In the absence of fossil evidence, however, no reliable conclusions as to the origin and early phylogeny of the family can be drawn.

Pollen similar to that of the fossil genus *Clavatipollenites* from the Lower Cretaceous Period shows some tenuous resemblances to that of the family Chloranthaceae. The fossil history of other families is virtually unknown.

In view of the present paucity of information on the paleontology of the order, evolutionary interpretations must of necessity be based on comparative morphological studies of present-day groups. The Austrobaileyaceae family is thus placed lowest in the evolutionary scale because of the preponderance of its apparently primitive vegetative and floral characters, such as the absence of companion cells and typical sieve tubes, the primitive cambium and wood anatomy, numerous spirally arranged flower parts, laminar stamens, and numerous ovules per carpel.

The Amborellaceae family also manifests many primitive features including vesselless xylem, spirally arranged perianth (sepals and petals), and stamens with a broad filament. At the same time, however, its members have advanced to the level of unisexuality. An uncertain affinity to the Monimiaceae family has been suggested.

Relationships of families within the order

Many authorities consider Lactoridaceae to be an isolated family. Some assign it a relatively primitive place, near the Winteraceae family (order Magnoliales), while others give it a relatively higher place, assuming that it had a common origin with the family Chloranthaceae. Still others suggest that its mode of growth, stipules, and anatomy all indicate a relationship with the families Piperaceae (order Piperales) or Polygonaceae (order Polygonales). But without adequate data none of these views can be supported.

Of the Trimeniaceae family still less is known; all that can be said is that it exhibits a mixture of primitive and advanced features. Some authors, however, suggest an affinity with the Monimiaceae family.

The families Monimiaceae, Gomortegaceae, Hernandiaceae, Lauraceae, and Gyrocarpaceae have been found to be more specialized and more closely related to one another than other Laurales families. Of these. it has been suggested that the Monimiaceae family, with the greatest range in floral structure, constitutes the parent family from which the others have evolved. Its protoxylem strand patterns (developing form of xylem, or woody tissue), wood anatomy, and embryology all are similar to those of the other four families. In the five-family group, Gyrocarpaceae seems to be the most highly specialized. Its vessel cell end walls are inclined less, its medullary ray tissue (wood rays) is more specialized, its floral structure is generally reduced, and the family exhibits a peculiar type of embryo-sac development.

The floral hypanthium, a saucer-like structure formed from the fused bases of the petals and sepals, is common to the less specialized Calycanthaceae family and the parent Monimiaceae family. The subfamily Siparunoideae of Monimiaceae also resembles Calycanthaceae in the absence of staminal appendages. The Calycanthaceae family also shares numerous anatomical features with the Lauraceae family, but there is little morphological evidence to support its suggested affinities to the Rosales or the Magnoliales orders. Its pollination by beetles has been considered a primitive feature only because other types of insects appeared later in the geological time scale.

The affinities of the Chloranthaceae family have also been much disputed. Some authors have placed it in the order Piperales, but the nodal anatomy, opposite stipulate leaves, united petiole (leaf stalk) bases, and inferior ovary with its solitary pendulous ovule all put this family outside the Piperales order. Those features plus staminal appendages indicate an affinity with the Monimiaceae family. Thus, with the possible exception of the families Austrobaileyaceae and Lactoridaceae, all other families in the Laurales order can probably be derived from a Monimiaceous stock.

Although the Laurales order is more advanced than the Magnoliales, it seems more probable that it originated independently, rather than from the Magnoliales order. Because the Lactoridaceae family possesses a combination of features of both orders, it has been suggested that both Laurales and Magnoliales had a common origin and had early diverged along separate evolutionary lines.

Classification. *Distinguishing taxonomic features.* As early as 1950, the woody plant families having ethereal oil cells and monocolpate (single-furrowed), derived polyporate, or acolpate pollen were divided into two groups—one with unilacunar and the other with trilacunar or multilacunar nodal anatomy. The Laurales order as presented here corresponds to the unilacunar group. Certain other features are also characteristic of several members of the order although a few families do not share them. These include the polygamous flowers, the presence of staminal appendages, anthers that open by means of valves, two pollen sacs per anther, and the single pendulous ovule per carpel.

Annotated classification. The scope of the order Laurales as presented here is a recent and authoritative treatment (see below, *Critical appraisal).* The delimitation of this order and the taxonomic position and delimitation of its families, however, differ in some respects in other classification systems.

ORDER LAURALES

Plants woody; ethereal oil cells present; pollen monocolpate, derived polyporate, or acolpate; nodal anatomy unilacunar; staminal appendages often present.

Family Austrobaileyaceae

Woody vines. Phloem devoid of companion cells and true sieve tubes. Leaves opposite, entire (smooth margined), simple, stipulate. Floral parts numerous, spirally arranged; stamens laminar. Ovules many. One genus, *Austrobaileya,* with two species, native to Queensland, Australia.

Family Amborellaceae

Vesselless shrubs. Leaves alternate, distichous (in two ranks). Flowers solitary, unisexual; perianth spirally arranged. Stamens with broad filament. Ovule orthotropous. One genus, *Amborella,* with one species restricted to New Caledonia.

Family Trimeniaceae

Trees or shrubs. Leaves exstipulate. Flowers polygamous, in panicles. Calyx 4 to numerous; corolla none. Stamens 4 to numerous. Carpels 1 or 2 fused, with a single ovule. Fruit drupaceous. Seed endospermous. Four genera with 15 species distributed in east Malaysia, east Australia, and on islands of the Pacific Ocean.

Family Monimiaceae

Trees and shrubs. Flowers solitary, or in cymes; usually unisexual; perigynous. Perianth inconspicuous or absent. Hypanthium present. Stamens numerous or few; glandular appendages usually present; anther dehiscence by valves or by longitudinal slits; pollen sacs 2 or 4. Carpels numerous each with a single ovule. Seed endospermous; embryo small. Twenty genera with 150 species distributed chiefly in southern tropical regions including Madagascar, Australia, Polynesia, New Zealand, Chile, New Guinea, and New Caledonia.

Family Gomortegaceae

Trees. Flowers bisexual, epigynous. Sepals 7 to 10, spirally arranged. Petals absent. Stamens 2 to 11 with glandular appendages; pollen sacs 2; valvular dehiscence. Carpels 2 to 3, united. Seed endospermous. One genus, *Gomortega,* with 1 species confined to Chile.

Family Hernandiaceae

Tropical trees or shrubs. Leaves with palmate venation. Flowers bisexual or unisexual, epigynous. Perianth 6 to 10, biseriate. Stamens 3 to 5, with staminal appendages; anther dehiscence longitudinal; pollen sacs 2. Carpel 1; ovule 1. Seed exalbuminous. Two genera with 50 species distributed in tropical Central America, the West Indies, the northern edge of South America, west central Africa, Madagascar, South Africa, and the islands of the South Pacific.

Family Chloranthaceae

Tropical and subtropical herbs, shrubs, or trees. Flowers unisexual or bisexual, often epigynous. Perianth none or minute calyx present in female flowers. Stamens 1 or 3 united with longitudinal dehiscence; staminal appendages present; pollen sacs 2. Carpel 1; ovule 1. Seed endospermous. Five genera with 65 species distributed from Jalapa, Mexico southward in South America.

Family Lactoridaceae

Shrubs. Leaves with large membranous stipules. Flowers unisexual or bisexual. Perianth 3. Stamens in 2 whorls of 3 each, laminar; connective projecting; pollen in permanent tetrads. Carpels 3, with 6 ovules in each. Seed endospermous. One genus, *Lactoris,* with 1 species native to the Islas Juan Fernández.

Family Calycanthaceae

Shrubs. Cortical vascular system present; flowers bisexual, perigynous, hypanthium present. Perianth of numerous spirally arranged tepals. Stamens 5 to 10; pollen sacs 2. Food bodies present on the tepals (sepals and petals) and stamens. Carpels numerous with 1 or 2 ovules in each. Seed exalbuminous. Embryo large. Two genera, with 7 species distributed in east Asia, northeast Australia, and North America.

Family Lauraceae

Tropical and subtropical trees and shrubs. Flowers bisexual, rarely unisexual. Perianth in 2 whorls of 3 each. Stamens in 3 or 4 whorls of 3 each, some of them modified into staminodes; glandular appendages present; 2 or 4 pollen sacs; valvular dehiscence; third whorl extrorse, others introrse. Carpel 1, ovule 1. Seed exalbuminous. Embryo large. Thirty genera with 2,500 species distributed largely in tropical regions, most abundantly in Southeast Asia from Australia to Japan and in tropical America.

Family Gyrocarpaceae

Small trees or shrubs. Inflorescence a thyrse (*i.e.*, the main axis is a raceme, or long central stem bearing flowers on short side stems, the secondary axes are cymes). Flowers bisexual or unisexual. Perianth 8. Stamens 3 to 5; glandular appendages present; valvular dehiscence; pollen sacs 2; 3 or 4 staminodes sometimes present. Carpel 1, ovule 1; gynoecium inferior; fruit a samara (a dry one-seed winged structure). Seed exalbuminous; embryo large. Two genera with 22 species distributed in tropical and subtropical regions of the world.

Differing viewpoints concerning the limits of the Laurel order

Critical appraisal. The scope of the Laurales order varies in other systems of classification. Some authorities include the Myristicaceae family and exclude the Calycanthaceae and Chloranthaceae families, placing Calycanthaceae in the Rosales order and Chloranthaceae in Piperales. Others include the Calycanthaceae family. The Lactoridaceae family has often been placed in the order Magnoliales near the Winteraceae family. Similarly, earlier authors have included all these families in the order Ranales or Magnoliales along with other woody families. Recently, however, most taxonomists have included the families Monimiaceae, Lauraceae, Gomortegaceae, Hernandiaceae and Gyrocarpaceae in the Laurales order as distinct from the Magnoliales order.

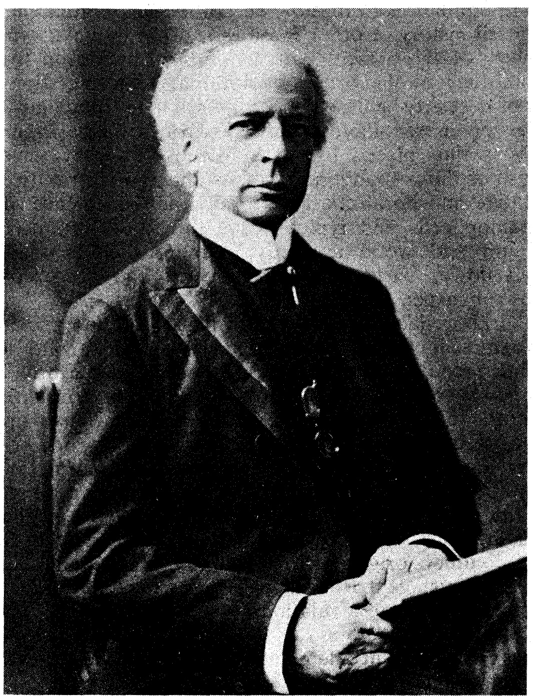
There is some disagreement among taxonomists concerning the delimitation of certain families. The parasitic genus *Cassytha* is sometimes segregated from the Lauraceae family into a separate family Cassythaceae, but recent morphological studies have lent no support to this treatment. Some systematists have removed the genera ***Atherosperma*, *Siparuna*, *Laurelia*, *Daphnandra*** and ***Doryphora*** and a few others from the Monimiaceae family into a separate family Atherospermataceae. The families Hemandiaceae and Gyrocarpaceae were merged in earlier systems but their separation now seems justified.

BIBLIOGRAPHY. A.J. EAMES, ***Morphology of the Angiosperms*** (1961), an advanced textbook that includes a discussion of the morphology and phylogeny of many of the concerned families; R.L.N. SASTRI, "Comparative Morphology and Phylogeny of the Ranales," ***Biol. Rev.***, 44:291-319 (1969), an objective assessment of the relationships and phylogeny of ranalian plant families on the basis of comparative morphology; J.C. WILLIS, ***A Dictionary of the Flowering Plants and Ferns***, 7th ed. rev. by H.K. AIRY SHAW (1966), a classical reference work that also describes some of the little-known families of the order; A. TAKHTAJAN, ***Flowering Plants: Origin and Dispersal*** (1969; Eng. trans. from the 2nd Russian ed. of 1961), a recent and fairly satisfactory phylogenetic system; L.D. BENSON, ***Plant Classification*** (1957), a standard textbook that includes ecological data; L.S. COBLEY, ***An Introduction to the Botany of Tropical Crops*** (1956), a college-level account of the economic importance of the families; H.N. ANDREW~Jr., ***Studies in Paleobotany*** (1961), an advanced treatise giving a good account of angiosperm fossils.

(R.L.N.S.)

Laurier, Sir Wilfrid

The first French Canadian to serve as prime minister of the Dominion of Canada, Sir Wilfrid Laurier spent his 30 years as leader of the Liberal Party and his 15 years as prime minister in working, as he put it, "to promote unity and harmony and amity between the diverse elements of this country." Yet in the end, defeated by the covert alliance of French-Canadian nationalists and inflamed British imperialists, he left behind him a country internally still unreconciled.



Laurier.
National Film Board of Canada

Early life and education

He was born in the small farming town of Saint-Lin, north of Montreal, on November 20, 1841, of French-Canadian parents. He studied under a Scottish tutor in the nearby village of New Glasgow before being sent to the college at l'Assomption, where he received literary training under Catholic priests. He then studied law at McGill University in Montreal and was called to the bar in 1864. His bicultural education, most unusual at the time, may have played a part in his lifelong dedication to Canadian unity. While at McGill, he became a leading member of the Institut Canadien, a political club of advanced liberals (Les Rouges) with anticlerical and republican views. Later he joined the law offices of one of the leading Rouge politicians and contributed a number of articles to radical newspapers, one of which he edited for a few months in the mid-1860s.

In 1866, for reasons of health, Laurier moved to Arthabaska, where he opened his own law practice. In 1868 he married Zoe Lafontaine of Montreal, and despite a long relationship with Emilie Lavergne, his law partner's wife, his childless marriage seems to have been a happy one. In 1871 he was elected to the opposition benches of the provincial legislature of Quebec, where his first speech, an eloquent plea for educational reform, attracted much attention. In 1874 he was elected to the Canadian House of Commons, of which he was to be a member until his death.

Rise to leadership

As Laurier gradually rose to become minister of internal revenue (1877–78) and eventually to leadership of the opposition Liberal Party in 1887, he persistently sought to bring together his countrymen on the issues that have since been recognized as the dominant themes of modern Canadian politics: the relations of church and state, the bicultural entente between French and English-speaking Canadians, and the country's association with the British Empire and relations with the United States. One of the political highlights of these years for Laurier was his famous speech on Liberalism delivered in 1877 in the city of Quebec. In that speech he set himself against both the Quebec politicians who attempted to form a Catholic party and the extremist elements in his own group who sought to exclude the clergy from all political activity. Because of his skillful statesmanship, the cold antagonism between conservative churchmen and liberal politicians gradually began to thaw; after 1896 no anticlerical ever attained important public office and no cleric officially interfered in politics.

In 1885 Laurier became a national figure when he delivered a moving plea of clemency for Louis Riel, who had led a rebellion of the Métis (people of mixed French and Indian extraction) in Manitoba and whose death sentence provoked violent outbursts between the French Catholic nationalists in Quebec and the Britannic groups in Ontario. Showing great courage, Laurier, though not condoning Riel's actions, charged the government with mishandling the rebellion. Although he did not succeed in saving Riel, he established his reputation as a man of principle and high ideals. Throughout his political life, he emphasized moderation and compromise and gradually became recognized as the only leader able to effect a national reconciliation.

At the same time, he was turning his personal magnetism into a valuable political weapon. Between 1887 and 1896, he perfected his party's organization, refined Liberal strategy, made political alliances, assessed local partisans, and judiciously applied his personal charm to winning over Conservative adversaries and dissident Liberals. He infused new life into his party, for instance, by campaigning vigorously for unrestricted reciprocity, the grant of mutual commercial privileges, with the United States. After the policy had served its purpose, however, he dropped it from his platform in 1893. Between 1895 and 1896 he spoke at between 200 and 300 meetings, thus personally reaching some 200,000 voters. In mid-1896, with the Conservative government divided and disorganized, he easily carried the Liberal Party to victory in the general election.

Laurier's "national policy"

Intent on heading an administration of national unity, Laurier attracted to his first cabinet men who had won distinction in their own provinces. His "national policy" consisted of protection for Canadian industries, the settlement of the west, and the building of an effective transportation system. The years between 1896 and 1911 became a boom period for which the Prime Minister himself provided the slogan: "The Twentieth Century belongs to Canada." The budget of 1897 lowered tariffs but established a protection policy that lasted until 1911. Laurier's land and emigration policy perhaps remains as the basic achievement of his government. During 15 years more than 1,000,000 people moved into Manitoba and into the western territories, which in 1905 became the provinces of Saskatchewan and Alberta. Wheat became the major product of the new Prairie Provinces; towns and ports sprang up; railroads flourished; and in 1903 Laurier announced that a second transcontinental rail system would be built: the Canadian west had become the granary of the world.

Meanwhile, the Prime Minister's attention had been diverted to external affairs. In 1897, 1902, 1907, and 1911 he attended Imperial Conferences at which he steadily resisted British proposals for closer ties that might commit Canada to defense responsibilities. He sincerely admired the institutions and liberal policies of Great Britain —he accepted a knighthood (1897) and once declared that he would be proud to see a Canadian of French descent affirming the principles of freedom in the British Parliament—yet he would never agree to any dilution of Canadian autonomy. Thus, from his policies there began to emerge the modern concept of a British Commonwealth of independent states.

Britain's South African War of 1899 marked the start of Laurier's decline. Quebec nationalists denounced his decision to send a force of 1,000 men, while English Canadians thought the number insufficient. Then, a series of invidious disputes—over denominational schools in the Northwest, Sunday observance laws, the restrictions of French linguistic rights in Manitoba and Ontario—kept widening the rift between the nationalities in the east and new Canadians in the west and between Laurier and his cabinet. As the election of 1911 approached, the Prime Minister attempted to reunite his factious party by negotiating a treaty of reciprocity with the United States, but he failed. Reciprocity did not distract Quebec from the convincing argument that each one of Laurier's compromises was a surrender of French Canada's fundamental rights. Among the Britannic Canadians, reciprocity seemed an opportunistic capitulation to the United States, the first step toward annexation. In a month of bitter campaigning in 1911, the 70-year-old Prime Minister delivered more than 50 speeches yet could not overcome the powerful combination of Imperialist business interests and bigoted nationalism. He retired with the dignity Canadians had learned to expect of him and spent his remaining years as leader of the opposition. He died in Ottawa on February 17, 1919.

Assessment

To his faithful followers, especially in Quebec, where his surname is used as a first name by many other Canadians, Laurier is a charismatic hero whose term of office was a happy time in Canadian history. He worked all his life for cooperation between French and English-speaking Canadians while he strove to keep Canada as independent as possible from Britain. His personal charm and dignity, his great skill as an orator, and his great gifts of intellect won the admiration of all Canadians and non-Canadians alike.

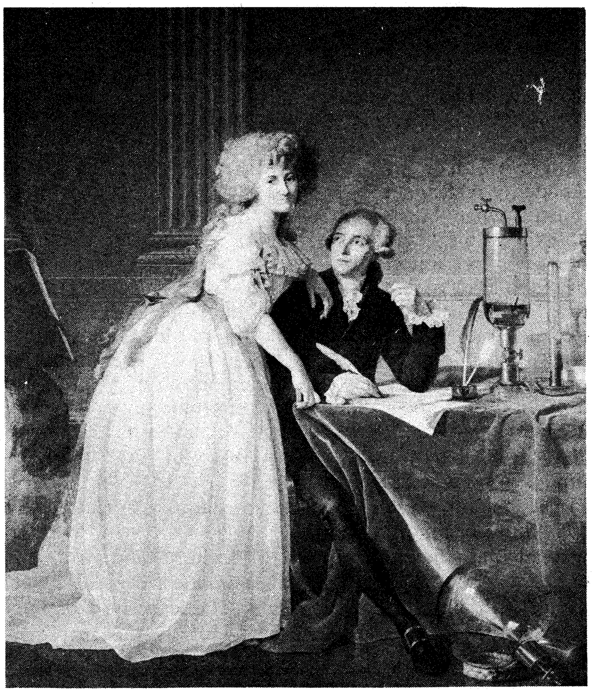
BIBLIOGRAPHY. Authoritative biographies of Laurier are L.O. DAVID, *Laurier: sa vie—ses oeuvres* (1919); and O.D. SKELTON, ***The Life and Letters of Sir Wilfrid Laurier, 2*** vol. (1921, reprinted 1965). These were written by friends and/or close political associates of the subject and naturally treat him most sympathetically. Also laudatory is SIR JOHN WILLISON, ***Sir Wilfrid Laurier and the Liberal Party: A Political History,*** 2 vol. (1903). A number of younger historians have begun a reappraisal of Laurier's life and work; among these, PAUL STEVENS, "Wilfrid Laurier: Politician," in M. HAMELIN (ed.), ***The Political Ideas of the Prime Ministers of Canada*** (1969), is of great interest.

(J.Mo.)

Lavoisier, Antoine-Laurent

A French chemist and the father of modern chemistry, Antoine-Laurent Lavoisier was a brilliant experimenter and many-sided genius. He was born in Paris on August 26, 1743. His father, an *avocat au parlement* ("parliamentary counsel"), gave him an excellent education at the Collège Mazarin, where he studied mathematics, astronomy, chemistry, and botany. In 1766 he received a gold medal from the Academy of Sciences for an essay on the best means of lighting a large town. Among his early work were papers on the analysis of gypsum, on thunder, and on the aurora, and a refutation of the prevalent belief that water is converted into earth by repeated distillation. He also accompanied the geologist J.-E. Guettard (1715–86) on a long geological trip and assisted him in preparing his mineralogical atlas of France. In 1768 he

By courtesy of The Rockefeller University, New York



Lavoisier, with his wife, oil painting by Jacques-Louis David, 1788. In the collection of the Rockefeller University, New York.

was nominated *adjoint-chimiste* ("associate chemist") to the academy. He passed through all the grades in the academy, was director in 1785, and was made treasurer in 1791. His father bought him a title of nobility in 1772.

Career of public service. In 1768 he became an assistant in one of the revenue-collecting departments of the government, subsequently becoming a full titular member of the Ferme Générale, the main tax-collecting agency. Appointed *régisseur des poudres* (a "director of gunpowder") in 1775, he abolished the vexatious search for saltpetre in the cellars of private houses, increased the production of the salt, and improved the manufacture of gunpowder. In 1778 he started a model farm at Fréchines, where he demonstrated the advantages of scientific agriculture. In 1785 he was nominated to the committee on agriculture and as its secretary drew up reports and instructions on the cultivation of crops, promulgating various agricultural schemes. Chosen a member of the provincial assembly of Orléans in 1787, he planned the improvement of the social and economic conditions of the community by such means as savings banks, insurance societies, canals, and workhouses. He advanced money without interest to the towns of Blois and Romorantin for the purchase of barley during the famine of 1788. He was associated with committees on hygiene, coinage, the casting of cannon, and public education. Lavoisier was a member of the nobility of the Blois district. In 1789 he drew up the code of instructions *(cahier)* for the guidance of the deputies to the States General, to which he was elected alternate deputy. He was secretary and treasurer of the commission appointed in 1790 to secure uniformity of weights and measures throughout France. This led to the establishment of the metric system.

Lavoisier's fate in the Revolution

Lavoisier's membership in the Ferme Générale was alone sufficient to make him an object of suspicion to the authorities during the French Revolution. The extremist revolutionary Jean-Paul Marat accused him of putting Paris in prison and of stopping the circulation of air in the city by the wall erected at his suggestion in 1787. In August 1792 he had to leave his house and laboratory. In November the Revolutionary Convention ordered the arrest of the former members of the Ferme Générale, and in May 1794 they were sent to be tried by the revolutionary tribunal. The trial lasted only part of a day. Lavoisier and 27 others were condemned to death. That same afternoon, May 8, he and his companions, including his father-in-law, were guillotined at the Place de la Révolution (now Concorde). His body was thrown into a common grave.

Scientific achievements. Lavoisier's name is indissolubly associated with the overthrow of the phlogistic doctrine, which had dominated the development of chemistry for more than a century, and with the establishment of the foundations upon which modern science rests.

Work on combustion

On November 1, 1772, Lavoisier deposited with the Academy of Sciences a note stating that sulfur and phosphorus when burned increased in weight because they absorbed "air," while the metallic lead formed from litharge by reduction with charcoal weighed less than the original litharge because it had lost "air." The exact nature of the airs concerned in the processes he did not explain until after the preparation of "dephlogisticated air" (oxygen) by Joseph Priestley in 1774. Then, perceiving that in combustion and the calcination of metals only a portion of a given volume of common air was used up, he concluded that Priestley's new "air" was what was absorbed by burning and that "nonvital air," azote, or nitrogen, remained behind.

In a memoir presented to the academy in 1777, read in 1779 but not published until 1781, he assigned to dephlogisticated air the name oxygen, or "acid producer," on the erroneous supposition that all acids were formed by its union with a simple, usually nonmetallic body. Combustion was explained by Lavoisier not as the result of liberation of hypothetical phlogiston but as the result of the combination of the burning substance with oxygen. On June 25, 1783, he announced to the academy that water was the product formed by the combination of hydrogen and oxygen; in this, however, he had been anticipated by the English chemist Henry Cavendish. From his knowledge of the composition of water, Lavoisier was led to the beginnings of quantitative organic analysis. He burned alcohol and other combustible organic compounds in oxygen and from the weight of water and carbon dioxide produced calculated their composition.

The spread of Lavoisier's doctrines was greatly facilitated by the defined and logical form in which he presented them in his *Traité élémentaire de chimie* (1789), and eventually they were adopted universally.

In addition to his purely chemical work, Lavoisier, mostly in conjunction with the mathematician and astronomer Pierre Laplace, devoted considerable attention to physical problems, especially those connected with heat. The two carried out some of the earliest thermochemical investigations, devised an apparatus for measuring linear and cubical expansions, and employed a modification of Joseph Black's ice calorimeter in a series of determinations of specific heats. Regarding heat (*matière du feu*) as a peculiar kind of imponderable matter, Lavoisier held that the three states of aggregation—solid, liquid, and gas—were modes of matter, each depending on the amount of *matière du feu* with which the substances concerned were associated. He also worked at fermentation, respiration, and animal heat, looking upon the processes concerned as essentially chemical in nature. A paper discovered many years after his death showed that he had anticipated later thinkers in explaining the cyclical process of animal and vegetable life.

BIBLIOGRAPHY. *Oeuvres de Lavoisier*, 6 vol. (1864–93), is a collected edition of Lavoisier's writings that includes all his main works and memoirs from the academy volumes as well as numerous letters, notes, and reports. D. MCKIE, *Antoine Lavoisier: Scientist, Economist, Social Reformer* (1952), is an authoritative biography that contains all relevant details of Lavoisier's life and work. D.I. DUVEEN and H.S. KLICKSTEIN, *A Bibliography of the Works of Antoine Laurent Lavoisier, 1743–1794* (1954) together with the *Supplement* (1965), is definitive and contains a description of all Lavoisier's writings together with copious explanatory annotations.

(D.I.D.)

Law, Western Philosophy of

Philosophy of law is concerned with the formulation of concepts and theories to aid in understanding the nature of law, the sources of its authority, and its role in society. In English-speaking countries the term "jurisprudence" is often used synonymously and is invariably used in reference to particular subdivisions of the field. (For material on non-Western law, see ISLAMIC LAW; CHINESE LAW.)

To the extent that it implies some necessary link or coincidence with general philosophy, the phrase "philosophy of law" may be somewhat misleading, for philosophy of law is mostly untouched by the conflicts of different philosophical schools, and its practitioners may without incongruity draw on diverse philosophical outlooks—on Logical Positivism for some analytical problem of the structure of legal orders and, simultaneously, on Existentialism for a problem of sociological jurisprudence or justice, for example—without commitment in any such instance to an entire philosophical outlook.

Nor can one treat philosophy of law as a specialized branch of philosophy such as ethical or political philosophy, epistemology, or logic, for in philosophy of law all these branches may make contributions. Ideas that may illumine jurisprudential problems must indeed be sought not only in philosophy but in all systematic bodies of thought. Only if "philosophy" is interpreted in its least technical and broadest sense does "philosophy of law" cease to be a misnomer.

PROBLEMS OF THE PHILOSOPHY OF LAW

Three major subdivisions of philosophy of law

Various approaches to the philosophy of law. For practical reasons, such as to avoid overlappings, it is convenient to organize jurisprudence into three principal branches only: analytical jurisprudence, sociological jurisprudence, and the theory of justice.

Analytical jurisprudence. The analytical questions in jurisprudence are concerned with articulating the axioms, defining the terms, and prescribing the methods that best enable one to view the legal order (or part of it) as a self-consistent system and that maximize awareness of its logical structure. Perhaps the most rigorous solutions are those which, like that of Hans Kelsen, a contemporary Austrian–American legal philosopher, attempt to identify structural or relational features as being necessarily entailed in the meaning of legal norms or in lawyers' intellectual operations with them (see below ***Pure theory* of *law***). Alternatively, the basis for logical structuring may be found in some imputed attribute of law not itself inherently structural. The 19th-century English legal philosopher John Austin, for example, thought it an essential preliminary to his quest for a logical system in law to clarify what was involved in his assumption that law always consists of "commands." This clarification is important, but the claim that such a clarified version of a common assumption necessarily amounts to an analytical model of law seems unwarranted.

On more modest levels, the analyst may seek to infuse clarity and orderliness into some particular branch of a legal system or even into the applications of some particular rule. Such work shades over, on the one hand, into ordinary legal analysis and, on the other, into jurisprudential efforts to clarify the meaning of particular legal terms. Analysis of a particular word or even a particular branch of law will usually, by necessity, be particularist in the sense that it works upon legal materials found only in one particular legal system. In between particularism and universalism lie what are sometimes called comparative approaches, in which analytical jurisprudence is applied to materials drawn from more than one (but not from all) legal systems.

Sociological jurisprudence. The sociological questions in jurisprudence are concerned with the actual effects of the law upon the complex of attitudes, behaviour, organization, environment, skills, and powers involved in the maintenance of a particular society. Conversely, sociological jurisprudence is also concerned with the effects of social phenomena on both the substantive and procedural aspects of law, as well as on the legislative, judicial, and other means of forming, operating, changing, and disrupting the legal order. The fact that men in a given time and place hold particular ideas and values, including ideals of justice, is itself a fact the relation of which to law must be studied; but the focus is sharply different from that in the study of theories of justice. Its focus is descriptive, not normative; it is concerned with what is or with what goes on, not with what ought to be or ought to go on.

The theory of justice. The theory of justice is concerned with the evaluation and criticism of law in terms of the ideals or goals postulated for it. This involves the identification and articulation of the values that the legal order seeks to realize. This aspect of jurisprudence is inextricably interwoven with ethical and political philosophy, and theories of justice thus tend to parallel the full range of ethical and political philosophies.

Law, morality, and natural law. A consideration of fundamental importance in the philosophy of law is that of the distinction between law and morality. The importance of the distinction is illustrated by the main questions to which it gives rise: (1) How far and in what sense should the law of a community seek to give effect to its morality? (2) Is there a moral duty to obey the law even when it does not embody morality, and, if so, are there any limits to this duty? (3) When a legal rule directs conduct that morality forbids, which should the citizen obey? (4) Is there ever (and, if so, when is there) a duty to overthrow an entire legal system because of its conflict with morality?

Definitions of morality

In all these questions the word "law" refers to the specialized form of social control familiar in modern, secular, politically organized societies. The word "morality" in the four questions may, however, refer to any of the following: (1) the community's relevant factual behaviour patterns (its mores); **(2)** its socially approved behaviour patterns, as sanctified by some widely held rational or religious ideal, whether observed in practice or not (social morality); or (3) the moral ideals accepted by each individual as binding on himself and on others, whether or not those others agree (individual morality). All these, like law, are means of controlling human conduct by setting normative standards; and all three have a constantly changing interaction with each other, as well as with law.

The fact that legal and moral norms vary from place to place and from one historical period to another lies in part behind a persistent theme in the philosophy of law: the search for unchanging norms that are universally valid. Clearly, the most certain way of establishing such norms would be to base them on widely observed facts, such as man's social propensities or the ubiquitous importance of kinship in social organization, which supposedly reveal something fundamental about the nature of man and his adjustment to the world. The attempt to base norms on some such category of facts has for two millennia been associated with the concept of natural law (*q.v.*). This concept has many versions, the principal of which are outlined in the historical survey below, but the significance of the topic merits some separate preliminary discussions.

Aristotle's conception of natural law

It has always been possible to trace a mainstream of natural-law thought. flowing from Aristotle's premise that the "nature" of any creature, from which obligations must be derived, is what it will be in its fullest and most perfect development. For man, this means what he is when the powers and qualities distinguishing him from other creatures, namely, his reason and his impulse to social living, are fully developed. Natural law embodies those obligations that will appear if mankind's reason and sociality are fully unfolded.

A major difficulty presented by this attempt to develop normative standards appears to be that it is very difficult to demonstrate, let alone create a sense of obligation toward, values that are only immanent. All theories of natural law, moreover, have found it necessary to rely on what are essentially intuitions or preconceptions as to what man's true nature is. All such theories acknowledge, for instance, that the full development or fulfillment of an entity is not the same as its mere continued existence, that there may be a "warping" or "impeding" of the natural tendencies, so that what exists may then "be said to be unsound or incorrect." Thus, mere factuality is not a sufficient source of obligation. Similarly, St. Thomas Aquinas himself, in identifying the "inclinations" from which men may learn natural law, found it necessary to order these in grades of inclination, so that those inclinations most closely related to reason and sociality take priority over those concerned (for example) with procreation and self-preservation. The criteria by which such a hierarchy is ordered must be drawn from sources other than the factual inclinations themselves. The "lower" grades (such as self-preservation) may well be based on something like instinct; but the question arises at the higher grades whether there is any comparable instinct by which men seek to find moral precepts binding all of them in common. Aquinas here appealed to synderesis, a kind of sympathetic understanding found in men, a disposition (habit) of the practical intellect inclining them to the good and murmuring against evil.

To derive from this synderesis a universal natural law, however, it would be necessary to demonstrate some "universal conscience" of all mankind. But natural lawyers faced with the fact that men's consciences do not coincide explain that conscience may err and reason be corrupt. Invocation of synderesis is in fact helpful not as an account of how one may arrive at factually based normative standards but as an illustration of the psychological tendency of men to assert values.

HISTORICAL SURVEY OF LEGAL THEORIES

The ancient world. ***Greek thought.*** The major contribution of Greece was a body of philosophical and cosmological ideals about justice, more apt for orators' appeals to popular assemblies than for preceptual application to day-to-day life situations.

Early Greek cosmologies, embedded in some of the earliest myths, had seen the individual as held within a kind of transcending harmony of the universe, emanating from the divine law (logos) and expressed in relation to human life in the law (nomos) of the polis, the city-state. The later Sophists, however, who examined critically all assumptions relating to life in the city-state, pointed to the wide disparities in human law and morals and rejected the claim that this human law (nomos) necessarily reflected any universal law (logos). Taking man as "the measure of all things," they rejected any claims of his law (nomos) to absolute value and saw law and justice and values generally as created by men's reasons, in their multitudes and generations, in all their individuated, relativistic, and historically changing dimensions.

Plato's idea of law

In the restless intellectual and political climate of 5th-century Athens, Plato was concerned to redefine the nature of justice by relating it to something far more permanent and absolute than the nomos of the city-state. He assigned "reality" to the unchanging archetypal forms—*i.e.*, the ideas--of things rather than to the ephemeral phenomena as superficially and confusedly perceived by individual men unenlightened by philosophy (see PLATO; IDEALISM). In the utopia described in The Republic, Plato defines justice in an architectonic sense: justice prevails when the state is ordered in accordance with the ideal forms ascertained by its philosopher-kings and is thus unrelated to the nomos of the city-state. There is no need for human law, since transcendental knowledge rules. In his later thought, however, as revealed in *Politicus* (the "Statesman") and the Laws, when he is concerned to describe a more practicable but nevertheless "second best" state, Plato assigns to law a role almost as important as that of knowledge in The Republic. A famous classification of states given in Politicus is indeed based on the criterion of whether or not they are ruled by law. The law as Plato here conceived it, however, was not mere convention or the imperfect individual judgments of men but a reflection of the common human reason in its full development. To this extent the rule of law might approximate the ideal rule of knowledge envisaged in The Republic, for in the inherited law of men is crystallized that much wisdom of which they are capable.

And yet it was difficult for Plato to find justification for such an argument in his basic philosophical position, with its emphasis on the contrast between the mere opinion of ordinary men and the transcendental knowledge of the philosopher. Aristotle, who in common with Plato held a view of nature or reality that transcended the variability of things as perceived by the senses, was, however, able more successfully to defend the validity of a law resulting from the practice of ordinary men. For Aristotle's transcendental reality is more firmly related to things as they are: it comprises that which they will become as their potentialities unfold in nature toward the end that is theirs in nature. Man, in his nature, is moral, rational, and social, and his law may be judged by the extent to which it facilitates the development of these innate qualities.

The Stoic conception of natural law

The Greek conception of natural law underwent further refinement by the Stoic school of philosophy, which became active toward the end of the 4th century BC. The Stoics posited the existence of a natural law, the *jus naturale*, which was an emanation of the lex *aeterna*, the law of reason of the cosmos. The existence of an innate reason in men linked everyone with the cosmic order and subjected all to a universally valid moral law. This latter concept thoroughly infused Roman thinking, largely as a result of the influence of Stoic philosophy on Rome.

Roman thought. Greek law scarcely survived as a system, because it never developed a class of legal specialists or abandoned its lay administrators or its popular tribunals of grotesque size. Roman law, on the other hand, developed through the efforts of expert jurisconsults (learned lawyers) and praetors (magistrates) into a permanent heritage of Western society. By its adoption into works such as Cicero's De re *publica* as well as in the work of the great jurisconsults, Stoic speculation concerning reason and nature was brought onto the level of precepts for concrete problem solving. The crude, tribal *jus* civile ("civil law") of the Romans was thus transformed into a natural-law-based jus *gentium* (law applying to all people), a set of principles common to all nations and appropriate, therefore, for application to foreigners as well as Romans.

Hebrew *thought.* In the Talmud there is an assertion that "Whatever decision of a mature scholar in the presence of his teacher will yet derive from the Law (Torah) that was already spoken to Moses on Mount Sinai." In theory, this presupposed that the Oral Law must respect every jot and tittle of the revealed written law. Yet the richness, ambivalences, and silences of what was written, in relation to a changing world, still left the widest freedom to the scholarly reason of the rabbinical exegetes into whose care both the written law and the Oral Law finally came.

The operations of the rabbinical schools and courts over many creative centuries, especially during and following the first Babylonian Exile, resembled those of the great Roman jurisconsults and the great judges of the common-law tradition. One Talmudic story tells of a doctrinal rift between the majority of a rabbinical court led by a great rabbi and a dissenting but no less great rabbi, in which the dissenter successfully summoned the authentic voice of God onto his side of the argument. To this intervention the majority of the court responded: "The law is not in heaven, the law has been handed down to us on earth from Mount Sinai, and we no longer take notice of heavenly voices. . . ." And the story relates that, at that point, God said with a smile to Elijah the prophet, with whom he was walking: "My children have defeated me, my children have defeated me." But this was an indulgent ratification, not an implacably cruel wrath such as the Greek god brought down on the head of Prometheus. Thus, even against divine intervention, the learned stood their ground, relying for the interpretation of the law on their own wisdom and reason.

The Middle Ages. Augustine. St. Augustine of Hippo, in attempting to refute the pagan assertion that Christianity was responsible for the decline of Roman power, reintroduced Stoic philosophy alongside Judeo-Christian thought into the stream of modern jurisprudential speculation. He placed God's reason beside God's will as the highest source of the unchangeable, eternal, divine law binding directly on man and all other creatures. The divine law was thus accessible to both man's reason and his faith and was not, as St. Paul had largely concluded, the product of his will alone and hence not rational in terms of human as opposed to divine reason.

At a second level, Augustine placed the no less unchangeable natural law, being the divine law as man is given the reason, heart, and soul to understand it. The third level, of temporal, or positive, law (for him, the Roman law of the Christian Roman Empire), was warranted by the eternal divine law, even though it changed from time to time and from place to place, so long as it respected the limits laid down by the divine and natural law. This rationale of secular power, some have thought, preserved the idea of government under law through the disintegration of the ancient world, for recultivation in the revival of learning of the 12th and 13th centuries.

Scholasticism. Aquinas, like Augustine long before, succeeded in quieting momentarily the competing claims of the will against the reason of God, the struggle between "voluntarism" and "rationalism," as the underlying basis of the eternal and natural law. Aquinas, like Augustine, gave a plausible place to both natural law and temporal (or positive) law under the eternal law. Human, or positive, law is a creation of human reason for the common good, within limits that natural law prescribes, so that even this proceeds from right reason and therefore from the eternal law. Such positive law as violated the natural and thus the eternal law "was not law" or merely was not binding "in conscience."

Voluntarism of Duns Scotus

The tendency to make reason prevail over will (as in Plato's call for philosophers to be kings or the Arab Averroes' call for philosophers to interpret what is revealed) was challenged by a voluntarist countermovement at Paris and Oxford in the quarter of a century after Aquinas' death in 1274. A Franciscan, John Duns Scotus, insisted on the uniqueness of all beings as finally traceable to the uniqueness of God's will. All precepts, even of the divine law, depend on the single precept "Love God," and, since not reason but will gives access to this, there is no natural law accessible to man's reason. All that can be required of human, or positive, law is that it must be "consonant" with the precept "Love God," or with any other precept willed by God.

The Renaissance period to the 18th century.

Machiavelli. Machiavelli presented himself (on one interpretation, at least) as seeking to escape from both transcendent will and transcendent reason into the empirical, into life as it is, observed through the eyes of a worldly man whose mind is uncluttered with philosophical and theological preconceptions. He can be understood, in his own words, to be seeking "what a principality is, the variety of such States, how they are won, how they are held, how they are lost." This conception was the more remarkable in 1513, since such an approach had then barely been promulgated for study of the physical world. It had still, indeed, to await its major manifesto in that sphere until Francis Bacon's ***Advancement of Learning*** at the end of the century.

The normative element in Machiavelli's thought

Even on the more favourable view of Machiavelli's aim —*i.e.*, as describing, rather than prescribing, political behaviour—it remains true that he saw this description as ancillary to the art of maintaining the state and its ruler, so that this maintenance is a kind of end in itself. The omnipotence—unrestrained by law or morality—that he both ascribes and prescribes to the prince is thus a product not so much of his scientific detachment as of his tendency to view political power as a value, as an end in itself.

Natural law and social-contract theory. The supremacy of the human lawgiver, as posited by Machiavelli and in their diverse ways also by the French and English political theorists Jean Bodin and Thomas Hobbes and others, interwove in the following centuries with continued insistence by Grotius and others on the dominance of the divine reason and man's participation in it, by which he has access to the natural law.

The views of Hugo Grotius

The Dutch political and legal philosopher Hugo Grotius, amid the political expediencies and anarchy of the Thirty Years' War (1618–48), sought to introduce a degree of normative restraint among the monarchical rulers of the newly emerged sovereign states of Europe and to establish a basis in natural law for a rejection of *raison d'état* as a just cause for war, as well as for legal limits on the means and modes of violence in war. Even if the wills of sovereign states form the basis of the international order, Grotius argued, "the totality of the relations between States" is still "governed by law." That law he found in an updated version of the Stoic natural law, as naturalized into Roman law and Christian theology.

With Grotius, as with the Stoics, the normative or moral power of the natural law derives from the fact that man's innate nature (itself part of the nature of the cosmos) and his propensities are viewed as ideal or inherently good. In Grotius' own time, however, there arose a skepticism toward such unfounded optimism, a skepticism that underlies the thought of Hobbes.

With Hobbes (1588–1679), as with the Greek Sophists, the nature of man is not the ideal nature of Grotius and the Stoics. It is rather man's supposed actual nature, before sociality and authority have tempered it. Man, in a state of nature, is motivated by desire and aversion and most of all by the desire to preserve his biological existence. This need for security is best met by all men vesting their rights of self-help in a sovereign, whether that sovereign be a single man or an assembly of men, and subjecting themselves to the laws of that sovereign, or "great Leviathan."

The reason why men must obey the law of the sovereign state, which is the only institution capable of protecting men against each other, is thus based firmly in Hobbes's conception of man's nature, albeit a very different conception from the idealist premises of earlier theories of natural law. Natural-law theorizing after Hobbes is thus divided into two streams.

Judicial supremacy. By the beginning of the 17th century the idea of applying natural law as a test of the validity of the positive law (the law as it is found within a jurisdiction) had passed from the province of speculative writers to courts of law. The English jurist Lord Coke, in Bonham's case in 1610, was already referring to the tradition that "when an act of Parliament is against common right or reason or repugnant or impossible to be performed, the common law will control it, and adjudge such act to be void." About a century before that, an English treatise known as "St. Germain, Doctor and Student" had already presented a three-tier hierarchy of the law of God, natural law (the law of reason), and human (positive) law, obviously deriving from Augustine and Aquinas.

Natural law and American constitutionalism

In the United States in the next century, constitutional theory became highly infused with ideas of natural rights. The Declaration of Independence, with its assertion of the self-evident rights of life, liberty, and the pursuit of happiness, marked the beginning of a continuing natural-law influence on American constitutional development. The power of the judiciary to "review" legislation for consistency with a written constitution was taken in the United States to import the power to declare it void, constitutional law being analogized to natural law. Indeed, American judicial statements of 1814, 1822, and 1831 asserted the power of the judiciary to strike down statutes for violation not only of explicit constitutional restraints but also of "eternal principles of justice which no government has a right to disregard." The analogy of constitutional and natural law did not necessarily require that the power to strike down legislation should be a judicial power: this was not so in ancient Rome, nor is it always so in modern civil-law countries. It is arguable that such a judicial repository of the power of final review is unavoidable, since the legislature cannot be expected to annul its own acts; and the executive, even if it were not a party to such acts, is scarcely equipped for the tasks of objective interpretation involved. Yet there are real difficulties of policy and principle raised by giving the judiciary the final word. A distinction must first of all be made among diverse constitutional restraints. Safeguards for such rights as free speech and assembly and access to courts, which help to assure the responsibility of rulers and to prevent the fall of democracy into tyranny or demagoguery, may well be placed in the final custody of judges. But, beyond this point, others have argued, judicial supremacy, in enforcing restraints laid down by the Founding Fathers of an earlier generation, may clearly constitute an obstacle to the implementation by the courts of a society's present convictions.

Decline of natural law. If man is the measure of all things, as the Sophists taught, then a given society of men is the measure of its culture, including its moral and legal standards. In the modern period the French jurist and political philosopher Montesquieu's ***De l'esprit des lois*** (1748) and ***Lettres persanes*** (1721) offered the thesis that a people's law and justice are determined by the particular factors and environment that operate upon them. They thus could not, as the natural-law theory of the time held, be unchanging from age to age and from people to people. The French sociologist Auguste Comte's ***Système de philosophie positive*** (1851–54), which set out to explain positive laws, like other social facts, by reference to verified hypotheses concerning cause and effect and interaction, was similarly antithetical to natural-law theory as it had so far developed. To Comte, metaphysical concepts about such abstractions as ideal essences belonged to a past stage in man's intellectual development. And Darwin's ***On the Origin of Species by Means of Natural Selection*** (1859), the English philosopher Herbert Spencer's positivism, and other related thinking of the period provided a biological model of self-development of organisms and institutions through a struggle in which survival was a function of challenge and response in the given environment. Change and adaption, rather than constancy and inviolability, were thus at the heart of their system.

Under the leadership of anthropologists, analyses of man's internal process of response to the exigencies of existence within a particular culture—to conscious and subconscious psychic drives and motivations—deeply affected the jurisprudential study of law and society and helped to bring natural-law thinking to a 19th-century nadir. In the anthropologist Bronislaw Malinowski's most mature statement on the matter, he distinguished four major meanings of the word law as important in understanding the growth of civilization. They included "laws of nature" in the scientific sense of rules governing men's conscious adaptations to the environment; rules of "efficiency" and "convenience" according to which the group lives; rules for conflict adjustment; and rules about enforcement of the last two. No conception of natural law, which had engaged earlier thinkers for two millennia and more, was included.

Idealism and justice. Another line of thought, which was also divorced from natural-law concepts, was contained in the Idealist philosophy of Immanuel Kant. Fundamental to Kant's ethical and jurisprudential reasoning is the premise that all moral concepts have their basis wholly in a priori thought, that they can be arrived at by reason alone, without reference to experience or recourse to intuition of rules alleged immanent in experience (see KANTIANISM). Man, furthermore, is a free agent whose actions are determined by aims that he is at liberty to select. From such premises Kant deduced the nature of an ideal law, in which is implicit a theory or criterion of justice. This ideal law comprises the conditions under which all members of society can enjoy the maximum freedom from subjection to the arbitrary will of others.

Kant's ideal law

But Kant's supposedly a priori concepts are in fact as transcendental as anything natural lawyers have offered. It is thus not surprising that later thinkers, such as Johann Fichte, Kant's Idealist successor, had little difficulty in putting the new Kantian wine into natural-law bottles.

The 20th century saw a fresh attempt at the Kantian approach in the work of the German legal philosopher Rudolf Stammler. Adopting the Kantian position that knowledge is independent of sensory experience, Stammler set out to discover pre-experiential categories, or "pure forms," of thinking about law. Stammler arrived at a social ideal of a "community of free-willing men," an ideal that he claimed to have universal validity because of its supposed a priori basis. Having thus arrived at a "pure" ideal of society, untainted by empirical content deriving from sense perception, he felt able to formulate equally pure principles for just law that would regulate his ideal society. Stammler's pure idea of society comprised the harmony of individual and common purposes: his pure idea of just law thus comprised those principles conducive to such harmony—the mutual respect of individuals for each other's purposes and the participation of all in the achievement of the common purposes.

The different stream of Idealism flowing from Hegel's philosophy of history (see HEGELIANISM) was fed into jurisprudence by Josef Kohler, Stammler's close predecessor in that subject in the Berlin University. His work is still another effort to relate social facts and the norms of justice by exposing the immanence of values in facts—in "civilization" in Kohler's case.

In perspective, these idealisms, despite their formal or philosophical antagonism to "rationalism" and natural-law thinking, seem to have reinforced in the age of the Industrial Revolution the individualist and libertarian trends that natural law had built up successively against medieval church and empire, the shackles of medieval social, political, and economic organization, and 18th-century despotism.

The 19th and the 20th centuries. ***Analytical Positivism.*** The close of the 18th century witnessed a reaction against both Kantian Idealism and iusnaturalism (natural-law theorizing). The scientific temper of the age, reflected in the practical achievements of the early decades of the Industrial Revolution, was not conducive to deductive reasoning from a priori hypotheses, which appeared an impractical method of solving the problems of complex societies. Such problems might better be approached via a thorough analysis of existing law and institutions. This new climate of opinion came to be known as Positivism.

Meaning of Positivism in jurisprudence

Among the chief meanings of Positivism in the legal-analytical sphere are the separation of law as it is and law as it ought to be, stress on the analysis of legal concepts, reliance on logical reasoning in the search for applicable law, and denial that moral judgments can be based on observation and rational proof. Anglo-Saxon analytical Positivism has directed itself mainly to the logical dissection, appraisal, and clarification of the precept element of law, ignoring the elements consisting of lawyers' traditional techniques and received ideals. By the nature of its tasks, analytical jurisprudence does not concern itself with either the facts surrounding or the consequences flowing from legal precepts or with their ethical evaluation, though particular analysts may also be interested in those matters.

Analytical Positivism in England began with the work of the philosopher and legal reformer Jeremy Bentham. His work influenced John Austin, the most outstanding figure in English jurisprudence, who set out to analyze the notions pervading English law. In order to delimit his subject, he defined positive law as the commands of a sovereign addressed to political inferiors and backed by threats of evil in the event of disobedience. Positive law might well be derived from moral precepts and other sources, but such precepts become law only when commanded by a sovereign.

The analytical-Positivist attitude has continued to influence thinkers, although the particular approach of Austin is now of historical interest only. Logical analysis is clearly a tool that may be employed in many spheres of jurisprudence, and its importance thus transcends the limits of any one school. Analysis means little in itself; its value depends largely on the validity of the premises from which the argument is made and on the relevance of the subject matter that is chosen for analysis. The paramountcy for the analytical Positivist of questions of logical order and consistency represents a permissible deviation, yet still a deviation, from the wider concerns of ordinary lawyers and students of law and society generally. This is not to say that there is a necessary conflict between the requirements of justice among changing social facts and those of logical consistency of precepts. It means only that there is surely no necessary (nor indeed usual) coincidence between them.

Historical Positivism. In discarding speculative cosmology, the a priori, and the self-evident, 19th-century historical jurisprudence opened the way for the search for the realities of law through empirical observation—for a sociodescriptive rather than a logico-analytical-Positivist jurisprudence.

The leading figure in the historical school was the German jurist Friedrich Karl von Savigny, who confronted the natural-law aspiration for a universal human code with the singularity of the law of particular peoples resulting from their unique sociocultural experiences.

For Savigny, law rests on the ***Volksgeist,*** or innate popular consciousness; law ***par excellence*** is customary law. He recognized, of course, that the details of a developed legal system do not spring from simple group intuition. With maturity, both life and law become more specialized and artificial, creating a dualism in more mature law. Part of such a system still rests directly on the popular consciousness and way of life ("the political element"); but this becomes elaborated by jurists, be they Roman jurisconsults or common-law judges, who in this respect represent the community ("the technical element").

Savigny's emphasis on the need of legal change to respect the continuity of the ***Volksgeist*** offers a pre-Darwinian concept of juristic evolution. The ***Volksgeist*** corresponds to modern notions of social rather than biological inheritance. Savigny's sense of the impotence of legislatures in the face of the restraints imposed by the ***Volksgeist*** foreshadows modern recognition of the social and psychological limits of effective legal action.

The English legal historian Sir Henry Maine's dual academic concern with both English law and Roman law challenged him to explain their independent yet often parallel growths and may well have redeemed him from Savigny-like overemphasis of national uniqueness. His concern led him to a comparative historical jurisprudence seeking hypothetical "laws" of development controlling all legal systems. He saw changes in substantive law and in the machinery and modes of legal enforcement and growth as moving in pace with certain recognizable stages in social growth, from primitive, kin-organized society to the mature, complex commercial and industrial societies of Europe.

Maine's experience in India after the publication of his *Ancient Law* in 1861 broadened his interests so that he embraced less well known and less developed systems, such as the Brehon, Hindu, Welsh, Germanic, Anglo-Saxon, and Hebrew. His breadth of interest matched the concurrent growth in anthropological study of primitive peoples.

Maine's work shows the strong combined influence of the analogy of biological to social evolution and of the Hegelian philosophy of history. The consequent, somewhat mechanistic tenor of his interpretations resulted in his being accused by many anthropologists and legal historians of making false assumptions concerning the pattern and sequence of social development.

The Marxist conception of law

Economic interpretations. Certain residues of the Marxist economic interpretation of history have won a central place in sociological jurisprudence (see below ***Sociological jurisprudence),*** as indeed in most branches of social science. One such persistent trend of thought is the close interrelatedness of legal, ethical, economic, and psychological inquiries; another is the pre-eminence among these of economic factors. According to Marxist doctrine, the political and judicial systems—the state and the law—represent the superstructure of society, their nature being determined by the economic base—the mode of production and exchange. The state and its repressive law are but instruments of class domination, becoming redundant under Communism, which has no need of coercion. During the transition to full Communism, they would "wither away." There were, of course, softenings of this bold doctrine in its original authors, with admissions that the ethical or legal superstructure should not be seen as a merely passive effect; and Lenin himself pressed to extremes both the passion of the original thesis and its qualifications. Lenin, indeed, saw state power as an essential weapon of the proletarian dictatorship until the movement to a full Communist society should be completed.

The first half-century of the Soviet Union, with its steady consolidation of state power and its attendant law, has imposed the severest strains on the withering-away prediction. The general tenor of explanation is that the "law" the disappearance of which is prophesied refers only to the kind of coercive order manifest in such instrumentalities as the courts, police, and jails of capitalist countries.

Within these sweeping theses of Marxist thinking, more modest subtheses have played a valuable part. The Socialist jurist Karl Renner, for example, in his ***Rechts-institute des Privatrechts und*** *ihre soziale Funktion* (1929), was concerned to show that the legal conception of ownership, formulated in early economies, had profound new effects when continued as an institution of the 19th-century economy. It then, through the law of property and contract, alienated into private hands great segments of what should be in the public domain.

Even more notable are the German sociologist Max Weber's studies of the correlations of socioeconomic and ethicojuristic change, freed of the straitjacket of economic determinism. In these, the impact of unique factors or combinations of factors in particular civilizations is taken into account, including the existence of accepted systems of values, immediate and ultimate, which may (and in Weber's view did) have a decisive effect on the emergence of the Western capitalist system.

Sociological jurisprudence. The historical jurisprudence of the earlier part of the 19th century became subject to the influence of the developing social sciences, which attempted to explain law in its social context. The result was the emergence of a sociological school of jurisprudence.

The early decades of sociological jurisprudence combined 19th-century faith in progress, social evolution, rationalism, humanitarianism, and political pluralism with a sanguine belief that the Newtonian model of natural science would also hold for the social sciences. It was affected by questions of whether the social sciences are truly sciences, what their mutual boundaries are, and whether they can be integrated or somehow transcended by some subject such as sociology or anthropology.

An outstanding figure of the early sociological school was a German, Rudolf von Jhering, who in the 1860s contributed to the intellectual stream a theory of justice predicated on a view of law as a social phenomenon. He saw law as an outcome of the struggle of men to fulfill their purposes and of the force that they marshal behind this. Another historical jurist, the German Otto von Gierke, stirred a related interest with his emphasis on the importance of the inner life and activities of groups and associations as sources of binding social norms. This opened up jurisprudence to some psychological issues. Gierke's work also contributed to the later American Neorealism through its influence on Oliver Wendell Holmes, Jr., and to the theory of the "living law" of the Austrian jurist Eugen Ehrlich, in the first decade of the 20th century. Ehrlich insisted on the profuse norm-creating activities of the countless associations in which men are involved.

Early 20th-century views

At the beginning of the 20th century a great variety of psychological hypotheses were brought to bear on law. A theory of dynamic psychic drives, for example, was propounded by an American sociologist, Lester F. Ward, who argued that such drives could be utilized in social planning. Freud's exploration of psychic activity on a subconscious level, as well as studies of the nonrational and the irrational in the social process by the Italian and German sociologists Vilfredo Pareto and Max Weber, were also profoundly influential.

Revival of natural-law theories. Iusnaturalism, in the sense of the assertion of an order of norms for human conduct transcending human will, to which the validity of positive law is subjected, has certainly experienced a 20th-century revival. The massive human delinquencies of the century, such as those of the Nazis, have been important in stimulating these modern natural-law yearnings. The revival, indeed, has rarely overthrown dominant Positivist positions, but it has certainly reopened some questions that Positivists have not adequately faced.

Contributions to this re-emergence have come from varied directions, rather than from a single intellectual movement. They have often avoided explicit reference to natural law and have even expressed hostility or ridicule toward it. The German Stammler and the French jurist François Gény were certainly among its pioneers. Gény's ***Méthode*** *d'interprétation* (1899) displayed the inescapably creative (or lawmaking) role of the judiciary even under a comprehensive code such as the Code Napoléon. It led him to the questions of what are "the sources of law" and where does the legislator's prescription fall short. Answers to such questions must be based on the facts of each particular situation to be adjusted—the legislator cannot impose his view on the court. This line of thinking foreshadowed a variety of doctrines about "the nature of things" or "the nature of facts," all of which shared the idea that the decisive nature of a situation has its base in the facts for which men seek governing law. The properties and circumstances of these facts themselves afford immediate guidelines for just regulation. The fact situation, if only its essence will be perceived, has the superior applicable norms immanent within it.

In his ***Lehre von*** *dem* ***richtigen Recht*** (1902), Stammler sought, as described above, the a priori social principles of just law concerning respect for and participation by all members. His call for "natural law with a changing content" based on these a priori principles quickly became a 20th-century slogan.

Even as this express reinvocation of natural law was proceeding, the French public lawyer Léon Duguit was expressly denouncing it. Duguit's concern was to place law and lawyers within what he saw to be the correct frame. This he found in Émile Durkheim's Positivist sociology. This led him, with some paradox for a contemner of natural law, to insist that law is but *"le produit spontané des faits"* ("the immediate result of the facts"). The observed "facts" of social solidarity arising from economic specialization of functions generated, Duguit argued, the society's norms. Breach of these norms causes social disorder and a spontaneous movement toward readjustment. Even a supreme legislator was bound (Duguit affirmed) by this objective "rule of law," so that his acts violating it are void, even apart from any other constitutional restraint. All this bears the clear iusnaturalist mark of the assumed immanence in observed facts of a transcending and overriding order. It pays cryptic homage to a natural law, fealty to which Duguit denied.

The German legal philosopher Gustav Radbruch's turn toward natural law at the end of a life of great contributions to democratic legal relativism and Positivism was very different. Positivism, Radbruch argued, had encouraged German lawyers to stand by at Nazi barbarism, declaring *"Gesetz ist Gesetz"* ("Law is Law"). Nor was Radbruch's turn to natural law in any way cryptic. He came to declare quite openly that:

> where justice is not even striven for, where equality which is the core of justice is constantly denied in the enactment of positive law, there the law is not only "unjust law" but lacks the nature of law altogether.

The linkage with the revived natural law of the legal institutionalism of the French legal philosopher Maurice Hauriou and the writer and historian Georges Renard is different again. As with Duguit, the linkage is not proclaimed, but no overt hostility disguised their obvious sympathy for Thomist positions. Theirs is a Catholic version of institutionalism (which regards social institutions such as the family or the corporation as expressing the social reality underlying the law). The natural-law assumptions are apparent in the insistence on "the principles of organization," the "communion" of members in realizing "durable ideals," and the placing of men's powers of organization into the service of such ideals, as essential elements of any institution. For them, as for Duguit, the principles of justice were principles of social organization, immanent and self-evident.

It has been tempting for many to seek kinships between natural law and Existentialism, as was attempted by the German legal philosopher Werner Maihofer. Such efforts seem, however, destined to denature either Existentialism or natural law itself. Even in all their varieties, Existentialist positions approach no nearer to natural law than to assert that the traumas, anxieties, and demands of mere "existence" confront men with fateful value choices. Yet this is far short of asserting that any transcending principles of harmony may be discoverable.

Abstract symbols such as "social solidarity," "the principles of social organization," or "immanence in the facts of social life" are by virtue of their ambiguity susceptible to misappropriation by absolutist governments. The same may be said of Savigny's *Volksgeist* notion, as witness its affinity to the racialism of Nazi law. Thus, while the modern revival of natural law has been in part a revulsion from totalitarianism, it can also be exploited to rationalize totalitarianism.

There is another paradox also: The growth of the social sciences has invited restatement of natural-law traditions in terms of social ideals. Yet the very complexity of the social and economic orders and of their attendant sciences has placed forbidding barriers before the aspiration to base justice or other values on "objective" knowledge. Some have been tempted to hope that natural law may somehow overleap such barriers.

Pure theory of law. In part, at least, the influence of the distinguished legal philosopher Hans Kelsen's "pure theory of law" reflects early 20th-century skepticism about natural law and sociology, to both of which Kelsen opposed his claimed purity of method; *i.e.,* a method free from contamination by values of any sort.

He asserted, first, that legal theory was properly a science in the sense of an uncommitted, value-free, methodical concern with a determined object of knowledge. Second, he argued, legal theory must be isolated from psychological, sociological, and ethical matters. Third, purity of method permits the analyst to see that every legal system is in essence a hierarchy of norms in which every proposition is dependent for its validity on another proposition. The justification for describing any particular rule as law thus depends on whether there is some other proposition standing behind it, imparting to it the quality of law. This regression is continued until the *Grundnorm,* or "basic norm," is arrived at. The basic norm derives its validity from the fact that it has been accepted by some sufficient minimum number of people in the community.

Kelsen's hierarchy of norms

Kelsen's assertion that norms can spring only from other norms seems but another way of stating his rejection of the relevance of facts to values and, therefore, of iusnaturalism and sociology to his pure science of law. Yet finally it has seemed to many dubious whether the Kelsenite theory itself escaped the *liaison fatal* between facts and norms; for, if all legal norms must finally hang on the basic norm, then whatever it is that the basic norm hangs on must be nonlaw. And whether the basic norm hangs on "habitual obedience to determinate persons," as the English legal philosopher John Austin in effect proposed a century before, or on "efficaciousness," as Kelsen proposed, what it hangs on is fact rather than norm. Critics have complained that, at most points in the creation of norms in Kelsen's system, what is decisive is the intervention of acts of will of persons endowed by higher norms with norm-making authority. The determination whether such acts of will have occurred is a factual inquiry, to the decisiveness of which Kelsen's pure theory gave little weight.

Modern schools of realism. The American jurist Oliver Wendell Holmes's description of law in 1897 as "what the courts will do in fact" and of the "real ground" of decisions as resting often in some "inarticulate major premise" rather than in expressed reasons gave 20th-century legal realism its central theme.

Certain features are common to the "realist" jurists. They include (besides the above-mentioned concern with "the law in action") stress on the social purposiveness of law, on the endless flux in both society and law, on the need to divorce the "is" and the "ought" for purposes of study and to question all orthodox assumptions made by lawyers, and in particular on the need to substitute more realistic working categories for current lawyers' generalities. Among the orthodoxies thus challenged, these writers tended to include the works of early sociological jurisprudence. Yet it is clear, from the present perspective, that the concerns common to the realists and the more orthodox sociological jurists were far more important than the ephemeral if bitter conflicts that at first flared up between them. The American realists in their important surviving contributions have for the most part reinforced, clarified, and elaborated a number of main insights, notably about rule uncertainty and fact uncertainty, which they shared with sociological jurisprudence.

Scandinavian realism

Scandinavian realists, while temperamentally akin to their American colleagues, were rather different in intellectual concerns. Methodologically, they invoked a somewhat gross empiricism, leading them to deny that the law could be the subject of scientific inquiry at all, since its concepts and principles are not founded on spatial and temporal data of experience. Taking lawyers' talk of the will of the sovereign very literally, they were concerned to show that there is no such will of common content and that even legislators who enact a code are merely rubber stamping what others drafted.

The Swedish jurist Axel Hagerstrom insisted that the idea of rules of law as commands is an idea not corresponding with facts. His disciple Karl Olivecrona added that this false idea results from the syntactical imperative form used in modern legislation. Such rules, he urged, were commands only in a depersonalized sense. He preferred to describe them as "independent imperatives." Such "imperative statements about imaginary actions, rights, duties" may not be directed to any particular persons. Yet, even if some legal rules are directed "so to say, into the air," others are certainly directed to particular persons. If any form of imperative notion is to be preserved, it should be one that accommodates both situations.

Some of the problems that these writings address are rather tied to the special experience of their authors' own legal cultures. Others reach out independently toward truths already reached earlier in Anglo-American jurisprudential scholarship, especially as to the merely noetic and conceptual (rather than physical or psychological) nature of rights, duties, and liabilities.

THE STATUS OF CONTEMPORARY PHILOSOPHY OF LAW

By the middle of the 20th century, serious scholars no longer argued for or against the exclusive imperium of either the analytical logical, the justice-ethical, or the sociological approach. Whether jurisprudence is a single field in some scientific sense or whether its unity lies in the need to serve the intellectual needs of those concerned with making, applying, improving, or generally understanding law, all the above areas are included within it.

The "revolt against formalism"

A characteristic feature of contemporary jurisprudence is what has come to be known as "the revolt against formalism"; that is, against preoccupation with the technical and logical aspects of law. It can be traced back to Savigny's early-19th-century reaction against natural law, to Jhering's attacks on the German Pandectists (commentators on Roman law), and to Maine and the work of the anthropologists and early sociological jurists. Its early pressure was toward broader and deeper history, toward recognition of the organic nature of the processes of cultural growth, and toward problems of social action and the value choices therein entailed.

In the United States the legal philosopher Morton White identified five later contributing strains of thought, including the pragmatism of John Dewey; the economist Thorstein Veblen's institutionalism, rejecting both the abstractions of classical political economy and the fatalism of the Marxist interpretation of history; the revolt within jurisprudence of the American legal realists already described; and the approach to history as no mere chronicle of kings and battles but rather as a product of underlying economic forces and a guide to present and future civilizations.

The sometimes-overhasty iconoclasms of this revolt have proved less important than its positive affirmations. It has affirmed, for example, that the evaluating activities of justice must somehow move alongside the describing activities of sociological jurisprudence, that the choices of ethics, social policy, and justice still remain to be made when all the empirics of social science are done. The central question includes not merely what are the facts but also what should be done about the facts. These affirmations reject any regression to simple amoralism, stirring new temptations to return to natural law or other intuitive absolutes.

There are important advantages in the drive, characteristic of much contemporary social science, for overall cognition of the social and legal orders and the identification of key points for social action within them. But there are also dangers, for, especially with subject matters such as the law, systematic theory and overall cognition can rarely be of aid save in the rather long run, for which present decision makers cannot usually wait.

Growth of the sociological school. The most eminent pioneers and champions of modern sociological jurisprudence were Roscoe Pound in the United States and Hermann Kantorowicz in Europe. For both, the task of sociological jurisprudence, though orientated mainly to practical administrative or legislative problems, included that of framing hypotheses (as to the limits of effective legal action, for example) on which to base general laws of the operation of law in society.

Methodology of sociological jurisprudence

As with the social sciences, the principal methods available to sociological jurisprudence are those of survey, statistical analysis, comparative observation, and experimentation. The controls and corrections available usually fall far short of those of the natural-science models. Much work in sociological jurisprudence has merely brought to bear upon the law relevant findings from other social sciences. But it may also generate its own findings, as it has done in relation to traffic laws, control of moneylending, credit unions, bankruptcy laws, the effect of antitrust practices or of poverty on legal rights, the theory of appellate judicial decision making, and a host of other matters. Examinations of the prehistory and aftercareers of convicted criminals and of persons on probation or parole, probings of family and environmental influences bearing on potential deviance, and attempts to identify decisive factors predictive of future deviance are among the staples of sociological jurisprudence.

Sociological jurisprudence is confronted by the questions whether (and, if so, how and how far) it is possible through empirical methods to approach central issues of social action that involve value judgments. The fact that lawyers are necessarily involved with ideas of obligation, values, and norms sharpens this confrontation. A second group of problems arises from the high level of individuality of men, groups, and societies, from the unending variety of their emotions, roles, and expectations, and from the feedback effects on human behaviour that the empirical observation and testing of that behaviour brings about.

These problems give central importance to efforts to develop frames of social knowledge that give due place to both facts and values. Such inquiries show the great complexities of values held and their intricate and dynamic relation to the physical and cultural environments.

The study of law in society thus shares with anthropology and other social sciences a central interest in roles and functions as basic meaningful categories and in certain mechanisms and channels whereby conduct is thought to become socially meaningful. These notions are thought to permit the analysis of complex social situations into more refined terms, such as constituent goals, tasks, expectations, and allocated rights, powers, and duties.

As to the mechanisms or channels through which conduct becomes socially meaningful, earlier thought tended to explain social norms as built up from individual instances through group usages and mores that then crystallize in institutions such as law. Insofar as this suggests a cumulative movement or process, current thought would regard it as oversimplified. The growth of socio-ethical convictions is rather to be seen in terms of symbolic interaction between individuals. A particular society may be seen, in this light, as a collection of individuals with a culture that has been learned by symbolic communication from other individuals back through time, enabling members to gauge their behaviour to each other and to the society as a whole.

The future of sociological jurisprudence. In his famous program of 1911–12, Pound formulated a series of rather practical objectives for the movement, including making studies of the law in action, of the means of more effective legislation and law enforcement (by creation of ministries of justice, for example), of legal and judicial reasoning, of legal history in its social context, and of the role of the legal profession. An early quip against the sociological school was that it was like a great orchestra constantly tuning its instruments but never actually playing. Yet many practical tasks have been performed, and the school continued to show a gathering momentum and a widening range of concerns.

The maladjustments and inadequacies of the law gave to early sociological jurisprudence an intensely activist drive, directed to ad hoc remedies, and a great deal of the relevant work is still of this nature. Especially since 1945, however, juristic work on the relations of law and society has come into more fruitful contact with other social sciences, leading in turn to greater stress on cognition of the social and economic orders in their complex unity. Whatever the difficulties of designs for an overall analysis of the social system, some adjustment toward them is inevitable for sociological jurisprudence. This is in part, no doubt, a result of the waning of interest in many of the kinds of ad hoc problems with which it was initially concerned. But the interest in sociological theory also results from growing awareness that some problems require to be approached on a wider basis. This has created new stirrings of the turn-of-the-century ambition that the study of law in society become a specific branch of social science, concerned with framing and testing general laws governing law as a social phenomenon.

BIBLIOGRAPHY

Leading treatises in English: A. ROSS, *Om ret og retfaerdighed (1953;* Eng. trans., *On Law and Justice, 1958);* W. FRIEDMANN, *Legal Theory,* 5th ed. *(1967);* H. KELSEN, *General Theory of Law and State (1945);* G.W. PATON, *A Text-Book of Jurisprudence,* 3rd ed. by D.P. DERHAM (1964); R. POUND, *Jurisprudence,* 5 vol. *(1959);* J. STONE, *Legal System and Lawyers' Reasonings (1964); Human Law and Human Justice (1965); Social Dimensions of Law and Justice (1966).*

Notable works on the history of philosophy of law: E. WOLF, *Griechisches Rechtsdenken,* 3 vol. *(1950–54);* A. VERDROSS, *Abendländische Rechtsphilosophie,* 2nd ed. *(1963);* R. MARCIC, *Geschichte der Rechtsphilosophie (1971);* G. FASSO, *Storia della filosofia del diritto,* 2 vol. (1966–68; 2nd ed., vol. *1, 1970).*

Basic works in the main areas: (Analytical jurisprudence): J. AUSTIN, *Lectures on Jurisprudence (1832);* H.L.A. HART, *The Concept of Law (1961);* W.N. HOHEELD, *Fundamental Legal Conceptions As Applied to Judicial Reasoning (1923);* H. KELSEN, *Reine Rechtslehre,* 2nd ed. *(1960;* Eng. trans., *Pure Theory of Law, 1967);* A. KOCOUREK, *Jural Relations,* 2nd ed. *(1928);* I. TAMMELO, *Outlines of Modern Legal Logic (1969). (Ethical jurisprudence):* C.K. ALLEN, *Aspects of Justice (1958);* J. BENTHAM, *Introduction to the Principles of Morals and Legislation (1780);* E. BODENHEIMER, *Treatise on Justice (1967);* H. KELSEN, *What is Justice? (1957);* C. PERELMAN, *Justice (1967);* G. RADBRUCH, "Legal Philosophy," in K. WILK (ed.), *The Legal Philosophies of Lask, Radbruch and Dabin,* pp. *47–226 (1950). (Sociological jurisprudence, including historical jurisprudence):* W. AUBERT (ed.), *Sociology of Law (1969);* E. EHRLICH, *Grundlegung der Soziologie des Rechts (1913;* Eng. trans., *Fundamental Principles of the Sociology of Law, 1936);* L.M. FRIEDMAN and A.S. MACAULAY (eds.), *Law and Behavioral Sciences (1969);* G. GURVITCH, *Sociology of Law (1942);* K.N. LLEWELLYN, *Jurisprudence (1962);* H. MAINE, *Ancient Law (1861);* F. VON SAVIGNY, *Vom Beruf unsrer Zeit fur Gesetzgebung und Rechtswissenschaft (1830;* Eng. trans., *Of the Vocation of Our Age for Legislation and Jurisprudence, 1831);* E.M. SCHUR, *Law and Society (1968);* J. STONE, *Law and the Social Sciences in the Second Half-Century (1966).*

Notable recent developments: (Controversy on iusnaturalism and legal positivism): N. BOBBIO, *Giusnaturalismo e positivism~giuridico (1965);* L.L. FULLER, "Positivism and Fidelity to Law," *Harvard Law Review,* 71:630–672 *(1958);* H.L.A. HART, "Positivism and the Separation of Law and Morals," *ibid., 71*:593–629 *(1958). (Non-stringent legal reasoning and theory of justification):* C. PERELMAN, *De la justice (1945;* Eng. trans., *The Idea of Justice and the Problem of Argument, 1963);* T. VIEHWEG, *Topik und Jurisprudenz,* 4th ed. *(1969);* R.A. WASSERSTROM, *The Judicial Decision: Toward a Theory of Legal Justification (1961).*

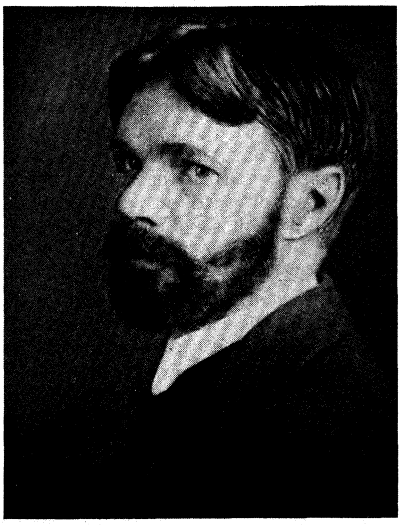
Quantitative analysis of judicial behaviour: R.C. LAWLOR, "What Computers Can Do: Analysis and Prediction of Judicial Decisions," *Journal of the American Bar Association,* 49:337–344 *(1963);* G.A. SCHUBERT, *Quantitative Analysis of Judicial Behaviour (1959); The Judicial Mind, the Attitudes and Ideologies of Supreme Court Justices, 1946–1963 (1965);* and with D.J. DANELSKI (eds.), *Comparative Judicial Behaviour (1969);* J. STONE, "Man and Machine in the Search for Justice," *Stanford Law Review,* 16:515–560 *(1964).*

(Ju.S.)

Lawrence, D.H.

One of the most original and controversial English writers of the 20th century, D.H. Lawrence revealed the instinctual forces in human nature, presenting a bitter critique of modern industrial society.

Life and works. David Herbert Lawrence was born on September *11, 1885,* at Eastwood, Nottinghamshire. He was the fourth child of a coal miner; his mother had been a teacher and had written poetry. After winning a scholarship to Nottingham High School (1898), and a short time as a clerk, and as a pupil–teacher at Eastwood, at *21* he went to University College, Nottingham. In *1908* he left Eastwood for Croydon to teach at the Davidson Road School.



D.H. Lawrence.
Elliott and Fry

Lawrence had already begun writing, his first published work being a story in the *Nottinghamshire Guardian.* His first literary friend was Ford Madox Hueffer (Ford Madox Ford), who placed his first novel, *The White Peacock* (1911), and published his work in the *English Review.* Hueffer did not like *The Trespasser* (1912), Lawrence's second novel, based on a visit to the Isle of Wight with a friend he calls "Helena"; but Edward Garnett (then reader to the publishers, Duckworth and Co.) did, and Garnett became for a time Lawrence's most helpful literary friend.

Early works and recognition

An epoch in Lawrence's life that was to cease only with his death began when he met Frieda Weekley in April *1912.* She was the aristocratic German wife (*née* von Richthofen) of the professor who had taught Lawrence at Nottingham. Lawrence left with her for Germany in May *1912.* So began a restless, nomadic, gypsying life, and a difficult but intensely intimate relationship that was to form the underlying theme of much of Lawrence's later fiction. While abroad Lawrence was working on *The Rainbow.* He returned with Frieda to England in June *1913* and while there met the critic John Middleton Murry and Katherine Mansfield, the writer. After living for a time in Italy, Lawrence and Frieda returned to England and were married in July *1914.* Meanwhile *Sons and Lovers* (*1913*), his first mature novel, was not selling widely, but it made him well known. His first volume of short stories, *The Prussian Officer and Other Stories,* was published in *1914;* his first volume of poems, *Love Poems and Others,* had been published in *1913.*

The years of World War I were a dark period in Lawrence's life. Wandering restlessly about England, he dreamed of an ideal community, "Rananim," into which to escape from the horrors of the contemporary world and intermittently urged his friends to join him in it, wherever it should be; unfortunately, these friends included many talented but egotistic people who disliked each other, as their later memoirs have shown, One of them at this time was Lady Ottoline Morrell, the wife of a member of parliament, who lived at Garsington Manor, near Oxford. Flamboyant, stately, and eccentric, with a beautiful home and considerable social power, she was a devoted and kindly patron of literature and the arts. Her striking personality emerges clearly in a collection of her writings, *Ottoline (1963,* ed. by R. Gathorne-Hardy). In *1915* Lawrence earnestly hoped that she would assist him in forming a new community. He met Bertrand Russell at Garsington; they planned a series of lectures together and exchanged many letters but soon quarrelled. The "Breadalby" chapter of *Women in Love* shows Lawrence's revulsion from the Garsington milieu. Meanwhile *The Rainbow,* which had already caused a rift between Lawrence and Edward Garnett, was officially banned as obscene. Lawrence, in despair with Lady Ottoline, turned to Murry; they collaborated in a short-lived magazine, *The Signature,* but the collaboration was not wholehearted, and Murry and Katherine Mansfield disliked *The Rainbow. A* planned emigration to Florida was postponed, partly because of Lawrence's health and partly because of difficulties over Frieda's passport; so in 1916 the Lawrences went to live in a cottage at Zennor, near St. Ives, Cornwall, and the Murrys came to live nearby. But there was conflict between the two couples, which may be reflected in *Women in Love,* in which "Gudrun" suggests traits of Katherine Mansfield. After conscription came into force (May 1916), Lawrence underwent several unpleasant experiences at the hands of the military authorities. His first "medical" and other wartime ordeals are vividly described in the "Nightmare" chapter of *Kangaroo* (1923). During 1917 an atmosphere of suspicion and hostility grew around the Lawrences in Cornwall; Lawrence was distrusted because he wrote and had a beard and because his wife was German. Ordered by the police to leave in October 1917, they spent the rest of the war in London and Derbyshire.

Postwar years

After the war Lawrence resumed his friendly relationship with Murry, who accepted an article of his for *The Athenaeum*, but the rejection of a second article led to another quarrel, and the two men did not meet for some years. Lawrence then thought of going to the United States, and in order to raise money from an American publisher wrote his brilliant, if idiosyncratic, *Studies in Classic American Literature* (1923). Meanwhile, Frieda left for Germany in October 1919; Lawrence went to Italy, where he met Norman Douglas at Florence, and Douglas' friend Maurice Magnus, thus beginning a tragicomic relationship described in Lawrence's introduction to Magnus' *Memoirs of the Foreign Legion* (1924). The Lawrences later took a house near Taormina in Sicily. Lawrence's trip to Sardinia at this time led to his travel book *Sea and Sardinia* (1921). During 1920 his financial position improved; his work began to sell in the United States, and he found a London publisher (Martin Secker) who was willing to reissue *The Rainbow* and to bring out *Women in Love* (1921). One of the books he sold to Secker was *The Lost Girl* (1920). In the summer of 1920 the Lawrences went to stay with Frieda's mother in Germany, and there Lawrence wrote *Fantasia of the Unconscious* (1922), a treatise on his psychological theories; shortly afterward he completed *Aaron's Rod* (1922).

Throughout 1921 Lawrence's desire to leave Europe was growing. Wandering about Italy, Germany, and Austria, with Frieda or alone, he grew restless, and his disgust with Europe was accentuated by trouble over *Women in Love,* which had been attacked in the London gutter press as "a loathsome study of depravity." He therefore accepted an invitation from an American couple, Earl and Achsah Brewster, who were studying Buddhism in Ceylon, to take a bungalow there with Frieda, But, soon repelled by Buddhism and Ceylon, the Lawrences went on to Australia. The novel that resulted from this visit, *Kangaroo,* shows Lawrence momentarily weary of his mystic doctrines of sex and power, and the messianism of *Aaron's Rod,* and conscious of the consequences to himself and Frieda of their mode of life. In the summer of 1922 Lawrence at last decided to accept an invitation from Mabel Dodge, a rich American woman who had admired *Sea and Sardinia,* to join her at Taos, New Mexico.

In a singular work, *Lorenzo in Taos* (1932; Lawrence was often called Lorenzo by his friends), Mrs. Dodge (Luhan) has described the difficult relationships between herself, Lawrence, Frieda, and others who settled there. In March 1923 Lawrence left for Mexico, where he became fascinated by Aztec civilization; retiring to a watering place on Lago de Chapala, he began *The Plumed Serpent.* He left for New York in July, planning to return to Europe but, homesick for Mexico, decided not to go; his indecision exasperated Frieda, and she went off to Europe alone. In December 1923 Lawrence arrived in London and was met by Frieda and Murry. During their estrangement, Murry had adversely reviewed Lawrence's major novels; moreover, Lawrence had quarrelled with Katherine Mansfield. After her death (January 1923) Lawrence and Murry were reconciled, but Lawrence was not enthusiastic about Murry's new paper, *The Adelphi,* and it appears likely that he was increasingly unhappy about the friendship developing between Frieda and Murry. Soon Lawrence was sounding out his friends to find out which of them would return with him to Mexico to form the nucleus of a new society: Catherine Carswell, in *The Savage Pilgrimage*, has described the famous Café Royal dinner at which Lawrence made this appeal. In March 1924 the Lawrences, with their aristocratic friend Dorothy Brett, left London for Taos, where opposition between Mabel Dodge and Frieda was complicated by "Brett's" presence. During the winter of 1924–25 Lawrence was seriously ill, and for the rest of his life his health became increasingly bad.

Lawrence returned to Europe in September 1925. In the same year his fine short novel *St. Mawr* was published (in a volume also containing "The Princess"), and his Mexican novel, *The Plumed Serpent,* appeared in 1926. While the Lawrences were living at Spotorno, near Genoa, the final breach occurred with Murry, because (according to Murry) he would not accept Lawrence's invitation to join the Lawrence circle there, which included Lawrence's sister, Frieda's daughter and the faithful "Brett." in April 1926 the Lawrences moved to Florence; Lawrence's interest in the ancient Etruscans bore fruit in his posthumously published *Etruscan Places* (1932). During 1926 Lawrence and Aldous Huxley became close friends. Lawrence's last years were chiefly spent in painting pictures and in writing and rewriting his last novel, *Lady Chatterley's Lover,* to discharge his accumulated bitterness about society and to assert his belief in the healing possibilities for civilization of a new relationship between men and women. Published in a limited edition in Florence (1928) and in Paris (1929), this appeared in an expurgated version in 1932. The full text was only published in 1959 in New York and in 1960 in London, when it was the subject of a sensational court case *(Regina* v. *Penguin Books Limited)* that turned largely on the justification of the use in this novel of "tabooed" sex words.

Friendship with Aldous Huxley

Lawrence's health grew rapidly worse while he was working on *Lady Chatterley,* and his *Pansies* (1929), light verses thrown off during this period, shows his social mortification also. At this time he also composed the poignant story *The Man Who Died* (1931). The pirating of *Lady Chatterley* in the United States and the closing by the police of an exhibition of his paintings in London (1929) angered him, and he relieved his anger through more sarcastic verses: *Nettles* (1930) and *More Pansies,* included in *Last Poems* (1932). From June 1928, when he left Florence, until his death, he went from place to place in a desperate struggle for life. He worked until the end, completing *Apocalypse* (published in 1931), a commentary on the Book of Revelation.

Lawrence died of tuberculosis on March 2, 1930, at a sanatorium in Vence, near Antibes.

Personality and genius. The fascination of Lawrence's personality is attested to by all who knew him, and it abundantly survives in his letters, his fiction, his poetry, his numerous prose writings, and in biographical accounts. Judgments on him range from Catherine Carswell's picture in *The Savage Pilgrimage* (1932) as charming, wise, and sympathetic, to Murry's in *Son of Woman* (1931; reprinted with new introduction, 1954), as tortured, sick, and filled with hatred. The only certain conclusion is that he was a many-faceted and complex being, the understanding of whose thought, "message," and art cannot for long be separated from investigation of his own psychological and social problems. But above all he was an original genius who achieved at least a considerable realization of his power in imaginative terms.

Technique and vision as novelist

As an imaginative writer, Lawrence is a remarkable compound of the naturalistic novelist and the poet, the realist, and the mystic. But on the whole he ranks with Melville and Emily Bronte rather than with Balzac and Tolstoy; he communicates a vision, rather than creates an autonomous world of characters. His unmistakable qualities are a flamelike intensity, an elevation, an urgent movement; he transmits a feeling of inward sympathy with all he depicts, whether it belongs to the human world or to the world of animal life or of inanimate matter.

Lawrence's novels do not give fully harmonious expression to his genius. The greatest of them, and the most representative of his powers as a novelist, are *The Rainbow* and *Women in Love*—large canvases in which much of the detail is blurred or drawn hastily but which as a whole have a sweeping imaginative effect. His subject here is the relations between men and women in marriage. In all his best work the depths of sexual relationship are sounded, its reality is illuminated, its meaning is restated with audacious courage. Lawrence's characters are not heroes of romance but "ordinary people"; the farm people of *The Rainbow,* the rootless Bohemians of *Women in Love,* belong to a recognizable social world. But these novels lack realism in that they do not bring out the pressure of social forces that, even had the lovers ignored them, would have made themselves felt. Lawrence's "world" is intensely subjective; the world outside the lovers is excluded or rejected with scorn. Although in his later novels the sexual theme is connected to other aspects of reality—in *Aaron's Rod,* to the communist upheaval in Italy after World War I; in *Kangaroo,* to an (imaginary) fascist movement in Australia; in *The Plumed Serpent,* to the struggle between Catholicism and the insurrectionary peasants of Mexico; in *Lady Chatterley's Lover,* to industrial conditions in the English midlands—yet the fiery intensity of *The Rainbow,* because of Lawrence's concentration on the sexual theme, is missing from these later works.

Characters of the novels

Lawrence's male characters are not as a rule fully defined. It is often hard to distinguish between a Will Brangwen and a Skrebensky *(The Rainbow),* a Rupert Birkin and a Gerald Crich *(Women in Love),* in their states of esoteric stress and turmoil. Don Cipriano of *The Plumed Serpent* seems to think and feel at times remarkably like the mysterious gamekeeper (Mellors) of *Lady Chatterley,* or the restless novelist-hero of *Kangaroo.* Lawrence's women are more distinct personalities, but they all have a noticeable family resemblance; the dominating willpower of Lawrence's mother, the imaginative sympathy of "Miriam" *(Sons and Lovers),* the tenacious feminine resistance of Frieda, reappear in various guises. Lawrence's novels describe the continuous history of a relationship between a single pair of people; sometimes the fictional dress is transparent, as in "The Captain's Doll." But generally in Lawrence's shorter fiction, his "tales," he achieves a greater impersonality and a more manifest creative power; stories as different as "England, My England," "The Horse Dealer's Daughter," or "The Woman Who Rode Away" testify to an imaginative range not always predictable from his novels.

Criticism and books of travel

Powerful and audacious as Lawrence's work in fiction is, it does not always create in the reader a sense of irresistible reality. Lawrence's experience and his sympathies were wide, but his writing shows a certain fixity of outlook, a narrow and obdurate concentration of vision. His later work sometimes shows a strain, a didactic harshness, an arbitrariness in the making of points or the development of situations. Some of this may be due to the lung disease that sapped his later years and that may have contributed to the moods of anger, sarcasm, or misanthropy recurring so frequently after 1920. But Lawrence's books as a whole show a remarkable breadth and generosity of interest. His work includes (besides fiction) poetry, drama, criticism, books of travel, and translations. His marvelous descriptive gifts record memorably his reactions to places and people all over the world, in England, Italy, Mexico, and Australia, and he gives a new form and importance to the journal of travel. In *Studies in Classic American Literature* he practices an equally unprecedented mode of criticism, studying the strange emergence of the new American "consciousness" and the American world vision. Even his excursions into metaphysics in works such as "The Crown" (1925), or *Fantasia of the Unconscious,* are full of his characteristic poetry and lit by glints of his attractive humour. The mystery and wonder of *Twilight in Italy* (1916), the power and wit of the *Studies,* the occasional profundity of one of his more neglected works, the "Study of Thomas Hardy" (in *Phoenix,* 1936), prevent these works from being relegated to a secondary place.

Lawrence's poetry

Lawrence's poetry deserves separate mention. In his early poems, except when he is writing in the dialect of his youthful environment, his touch is often unsure, he is too "literary," he is not fully the master of conventional forms. But by a remarkable triumph of development, he evolved a mode of writing that in its freedom and spontaneity allowed him to express what he called the palpitating moment of being and the naked self. His poetry is often of great biographical interest, as in the series of poems *Look! We Have Come Through* (1917). But his most original contribution to the art of verse is *Birds, Beasts and Flowers* (1923), in which he creates an unprecedented poetry of nature, based on his experiences of the Mediterranean scene and the southwestern United States.

No account of Lawrence's work, however brief, can omit his unsurpassable letters. In their variety of tone, vivacity, and range of interest, they convey a full and splendid picture of himself, his relation to his correspondents, and the exhilarations, depressions, and prophetic broodings of his wandering life. They form perhaps the best introduction to his work as a whole.

Style. Lawrence, though a great writer, was not always the most careful or exact. Emotionalism sometimes blurred his vision, and repetitiveness marred his style. Of this last mannerism he himself was fully conscious, and it is only fair to quote his own defense of it in a special preface that he wrote (September 1919) for the American edition of *Women in Love:* ". . . fault is often found with the continual, slightly modified repetition. The only answer is that it is natural to the author; and that every natural crisis in emotion or passion or understanding comes from this pulsing, frictional to-and-fro which works up to culmination." Lawrence's language in all his writings expresses the immediate feeling of life. Sometimes he falters, trying to seize some sensation or intuition too rapid and intangible to be caught; sometimes he overinsists, repeats himself, falls into jargon. But his cultivation of the "organic" style of composition, his refusal to isolate the reason from the passions and the nerves, can lead to great triumphs of symbolic art, which make the work of more deliberate artists seem contrived and cold; the scenes among the sheaves and the dunes in *The Rainbow,* the episode in *Women in Love* where Birkin shatters the image of the moon in water, are among these triumphs. Always he seeks to express, whether in novels, poems, or essays, the wholeness of man; and whatever he evokes becomes rich with the feeling of the deepest impulses of life. In the modern world, which he found turbid and confused, he worked through his art to discover form and coherence.

MAJOR WORKS

NOVELS: *The White Peacock* (1911); *The Trespasser* (1912); *Sons and Lovers* (1913); *The Rainbow* (1915; expurgated ed., 1916); *Women in Love* (completed 1916, privately printed 1920; first Eng. ed., 1921); *The Lost Girl* (1920); *Aaron's Rod* (1922); *Kangaroo* (1923); *The Boy in the Bush* (1924), in collaboration with M.L. Skinner; *St. Mawr* (1925), also contained the story, "The Princess"; *The Plumed Serpent* (1926); *Lady Chatterley's Lover* (privately printed 1928; expurgated ed., 1932; full text, 1959); *The Escaped Cock* (1929; published as *The Man Who Died,* 1931); *The Virgin and the Gipsy* (1930).

POETRY: *Love Poems and Others* (1913); *Amores* (1916); *Look! We Have Come Through* (1917; new ed., with additions, 1958); *New Poems* (1918); *Bay: A Book of Poems* (1919); *Tortoises* (1921); *Birds, Beasts and Flowers* (1923; new ed., with prefatory notes to each section, 1930); *Collected Poems,* 2 vol. (1928); *Pansies* (1929); *Nettles* (1930); *The Triumph of the Machine* (1930); *Last Poems,* ed. by R. Aldington and G. Orioli (1932).

SHORT STORIES: *The Prussian Officer and Other Stories* (1914); *England, My England and Other Stories* (1922), including "Samson and Delilah" and "The Horse Dealer's Daughter"; *The Ladybird* (1923; published the same year in the U.S. under the title *The Captain's Doll),* including "The Fox" and "The Captain's Doll"; *Sun* (1926); *Glad Ghosts* (1926); *Rawdon's Roof* (1928); *The Woman Who Rode Away and Other Stories* (1928); *Love Among the Haystacks, and Other Pieces* (1930).

PLAYS: *The Widowing of Mrs. Holroyd* (1914); *Touch and Go* (1920); *David* (1926).

TRAVEL BOOKS: *Twilight in Italy* (1916); *Sea and Sardinia* (1921); *Mornings in Mexico* (1927).

OTHER WORKS: *Movements in European History* (1921); *Psychoanalysis and the Unconscious* (1921); *Fantasia of the Unconscious* (1922); *Studies in Classic American Literature* (1923); *Reflections on the Death of a Porcupine and Other Essays* (1925); *Pornography and Obscenity* (1929); *Assorted Articles* (1930); *Apocalypse* (1931); *Phoenix,* ed. by Edward D. McDonald (1936, 1961); *Phoenix* 11, ed. by Warren Roberts and Harry T. Moore (1968).

BIBLIOGRAPHY. EDWARD NEHLS (ed.), *D.H. Lawrence: A Composite Biography,* 3 vol (1957–59), is of unique value as a detailed record of Lawrence's life compiled from various sources. Among the most important of the personal accounts by people who knew Lawrence are JESSIE CHAMBERS (E.T.), D H. *Lawrence. A Personal Record*, 2nd ed. (1965); FRIEDA LAWRENCE, *Not I, But the Wind* (1934); CATHERINE CARSWELL, *The Savage Pilgrimage* (1932); and JOHN MIDDLETON MURRY, *Son of Woman* (1931). RICHARD ALDINGTON, *Portrait of a Genius, But . . .* (1950); and HARRY T. MOORE, *The Intelligent Heart* (1954), are comprehensive biographical studies.

Lawrence's *Letters,* ed. by A. HUXLEY (1932) and by HARRY T. MOORE, 2 vol. (1962), are an indispensable asset to the study of his life and works. as are his *Letters to Bertrand Russell,* also ed. by MOORE (1948). Of the many modern critical accounts, pride of place should still be given to F.R. LEAVIS, *D.H. Lawrence: Novelist* (1955). A different critical emphasis is to be found in GRAHAM HOUGH, *The Dark Sun* (1956); ELISEO VIVAS, *D.H. Lawrence: The Failure and the Triumph of Art* (1960); and two valuable recent studies, JULIAN MOYNARAN, *The Deed of Life* (1963); and *HM.* DALESKI, *The Forked Flame* (1965).

Lawrence's own literary criticism is widely represented in *Selected Literary Criticism,* ed. by A. BEAL (1955).

(W.W R.)

Lawrence, T.E.

Thomas Edward Lawrence, an archaeological scholar turned war hero (as "Lawrence of Arabia") in 1914–18, became a myth in his own lifetime even before he published his own version of his legend, *Seven Pillars of Wisdom,* which is one of the greatest modem epics in the English language. Although it is Lawrence as myth and enigma who continues to inspire interest in the man and his work—for he renounced the concomitants of fame to enlist in the ranks—his accomplishments themselves were solid enough for several lives. More than a military leader and inspirational force behind the Arab revolt against the Turks, he was a superb tactician and highly influential theoretician of guerrilla warfare. His *Seven Pillars of Wisdom* is a stylistically self-conscious yet masterly memoir, unique in his age, while his sharply etched service chronicle, *The Mint,* and his mannered prose translation of the *Odyssey* have added to a literary reputation further substantiated by the immense and fascinating correspondence that establishes him as one of the major letter writers of his generation.

Early life. Lawrence was not his legitimate name, and he once wryly told a friend that his *Odyssey* might well be called "Chapman's Homer," for he was born in Wales (at Tremadoc, Caernarvonshire) on August 15, 1888, to Sir Thomas Chapman and Sara Maden, the governess of Sir Thomas' daughters at Westmeath, with whom he had escaped from both marriage and Ireland. As "Mr. and Mrs." Lawrence, the couple had five sons (Thomas Edward was the second) during what was outwardly a marriage with all the benefits of clergy. In 1896 the family settled in Oxford, where T.E. (he preferred the initials to the names) attended the High School and Jesus College. Medieval military architecture was his first interest, and he pursued it in its historical settings, studying crusader castles in France and (in 1909) in Syria and Palestine, submitting a thesis on the subject that won him first-class honours in history in 1910. (It was posthumously published, as *Crusader Castles,* in 1936.) As a protégé of the Oxford archaeologist D.G. Hogarth, he acquired a demyship (travelling fellowship) from Magdalen College and joined an expedition excavating the Hittite settlement of Carchemish on the Euphrates, working there from 1911 to 1914, first under Hogarth and then under Sir Leonard Woolley, and using his free time to travel on his own and get to know the language and the people. Early in 1914 he and Woolley, and Capt. S.F. Newcombe, explored northern Sinai, on the Turkish frontier east of Suez. Supposedly a scientific expedition, and in fact sponsored by the Palestine Exploration Fund, it was more a map-making reconnaissance from Gaza to Aqaba, destined to be of almost immediate strategic value. The cover study was nevertheless of authentic scholarly significance; written by Lawrence and Woolley together, it was published as *The Wilderness of Zin* in 1915.

Archaeological studies

T.E. **Lawrence.**
By courtesy of Lowell Thomas and Harry A. Chase

Service during World War I. The month the war began, Lawrence became a civilian employee of the Map Department of the War Office in London, charged with preparing a militarily useful map of Sinai. By December 1914, he was a lieutenant in Cairo. Experts on Arab affairs—especially those who had travelled in the Turkish-held Arab lands—were rare, and he was assigned to intelligence, where he spent more than a year, mostly interviewing prisoners, drawing maps, receiving and processing data from agents behind enemy lines, and producing a handbook on the Turkish Army. Although restless and vaguely ambitious, he enjoyed the intrigue and confusion common to all headquarters staffs; in mid-1915, however, his brothers Will and Frank were killed in action in France, and T.E. was reminded cruelly of the more active front in the West. Egypt at the time was the staging area for Middle Eastern military operations of prodigious inefficiency—a sideshow of wasteful desert campaigns and a bloody, bungled frontal assault on the Dardanelles at Gallipoli. A trip to Arabia convinced Lawrence of an alternative method of undermining Germany's Turkish ally. In October 1916 he had accompanied the diplomat Sir Ronald Storrs on a mission to Arabia, where Husayn ibn ʿAlī, *amīr* of Mecca, had the previous June proclaimed a revolt against the Turks. Storrs and Lawrence consulted with Husayn's son Abdullah and Lawrence received permission to go on to consult further with another son, Fayṣal, then commanding an Arab force southwest of Medina. Back in Cairo in November, Lawrence urged his superiors to abet the efforts at rebellion with arms and gold and to make use of the dissident *shaykh*s by meshing their aspirations for independence with general military strategy. He rejoined Fayṣal's army as political and liaison officer.

Guerrilla leader

Lawrence was not the only officer to become involved in the incipient Arab rising, but from his own small corner of the Arabian Peninsula he quickly became—especially from his own accounts—its brains, its organizing force, its liaison with Cairo, and its military technician. His small but irritating second front behind the Turkish lines was a hit-and-run guerrilla operation, focussing upon the mining of bridges and supply trains and the appearance of Arab units first in one place and then another, tying down enemy forces that otherwise would have been deployed elsewhere, and keeping the Damascus-to-Medina railway largely inoperable, with potential Turkish reinforcements thus helpless to crush the uprising. In such fashion Lawrence — "Amir Dynamite" to the admiring Bedouins — committed the cynical, self-serving *shaykhs* for the moment to his king-maker's vision of an Arab nation, goaded them with examples of his own self-punishing personal valour when their spirits flagged, bribed them with promises of enemy booty and English gold sovereigns.

Capture by the Turks

Aqaba — at the northernmost tip of the Red Sea — was the first major victory, seized after a two-month march on July 6, 1917. Thenceforth, Lawrence attempted to coordinate Arab movements with the campaign of Gen. Sir Edmund Allenby, who was advancing toward Jerusalem, a tactic only partly successful. In November Lawrence was captured at Darʿā by the Turks while reconnoitring the area in Arab dress and was apparently recognized and homosexually brutalized before he was able to escape. The experience, variously reported or disguised by him afterward, left real scars as well as wounds upon his psyche from which he never recovered. The next month, nevertheless, he took part in the victory parade in Jerusalem and then returned to increasingly successful actions in which Fayṣal's forces nibbled their way north, and Lawrence rose to the rank of lieutenant colonel with the Distinguished Service Order (DSO). By the time the motley Arab army reached Damascus in October 1918, he was physically and emotionally exhausted, having forced body and spirit to the breaking point too often. He had been wounded numerous times, captured, and tortured; had endured extremities of hunger, weather, and disease; had been driven by military necessity to commit atrocities upon the enemy and even murder his own wounded to prevent the Turks from doing worse; and had witnessed in the chaos of Damascus the defeat of his aspirations for the Arabs in the very moment of their triumph, their seemingly incurable factionalism rendering them incapable of becoming a nation. (Anglo-French duplicity, made official in the Sykes-Picot Agreement, Lawrence knew, had already betrayed them in a cynical wartime division of expected spoils.) Disillusioned and distinguished, Lawrence left for home just before the Armistice and politely refused, at a royal audience on October 30, 1918, the Order of the Bath and the DSO, leaving the shocked King George V (in his words) "holding the box in my hand." He was demobilized as a lieutenant colonel on July 31, 1919, afterward calling his war's end status as a colonel "temporary" and "acting" (to expedite travel from Cairo to London).

Postwar activities. A colonel at 30, Lawrence was a private at 34. In between he lobbied vainly for Arab independence at the Paris Peace Conference in 1919 (even appearing in Arab robes) and lobbied vainly against the detachment of Syria and Lebanon from the rest of the Arab countries as a French mandate. Meanwhile he worked on his war memoir, acquiring for the purpose a research fellowship at All Souls College, Oxford, effective (for a seven-year term) in November 1919. By that time his exploits were becoming belatedly known to a wide public, for in London in August 1919 an American war correspondent, Lowell Thomas, had begun an immensely popular series of illustrated lectures, "With Allenby in Palestine and Lawrence in Arabia." The latter segment soon dominated the program, and Lawrence, curious about it, went to see it himself.

Adviser on Arab affairs

Lawrence was already on a third draft of his narrative when, in March 1921, he was wooed back to the Middle East as adviser on Arab affairs to the colonial minister, then Winston Churchill. After the Cairo political settlements, which redeemed a few of the idealistic wartime promises Lawrence had made, he rejected all offers of further positions in government; and, with the covert help of his wartime colleague, Air Marshal Sir Hugh Trenchard, enlisted under an assumed name (John Hume Ross) in the Royal Air Force on August 28, 1922. He had just finished arranging to have eight double-columned copies of the revised and rhetorically inflated 330,000-word text of ***Seven Pillars of Wisdom*** run off by the press of the ***Oxford Times*** and was emotionally drained by the drafting of his memoir. Now he was willing to give up his £1,200 Colonial Office salary for the daily two shillings ninepence of an aircraftman, not only to lose himself in the ranks but to acquire material for another book. He was successful only in the latter. The London press found him at the Farnborough base, the ***Daily Express*** breaking the story on December 27. Embarrassed, the RAF released him early the next month.

Finding reinstatement impossible, Lawrence looked around for another service and through the intervention of a War Office friend who had served in the desert war, Sir Philip Chetwode, was able to enlist on March 12, 1923, as a private in the Royal Tank Corps, this time as T.E. Shaw, a name he claimed to have chosen at random, although one of the crucial events of his postwar life was his meeting in 1922, and later friendship with, George Bernard Shaw. (In 1927 he assumed the new name legally.) Posted to Bovington Camp in Dorset, he acquired a cottage nearby, Clouds Hill, which remained his home thereafter. From Dorset he set about arranging for publication of yet another version of ***Seven Pillars,*** this one based upon editorial suggestions made by friends among whom he circulated five copies of the Oxford text—notably George Bernard Shaw, who suggested lowering the emotional pitch and blue-pencilling the possibly libellous and politically questionable passages. (It may have been the puritan in Shaw that resulted, too, in excision of the most gamy homosexual passages about the behaviour of the Arab troops when without women.) About 13 percent of the text was pruned for the famous 128-copy subscription edition of 1926, sumptuously printed and bound and illustrated by notable British artists commissioned by the author.

Major literary works

Seven Pillars of Wisdom (posthumous trade edition 1935, with subsequent editions since) remains one of the few 20th-century works in English to make epical figures out of contemporaries. Overpopulated by adjectives, often straining for effects and for "art," it is, nevertheless, an action-packed narrative, replete with incident and spectacle, filled with rich character portrayals and a tense introspection that bares the author's own complex mental and spiritual transformation. Confessedly inexact and subjective, it combines the distances of heroic epic and the closeness of autobiography.

To recover the costs of printing ***Seven Pillars,*** not even possible through the 30-guinea subscription price, Lawrence agreed to a trade edition of a 130,000-word abridgment, ***Revolt in the Desert.*** By the time it was released in March 1927, he was at a base in India, remote from the publicity both editions generated; yet the limelight sought him out. Unfounded rumours of his involvement as a spy in Central Asia and in a plot against the Soviet Union caused the RAF (to which he had been transferred in 1925 on the intervention of George Bernard Shaw and John Buchan with the prime minister, Stanley Baldwin) to return him to England in 1929. In the meantime he had completed a draft of a semifictionalized memoir of Royal Air Force recruit training, ***The Mint*** (published 1955), which in its explicitness horrified Whitehall officialdom and which in his lifetime never went beyond circulation in typescript to his friends. In it he balanced scenes of contentment with air force life with scenes of splenetic rage at the desecration of the recruit's essential inviolate humanity. He described dehumanization as an ironic good, the metaphor of minting conveying the concept of men being stamped into the uniform image necessary for discipline and interchangeability; and he aimed at communicating the feel of life at a training camp and at a working station, using short-journal-entry chapters, which ranged from the oppressive and humdrum, as well as the sentimental and lyrical, to the stark, the brutal, and the obscene. Published a generation late, its improprieties of language and its experiments in the surreal, the sadistic, and the grotesque lost their force. He had also begun, on commission from the book designer Bruce Rogers, a translation of Homer's ***Odyssey*** into English prose, a task continued at various RAF bases from Karāchi in 1928 through Plymouth in 1931. It was published in 1932 as the work of T.E. Shaw, but posthumous printings have used both his former and adopted names. The work remains an effective prose translation of Homer, blending simplicity and archaic affectation and using to advantage the know-how gained in primitive desert conditions and acquired as aircraft mechanic and boat handler to enrich the language with unhackneyed accuracy.

No additional book by Lawrence was published in his lifetime other than a minor translation he had done under his J.H. Ross pseudonym, Adrien le Corbeau's *Forest Giant* (1924), although his wartime *Arab Bulletin* dispatches to Cairo (adapted into *Seven Pillars)* appeared as *Secret Despatches from Arabia* and a miscellany as *Oriental Assembly* (edited by A.W. Lawrence), both posthumously in 1939. His first postwar writings, including a famous essay on guerrilla war and a magazine serial version of an early draft of *Seven Pillars*, have been published as *Evolution of a Revolt* (edited by S. and R. Weintraub, 1968). *Minorities* (1971) reproduced an anthology of more than 100 poems Lawrence had collected in a notebook over many years, each possessing a crucial and revealing association with something in his life.

Last years. Lawrence's last years were spent among RAF seaplanes and seagoing tenders, although officialdom refused him permission to fly. In the process, moving from bases on the English Channel to those on the North Sea and leading charismatically from the lowest ranks as Aircraftman Shaw, he worked on improved designs for high-speed seaplane-tender watercraft, testing them in rigorous trials and developing a technical manual (*The* 200 *Class Royal Air Force Seaplane Tender,* 1932) for their use. A stencilled and duplicated production, it was as unpretentious as his first book had been opulent.

Retirement and death

Discharged from the Royal Air Force on February 26, 1935, Lawrence returned to Clouds Hill to face a retirement, at 46, filled alternately with optimism about future publishing projects and a sense of emptiness. To Lady Astor, an old friend, he described himself as puttering about as if "there is something broken in the works . . . my will, I think." A motorcycling accident on May 13 solved the problem of his future. He died six days later without regaining consciousness.

Lawrence found despair as necessary as ambition. He lived on the masochistic side of asceticism, and the nonphysical part of his self-punishment involved creating within himself a sense of deep frustration to immediately follow, and cancel out, high achievement by denying to himself the recognition he had earned. At its most extreme it involved a symbolic killing of the self, a taking up of a new life and a new name. Under whatever guise, he was a many-sided genius who by his achievements denied himself the privacy he constantly sought and by the manufacture of his myth, however solidly based, created in his own person a characterization rivalling any in contemporary fiction.

BIBLIOGRAPHY. RICHARD ALDINGTON, *Lawrence of Arabia: A Biographical Enquiry,* new ed. with introduction by CHRISTOPHER SYKES (1969), a searching, often scurrilous attack on Lawrence; DAVID GARNETT (ed.), *The Essential* T.E. *Lawrence* (1951), the only anthology of Lawrence's writings, comprising in effect, a brief autobiography—includes extracts from *Leaves* in *the Wind* (a probable continuation of *The Mint,* which never went beyond a few pages) and *The* 200 *Class Royal Air Force Seaplane Tender;* ROBERT GRAVES, *Lawrence and the Arabs* (1927), a book vetted by Lawrence; PHILLIP KNIGHTLEY and COLIN SIMPSON, *The Secret Lives of Lawrence of Arabia* (1970), an expansion of their 1968 London *Sunday Times* articles including much new material but also some indefensible surmises; A.W. LAWRENCE (ed.), *Letters to T.E. Lawrence* (1962), does not include the correspondence of Mrs. Bernard Shaw, who wrote Lawrence hundreds of letters; T.E. LAWRENCE, *T.E. Lawrence to His Biographer, Robert Graves* (1938), and *T.E. Lawrence to His Biogrupher, Liddell Hart* (1938), both reprinted in one volume, 2nd ed. (1963), Lawrence's letters, notes, and interlinear emendations concerning both men's biographies, with their explanatory interpolations; *The Letters of T.E. Lawrence,* ed. by DAVID GARNETT (1938), and *The Home Letters of T.E. Lawrence and His Brothers,* ed. by A.W. LAWRENCE (1954), two collections that have assembled a substantial sample of his extant correspondence; A.W. LAWRENCE (ed.), *T.E. Lawrence, by His Friends* (1937), a mostly adulatory but biographically important collection, with an appendix listing the books in Lawrence's Clouds Hill library at the time of his death, with transcriptions of flyleaf inscriptions where they appear; B.H. LIDDELL HART, "T.E. Lawrence," in *Arabia and After* (1934), a biography written by the foremost writer on military tactics of his time; SULEIMAN MOUSA, *T.E. Lawrence: An Arab View* (1966; orig. pub. in Arabic, 1962), denigrates Lawrence's veracity and impact on the Arab revolt, and considers him a less-than-covert Zionist; ANTHONY NUTTING, *Lawrence of Arabia: The Man and the Motive* (1961), builds a portrait of Lawrence as masochist; TERENCE RATTIGAN, *Ross: A Dramatic Portrait* (1960), a play with acute psychological insight into Lawrence, but not dependable as history or biography; VYVYAN RICHARDS, *T.E. Lawrence* (1939), an admiring biography by a wartime friend; LOWELL THOMAS, *With Lawrence in Arabia,* new ed. (1967)—Thomas was a war correspondent in Arabia when he first met the subject of his famous lecture; JEAN BERAUD-VILLARS, *Le Colonel Lawrence; ou, La recherche de l'absolu* (1955; Eng. trans., *T.E. Lawrence: or, The Search for the Absolute,* 1958), one of the best studies of Lawrence—sound scholarship where it does not exaggerate Lawrence's Francophobia; STANLEY WEINTRAUB, *Private Shaw and Public Shaw: A Dual Portrait* of *Lawrence of Arabia und G.B.S.* (1963), Lawrence as seen from the perspective of his crucial postwar friendship with both Shaws—Bernard and Charlotte.

Bibliography: ELIZABETH W. DUVAL, *T.E. Lawrence: A Bibliography* (1938), out of date but still useful; J.M. WILSON, *A Bibliography of T.E. Lawrence* (1972), not a full-scale descriptive bibliography, but a checklist, with notes, of publications and textual variants, as well as a list of primary biographical sources.

(S.We.)

Lead Products and Production

Lead, a heavy, malleable, durable metal with a low melting point, has found a wide range of applications throughout history and in terms of world tonnage today remains one of the most important metals produced.

History. The Egyptians may have used lead as early as 5000 BC. Applications in the ancient world included use as weight standards, in coinage, for ornamental objects, for refining silver, as solder, and as a glaze for pottery. The Egyptians used red lead, an oxide of lead, as a cosmetic pigment. A lead figure dating from 3800 BC was found at the temple of Osiris, on the site of ancient Abydos, and is now in the British Museum. About 2000 BC the Phoenicians were mining lead deposits in Spain, and it was at about the same period that the Chinese first employed lead in their monetary system. Later, when silver coins replaced lead, the mining of lead was prohibited to prevent counterfeiting, but lead became and remained the principal metal used to debase silver for counterfeiting purposes into modern times.

Lead was well known in biblical times and is used metaphorically in Exodus in noting that the pharaoh's host "sank as lead" in the Red Sea. The Hanging Gardens of Babylon, built about 600 BC, are said to have had floors sheathed with sheets of lead soldered together to hold the soil and moisture.

Lead in Roman times

The lead mines at Laurium, Greece, flourished in the 5th century BC and in the 3rd century BC the Romans operated mines in the Rio Tinto region of Spain. The Romans used lead pipe extensively, and fabricated it in 15 standard-sized diameters, all ten feet (3 metres) long. Many sections of pipe in almost perfect condition have been unearthed at Pompeii, Rome, and Roman sites in England. Indeed, the Latin word for lead was *plumbum,* denoting water conductors or spouts; from it is derived the word plumber. The Romans also made wide use of lead in soldering.

Lead applications multiplied in the Middle Ages; lead "cames" held the fragments of stained-glass windows together, and lead gutters, drainpipes, and interior pipes were fashioned for churches, monasteries, castles, and houses.

Almost since the beginning of warfare, lead has been important in missiles because of its low melting point (which made it easy to fabricate) and its high density. Lead chunks were hurled with slings, and molten lead was poured on attackers from fortress walls. Lead bullets and shot have been used since the inception of firearms.

The medieval alchemist, who preceded today's chemist and metallurgist, considered lead the father of all metals, and its low melting point was a factor that led him to try to convert it to a substance that would transform other base metals into gold.

In the United States, now the world's major lead-producing country, lead mining and smelting began in Virginia in 1621. Lead ores from the Mississippi Valley were used from 1700 to 1867, when the first of the great western mines was opened.

Lead ores: galena. Galena (lead sulfide), by far the most important primary lead ore mineral, frequently contains as impurities zinc, copper, cadmium, bismuth, arsenic, antimony, and silver; often the silver value is greater than that of the lead, in which case it is a true silver ore. A primary mineral is formed by natural processes and unaffected by subsequent events, while a secondary mineral is affected in some way; for example, by weathering. Other primary lead minerals are clausthalite, altaite, jordanite, kobellite, boulangerite, aikinite, jamesonite, and bournonite. Carbonates and sulfates, formed by weathering, are the chief secondary minerals. The most common, cerussite (lead carbonate), is formed by the action of carbonate groundwaters on galena. Anglesite (lead sulfate) is derived from galena that has been contacted by sulfate solutions generated from the oxidation of sulfide minerals. Other oxide minerals of lesser importance are wulfenite, vanadinite, and pyromorphite. Lead is rarely found in its elemental form.

Over 95 percent of mined lead is from primary ores in which galena is the preponderant lead mineral. Galena is usually intermixed with other sulfides, chiefly sphalerite (zinc sulfide) and the sulfides of iron: pyrite and marcasite, less commonly with chalcopyrite (copper and iron sulfides). Lead ores constitute the chief source of silver, usually in solid solution with galena.

PRODUCTION PROCESSES

Mining. The widely varying characteristics of lead ores have given rise to many mining techniques, of which underground methods predominate. Most lead deposits occur at moderate depths in narrow channels called veins or in deposits of roughly spherical shape (lodes). Both veins and lodes are of varying thicknesses. Holes are drilled for blasting, often by large mounted diamond drills or carbide percussion drills. After blasting, the broken ore is sorted and loaded onto cars or continuous belts for transportation to the surface. Most modern lead mines are highly mechanized.

Ore treatment. Although galena contains 86.6 percent lead, it usually occurs dispersed with other minerals, and in some places, ores with as little as 3 percent to 30 percent lead are mined.

Preparation for roasting. Lead ore is prepared for roasting by grinding, flotation, and concentration. First, the ore is ground and classified by particle size: approximately 70 percent is less than 0.0017 inch in diameter, and 90 percent less than 0.0029 inch; some particles are less than 0.0003 inch. The ground and classified ore, which is a mixture of lead sulfide and worthless rock called gangue, passes to large tanks where flotation takes place.

Flotation separation process

Flotation, which has supplanted earlier methods of mechanical and gravity separation, takes advantage of the fact that water with a suitable wetting agent added to it, wets gangue but not lead sulfide. In the flotation tank, the ground and classified ore is mixed with water and some flotation agent such as oil. The lead sulfide attaches itself to the oil, which is turned into froth by bubbling air into the mixture, and then is floated off the top. The wet gangue sinks to the bottom of the tank and is discharged.

Concentration involves aggregating the tiny particles of lead sulfide; this is done with the aid of an agent such as lime or alum. Filtered to reduce its water content to about 10 percent, the concentrate then passes to the smelter. Concentrates include 45 to 60 percent lead, along with varying amounts of impurities: 10 to 30 percent sulfur, up to 15 percent zinc, 8 percent iron, 3 percent calcium oxide, 3 percent copper, 2 percent antimony, and small amounts of arsenic, bismuth, gold, silver, and other elements.

Roasting. Before lead concentrates can be added to the furnaces, they must be treated to lower the sulfur content because the ordinary furnace does not reach a high enough temperature to reduce lead sulfide. Substances such as limestone, iron ore, silica, coke, etc., are mixed with ore concentrate; these fuse with the sulfur in the lead sulfide to form compounds. The ore mix is laid on a moving grate; air is blown through the mass; sulfur acts as the main fuel, forming sulfur dioxide, which is usually recovered for sulfuric acid manufacture. The mix is fused into a strong, homogeneous material called sinter, comprised primarily of oxides of lead, iron, and zinc, with substantial amounts of metallic lead and a little sulfur. The porous sinter mass is broken into lumps as it is discharged from the moving grate. Sinter clinkers up to five inches (125 millimetres) in size are passed over grates to remove the fine particles, which are stored for blending with other materials prior to reduction in the blast furnace. Reduction is the removal of oxygen from a compound; for example, reduction of lead oxide yields metallic lead.

Smelting. Smelting is the continuous reduction of ore constituents to metal, generally carried out in a blast furnace, which consists mainly of a water-jacketed shaft or column up to 24 feet (7.2 metres) high with a rectangular cross section and is loaded from the top. The lead or charge consists of properly proportioned sinter, limestone, coke, silica, and other fluxes. Orifices called tuyeres, through which air is blasted, are positioned just above a crucible at the bottom of the shaft. The crucible collects molten lead and is arranged with a suitable siphon well so that the molten lead can be removed from the furnace without affecting the blast pressure within the furnace. Ports to discharge slag and metallic waste (matte) are located near the top of the crucible. Sight glasses that allow for visual inspection of the smelting reaction are located in the tuyeres. Air is injected to aid combustion and to complete the formation of metallic oxides, which are then chemically reduced to metal by the coke and carbon monoxide present. Nonmetallic waste forms a slag with the fluxing materials. A balanced smelting charge of iron, limestone, and silica yields a slag that is fluid. When the coke and sinter ratio is not correct, there is an improper reduction of ore to metal.

Blast-furnace operation

Reverberatory furnace smelting, in which a furnace charge is held in a basin-like vessel and heated by flames on the surface to effect the reduction, is rarely practiced in lead refining as it requires that the ore contain 70 percent or more lead.

The output of the blast furnace is lead bullion, which contains dissolved metallic and nonmetallic (oxide and sulfide) impurities. Slag, matte, and other products of the blast furnace are either further treated to recover metal values or scrapped.

Drossing. Lead bullion is maintained in a molten state just above its melting point in a drossing kettle, which contains up to a hundred tons. (The term dross refers to any solid scum floating on top of a metal bath comprised of metal oxides, sulfides, metallic compounds, etc.) At temperatures of about 330" C, remaining impurities are brought to the surface by stirring, by the addition of small amounts of sulfur, and by agitation with submerged air lances and are removed by skimming.

Refining. The bullion is transferred to the refinery either as a pumped liquid or as large cast blocks, weighing as much as ten tons.

Electrolytic refining. Electrolytic refining involves purifying the bullion by means of electricity. If electrolytic refining is to be practiced, the bullion is cast into anodes (or plates) weighing several hundred pounds. These are hung in electrolytic tanks where they dissolve; pure lead is deposited on a thin sheet of lead that serves as the cathode. Impurities left behind can be recovered by many complex operations.

Kettle ***refining.*** Kettle refining of bullion is complex, but the recovery of by-product silver and gold is simpler and faster. All of a number of kettle-refining processes or modifications have the following steps: softening, desilverizing, dezincing, and debismuthizing. Softening refers to removal of impurities that harden lead, such as antimony, arsenic, tin, zinc, and others. Large-capacity steel pots or reverberatory furnaces are used to hold the bullion at red heat. Blowing with air or applying suitable oxidizing reagents oxidizes impurities. The oxidation is selective and oxide drosses or slags in which each element is concentrated can be removed by skimming.

Oxidation of impurities

Silver is removed from softened lead by adding zinc and raising the temperature sufficiently to melt it (about 540" C). With stirring, the molten zinc reacts, forming compounds with gold and silver, plus traces of copper and nickel. 'These zinc compounds are lighter than the metallic lead; on cooling, they form a crust that is removed. The gold and silver are recovered in a parting plant. The lead is now treated for removal of the remaining zinc by a vacuum process. The zinc recovered from the vacuum retort. along with that from the parting plant, is reused.

If the bismuth content is too high, the softened lead is next debismuthized. The operation is similar to desilverizing, except that a combination of calcium and magnesium metals are used as the reagents to form high-melting-point intermetallic compounds with bismuth, which are skimmed off and treated for recovery of bismuth.

The refining processes yield pig lead of good marketable uniformity and purity. Large quantities of pig lead are available, 99.999 percent pure.

MAJOR USES

Lead is widely used both in the metallic form and in chemical compounds.

Storage batteries. The lead-acid storage battery, the largest single use for lead, remained in the early 1970s the most dependable apparatus by which electrical energy can be conveniently stored for future use. Economy, reliability, and the ability to deliver large amounts of energy in a short time have made it an essential part of every automobile. Storage-battery grids require a metal with a low melting point, high corrosion resistance, and certain electrical properties. The modern automotive SLI (starting, lighting, ignition) battery consists of about one-third lead metal (grids, posts, and connectors) and two-thirds litharge (lead monoxide). The lead metal, 6 to 12 percent antimonial lead plus a small amount of tin, is almost wholly recovered by smelting secondary materials, mostly scrap storage batteries. Litharge mixed with sulfuric acid to a paste consistency for incorporation into the metal grid is usually made from primary (mined) lead because impurities such as silver and bismuth cannot be removed economically during refining of secondary (salvaged) lead.

Storage batteries are also used as a standby source of power for telephone exchanges, power companies, hospitals, submarines, etc. The thicker grids used for this type of service do not require the casting properties of the SLI batteries. In order to minimize the self-discharge caused by antimony-containing grids, pure lead or calcium-lead alloy is used. The calcium content of 0.06 to 0.08 percent in lead yields an alloy of essentially the same physical properties (primarily improved mechanical strength) as 4 percent antimonial lead, necessary characteristics in battery production and use. Large industrial batteries supplying energy for electric locomotives, industrial trucks, and some experimental electric automobiles usually have antimonial lead grids.

Chemical lead. Lead is used extensively in the construction of chemical-plant equipment. If unalloyed lead is used it is generally of the grade commonly known as chemical lead, sometimes termed acid lead, or copper lead. The term chemical lead denotes a metal of very high purity: standards vary slightly from country to country. Chemical lead forms a tightly adherent, insoluble lead sulfate film in sulfuric acid solutions. As sheet, pipe, rod, etc., it is readily welded (burned) into complex structures for holding and conveying sulfuric acid mixtures. Other materials, such as plastics, rubber, etc., have been used for similar operations, but lead's demonstrated corrosion resistance over a wide range of conditions, its ease of repair, and, after a long useful life, its high salvage value have enabled it to retain its position as a material of construction in the chemical industry. A 6 percent antimony-chemical lead alloy is used where greater strength or fatigue resistance is required. The use of 6 percent alloy is restricted to temperatures below 125° C because of its lower melting point. Similar alloys are used with phosphoric acid, hydrofluoric acid, chromic acid, and other corrosive environments where insoluble protective films are formed. Each different smelter uses a proprietary alloy as insoluble anode in chromium plating, and electrowinning (recovery by electrolysis) of copper. To achieve a higher order of strength, lead-lined and lead-covered steel pipe, fittings, valves, and vessels are used with lead metallurgically bonded to the substrate to maintain suitable heat transfer.

Other uses. The electric cable in industry is a major user of lead. Lead alloys, usually 1 percent antimonial lead, can be easily extruded, forming an impervious corrosion-resistant cover on telephone and power cables. When the cable is replaced, the sheathing can be recovered by a simple stripping operation, refined, and used to cover new cable.

Lead has a longer record of satisfactory performance than any other water-service pipe material. The flexibility of lead pipe allows adjustment to ground settlement without damage, and minimizes the number of joints. Lead is also used for caulking joints in cast-iron water mains and sewer and soil pipes. Because it withstands the effects of polluted air and water, lead is useful in drainage and venting systems.

Sheet-lead roofing and flashing have long been used on buildings because of their durability under all types of climatic conditions. The development of special alloys capable of resisting buckling due to alternating high and low temperatures (thermal cycling) has caused a renewed interest in the use of lead in modem architecture. Special joint designs have also helped in the construction of essentially maintenance-free roofs, reflecting pool liners, etc., and soft lead cames are easily worked to form leaded art-glass windows.

Architectural applications

Ammunition, both military and sporting, accounts for a large use of lead. Spherical shot for shotgun use is made by pouring molten 2 to 6 percent antimonial lead, with up to 1 percent arsenic added, through suitable sieves at the top of a 125-foot tower. The molten alloy, while dropping, forms a true sphere before solidifying near the bottom of its fall. The shot is collected in water, removed, dried, and sorted for size and sphericity. Military ammunition is formed into bullets by swaging (cold-shaping) extruded wire in suitable dies.

The densest of the common metals, lead is widely used for radiation shielding. It prevents the penetration of gamma rays and neutrons, and since it does not itself become radioactive, it may be used continuously. High density glasses, containing as much as 80 percent lead oxide, are used as windows in radiation installations.

Due to its limpness and high density, lead is finding expanding use in the control of noise pollution. Thin-cast sheet lead is used for encasing noisy machinery, or in room partitions for keeping noise out. Jet airliners use a lining of an adhesive-backed layer of vinyl plastic compounded with lead powder to keep the noise level in the passenger compartment to a minimum.

LEAD ALLOYS AND COMPOUNDS

Lead alloys. The mechanical properties of pure lead are quite poor; therefore, when more strength is required, lead is alloyed with other elements.

Antimony. The most common metal used is antimony, which on additions up to 12 percent by weight not only increases the tensile strength of the alloy as much as four times but also lowers the melting point. As noted earlier, antimonial lead is used as a storage-battery grid metal.

Tin. Tin alloyed with lead in amounts from 2 to 70 percent by weight forms the solders widely used in industry. Not only is the strength of the alloy increased with higher tin content but the melting point decreases from 327° C for lead to 183° C for 63 percent tin-37 percent lead. Solders of 2 percent tin-lead are used for automatic soldering of tin-coated steel cans. Industrial solders, used for such applications as automotive radiators, contain from 5 to 30 percent tin. The electronic solders are generally of the lowest melting point (63 percent tin), so that the temperature-sensitive components will not be damaged.

Industrial soldering alloys

Casting alloys. Lead alloys containing antimony and tin form a series of low melting metals with excellent casting properties and the ability to produce intricate castings with excellent detail. Linotype metal with 4 percent tin and 12 percent antimony, used for casting the finest printer's type, exhibits suitable strength for use on a press. Electrotype and stereotype metals used by printers also have excellent casting properties. Ornamental castings, which are usually plated with silver or other metals, are made with 11 percent antimonial lead, with about 0.5 percent arsenic added to increase the strength and to yield better detail and a smooth-surface finish. Lead alloyed with varying combinations and amounts of bismuth, tin, cadmium, antimony, indium, or gallium yields a series of metals that have very low melting points, culminating in one so low that it can be melted in the palm of the hand. Such alloys, rated as to melting point, are widely used for equipment of very high reliability.

Bearing metals. Lead-base bearing metals are usually antimonial lead with additions of tin or copper or both and arsenic to increase the compressive strength. The ability of these alloys to maintain a thin film of lubricating oil at relatively high loads and speeds prevents wear on the bearing-metal surface. The alloys are sufficiently soft so that even in case of lubrication failure no damage is done to the rotating shaft. The use of thin layers of bearing metal, with a stronger backing material reinforcing the lead alloy, increases the load-carrying capacity of such bearings.

Lead **compounds.** At least a third of all lead produced is used in chemical compounds. Most of the uses of lead discussed thus far are not consuming uses, in that the metal eventually returns as scrap for secondary refining and reuse. The chemical compounds, which generally use corroding lead (lead refined to a high degree of purity) as a base, are generally dissipated into the atmosphere or elsewhere so that recovery is impossible.

Gasoline additives. A major use of lead compounds in the 1960s was in gasoline additives, such as tetraethyl lead to form the antiknock, high-octane gasoline required by high-compression internal combustion engines. This use was declining in the 1970s owing to air-pollution problems.

Paint pigments. The earliest known and one of the most widely used lead paint pigments is white lead (basic lead carbonate). A process for making white lead starts with metallic lead blown with steam or air to produce a fine powder, which is then treated with acetic acid, air, water, and carbon dioxide gas in long, revolving wooden cylinders. The process requires about 12 to 14 days for conversion. Recent modifications of the process use litharge as the starting reagent to shorten the production time. The outstanding characteristic of white lead is its chemical affinity for paint vehicles, becoming an integral part of the paint film. This reinforcement results in a tough, flexible film, unusually resistant to embrittlement.

Basic lead sulfate, also called basic sulfate white lead, is a pigment used in mixed paints with the same properties as basic carbonate white lead.

Red lead (Pb_3O_4) is manufactured in furnaces by heating litharge at a carefully controlled low temperature. Litharge takes up more oxygen to form red lead. **A** part of the litharge usually remains unconverted in the core of red lead particles. **A** 98 percent true red lead is used for paint.

Litharge. Litharge (lead monoxide) is one of the most important and widely used metallic compounds. It is produced by melting pig lead in the presence of air, or occasionally by burning lead to a fume in the presence of oxygen. It is possible to produce a specific type of litharge for almost any use, and the various commercial grades are generally known by the uses to which they are put. Storage batteries account for the largest consumption of litharge. Battery litharge is very pure and may be blended with some red lead and other additives to form the active material of the plates. Glassmaker's litharge is a high-purity flake or powder, which must be free of iron or other impurities that would discolour the glass. Similarly, pottery maker's and enameller's litharge must be pure. Porcelain-enamelled aluminum sheets have litharge as a major constituent in the coating. Colourmaker's litharge is reacted with acetic acid to form lead acetate. Various compounds added to the lead acetate precipitate a number of pigments, such as chrome yellow, chrome red, and chrome green.

Other compounds. Lead arsenate, an insecticide, can be prepared by dissolving litharge in arsenic acid, with acetic or nitric acid as the catalyst, or litharge may be converted to lead acetate or nitrate and allowed to react with sodium arsenate.

Lead azide is an initial detonating agent for high explosives, such as TNT. Crystalline lead azide is prepared by the slow addition of a dilute solution of sodium azide to a constantly stirred dilute solution of lead acetate or nitrate. The precipitated azide is washed with water and stored under water or a 50–50 mixture of water and ethyl or methyl alcohol. By varying procedures it can be prepared as a colloid or treated with starchy compounds called dextrins to achieve certain properties.

Lead dioxide (PbO_2) is formed electrolytically by anodic oxidation of metallic lead. In neutral or alkaline solutions the alpha crystalline form of lead dioxide is formed. In acid solutions the beta crystalline form results. Both forms have been identified in the positive plates of storage batteries.

Unless stabilizers are added, the widely used vinyl plastics (polyvinyl chloride) suffer a breakdown due to heat and light, rendering the plastic brittle and discoloured. Stabilizers take up hydrogen chloride released in the degradation. Theoretically, litharge would be a very efficient acid acceptor, but the nature of vinyl stabilization goes beyond mere acid acceptance, and various other chemical and physical actions must be incorporated into a desirable stabilizer or combination. A series of reaction products of litharge with various organic and inorganic acids are widely used as vinyl stabilizers.

Lead compounds as vinyl stabilizers

Lead oxide was used as a major constituent of glass by the Chinese as early as the 6th century. The glass was strongly discoloured by impurities such as copper. The use of lead oxide in glass to achieve clarity was unknown until the invention by George Ravenscroft of English crystal, or flint glass, in 1675. Lead oxide yields a glass with a high index of refraction, which accounts for the brilliance of cut lead crystal and for its use in optical lenses. It also lends weight and resonance to the glass.

Economic importance. On the basis of tonnage of metal refined, lead is the fifth most important metal, after iron, aluminum, copper, and zinc.

World consumption of primary lead for the mid-1970s exceeded 3,000,000 tons annually, of which about 40 percent was consumed by the U.S., a third by the Soviet Union, and the remainder by the rest of the world.

World Production and Consumption of Lead, Representative Years
(000 tons)

	world lead mine production		world refined lead production		world refined lead consumption	
	1964	**1975**	**1964**	**1975**	**1964**	**1975**
America	**936**	**1,379**	**1,331**	**1,572**	**1,274**	**1,417**
Europe	**376**	**422**	**996**	**1,173**	**1.308**	**1,267**
Africa	**237**	**152**	**100**	**109**	**39**	**76**
Asia	**107**	**143**	**151**	**221**	**264**	**316**
Australia	**412**	**395**	**244**	**194**	**70**	**84**
Total	**2,068**	**2,490***	**2,822**	**3,268***	**2,955**	**3,161***

*Detail does not add to total because of rounding.
Source: International Lead and Zinc Study Group, United Nations.

BIBLIOGRAPHY. LEAD INDUSTRIES ASSOCIATION, *Lead in Modern Industry* (1952), a comprehensive text treating the history, mining, production, and uses of lead; J.L. BRAY, *Non-Ferrous Production Metallurgy,* 3rd ed. (1947), a text designed for both students and engineers, covering the processes and materials common to all metallurgical operations; D.M. LIDDELL, *Handbook of Non-Ferrous Metallurgy,* 2nd ed. (1945), a description of specific operations for nonferrous metal from mine to market; E.R. THEWS, *Metallurgy of White Metal Scrap and Residues* (1930), a work on the chemistry and metallurgy of the treatment of scrap metals, excluding iron and steel; UNITED NATIONS, *Report of International Lead-Zinc Study Group* (annual), an international statistical report on production and use; UNITED STATES DEPARTMENT OF INTERIOR, BUREAU OF MINES, *Minerals Yearbook* (annual), a statistical report on U.S. production and use.

(E.J.M.)

Learning, Animal

Ways of defining learning and of interpreting observed behaviour depend, in a fundamental way, on the view taken of the basic nature of the learner. Some of the difficulties in definition are considered in the article LEARNING THEORIES.

All living organisms adjust or adapt to changes in the stimulation they receive from their environment. In other words, they receive and store information; this is in large measure equivalent to saying that they learn. In a very limited and special sense it can be said that plants learn (insofar as they receive and store information), and efforts to distinguish between animals and plants on this basis alone are not satisfactory. Indeed, certain groups of organisms—*e.g.*, slime molds and Euglena—can, at one stage of their life cycle, be regarded as plants and at another as animals. From the point of view of learning, however, the most important difference is that plants on the whole are fixed or sedentary, while animals generally are mobile and active.

Animals are in fact essentially predatory-behaviour machines that must go out and find and catch their food; or at least they must be equipped with special devices, as in so many marine organisms, for filtering food out of water, sweeping it toward themselves, and ingesting it in some form or other. This emphasis on the importance of movement in animals is absolutely fundamental. Once an animal starts to evolve movement, as distinct from being passively carried along by currents in the lake or sea or by wind, it may evolve in two directions: (1) it can either move in an entirely random manner, in which case the movement itself must bring it into contact with all that it needs (including, of course, food, oxygen, and water) and ensure normal development and survival of a sufficient proportion of the species; or (2) if such random movement is not sufficient, then there must be directed movement, which can be of many different kinds (see LOCOMOTION, ANIMAL). The first essential of directed movement is that there must be special devices for detecting objects in the environment that can give guidance, either direct or indirect. The more precise and elaborate these movements become, the greater will be the evolutionary pressure to develop more and more precise sense organs. If an animal is to move in a particular direction it will be able to perform much better by being elongate, perhaps to some extent "streamlined." This, of course, implies that the most acute sense organs, as well as the point at which food is absorbed, will tend to be at the front end—*i.e.*, the animal should have some kind of "head." Even below this level of evolution, however, animals use information collected about past events as a guide to the future, evidence of having learned some of the predictable regularities in their environments.

Much of the potential for animal behaviour seems so strongly ingrained and so very little influenced by past experience that it may be called inborn, unlearned, or instinctive; that is to say, many animals are innately programmed directive machines (see INSTINCT). It is when organisms can perceive and change their behaviour as a result of their perceptions that they are said to learn. Learning, then, is the organization of behaviour as a result of individual experience. This definition indicates the minimum qualities a phenomenon must have in order to be classed as an example of learning and avoids the confusion that is bound to result if an attempt is made to define the manner in which the phenomenon is brought about. Although the views as to the ways in which learning is mediated in an organism continuously change with the acquisition of new knowledge about structure and physiology, the basic idea of learning remains the same. That is to say, learning must first be defined behaviorally, after which it can be considered physiologically, when possible.

Definition of learning

It has often been suggested that learning in general is adaptive—it manifests itself by adaptive changes in individual behaviour as a result of experience. It is true that most learning usually is adaptive in the sense that it is a mode of adjustment that enables an animal to survive. This in fact has been included in the definition, because the word organization itself implies adaptiveness in some degree. Learning is also often regarded as something that is relatively long lasting, something that has a more or less enduring effect on behaviour. Again, this fits in with the general definition of learning, but the length of its effect may vary from changes of exceedingly brief duration to changes that last a lifetime.

The following articles are related to the subject of learning in animals, including man: BEHAVIOUR, ANIMAL; INSTINCT; LEARNING, PERCEPTUAL; LEARNING, PSYCHOMOTOR; LEARNING THEORIES; MEMORY: RETENTION AND FORGETTING. This article is divided into the following major sections:

- I. Types of learning
  - Habituation
  - Associative learning
- II. Behavioral and physiological aspects of learning
  - Stimulus–response and cognitive mechanisms
  - Neurophysiological mechanisms
- III. Evolution of learning ability
  - Learning in invertebrates
  - Learning in vertebrates

I. Types of learning

HABITUATION

It is almost impossible to divide the various types of learning into categories that are mutually exclusive, that can be sharply defined, and that cover the entire range of animal learning. It is convenient, however, to start with habituation, because this phenomenon appears to be the simplest form of learning. It can be defined as the relatively persistent waning of a response as a result of repeated or continuous stimulation that is not associated with any kind of reward or reinforcement. Habituation is further characterized as being specific for a particular stimulus, and its relative permanence also distinguishes it from such temporary manifestations as fatigue and sensory adaptation (*e.g.*, becoming oblivious to distracting sounds). Habituation thus implies a tendency of an animal merely to drop those responses that experience indicates are of no value in its life. As the simplest form of learning, habituation is quite effective in lower animals because organisms do not need more than one type of sensory mechanism to become habituated to a particular kind of stimulus. This rules out, of course, any learning by these organisms in which two or more different kinds of stimuli, received by two or more different kinds of sensory mechanisms, must be coordinated to produce either a new response or an old response in a new context.

Simplest form of learning

As a form of learning, habituation is known to occur in response to all types of stimuli and in animals of all grades of organization, including the most primitive one-celled organisms (protozoans). A classic example of habituation is provided by the snail Helix albolabris. When the surface on which the snail is crawling is subjected to a mechanical shock, the snail immediately withdraws its tentacles for a short time. If the mechanical shock is repeated at regular intervals and if the intensity of the stimulus is maintained at a constant level, the extent and duration of tentacle withdrawal steadily decrease until the animal has become completely indifferent (habituated) to the stimulus. If, during the course of the experiment, the intensity of the stimulus is increased, the response reappears for a short time, only to wane again on repetition. In due course the animal becomes so accustomed to the vibrations that a series of 50 to 60 of them will not produce any visible response; indeed, it may be difficult to cause complete tentacle withdrawal even with a very violent jolt.

The investigation with the snail provides ample evidence that the effects produced by habituation are not ordinary physiological fatigue. This is because a more intense stimulus tends to restore the response to the original stimulus, and "fatigue" that can be overcome by a more intense stimulus of the same kind is a contradiction in terms. If, on the other hand, the vibrations are interrupted by 30-second intervals during which the animal receives no stimulation, the response resumes after the jolting begins again. If the process is continued, however, a half-minute rest no longer suffices to restore the response, indicating that there are different degrees of habituation that could doubtless be graded according to the length of rest time necessary to reactivate response.

In contrast to the simple habituation just described, the waning of responses in higher animals incorporates a number of complicating features, even though the phenomena appear behaviorally identical to those in lower animals. Rats in which the spinal cord has been severed from the brain, for example, will continue to exhibit behavioral changes that are clearly habituatory in nature and that last for days. Moreover, although the orientation reflex in dogs (the innate pattern of reactions that occur in response to a stimulus) can be distinguished or changed by habituation, the process seems to indicate that some sort of complex reorganization in the nervous system is responsible for or superimposed upon the decrease in strength of the original response. It would appear, therefore, that the mechanisms underlying habituation become more elaborate as organisms become more complex.

ASSOCIATIVE LEARNING

Classical conditioning. Although the philosophical concept of associationism (the connection between a stimulus and response that had not existed before) has had a long and significant history traceable back to ancient times, and although the concept was applied to experimental studies of human learning in the 1880s, modern psychological study of associative learning in animals did not begin until just before the end of the 19th century. At that time a group of Russian physiologists became the first to offer some sort of physiological explanation of how a previously fixed and perhaps inherited behaviour pattern might be modified and adapted by becoming associated with new stimuli through learning.

Pavlov's *experiments*. During his studies of the physiology of digestion, Ivan Pavlov, who became by far the most influential of the Russian investigators at the end of the 19th century, stumbled on the fact that the quantity and timing of the secretion of salivary juices in the dog are controlled by the brain under the influence of sensory stimuli that shortly precede the release of the saliva. In an attempt to show the quantitative relationship between the stimuli and the subsequent flow of saliva, Pavlov devised an experiment in which he trained a dog to stand perfectly Still in a soundproof chamber. A special harness worn by the animal made it possible for Pavlov to measure the salivary flow when food, in the form of meat powder, was mechanically offered to the dog or placed directly on its tongue.

Conditioned reflex or classical conditioning

It was his study of salivation in dogs that led Pavlov to his theory of the conditioned reflex, or classical conditioning, as it is now usually called. Pavlov at first thought the mechanism that produced the flow of saliva as a result of the taste of meat was an inherited or fixed reflex (unconditioned) response. (A reflex is a simple involuntary response to a stimulus, such as the quick withdrawal of the hand when it touches a hot object.) He then found, however, that he was able to influence (condition) that response by exposing the dog, immediately before the meat powder (unconditioned stimulus) was offered, to such cues (conditioned stimuli) as a bell, a buzzer, or a flashing light. Thus, by his experimental method, Pavlov was deliberately eliminating, as far as possible, all general behaviour and behavioral reactions of the dog so as to isolate this one response, the conditioned reflex. Essentially, then, classical conditioning denotes the process by which an animal acquires the capacity to respond (*e.g.*, to salivate) to a given stimulus (bell, buzzer, etc.) with the same reflex action that can be elicited by another stimulus (the reward or reinforcement) when the two stimuli are applied concurrently (or in overlapping sequence) for a number of times.

Anticipatory behaviour

Limitations of classical *conditioning*. Pavlov's work led to an enormous spate of research in the physiology of behaviour, and for a long time the conditioned reflex provided the best technique for enabling at least a part of a learning process to be investigated quantitatively and to be subjected to an exact analysis. As work advanced over the years, however, it became increasingly evident that the classical-conditioning experimental technique artificially isolates, to an extreme degree, part of the total learning process of the higher animals—particularly that of mammals, such as rats and man. Moreover, the reflex relationship between stimulus and response is not a simple one. Indeed, because higher animals always respond in some degree to the conditioning situation as a whole, not just to the stimulus alone, it became apparent that the conditioned response in such animals is always in fact "anticipatory." The animals behave as if they were "expecting" the unconditioned stimulus rather than just associating the unconditioned response with the new stimulus. In other words, a dog does not salivate in response to a bell or other stimulus simply because it has been trained to do so (the conditioned response) but because it is anticipating the food that it will receive (the unconditioned stimulus). Furthermore, although the classical-conditioning procedure may involve the association of a new stimulus with the unconditioned response, it has been found that the intensity of the conditioned response is seldom the same as the unconditioned one—*i.e.*, a dog does not salivate as much in response to a buzzer as it does in response to food—a fact that Pavlov failed to observe adequately.

If the restrictions on the animal's freedom of movement usually imposed in conditioning experiments are removed, the conditioned response is then accompanied by behaviour that appears to the observer to be appetitive to the occurrence of the unconditioned stimulus. Still further, it has been shown that dogs, for example, will salivate to a conditioned stimulus even if, during the training trials, salivation to the unconditioned stimulus is blocked by the drug atropine. Thus, in higher animals at least, it now appears that conditioning occurs as a consequence of the simultaneous stimulation of two areas of the brain, for if one part of the brain is blocked, another part can still respond. At the level of such lower animals as the earthworm, on the other hand, conditioned behaviour can be regarded as pure classical conditioning because it seems to involve little more than a simple change in stimulus–response relations. Hence, as defined behaviorally, classical conditioning, like habituation, covers changes that differ greatly in complexity.

The pattern of activity in trial-and-error learning

Trial-and-error learning. The concept of trial-and-error learning resulted from watching animals obtain rewards (*e.g.*, food) upon the successful completion of certain tasks, such as rats finding their way through a simple maze or kittens escaping from an enclosed box by pressing a lever. In such situations the animal is always active, directing its attention first to one and then to another part of the experimental chamber; sniffing, pushing, pulling, biting, clawing, and, above all, running around until it more or less accidentally hits upon the action or system of actions that brings the desired reward. The action itself need not be and, indeed, seldom is immediately directed toward the achievement of a desired goal; the kitten in the box or the rat in the maze is not in a position to see, even if it could understand, the relationship between its actions and the acquisition of the desired objective. Thus, trial-and-error learning involves the use of a reinforcement (*i.e.*, a reward) to encourage the

The Two Types of Conditioning as Components of Trial-and-Error Learning

key: **C.A.$_i$—involuntary** element in consummatory act (*e.g.*, salivation); **C.A.$_v$—voluntary** element in consummatory act (*e.g.*, mastication); **Rsp.$_v$—voluntary** act as part of appetitive behaviour giving **St.$_p$**; St.,—external stimulus (*e.g.*, buzzer); **St.$_p$—stimulus** arising within the animal (proprioceptive)

	initial state	training process	result
Classical conditioning	$(\text{reward}^* \dashrightarrow \text{C.A.}_i)$	$\text{St.}, \dashrightarrow (\text{reward} \dashrightarrow \text{C.A.}_i)$	$[\text{St.}, \dashrightarrow \text{C.A.}_i]$
Operant **conditioning**	$\left(\text{reward} \dashrightarrow \begin{pmatrix}\text{C.A.}_i\\ \text{C.A.}_v\end{pmatrix}\right.$	$\begin{pmatrix}\text{Rsp.}_v\\ +\\ \text{St.}_p\\ +\\ \text{St.},\end{pmatrix} \dashrightarrow (\text{reward} \dashrightarrow \text{C.A.}_v)$	$[\text{St.}, \dashrightarrow \text{Rsp.}_v] \xrightarrow{\text{anticipates}} \text{reward}$
Trial-and-error	$\left(\text{reward} \begin{matrix}\dashrightarrow \text{C.A.}_i\\ \searrow \text{C.A.}_v\end{matrix}\right)$	$\text{St.}_e \dashrightarrow \begin{pmatrix}\text{Rsp.}_v\\ +\\ \text{St.}_p\end{pmatrix} \dashrightarrow \left(\text{reward} \begin{matrix}\nearrow \text{C.A.}_i\\ \searrow \text{C.A.}_v\end{matrix}\right)$	$\left[\text{St.}_e \begin{matrix}\nearrow \text{C.A.}_i\\ \searrow \text{Rsp.}_v\end{matrix}\right] \xrightarrow{\textbf{anticipates}} \textbf{reward}$

'That is, food.

selection of a proper response from a variety of behaviours in a given situation, until eventually an association is established between the stimulus or situation and the one correct response for obtaining the reinforcement. In such cases the reinforcement is preceded by both the stimulus and the requisite response, which, unlike classical conditioning, is not the inevitable inherited response to the reinforcement. Many other names are used to describe this type of learning, among which operant conditioning, instrumental conditioning, and action conditioning are the best known.

Operant conditioning methods. The relation between trial-and-error learning and operant conditioning is best explained by means of an experiment that also serves to indicate the difference between the classical-conditioning procedure and the action-conditioning procedure. In this experiment a guinea pig is placed in a comfortable harness, after which a buzzer is turned on and continues to buzz until the animal accidentally makes a particular movement that has been selected by the experimenter. When this action is performed, which may be a turning of the head—say, to the right—the buzzer is immediately stopped, and a portion of food (*e.g.*, a carrot) is delivered automatically to a food receptacle in front of the animal. Hence, in this case, the stimulus is the buzzer, the reward or reinforcement is the food, and the correct response is the turning of the head to the right. As head movements become more frequent, the animal soon learns to obtain food by turning its head only to the right. Eventually, the animal learns that it can ignore the external stimulus (the buzzer) entirely and associates its own voluntary act (turning of the head to the right) with the coming of a reward, an action that it will then continue to perform, even in the absence of the buzzer, in anticipation of obtaining food.

Types of operant-conditioning responses

Two types of responses involved in operant conditioning have been recognized: respondents, those responses that are immediately elicited by known stimuli; and operants, those responses that, prior to conditioning, are not necessarily correlated with any identifiable stimulus. The knee jerk in response to a tap on the knee is a respondent; the pressing of a lever to obtain food is an operant. In a typical operant-conditioning experiment, an animal is first trained to operate a mechanism by an appropriate response, such as the pressing of a lever by a rat or the pecking of a disk by a pigeon in the manner already described. Then the operant so selected can be used in isolation to study the effect of different programs or schedules of reinforcement. A suitable reinforcer is delivered to the subject either with every response or on a previously arranged schedule—*e.g.*, with every nth response or on the first response after an interval of *n* seconds from a previous response or reinforcement. This "schedule of reinforcement" is of crucial importance in determining the subsequent rate and pattern of responding, since it has been found that there is a decrease in the rate of responding—that is, an extinction of the response—by withholding reinforcement.

Nature of trial-and-error learning. As the above examples indicate, trial-and-error learning contains elements of both classical conditioning and instrumental, or operant, conditioning. Indeed, under natural conditions with normally behaving higher animals, instrumental, or operant, conditioning is only an isolated part of the total learning process, just as classical conditioning is only a part. Normally, the two are combined in varying proportions and adjusted in varying ways (see the Table).

Again, as has been the case for the different types of learning already described, the complexity of the behaviour that results from trial-and-error learning varies with the evolutionary level of the organism. Moreover, the criteria of what constitutes trial-and-error learning may embrace phenomena of enormous differences in complexity. There are fundamental differences, for example, between the learning of a maze by insects and by rats. The insect in the maze first masters each turn separately and later exhibits the habits thus formed in a stereotyped way. Starting with the first run, the rat, on the other hand, integrates its learning of each turn with its overall mastery of the maze pattern. Once the maze has been mastered, the rat can readily learn to run it in the reverse direction or under changed conditions, whereas insects have only a limited ability to do this. Thus, it is evident that in trial-and-error learning an integration of many different stimuli is proceeding all the time––in other words, the animal is engaged in a continuous process of perceptual organization.

Latent learning. Latent learning is a type of associative learning that takes place in the absence of any patent reward. Rats that have been allowed to run through a maze unrewarded for a period of several successive days will learn the pattern faster and with fewer errors when finally rewarded with food than will other rats that have never run the maze before but received food from the start of the learning sessions. Obviously, the initially unrewarded groups must have learned something during their early trials; their learning is latent because it did not become evident until activated by the reward.

As is described later, latent learning appears to be of great importance in the normal behaviour of higher animals in the wild. In itself, however, latent learning is not a hard-and-fast category; it is very much like trial-and-error learning, some of which can also occur in the apparent absence of reward. As a matter of fact, it is because learning can take place without any obvious reward that the concept of reward or reinforcement in the learning process has become an extremely tenuous one. Rather, it would appear that latent learning, as well as the concept of insight and insight learning, both of which are described in the next section, depend to a considerable degree on the processes of organizing sensory perception.

Insight learning. The concept of insight in animal learning has been the subject of experiment and debate among psychologists ever since the term was first introduced in 1921. Even today its use still causes many difficulties, though much progress has been made in understanding the nature of insight learning. Evidence for the existence of insight was provided by a famous experiment with captive chimpanzees confined in a large, lofty room with walls that they could not climb. A favourite food

Experimental evidence for the existence of insight

(*e.g.*, a bunch of bananas) was hung from the ceiling, well beyond the reach of the animals, and scattered about the room were several wooden boxes. After a number of unsuccessful attempts to reach the bananas by jumping for them, the chimps would abandon their efforts and sit quietly, appearing somewhat dejected. Then, in a typical case, one of the chimps would look at the various boxes and glance upward at the bananas. Suddenly he would drag one of the boxes into the centre of the room, stand on it, and try to reach the fruit. When he found that the box was not high enough, he would get another box and stand it on the first, again without success. After putting a third box, often very precariously, on top of the other two, he was able to climb the rather unstable structure and reach the fruit.

It was concluded that the chimps had suddenly realized the potentialities of the boxes; having done so, they were able to proceed to an intelligent solution of the problem. The ability to achieve this solution, however, obviously depended on a new organization of the animals' visual, tactile, and spatial perceptions, such that they were able to embark on a new course of action appropriate to the situation. Although the chimps appeared to solve the problem without any apparent trial-and-error behaviour, some psychologists have suggested that trial and error may have been the way in which various ideas were considered and rejected mentally. On the basis of such experiments, insight learning has been defined either as the sudden production of a new adaptive response that is not arrived at by trial-and-error behaviour or as the solution of a problem by the sudden adaptive reorganization of experience.

Experimental situations were also devised to determine if rats could learn insightfully under systematically controlled laboratory conditions. Most of these tests concerned the ability of an animal to make a detour around an obstacle without any preliminary trial-and-error behaviour. The obstacle was placed in a maze that the rats had already mastered; the fact that they almost immediately successfully circumvented the obstacle by taking a new path through the maze indicated that rats can learn by insight. Moreover, the performance of animals in this kind of test provides evidence of what many would call "intelligence." It is extremely difficult, if not impossible, to arrive at a satisfactory general definition of intelligence relative to animal experiments because much depends both upon the innate capabilities of an animal and upon its previous experience, full details of which are often lacking. It is obvious, however, that when an animal arrives at an insightful solution, it does so by abstracting certain features of its experience and relating them. Thus, it can be said that the first thing on which intelligence depends is the degree of abstraction of which an animal is capable.

Imitation. Imitation has often been regarded as evidence of a high degree of insightful behaviour, but this by no means follows unless the term is used in a critical and restricted manner. In fact, a number of behaviorally diverse phenomena are often included under the heading of imitation.

Social facilitation. One type of imitation is social facilitation, which is a pattern of behaviour already in an individual's repertoire that he performs when he sees the same behaviour being performed by other individuals. Yawning is a common example of such behaviour in man. Because the effect is immediate, learning need not be involved; though, of course, the propensity to exhibit social facilitation may depend upon learning.

Yawning as an example of social facilitation

Local enhancement. Another phenomenon sometimes classed as imitation is local enhancement, which is an increased tendency to respond to an environmental feature as a consequence of another individual's response to it. For instance, the habit of removing caps from milk bottles probably spreads through a local population of titmice (*Parus* species) partly because of local enhancement. In another example the greenfinch (*Carduelis chloris*), having once learned to discriminate between palatable and unpalatable foods, may start to eat unpalatable substances again after seeing an untrained companion do so. The reverse also occurs, as when one bird learns to avoid a potentially dangerous situation by observing the behaviour of experimental birds.

True imitation. The ability of true imitation, which implies the copying of an otherwise improbable act or utterance, is in the main restricted to man and to such other primates as chimpanzees and capuchin, macaque, and rhesus monkeys. For example, cats, weasels, and other hunting animals learn the best methods of killing their prey by watching their parents doing so. Even more striking is the way in which the chimpanzee "Washoe" learned the gestures of the American sign language by watching her human teachers (see below). It should be emphasized that true imitation need not necessarily include auditory–vocal imitation of animals such as birds. If, as just defined, true imitation means the copying of a novel or otherwise improbable act or utterance, then true visual imitation becomes something that apparently involves self-consciousness and something of intent to profit by another's experience.

Because song and call notes play such an important part in the social organization of birds, the subject of vocal imitation must be handled with care; it may, in many cases, actually be a kind of trial-and-error phenomenon. This is because the action of uttering a sound is unique among all actions in one very important respect: whereas action of any other kind cannot be perceived by the actor in the same way in which it is perceived by fellow creatures and in which he perceives their actions, the sounds he and they utter are, presumably, perceived by him and by them in much the same way. In short, animals are unable to see themselves as others see them, but they may very easily be able to hear themselves as others hear them. A creature (a child or other animal) endowed by nature with a number of vocal mechanisms that enable it to utter a variety of sounds expressing various emotions or impulses is an example. Whenever any mechanism is activated, the creature hears its own voice uttering a corresponding sound; in consequence, this sense impression becomes associated with the particular mechanism, and the hearing of a sound will tend to activate it. When the creature hears the sounds uttered by another, they have the same effect; namely, insofar as the creature responds by vocal utterance, it will tend, by virtue of the preformed association, to utter the same sounds that it hears. Thus, the apparently very complex and elaborate imitation of sounds by birds may be explained on the basis of trial-and-error. Such vocal imitation, characteristically a phenomenon of bird life, is a major factor in the system of vocal communication developed by birds (see COMMUNICATION, ANIMAL).

Tool use and construction. It is often argued that the ability to use a tool is evidence of a high degree of mental development. If the use of tools is considered to be the manipulation of inanimate objects for a particular end, then the vast majority of birds that manipulate grass, stems, leaves, moss, and other materials to make their nests would be tool users. This view is not adopted here, however, because even though the material used in building a nest is manipulated during the construction process, it is not actually used as a tool to make the structure. Also ruled out are those cases in which predators break open the protective shells of prey by smashing them against a rock or stone or by dropping them from a considerable height. Similarly, the bearded vulture (*Gypaetus barbatus*) is excluded as a tool user, even though it drops bones from a great height so that it can pick out the marrow from the broken remnants.

Tool-using birds and mammals

Recognized by most investigators as a genuine case of an animal that uses a tool is the satin bowerbird (*Ptilonorhynchos violaceus*); it employs fibrous material as a brush with which to paint the sticks of its bower with a coloured substance made from berries or charcoal. Still more remarkable is the woodpecker finch (*Camarhynchus pallidus*) of the Galapagos Islands, which feeds by using a cactus spine or a twig as a tool with which to probe into bark crevices; an insect or a spider that runs out is seized by the bird and eaten. The Australian black-breasted buzzard (*Haemirostris melanosterna*) is said to drive emus from their eggs and then fly aloft with a stone in its claws that it drops on the eggs. Another feeder on eggs, the Egyptian vulture (Neophron percnopterus), picks up as large a stone as its bill can hold and then, standing very erect, throws the stone at an ostrich egg repeatedly until the egg shell breaks.

Among mammals, one tool user is the California sea otter (*Enhydra lutris* nereis), which feeds on large mollusks, such as the abalone. The animal dives to the sea bottom to pick up a small boulder or a large stone with one foreleg and an abalone with the other. Then, while floating on its back with the stone resting on its chest, the otter breaks the shell of the abalone by hammering it against the stone. The stone is dropped as soon as the meat is accessible; when the meal is finished, the otter returns to the bottom for another mollusk and may even come up with the same stone. Clearly in this case the otter intentionally selects a stone as a tool, and all the behaviour involved here is striking evidence of at least elementary planning and foresight.

In the above examples, objects, although chosen as tools, are not worked or modified in any way by the animals that use them. Among scholars, it has always been assumed that the ability to make tools is characteristic only of man.

This assumption is now in doubt, however, because chimpanzees in the wild have been observed making tools from sticks for use in capturing termites for food. A chimp first breaks off a suitable branch from a tree, after which it strips off the leaves and then thrusts the tool so formed into the entrance of a nest of termites, which cling to the stick. The chimp then draws the tool through its lips, wiping off the clinging termites. In another remarkable tool-making feat chimpanzees, during the dry season chew leaves and thrust wads of them into small holes in rocks in which water has accumulated. When the chewed leaves are withdrawn, the chimps suck the absorbed water from them. Other examples of what appear to be insightful learning behaviour are mentioned below in connection with particular animal groups.

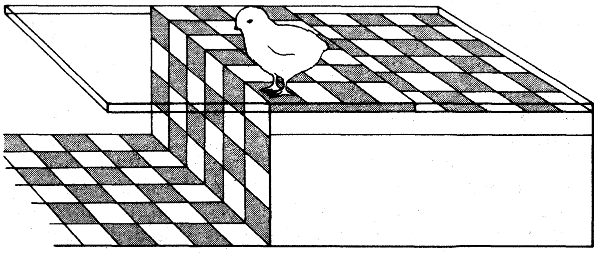
Perceptual learning. It has been emphasized that the categories of learning described above are not mutually exclusive and that each contains examples differing greatly in complexity. One important reason is the varying extent to which perceptual learning is involved (see LEARNING, PERCEPTUAL). In other words, an organism may modify the way it perceives stimuli as a result of continued or repeated exposure to them, even in the absence of any conventional reward. Thus, although rats reared in darkness during the first few months of life can learn later to discriminate between a circle and a triangle, other rats that have had early exposure to the same geometrical forms (but without any discriminating reinforcement) appear to learn faster how to distinguish between the figures when taught to do so. The effect, however, is dependent on the experimental conditions. When the early experience is with forms made from metal cutouts, subsequent discrimination learning is much more efficient than when the same forms are merely painted on a flat surface. It has been suggested that the difference lies in the attention-getting properties of the cutouts, which show depth at their edges.

The ability of rats to make abstractions

The ability to abstract certain properties of a stimulus has been studied with rats required to discriminate the magnitudes—successively, but in varying order--of brightness, area, and height and, also, the form of different stimuli. It has been found that although animals reared in the dark and given visual experience only during the testing periods showed a gradual improvement on the magnitude problems, their performance on all tests was inferior to that of light-reared animals. The difference in improvement in the magnitude problems, each of which was essentially a novel one, was interpreted as indicating that the animals responded not simply to a difference in brightness, size, or height but to differences in various properties of the stimulus itself. Similarly, when the performance of rats that had been reared in different ways was tested in a tactile–visual discrimination problem by exposure to three-dimensional forms (a cross and a triangle), the results supported the view that learning of various properties of the stimulus had been achieved. From these results it seems that the animal must first learn to differentiate between the various characteristics of the stimuli; that is, to differentiate tactual and visual cues related to the edges and angles of the forms. At the same time, they must also learn which of the objects is associated with food and which is not.

Many animals, such as chicks and rats, can respond to visual stimuli that indicate depth and solidity the first time they have an opportunity to use their eyes. One of the most striking of such responses is the reluctance to cross a visual cliff (see the Figure). This response may be retained, although it may become weaker, even if the animal is kept in total darkness for a considerable period. Such behaviour in mammals appears to be more sensitive to visual deprivation than that in birds.

From R.A. Hinde. *Animal Behaviour*, 2nd ed.; © 1970, reprinted by permission of McGraw-Hill Book Co.



Visual cliff.
The chick is reluctant to cross what appears to be a drop, although the unseen glass extension will provide support.

The development of correct visual responses in young kittens is greatly facilitated when active movement by the animal is part of the visual experience. When one kitten was maintained in a box mechanically linked to the movements of another kitten, so as to provide passive movement with the same visual experience as that gained by the active kitten, the inactive individual was inferior to the active one in learning movements. Such activity by the animal during the learning process is apparently another aspect of latent learning—namely, exploratory learning. It is the nature of most higher animals and a surprising number of lower ones, such as insects, actively to explore their environment and to learn its main features and their relationships—in other words, to "learn their way about."

Imprinting. Imprinting is an instructive example of the way in which the range of stimuli capable of eliciting a response can be limited and refined by experience. In 1873 an investigator observed that a newly hatched duckling or gosling will follow practically any moving object in the same way that it follows its own parent. This observation was not the first of its kind on this subject; the 1st-century Roman author Pliny the Elder wrote of "a goose which followed Lacydes as faithfully as a dog." A German ornithologist observed the same phenomenon in 1910 and christened it *Prägung*, which was translated into English as "imprinting."

Following behaviour in birds

Although imprinting is not fundamentally different from other forms of learning, it has certain properties that make it of particular interest. It is a process by which an animal—particularly the young of ground-dwelling birds, which leave the nest soon after hatching and so, being dependent on their parents for food or warmth, have to maintain contact with them by vision—may, accidentally, develop a preference for anything (animal or object) other than its own kind. Such a preference is shown by following behaviour. Although learning of the imprinting type is pre-eminently an avian phenomenon, following behaviour similar to that characteristic of imprinting in birds can be observed in young mammals that have been separated early from their parents. Such behaviour is much less common in mammals, however, because they are primarily creatures in which the sense of smell is highly developed, and it is this sense that is most likely to be the dominant one in establishing mother–offspring and sibling relations.

Sensitive period. One of the most remarkable characteristics of imprinting is its restriction to a limited sensitive period, which sometimes may be of only a few hours' duration. Among chicks and mallard ducklings, for example, imprinting is most effective from 13 to 16 hours after hatching. Thus, chicks will not follow a novel moving object (*e.g.*, a man) when they are only a few hours old nor when they are several days old but only during the intervening period. Moreover, the intensity of the sensitive period varies among animal species; even in animals of the same species, the sensitive period may be different for auditory and visual stimuli.

The period seems to begin as soon as the locomotory ability of the animal has developed sufficiently for following; the period is not always terminated, however, by a waning of the tendency to follow. Such birds as coots (Fulica atm), for example, will tend to follow one new object after another up to at least 60 days of age, but only when tested in an environment with which they are familiar. Some young birds will follow objects of almost any colour, shape, or size, from a matchbox to a man. There is much evidence that in some species the sensitive period is terminated by the establishment of other types of behaviour, fear responses in particular, which inhibit the tendency to follow.

Environmental influences on the sensitive period

Several experiments clearly indicate the influence of the environment on the nature and length of the sensitive period; for example, moorhens (*Gallinula* chloropus) and coots reared in groups in pens shielded from the general environment do not follow models to any extent; birds reared in more exposed pens, however, follow readily. In chicks the sensitive period is extended by rearing them in isolation; in ducks, by rearing them in diffuse light. Although many young birds are naturally more attracted to moving objects than to stationary ones, there is no absolute distinction between static and mobile stimuli: chicks may be attracted to static objects or to a general environmental quality or pattern in a manner akin to imprinting. In such cases the persistence with which chicks avoid a mobile object is reduced if the object is coloured in a pattern similar to that of the walls of the pen in which they have been reared. This indicates that, within a few days after hatching, chicks learn the characteristics of their pens well enough to be able to discriminate objects that are sufficiently dissimilar from those to which they are accustomed. This behaviour is particularly enhanced by the conspicuousness of the pattern with which the walls of the home pen are painted: the more conspicuous (and therefore presumably more easily learned) the home environment, the longer the chicks avoid differently painted models in the test pen.

Attraction of *conspicuous* objects. Imprinting raises the question of cause: why does the newly hatched bird approach conspicuous or moving objects in the first place? Parental calls may normally play an important role in the laboratory, but naive birds will approach silent objects equally well. Imprinting is not by any means peculiar to young birds, however, for the young of many animal species tend to approach a stimulus of low intensity and to flee from one of high intensity, which is more or less what happens in the imprinting situation. Thus, within certain limits it can be said that the more conspicuous an object, the more likely it is to be followed or approached in an initial exposure during the sensitive period.

It should be emphasized that even the newly hatched chick has a high level of behavioral organization and does not wait passively until exposed to a conspicuous object; rather, it shows behaviour that is likely to bring it into contact with the object. One investigator has found that chicks are more active a few days after hatching in a gray environment than in one painted with black and white stripes, even though both reflect the same total amount of light. This suggests an inverse relationship between activity and the complexity of the environment; that is, the active chick is likely to find less complex stimuli more appropriate for imprinting. It has also been shown in day-old chicks and ducklings that a conspicuous stimulus of the type used in imprinting experiments is reinforcing on the first occasion it is encountered; such birds rapidly learn to press a pedal that switches on a flashing light. There is also a sensitive period during which the light acts as a reinforcer; the readiness to learn the response decreases to a minimum about 48 hours after hatching, which is about the time when there is a corresponding peak in the persistence with which chicks in an imprinting situation avoid the models. Thus, during the course of the sensitive period, the birds develop a preference for the object with which they are rewarded. Imprinting and operant conditioning therefore proceed hand in hand, with imprinting narrowing the range of stimuli that can have reinforcing properties.

The role of familiarity in imprinting

Relationship to perceptual learning. It would appear that imprinting consists largely of the development of familiarity with a moving object. Specific rewards, such as food or contact with the moving object or model, are not necessary. Under ordinary barnyard conditions, chicks become familiar with at least two types of moving objects, the mother and the siblings; under experimental conditions, coots similarly learn to follow several different objects. If familiarity is the main issue, then imprinting would seem to be closely related to perceptual learning, for in both cases the responsiveness to a stimulus is influenced by a previous experience with the stimulus, independent of its association with any rewards. If this is indeed the case, then imprinting can be a valuable tool for investigating the genesis of the perceptual aspects of behaviour, for any learning during the sensitive period would be revealed in contexts other than that of the following response. Thus, the performance of chicks in the discrimination problem described above was improved when one of the patterns to be discriminated was the same as that on the walls of the home pen. This is similar to the test used for perceptual learning. It has also been found that early imprinting to a flickering light (visual flicker) made it easier to teach an animal to approach or avoid the same stimulus a month later. This effect was found, however, only in those individuals that exhibited the following response to the stimulus during the imprinting period, suggesting that mere exposure to the stimulus is not sufficient for the early perceptual learning to occur. Thus, it may be necessary for the stimulus to become associated with a response of some kind, although it does not have to be the same response that is used in subsequent testing to reveal any perceptual learning.

The conclusion that, during the sensitive period, conspicuous objects have a reinforcing effectiveness for naïve birds that seems to increase with familiarity is consistent with many of the findings of imprinting experiments. For instance, an imprinted bird shows a searching behaviour when the object is absent; in addition, the willingness to follow an object and to respond to it selectively tends to increase with experience. The evidence also indicates that reward is unnecessary for imprinting, that objects which at first invoke fear may later (after habituation and familiarity) elicit following responses, and that unfamiliar objects are more likely to elicit following if presented in the same situation in which the bird has previously been imprinted to other objects.

A feature of imprinting that once attracted much interest was the effect it was supposed to have on the subsequent social, sexual, and parental behaviour of the adult individual. Cases have been reported in which the sexual preference of the adult bird was more or less irreversibly determined by the types of companions or objects to which the bird had been imprinted in its early life. In some cases at least, this effect may be the result of an early imprinting experience that produces fear of the object, thus affecting later sexual responses.

II. Behavioral and physiological aspects of learning

What an animal will learn is not merely a matter of its learning capacity; there are many constraints and restrictions that impinge upon its basic learning ability—*e.g.*, the limitations of its sensory and perceptual mechanisms. Because of such limitations, an animal is more likely to react to some stimuli than to others; therefore, it learns

Constraints and restrictions to learning

more about some than about others. Furthermore, all stimuli to which an organism is responsive in some contexts may not be effective in producing learning in other contexts. Thus, an octopus compensates posturally for the weight of objects it picks up in its arms, but it is unable to learn to distinguish between objects that differ only in weight. Similarly, although bees can learn to orient themselves by complex shapes when steering a course to and from the hive, they cannot learn to recognize such shapes when they are used as signals to indicate the presence or absence of food on a tray.

Another important fact often not taken into consideration to account for differences in learning ability is that the reinforcing effect of such circumstances as food, water, or warmth is not an intrinsic property of the stimulus but of the animal; which means that their value to the animal must have been established by natural selection. Consequently, although the responses and the general "intelligence" of two related species may appear very similar, if not identical, in conventional laboratory tests in which food and water are employed as reinforcers, the learning capacities of the same two species may be very different when studied in other situations. In other words, under more natural conditions in which naturally occurring stimuli are used as reinforcers, one species may learn in a context in which the other does not, whereas both species had learned equally as well in the laboratory. Sunfishes (Centrachidae) of the Northern Hemisphere, for example, can orient their movements by the position of the sun when first exposed to it, even after having been reared in artificial light, provided the sun moves from left to right when they face it. If, however, the sun is moving from right to left when they face it, as it would in the Southern Hemisphere, they cannot orient their movements to it; hence, these fishes are restricted to northern waters. By contrast, a cichlid fish (Aequidens portalegrensis), whose range spans the Equator, can orient itself by either the northern or southern sun but not both, depending on which way the fish was facing when it first saw the sun.

The last example illustrates how the normal environment may restrict what is learned: in this case the mature cichlid fish can only respond to the sun in the Southern Hemisphere if its first experience was with a sun that appeared to move from right to left. Similarly, birds that leave the nest soon after hatching become imprinted to their parents rather than to other conspicuous objects because their parents are both available and moving, and the newly hatched offspring are sufficiently developed to be able to follow them. In contrast, birds that are reared for a time within the nest may well become imprinted to conspicuous objects in the environment, in addition to or instead of their parents. Thus, as all of the preceding examples indicate, all of the natural limitations and constraints on learning must by analyzed in order to identify the conditions necessary for learning to take place.

STIMULUS–RESPONSE AND COGNITIVE MECHANISMS

In the discussion of classical and operant conditioning, examples were considered of what may be called the stimulus–response (S–R) theory. In these examples a new S–R relationship was brought about by one or more procedures involving some kind of reward or reinforcement. It was once thought that classical conditioning applied only to involuntary visceral responses (*e.g.*, those of the salivary glands or the digestive system) and that operant conditioning applied only to voluntary responses (*e.g.*, purposeful movements of the limbs). This implied that if visceral responses could be modified only by classical conditioning, they could be learned and maintained only by using stimuli that had the ability to elicit the same response. If, on the other hand, visceral responses could be learned by operant conditioning, they could be acquired and maintained in a variety of ways. Indeed, in recent years it has been shown that operant conditioning is effective on a wide variety of visceral and other involuntary responses, such as heart rate, intestinal contractions, blood pressure, urine formation, and brain-wave activity. It has even been found that, in some cases at least, learning can take place while an animal is immobilized by a muscle-paralyzing drug (curare), and that anything learned by the animal while in this condition can be transferred to the nondrugged state. As a result of such findings, there has been a strong tendency for some investigators to subsume all learning under the S–R theory.

Modification of involuntary responses

On the other hand, another school of theorists strongly rejects the conclusion that all learning can be explained in terms of the S–R theory. These theorists believe that, in many cases, learning is often a restructuring of internal processes in the central nervous system, in a manner that is not necessarily the same as the mechanism involved in S–R learning. Thus, because those who hold this view are concerned with cognition (the internal acts or processes involved in knowing or perceiving), they are usually called cognitive theorists. A simple example serves to indicate the difference between the two. According to the S–R theory, a rat masters a maze by learning which responses must be made to which stimuli. The cognitive theory, however. might involve the assumption that the animal is learning a "cognitive map" of the maze in such a way that it acquires a knowledge of "what leads to what"—*i.e.*, that one particular stimulus will be followed by another one. Although the results of many learning experiments can be interpreted in terms of either theory, the difference between them is, nonetheless, an important one.

Role of reinforcement. Because the role of reinforcement in the formation of S–R relations is often a particularly crucial one, it merits further consideration. There have been many suppositions concerning reinforcement, one of which is the theory of contiguity. According to this theory, the only reinforcement needed for the association of stimuli and responses is that there be a certain temporal relationship between them. The problem with this theory is that not all stimuli or responses that occur in a time sequence become associated, hence it is still necessary to explain why some things are learned and others are not. Among the many attempts to overcome this difficulty is the suggestion that perhaps there is some motivational state that determines the stimuli to which the organism will attend; thus, the stimuli that become associated with each other or with responses govern what happens. In such a case, reward or reinforcement confirms an expectancy about the nature of the environment.

The theory of contiguity

Internal reinforcement. Many theories concerning the concept of internal reinforcement have been advanced. It has been proposed, for example, that such reinforcement can be a satisfaction of a biological need or drive (drive reduction), that it can come from the removal of stimuli associated with such a drive (drive–stimulus reduction), or that it can be the result of the appearance of stimuli previously associated with the drive (secondary reinforcement). Although such suggestions have been useful and evidence exists to support some of them, it now seems clear that they do not provide an answer in all instances. Thus, although drive reduction (*e.g.*, alleviation of thirst) may be reinforcing, reinforcement is not necessarily associated with it. And, although consummatory behaviour (*e.g.*, drinking) can be reinforcing, the consummatory response may not necessarily be the most important reinforcing event in that particular train of behaviour. Indeed, the kinds of stimuli that can be reinforcing seem almost infinite; rats and monkeys can learn to press a lever if rewarded by an increase in illumination. Yet other studies have shown the opposite effect; animals can learn to press a lever to decrease illumination. And in some experiments the rats wanted to turn the light both on and off; apparently they wanted novelty above all else.

Other significant examples providing evidence that internal reinforcement is widespread in cognitive learning concern searching image, reward expectancy, and selective attention. Much repetitive behaviour that involves a scanning of the environment until a particular set of stimuli is encountered can be understood on the assumption that the animal is using a "searching image." Thus, a bird such as the European jay (*Garrulus glandarius*) accustomed to feeding on plump caterpillars will fail to see so-called "stick" caterpillars, which have a close protective resemblance (camouflage) to small twigs, when such caterpillars are placed on the branches of trees in an aviary. Should the bird chance to touch one, however, and find that it moves and is edible, the bird will then examine actual sticks or twigs persistently for a while. If it finds no more "stick" caterpillars, the bird will stop searching for them and resume his search for the plump caterpillars. In other words, the bird behaves as if it were internally "expecting" a different kind of reward and exercises selective attention toward the situation that will provide that reward.

Novelty of environment. Many animals in captivity will work apparently to obtain novelty of environment. The reward value in the novelty of either increased or decreased illumination, for example, may really be related to the normal rhythm of light and darkness in the animal's life. Thus, when deer mice (*Peromyscus maniculatus*) are allowed to control their illumination in the laboratory by lever pressing, they tend to select light intensities similar to those they would experience at corresponding times of the day in the wild. Deer mice in captivity will also spend long periods of time running inside a turning wheel, but the reinforcing nature of wheel running is far from simple. If the speed or difficulty of the turning wheel are programmed in any way by the experimenter, the deer mice will invariably change the program when provided with levers that enable them to do so. It is as if the mice want to program their own running instead of having it programmed for them.

Exploratory behaviour in reinforcement

Learning also tends to be facilitated when animals have access to objects that can be played with; indeed, all sorts of situations that involve exploratory activity have been found to be reinforcing. Exploratory behaviour tends to cease, however, without any change in the outer environment (external stimulation), suggesting that the reinforcement of the exploratory activity must arise internally. In other words, because exploratory behaviour is elicited by novel stimuli, it comes to an end when, as a result of certain internal changes, the stimuli cease to be novel. This is in accordance with the restructuring of processes in the central nervous system as proposed by the cognitive theorists.

Conceptualization. The extent to which animals other than humans are capable of concept formation has been dramatically demonstrated. Pigeons have been trained to respond to the presence or absence of human images in photographs. The people depicted might be anywhere on the photograph; they might be clothed or nude, adults or children, in varied postures, of any race. Indeed, the pictures are so varied that simple stimulus characterization seems impossible; the results suggest remarkable powers of conceptualization by pigeons as to what is "human."

In another conceptualization experiment, monkeys have been trained to select the pictures of three different insects from among pictures of leaves, fruit, and branches of similar size and colour. In later tests the monkeys showed significant preferences for any insect pictures over those of other natural objects. Experiments with numbers (*i.e.*, "counting") done with birds and some nonhuman mammals provide striking evidence for the existence of number conceptualization in animals.

NEUROPHYSIOLOGICAL MECHANISMS

Memory and learning. It is naturally and correctly assumed that with all animals above the one-cell level of organization, some kind of nervous system is the primary mechanism for governing the changes involved in learning. It follows, therefore, that any system concerned with learning must also be concerned with memory, the storage of information (see MEMORY: RETENTION AND FORGETTING). Both psychologists and neurophysiologists have long been involved with what seems superficially to be a very clear-cut difference between two types of memory, short-term and long-term. This difference has been thought to indicate that there are two kinds of storage processes in the brain; one in which retention decays very rapidly and one in which retention is more or less permanent. Most people, for example, can remember a new telephone number only for a few seconds or at the most for a minute or two after first using it; in contrast, one brief experience of some other kind may become impressed, apparently indelibly, on the "consciousness." The distinction between the two types of memory is also based on the fact that a person who has suffered a brain concussion loses his memory for events that occurred just before the accident but not for earlier ones.

Considerable research on the function of memory in learning has been carried out with human subjects and with animals. It has been found, for instance, that when an electroconvulsive shock is administered to rats at various intervals following a training trial, the animals that receive the shock within a few minutes of the trial subsequently perform what they were taught less well than do those that receive the shock at later intervals. Similar evidence obtained by the use of certain anesthetics also supports the existence of short- and long-term memory. Moreover, the short-term storage system is held to be distinct from a very short-term preperceptual trace (one that decays in about half a second); in order for something to be recalled subsequently, subjects must respond to the material either during its presentation or within the half-second period afterward.

Suppositions concerning short-term and long-term storage systems

To account for the preperceptual trace, the short-term storage system, and the long-term storage system, it is assumed that the short-term system is mediated by neural mechanisms that decay autonomously with time and that are liable to interruption by many kinds of stimuli. The long-term system, on the other hand, is believed to be either permanent or nearly permanent, although the stored association may be subject to slow decay as a result of long disuse. It has also been assumed that long-term storage requires the preceding short-term neural activity for its fixation. This means that no long-term retention can develop if the formation of the short-term memory has been disrupted in any way, such as distractions of attention or electroconvulsive shocks. Much recent work on this subject, however, is causing drastic revision and rethinking of previous suppositions. For instance, there now seems little clear evidence that the disruptive treatment applied to the short-term memory system really affects storage; it could equally well impair recall.

The importance of time as a factor in fixing long-term memories has been demonstrated in the behaviour of such animals as the octopus. A blind octopus, for example, soon learns to reject inedible objects by means of tactual (touch) cues. If such an object is presented to the same arm briefly in a rapid series of trials, the object is rejected by that arm of the animal but not by the other arms. If, however, the trials are more widely spaced in time, the effect spreads to the other arms, all of which will eventually refuse the inedible object. This suggests that learning through one of the arms takes time to spread through the parts of the octopus brain that mediate the behaviour of the other arms.

Problems in explaining neurophysiological learning mechanisms. In considering the possible neurophysiological mechanisms of learning, there is at present no evidence that any single nerve cell (neuron) can perform the kind of operation necessary for comparing stimuli, as in the various kinds of associative learning. Even with habituation, which does not necessarily require simultaneous comparison, tnere is only scant evidence that one nerve cell is sufficient to bring about this simple form of learning. But even if one nerve cell can display the changes that appear to be associated with the simplest cases of behavioral habituation, this does not account for more complex examples of this phenomenon. In the closest link between habituation and association learning (the resemblance of the former to the processes of extinction and reflex inhibition that are so characteristic of the classical conditioning pattern, as well as of the more elaborate types of associative learning), there seem to be grave difficulties in supposing that behavioral habituation and extinction share the same underlying neural process, let alone that they are determined by the same neurophysiological mechanisms.

The problem of the relations between neurons

It follows that, if these conclusions are correct, investigators must look to the relations between neurons, often very large numbers of them, for any general physiological explanation of the mechanism of learning. Even then, given the many different kinds of neurons with their many different kinds of physiological properties and the many different kinds of junctions between them, there is no a priori possibility that the learning and memory functions of the nervous systems in both lower and higher animals can ever be explained. Indeed, theory has far outrun experimental evidence. There are innumerable examples of what can be called "wiring diagrams," according to which learning functions of nervous systems are explicable. It is even easy either to produce a mathematical model of relationships between nerve cells that would allow for pattern recognition or to argue plausibly that nerve nets, mainly random with regard to relationships between the individual neurons, could be efficient in the performance of many simple learning tasks. But, although elegant neural mechanisms can be imagined for learning form and pattern recognition and although impressive hardware models or mock-ups can be constructed that confirm theoretical conclusions, the physiological processes involved in learning and behaviour remain poorly understood.

The magnitude of the problems is apparent if the staggering number of stimuli to which an animal (including man) can respond selectively are considered. Obviously, the nervous system discriminates between a vast number of possible stimulus configurations, yet it is unlikely that there could be a separate mechanism for every object that an organism can recognize. A related problem is the question of the quantity of information that a nervous system of measurable size can store. The problem here is to determine to what extent memories are stored as discrete units or in complexes of associations. Many psychologists have thought that specific memories lose their identities by being organized into schemata, each experience being incorporated into an organization of experiences that are connected by a common experience. In such a case, recall would then be regarded as governed by generalization. Not until basic problems of this kind are better understood will it be possible to make reasonable estimates of the mental content of human brains.

Whatever the final answer may be, however, it is known that the memory store of some individuals is much larger than that of others. It was said of the 19th-century French philosopher Auguste Comte, for example, that, when writing a book, he never put a word on paper until he had composed the entire work, even the form of the sentences, in his head. Then, by an extraordinary feat of memory, he wrote the whole book and sent it to the printer without making any changes. Similarly, the Hungarian musician Franz Liszt could sight-read the most complex musical scores without error and apparently never forgot a work once he had seen and played it. For decades of life, until the impairments of aging become acute, the memory systems of such people seem impossible to satiate. Yet, although these and other problems are difficult to resolve and although much more experimental work remains to be done, the evidence is encouraging that neurophysiological changes are responsible for such behavioral adaptations as learning.

III. Evolution of learning ability

LEARNING IN INVERTEBRATES

For many years there has been controversy as to whether or not protozoans, such as the amoeba or paramecium, can "learn." Because some early observations with the amoeba failed to take into account certain changes in the physiological conditions of the animal, many experiments that were thought to have proved the existence of learning in these organisms were subsequently shown to be explicable on the basis of different and simpler hypotheses. Nevertheless, a number of instances indicate that both habituation and elementary associative learning do occur in some of the protozoans.

The nerve net in coelenterates

Evolution of neural structures. The coelenterates, a group of simple multicellular animals that includes sea anemones and jellyfishes, have specialized sense organs that are connected by a very diffuse kind of nervous system called a nerve net. Although there are no structures that remotely resemble a brain or spinal cord, the nerve net of a sea anemone contains about 100,000 cells and is able to mediate surprisingly complex behaviour, including oriented movement toward or away from chemical stimuli and such locomotory behaviour as rhythmic swimming movements. Sea anemones clearly exhibit habituation; more evidence is needed, however, before associative learning can be definitely attributed to these creatures, although it probably occurs. It has been suggested that the rhythmic nature of the activities in coelenterates, as well as the many cells in their nerve nets, may be the forerunners of the processes and structures out of which the central nervous activity of more complex organisms could have evolved.

The "memory molecule"

The free-living aquatic flatworms (*e.g.*, planarians) represent a very primitive stage in the evolution of a central nervous system, which in man and other higher animals consists of the brain and spinal cord. Groups of nerve cells (ganglia) in the planarian's head provisionally have been called a brain, and nerve cords extend along the entire length of the animal. After much controversy, it now seems safe to conclude that planarians do possess the capacity to learn; this includes both habituation and elements of associative conditioning. It also seems well established as a result of experiments with planarians that the substance ribonucleic acid (RNA) may transmit the effects of learning. When RNA is extracted from the bodies of flatworms that have learned to respond to paired light and shock treatment, it causes similar changes in behaviour in untrained planarians of the same species into which it is injected, thus confirming the assumption that RNA may behave as a "memory molecule." The mode of action of the injected RNA in flatworms is, however, completely obscure at present, and the early hopes of explaining memory storage in higher animals and in man in terms of DNA–RNA systems seem now to have faded nearly to the vanishing point.

The segmented worms or annelids—such as earthworms, sandworms, and leeches—have a much more highly developed nervous system than do the planarians; it includes a clearly defined brain and a double nerve cord with ganglia in each segment. Habituation and classical conditioning can be easily demonstrated in this group, and trial-and-error learning of a level sufficient to enable the animal to master a simple T-shaped maze has been shown. It seems, however, that although the annelid brain is implicated in some way in the learning process, it does not serve as a unique, discrete memory-storage centre.

Thus far, the evolution of anatomic complexity among animals has paralleled a fairly steady advance in learning capacity from simple habituation to associative learning, which appears first as classical conditioning and then as trial-and-error learning or operant conditioning. It is as if evolutionary progress in the powers of behavioral adjustment consists of elaborating and coordinating new types of responses that are built upon and act with the simpler ones of more primitive animals. That is, if habituation and classical conditioning are found low in the animal scale, they will also be found higher up, coexisting with more elaborate responses and with more highly developed sense organs and nervous systems.

Memory of the home range in arthropods and mollusks

Although the nervous system of arthropods (*e.g.*, insects, spiders, crabs) and mollusks (*e.g.*, snails, clams, octopuses) is of the same general type as that of annelids, it has much more complex sense organs—especially visual organs and proprioceptors (organs for sensing body position and movement). Such a big step in evolutionary development appears to have taken place at the stage at which it became necessary for an animal to find its way about or, in effect, to find its way back to a known home range or territory when displaced from it. Organisms with this ability, including the animals as low in the evolutionary scale as insects and snails, behave as if they had some kind of inner map of the region around whatever may be the centre of life—*e.g.*, a nest. Indeed, at first sight no fundamental advance in the ability to organize the perception of environmental features seems to exist between an insect such as a wasp and a more highly evolved vertebrate. Among the simpler mollusks, such as snails and slugs, even the more primitive forms behave as if they had an internal map of the environment that enables them to find their way home. Among the higher mollusks (*e.g.*, octopuses and squids), the neural structures, as well as visual and tactile sense organs, are as complex as those of many vertebrates.

Organizational patterns by neural elements. In addition to the highly organized and efficient memory of the home range found in some insects are the outstanding orientation abilities and direction-indicating powers of the honeybee. Experienced beekeepers know that, when a hive is moved to a new location, the worker bees coming out on foraging flights will pause and circle in increasing arcs around the new site for a few moments before flying off. It is clear that, in these brief survey flights, the insects learn the relative positions of new landmarks well enough to enable them to find their way back to the hive minutes or even hours later. Such behaviour raises the question of the number of nerve cells and the complexity of organization in the brain or other regions of the central nervous system required to mediate performances of this kind. It was this problem that prompted Charles Darwin to observe

> . . . it is certain that there may be extraordinary mental activity with an extremely small absolute mass of nervous matter: thus the wonderfully diversified instincts, mental powers, and affections of ants are notorious, yet their cerebral ganglia are not so large as a quarter of a small pin's head. Under this point of view, the brain of an ant is one of the most marvellous atoms of matter in the world, perhaps more so than the brain of man.

It is generally accepted that the outer shell (cortex) of the human brain contains about 10,000,000,000 neurons with an equal number of supporting, or glial, cells. In the octopus, which has behavioral achievements equal to those of many vertebrates, the nervous system is estimated to contain 5,000,000 neurons. It would seem that a bee's brain, despite its learning ability, certainly contains no more than 100,000 nerve cells, and that of ants considerably fewer. Yet 100,000 nerve cells is not vastly greater than the number found in the nerve net of a sea anemone. If "modelling" of the real world first appears somewhere in the scale between sea anemones and bees, it does not seem that this new faculty is the result of a vast increase in the number of nerve cells but rather that it results from changes in their pattern of organization.

Since a bee can do so many tasks with a brain of less than one cubic millimetre in volume, the functions of the remaining million or so cubic millimetres of the human brain are a matter for speculation. Whether in hunting wasps ***(Ammophila)*** or in man, the internal map or perceptual model of the real world is derived (abstracted) from changing patterns of sensory stimulation. This implies that within each organism, a process of abstraction from the highly varied patterns of individual stimuli occurs. Yet, however different the varied successive patterns of stimuli may be as perceived by wasps and man, they form a single coherent relationship that is determined by the same external conditions to which both are exposed. This suggests that the coherence and unity which our experience leads us to believe exist in the world of external phenomena also exist in the internal models created by the central nervous systems of different organisms. The ability of an organism to derive information from sensory stimuli (what might be called its level of abstraction), however, varies with the evolutionary grade of the organism.

Nature of intelligence. Despite the differences in neural structures of lower and higher organisms that are reflected in their ability to make abstractions, there appear to be certain, as yet unknown, properties controlling behaviour that are common to all organisms. Only when these common factors have been determined will it be possible to ascertain the true nature of intelligence, which obviously is not simply the result of more elaborate behaviour or of more highly refined nerve circuits. In order to understand the nature of intelligence, it is essential to understand first the way in which a species survives in the environment in which it lives and how it adapts to it. In order to measure an animal's intelligence, therefore, it is necessary to study a particular aspect of the animal's response to its environment and to devise tests to determine its learning ability as judged by its performance in that situation. But, because each species has different adaptations for coping with different situations, there is no single test of intelligence that is equally suitable for all species of animals. Even if a species appears to be deficient in some task as defined by a particular type of test, it is necessary, before the relevance of the result can be substantiated, to refer to the natural situation; in this way, a deficiency in one type of test may be compensated by superior performance in other types of tests. The ability to perform abstractions, "to conceptualize" (see CONCEPT FORMATION), is an important criterion in such tests, although this may largely reflect differences in sensory, perceptual, or motor functioning rather than differences in learning ability. The ability to reverse a learning process when required by changing circumstances is another important criterion. Experiments on learning in a large number of species indicate that, although the number of trials required for conditioned-reflex formation differs little between the goldfish and the chimpanzee, the rate of extinction (reversal) is slower or absent in the lower species. Within the vertebrates great differences exist regarding some aspects of this particular type of learning ability; small differences exist regarding others.

The relationship between intelligence and environmental factors

Among the invertebrates, an increase in the rate of learning and the complexity of problems that can be learned is clearly evident in the progression up the evolutionary scale from worms to ants; further increases are evident in lower vertebrates. Yet, in the entire animal kingdom only mammals (*e.g.*, rats and man) have been shown to be clearly better than insects in maze learning. This again illustrates the point that a fundamental step in learning ability is achieved whenever a species needs to find its way about. As mentioned previously in the discussion on trial-and-error learning, in the learning of a maze by an insect and a rat it has been found that the insect first masters the local choice points (where the maze branches), then later integrates the habits thus formed in a stereotyped way. By contrast, the rat starts both its local learning and its overall integration on the first run through the maze. Furthermore, once the maze has been learned, the insects have only a limited ability to learn to run it in the reverse direction or under changed conditions, but rats can do so readily.

Differences in ability to develop a learning set

Learning to learn. This would appear to indicate another difference in learning; namely, the ability to achieve a "learning set" or, as it is often put, "the ability to learn to learn." In other words, whereas rats seem to have the ability to learn any new behaviour required—*e.g.*, to run the maze in the opposite direction—the insects, once they have learned to run a given maze, are much more rigid in maintaining their behaviour. The insects may thus be said to be relatively inferior in developing a learning set. Rhesus monkeys appear to have a greater ability in learning to learn than do squirrel monkeys and marmosets. The learning performance of pigeons is comparable to that of cats; birds have a phenomenal ability for direction finding and orientation. Among mammals, the only reliable evidence for differences in learning ability between primates (*e.g.*, monkeys and man) and all the others is derived from "learning set" experiments.

Further discussion of the differences in learning capacity among lower animals will not be any more illuminating for, as suggested above, the situations under which such animals demonstrate their learning behaviour in nature have seldom been sufficiently analyzed. It should be of interest, however, to consider the higher animal groups (*i.e.*, fishes, birds, mammals) in terms of the exceptional features of their learning ability.

LEARNING IN VERTEBRATES

Fishes. Fishes are the dominant class among vertebrates, both in numbers of individuals and in numbers of species. There are probably more than 18,000 species, about 90 percent of which are bony fishes.

Significant behavioral characteristics. From the behavioral point of view three characteristics of fishes appear to be particularly significant. First is their elementary social behaviour, which expresses itself in their tendency to form schools or shoals. Second, the lack of prehensile (grasping) organs (*e.g.*, hands) and the small part that manipulation plays in the lives of most fishes might be expected to limit their powers of space perception. The third characteristic tends to counteract the second, however: most fishes actively move in a three-dimensional world for most if not all of their lives. Although there may be no obstacles present in the open sea, fishes that inhabit lakes, rivers, and coastal regions must constantly negotiate obstacles that they encounter. Such behaviour must aid substantially in the development of competence in spatial orientation and control.

Because many fishes form shoals or schools either with only their own species or with other species as well, they have ample opportunity to develop visual recognition of species by an imprinting-like process. Little firm evidence, however, exists of individual recognition among fishes. Moreover, because some species travel long distances during their life cycle, they also have ample opportunity to develop recognition of their local or home territory as well as to orient themselves over larger geographical ranges, as is observed in salmon returning from the sea to the breeding territories where they were born.

Exploratory learning by the goby fish

One of the most remarkable examples of exploratory learning is provided by the goby fish (Bathygobius *soporator*), a species that inhabits tidal pools. These fishes are so well oriented that they can jump from pool to pool at low tide without any significant risk of finding themselves on dry land. Yet it is known that the fishes are unable to see the neighbouring pools before leaping. The only apparent conclusion to account for such behaviour is that the gobies swim over rock depressions at high tide and thereby learn the general features and topography of the limited area around the home pool. Then, using the memory that they have acquired, they accurately jump from one pool to another at low tide.

Migratory behaviour. Although many laboratory experiments have been concerned with the ability of fishes to learn to make detours and to master simple mazes, the results of these tests are meagre compared with the well-documented migratory and homing performances of such fishes as salmon, trout, and white bass in the wild (see MIGRATION, ANIMAL). Most fish migrations involve the movement of larvae or young fishes from the places where they were hatched to their normal feeding habitats and a subsequent return journey to the spawning grounds. There is now indisputable evidence that a high proportion of some species of salmon and trout succeed in returning to the same streams in which they hatched. In transplantation experiments with the chinook salmon (*Oncorhynchus* tshawytcha), it has been found that, when the eggs were transplanted from one river to another, the fishes returned to the river in which they were hatched, not to the one in which the eggs were spawned. In another experiment, over a period of four years 97.9 percent of the steelhead trout (Salmo *gairdneri*) recovered had returned to their home stream, and only 2.1 percent of them went to a stream four miles distant; similar figures have been obtained for silver salmon (*Oncorhynchus milktschitsch*) in the same streams.

It is clear that the return of these fishes cannot be explained on the assumption that the intervening period of feeding and maturation has been spent in the neighbourhood of the native stream. Indeed, one species of chinook salmon (O. nerka) normally travels approximately 100 miles (160 kilometres) a day when returning from the ocean to its home stream to spawn; one salmon species (*Salmo salar*) is known to have travelled at the rate of 60 miles (100 kilometres) a day for 12 days for the same purpose. Among the large number of possible stimuli that may serve as guides or reference points during these astonishing homing achievements, water currents are among the more important. Fish also rapidly learn to keep a particular course and direction, employing orientation by the sun as a basis and then using many different clues for maintaining a steady course. Despite this ability, it seems clear that successful return in many cases cannot be the result of visual memory of the home or the route to it. This is because a young salmon descending a stream toward the ocean for the first time proceeds slowly, "playing" along near the shore and probably drifting passively downstream with the current for long stretches of the route. As a returning adult, the same fish swims strenuously upstream in deeper waters of the same river, often separated from its earlier downstream path along the shore by a distance that is greater than its visual range. Therefore, a possible alternative to visual memory is the assumption that the salmon has the ability to learn and to remember the chemical characteristics of both the water and the stream bed. This enables the animal to return to the stream of its nativity anywhere from two to six years after having left it.

Stimuli that serve as reference points for fishes

It has been demonstrated among migratory fishes that their chemical sense organs can detect small but probably constant differences in the chemical characteristics of water from different streams and ocean currents, including differences in salinity (see CHEMORECEPTION). Moreover, a learned olfactory preference for the waters of the home stream has been demonstrated experimentally in such fishes. Other cues for learning include differences in temperature, acidity, proportion of dissolved gases, density, and general turbulence. Added to these might be the recognition of characteristic sounds made by waterfalls and rapids, the memory of the general nature of the river bottom —which, of course, is partly visual—and perhaps the memory of the type of food obtained in the home stream. As for the route in deeper waters of the sea, there is the possibility that many fishes learn the contours of the bottom by making noises that echo back from the sea floor; or, for those fishes unable to make sounds, there is the possibility of learning contours through the resonant effects of surface wave noise on the bottom.

Translated into laboratory terms, the homing performance of a salmon is very much like learning to run a gigantic and complex maze in reverse after having gone through it only once and in the other direction two to six years before. Although the fish may not remember every detail of the hundreds of miles its journey may cover, it must be supposed that it remembers the characteristics of different sections, particularly those of the various choice points; *i.e.*, the junctions of the tributaries with the main rivers. Even given such assumptions in an attempt to simplify the problem faced by the returning fish, its performance is still an astonishing feat.

Birds. Orientation. Birds provide remarkable examples of long-distance orientation, particularly those species in which the young birds migrate to winter quarters before the adults do, thereby exhibiting an instinctive ability to fly in the appropriate direction. In these and other species the urge to migrate appears to be controlled by chemicals in the body called hormones. As a result of experiments in orientation cages with migratory birds, it has been established that many of them will flutter in the compass direction toward which they would be expected to fly at that time of year. Under these conditions such birds as starlings (*Sturnus* vulgaris) remain oriented (flutter in the correct direction) only as long as the sky and the sun are not obscured. If mirrors are used to change the apparent direction of the sun, the orientation of the birds changes accordingly. Although the position of the sun changes with the time of day, the direction of migration does not. Obviously, the bird must be able to correct for the movement of the sun by some sort of internal clock mechanism. This inner clock, which depends on the daily light–dark cycle, can be upset by providing an artificial cycle that is out of phase with the natural daily rhythms (see PERIODICITY, BIOLOGICAL).

Exactly how migrating birds use the sun is still somewhat obscure, although some species such as starlings can orient by using the horizontal (azimuthal) direction alone; others may be able to anticipate what the sun's position will be when it reaches its highest point in

The role of celestial bodies in long-distance orientation by birds

the sky. Because, in the Northern Hemisphere, this point is always directly south, it can serve as a fixed reference for the migrating animal. Naturally, nocturnal migrants have a different problem. When various warblers migrate in the autumn, they are able to maintain their direction at night even when they can only see the central part of the sky, but they become disoriented when the stars are hidden by clouds. In these cases it seems that the birds are responding to star patterns, especially those in the neighbourhood of the North Star. But in addition to the relatively simple problem of maintaining correct orientation, migratory flight also involves formidable sensory and perceptual problems to the bird which are not yet fully understood. Moreover, it has often been claimed that birds have a magnetic sense that serves as a compass, but the evidence for this is far from convincing.

In more complex orientational tasks, birds forcibly transported long distances from their nesting area during the breeding season are often able to return to the nest with such speed and reliability that random search for familiar landmarks seems unlikely to be the sole explanation. When Manx shearwaters (*Puffinus puffinus*), for example, were taken from their breeding burrows on the island of Skokholm off South Wales (Pembrokeshire) and transported in blacked-out boxes to Cambridge, England, where they were released, they covered the return journey to Skokholm (approximately 290 miles, or 470 kilometres) in as little as six hours. Because this species does not usually fly over land, most of the entire route must have been completely unfamiliar to them. Even more remarkable was a bird of the same species that was taken by airplane to Boston, Massachusetts. After being released, the bird returned to its Skokholm burrow in 13 days, having covered the journey of 3,050 miles (4,900 kilometres) at an average speed, assuming daylight flight only, of more than 20 miles (30 kilometres) per hour. In this case the bird could not possibly have been familiar with the east coast of North America and the western Atlantic Ocean because this area is well beyond its normal geographical range. In another of the many striking examples of experimentally relocated oceanic birds, the longest homing flight has been that of a laysan albatross (*Diomedea immutabilis*); it flew from the Philippines to Midway Island, a distance of 4,120 miles (6,630 kilometres), in 32 days. The fastest homing flight by a bird of this species was from Whidley Island, Washington: 3,200 miles (5,100 kilometres) in 10.1 days, or about 370 miles (600 kilometres) per day. In what is in some respects an even more astonishing feat, a much smaller bird, Leach's petrel, covered nearly 3,000 miles (5,000 kilometres) at an average rate of 217 miles (349 kilometres) per day.

Navigation. When birds are released in most experimental relocations, they set out immediately in the approximate direction of home. This would indicate that upon release the bird is already oriented toward home; it then performs as if it were fixing its present position on a flat map, calculating the course to steer, and effectively following it. This does not mean that the bird calculates in the same way as a human navigator. In fact, knowledge of the entire process is still so incomplete that it is not even yet known how the bird fixes its position on release. It has been established, however, that birds do not respond to forces resulting from the earth's magnetic field, nor is it likely that they respond to forces relative to the earth's rotation (Coriolis forces). Observation of the sky, on the other hand, seems important; many species are disoriented when the sky is overcast, and they "home" less effectively if they cannot see the horizon from the point of release than if the entire sky is visible.

Navigating by the sun

In contrast to the honeybee, birds do not have the ability to detect the pattern of polarised light reflected from the blue sky to aid them in their orientation flights. As a possible alternative, it has been suggested that day-flying birds orient themselves by observing the sun's arc at the location at which they are released and relating it to the characteristics of the sun's arc at home. For this system to be effective, the bird would have to observe the sun's movement over a small part of its arc, and then somehow compute the highest point of the arc to determine geographical south and local noon. Comparison of the remembered sun's noontime altitude at home with the observed noon altitude gives the difference in latitude, and comparison with the sun's noontime azimuth at home gives the difference in longitude. It seems more probable, however, that the difference in longitude is a direct time difference that is related to some kind of internal clock, the existence of which is now fairly well established. Astonishing though these conclusions may be, they seem to be inescapable on the basis of present information; in fact, this theory is simpler than many other hypotheses that have been advanced.

The hypothesis just described has been criticized on the grounds that it is improbable that a bird can make measurements of the required accuracy. Moreover, initial orientation sometimes appears to have occurred even before the sun has moved any appreciable distance along its arc. It seems, however, that a number of these objections are irrelevant. It has been shown, for example, that some species of birds have a tendency to orient in a particular direction, usually one that is characteristic of the stock or population, for a short while after release; during this time the position of the sun changes. Indeed, it has also been shown that pigeons can detect a movement as slow as that of the sun, as can crabs. There can be little doubt, therefore, that sun navigation occurs widely among birds, but exactly what is measured and with what degree of precision remain matters of dispute. However satisfactory the sun-orientation hypothesis may be, it does not account for the abilities of the many species that orient and navigate at night. Although birds can orient—reliably face an expected direction—in the night sky by using star patterns, there is no absolute evidence that they can actually navigate (*i.e.*, find their way about) by this method.

Nature of flight paths

The actual path followed by birds on homing flights has been studied in recent years by radio transmitters that are small enough to be attached to migrating birds without interfering with their flight abilities. The use of such tracking devices has revealed that the paths are often much straighter than investigators previously suspected. Still more extraordinary are the flight paths of some nocturnal spring migrants returning to their summer homes. In one such case a Swainson's thrush (Hylocichla *ustulata*) carrying a radio transmitter weighing only three grams (0.1 ounce) was tracked through the night by a receiver after it took off from Champaign, Illinois, until it landed eight hours later at its destination in Wisconsin, a point that was 450 miles (725 kilometres) away according to the map. The actual track flown measured only 453 miles (729 kilometres), a performance more remarkable than it may seem because the bird's slightly curved path was compensating for the curvature of the earth. Even if the bird had been flying a compass course, such accuracy would be the envy of a human navigator. In another case of a bird starting from Champaign, Illinois, a gray-cheeked thrush (Hylocichla minima) was tracked for 400 miles (600 kilometres) under a clear night sky as it flew at a speed of 50 miles (80 kilometres) per hour with the help of a 20-mile-per-hour (30-kilometre-per-hour) tailwind. After 140 miles (225 kilometres) its course took the bird over Lake Michigan; the plane following it, however, had to go overland for safety. After contact with the bird was renewed in Wisconsin, the plane had to turn back because of a thunderstorm; the thrush continued, despite the bad weather.

What do these results indicate in terms of known animal learning capacity? It must be accepted, of course, that birds have an initial innate ability to perceive and to coordinate readings from various stimuli in the environment, such as the rate of movement and altitude of the sun and possibly also of the stars. In the case of the thrushes, it also must be presumed that they had some visual knowledge of the environment and may have flown a similar route, at least in the opposite direction, a few months previously. With night fliers, however, it is doubtful that visual landmarks are of much help, particularly over a large body of water during a thunderstorm.

Aids to orientation and navigation. In homing experiments, in which a bird is transported in a closed box, it is apparent that the bird has no plan or map in its head based on previous experience to guide it. Although such birds clearly seem to have compass orientation to help them at the outset and landmarks near home toward the end of their flight, in between they appear to be truly navigating. With night fliers and often with day ones, celestial cues, whether from the stars or from the sun, frequently are lacking. Even when it seems possible that returning migrants might be orienting by recognizable landmarks, they often appear to disregard them as navigational aids. Thus, birds migrating over the Gulf of Mexico and its shores occasionally alter their direction, seemingly on the basis of geographic features, yet large rivers that course inland apparently are seldom employed as guides.

Use of winds

No doubt birds recognize many characteristics of the wind. Generally, they select those that are favourable for their goal; when forced to use unfavourable winds, however, they correct to a high degree for lateral drift. Migrants appear to select wind speeds within their flight capabilities; water fowl, for example, select wind speeds above 30 knots (nautical miles per hour) and passerine birds (*e.g.*, sparrows and robins) above ten knots. Because the birds appear to be able to determine the direction and strength of the wind, it is plausible that they may also use the wind as an orientation reference. It is known that winds blowing across terrain develop turbulence that migrating birds might be able to detect. It may well be that such turbulence serves as an additional means of orientation during migratory flights.

It is also important to consider other possibilities as aids to orientation and navigation. More needs to be known not only about the sensory physiology of birds but also about the atmospheric ecology of migration when neither sky nor ground is visible. It is possible, too, that birds might detect differential wind patterns experienced by upper and lower members of a flock during a flight and communicate this information to each other by certain calls. Many meteorological conditions may produce detectable updrafts that provide useful information about the direction of wind and wind patterns.

Oceanic aids

Over the oceans it is conceivable that birds may be able to detect local characteristics of air-temperature circulation, cloud formation, and the wave pattern of the water caused by the wind. Similarly, landfall on islands may be easier than supposed, because of the fact that the oceanic islands often have a characteristic cloud formation above them, clearly visible from 100 or more miles away. Moreover, because such islands in areas of steady winds may act as barriers that interrupt the wave pattern on the sea surface, these changes in pattern might be useful aids to landfall and navigation. It has also been found that even under cloud cover some homing pigeons can select the approximate homeward direction within a few minutes after release in unfamiliar territory. There is convincing evidence that the birds could not have seen the sun, suggesting that they oriented in some other fashion; perhaps their sensitivity to atmospheric pressure, which would help the birds to judge their altitude, could be a means by which they are able to determine the presence of updrafts or downdrafts and might provide directional guidance.

Finally, there is a structure (the pecten) in the avian eye that forms a large blind spot in the part of the visual field that usually views the sky. This structure might be useful for viewing the sun and perhaps for determining its altitude above the horizon. Because the pecten is heavily pigmented, it helps absorb the intense light of the sun; its rich blood flow may distribute the resulting heat, thereby avoiding any damage to the adjacent retina (the light-sensitive layer in the back wall of the eye).

Whatever the mechanisms birds employ for orientation and navigation, their performances on actual migratory and homing flights far surpass anything that has been revealed when their behaviour is subjected to laboratory experiments on latent learning, pattern recognition, perceptual organization, and exploratory learning. It is true, of course, that by simplifying the circumstances to which an animal is exposed in the laboratory beyond a certain point, it is impossible for that animal to show the real extent of its capabilities. To confine closely a highly organized animal may result in creating a moron. The principle of Ockham's razor, according to which the simplest explanation available is regarded as the most likely to be true, can be utterly misleading. Organisms often, perhaps one should say nearly always, turn out to be more complex than at first supposed.

Counting ability. The study of birds has also greatly enlarged man's understanding of the possibilities of abstraction and symbolization in animals. It is known that the human ability to symbolize in the sense of representing completely abstract or general ideas by words—which in themselves have nothing of the essential characteristics of the concepts that they denote—far exceeds in degree any similar capacity in animals below the level of the apes. It appears, nevertheless, that some animals, especially birds, are capable of some sort of numerical abstraction—*i.e.*, of behaving as if they understand number concepts. A raven (Corvus cornx), for example, learned to open a box that had the same number of spots on its lid as there were on a key card. Eventually, the bird was trained to distinguish among five groups of two, three, four. five, and six black spots on the lids of small boxes; the "key" was a group of objects lying on the ground in front of the boxes that were of the same number as the spots on the lids. The raven learned to raise only the lid that had the same number of spots as the "key" had objects. Despite efforts by the investigator to change various factors in a random manner from one experiment to the next—such as the positions of the five boxes and the key patterns as well as the size and form of the spots—the raven always chose the lid according to the only item that was not changed through all the experiments; *i.e.*, the number characteristic of the particular key pattern presented.

Sequential recall in birds

Birds also seem to have the ability to remember numbers of incidents in sequence, thus keeping in mind numbers presented successively in time, independent of rhythm or any other clue that might be helpful. Pigeons, for example, have been trained to eat only a specific number of grains out of many offered. They have also learned to eat only a specific number of peas rolled into a cup one after another at randomly varied intervals ranging from one to 30 seconds. Because the pigeon never sees more than one pea in the cup at a time, there is no visible clue for distinguishing the last "allowed" pea from rhe first "forbidden" one.

In another experiment birds have been taught to open the lids of boxes standing in a long row until a specific number of baits have been secured. Because the baits are arranged in the boxes in 20 or more different distributions that vary from one experiment to the next (some boxes were empty and some contained more than one bait), the number of lids to be opened constantly changed. Thus, a bird trained to take five baits (*e.g.*, peas) may have had to open any number of lids between one and seven in order to get the correct number of peas. Even more remarkable is the fact that birds can learn to master up to four problems of this kind at the same time; a jackdaw (Corvus *monedula*), for instance, learced to open black lids until it had secured two baits, green lids until it had three, red lids until it had four, and white lids until it had secured five.

Thinking unnamed numbers. Because all external clues seem to have been carefully excluded in the above experiment, some inner mechanism must have been responsible for the bird's stopping when the required number was reached. It would seem, therefore, as if the bird does some "inner marking"—*i.e.*, counting to itself—of the units it is acting upon, a possibility strengthened by the observation that supposed inward marks sometimes reveal themselves in external behaviour in the form of intention movements, such as nodding of the head once for the number 1, twice for the number 2, and so on. One investigator has called this phenomenon "thinking unnamed numbers."

A notable extension of these studies is an investigation of the number sense of a gray parrot (Psittacus *erithacus*); it was possible, after prolonged training, for the parrot to distinguish eight objects from seven objects. (Remarkably, it has not been demonstrated that any animal, including man, can immediately recognize the number of objects in a group beyond the point of about seven or eight.) The parrot learned to recognize that the successive presentation of a number of optic stimuli was a signal for a task involving the same number of actions. Thus, after having been shown four, six, or seven light flashes, the bird was able to take four, six, or seven pieces of irregularly distributed food out of a row of trays. Numerous random changes in the time sequence of the visual stimuli did not affect the percentage of correct solutions. After the bird learned this task, the signal of successive light flashes was replaced by successive notes on a flute. The parrot was able to substitute immediately without further training; the change from light flashes to flute notes had no effect on the number of correct solutions. The accomplishment was not hindered by completely arhythmic patterns or by a change of pitch. The parrot, however, was not able to accomplish a task that involved a combination of the two faculties of learning numbers; that is, it could not be trained to respond to visual numbers (*e.g.*, the number 5) when they were presented simultaneously with the same number of acoustic stimuli in series (*e.g.*, five beeps). Yet, when two different sounds were presented simultaneously (*e.g.*, a chord of two notes) or when a single note was presented, the parrot spontaneously opened a lid with two spots on it or a lid with one spot, according to the acoustic signal given. Although this work indicates that some species of birds have a counting ability in the form of two prelinguistic faculties—simultaneous and successive unnamed number sense—it is not true counting in the human sense.

Basic pre-linguistic faculties or "unnamed thinking"

In another area, however, the numerical ability of man and bird is approximately identical. If groups of objects are flashed on a screen long enough to be seen but too quickly to be counted, few persons will perceive as many as eight objects, and many perceive only five, which is the average for pigeons when given the same test. There thus seems to be no basis for expecting that man can achieve better results than birds if verbal counting (*i.e.*, counting by name) is excluded. The similarity in the limit of number perception in six bird species and man suggests that basic prelinguistic faculties, which may be described as "unnamed thinking," may be common to both birds and man. Because similar results have been achieved with squirrels and apes, it is clear that many birds and at least some subhuman mammals behave as if they can learn the concept of number even without being able to name numbers.

Mammals. The outstanding features of mammalian learning achievement are intimately related to the establishment and development of social life. Bodily contact—warmth, rocking, and body positioning, for example--enables the infant monkey to learn to identify its own mother from all other adults and to begin to understand the meaning of the mother's gestural communication. Although less is known about paternal affection—which expresses itself as protection against predators, protection of infants from aggression by others both without and within the social group, the social ordering of females and infants who are members of the in-group structure, and even a paternal interest in and play with infants—its effects appear to be important in the development of social behaviour.

Importance of play

Development of social behaviour. One important long-term socializing mechanism associated with early parental contact is the basic security and trust that it engenders. In all primates, including man, basic trust is often of crucial importance for the development of both exploratory and playful activity, particularly the latter. Play is the exercise of basically adaptive patterns of bodily action, either instinctive or learned or both combined, in circumstances where the normal objectives of such action (*e.g.*, hunting, food getting, aggressive and submissive display, etc.) are not sought. Thus the actions are incomplete in that they achieve no normal goal and therefore strike the observer as "not in earnest" (*e.g.*, the play of puppies, kittens, etc.). Play, particularly in young animals, leads to the development of motor skills and of perceptions, to the widening of perceptual horizons, and, thus, to the development of the exploratory drive. When placed in a secure environment, primate infants first explore and manipulate inanimate playthings and then turn their attention to social playmates, which are primarily other youngsters. Play automatically and inevitably results, both in man and monkeys, if the proper provisions are met—*i.e.*, associated playthings, agile and able playmates, and adequate protection. As the animals mature, play passes through increasingly complicated stages, stages that are obviously vastly more complex and more varied in humans than in other primates. One reason for this is the human capacity for speech.

Play appears to be the primary activity for successful social development in mammals. It involves, for example, the process whereby indiscriminate mingling with peers is transformed into heterosexuality. Homosexual play is practically universal in the early stages of life in mammals and in many birds. Among a great many mammals, and a few birds (*e.g.*, geese), such peer and playmate "love" patterns may continue into the heterosexual stage so that the animal behaves bisexually. The social–sexual bond uniting male and female members of a primate group may be profound and prolonged or perfunctory; there seems to be no consistent connection between the degree of evolutionary development and the length or intensity of heterosexual bonds. Indeed, among monkeys at least, the early peer-love relationship seems to be so important in the development of sexual behaviour in primates that its deprivation severely impairs adult heterosexuality.

Effects of social isolation. Social isolation, of which there are two qualitatively different types, is most devastating in its effect if instituted at birth before any affectional bonds have been formed. In studies of partial social isolation, individual rhesus monkeys have been placed in adjacent open cages so that they can see and hear but not touch each other. Total social isolation has been achieved by raising individual monkeys in enclosed cages in which all visual and manual contact with other animals is denied. Monkey infants subjected to partial social isolation exhibit excessive clinging and sucking behaviour. Moreover, locomotion and exploratory activity fail to develop normally; repetitive (stereotyped) behaviour, such as rocking, cage swinging, pacing, and circling, emerges instead. As such a monkey grows older, aggressive responses—both socially-directed and self-directed—develop, and eventually the frequency of both the abnormal and the normal patterns of locomotion and exploration decline. By the end of a decade, adult partially isolated monkeys do little but sit at the front of their cages and stare vacantly outward. If stressed by external stimuli, however, they may break into frantic frenzies of bizarre stereotyped activity or extreme self-aggression.

Sexual behaviour

Monkeys reared in partial social isolation for the entire first year of life show serious deficits in their social, sexual, and maternal behaviour. They rarely initiate interactions with others, and they become withdrawn and engage in disturbed irrelevant behaviour. If repeated social interactions are allowed after the first year, these monkeys ultimately display rudiments of play with other isolates; their sexual behaviour, however, is at best incompetent and often absent. Sexual drive may be evident (*e.g.*, through posturing and the rudiments of copulatory technique), but, even in the presence of a cooperative and experienced female, the behaviour of such males is little more than a caricature of normal sexual activity. Moreover, few female partial isolates become pregnant by natural means and, even then, very rarely by male partial isolates. If they do bear infants, they are hopelessly ineffective mothers, often indifferent or brutal, frequently to the point of killing the offspring.

By contrast, total social isolation only during the first three months of life merely leaves the monkey in a state of emotional shock. During this period the monkeys are denied both normal love by the mother and normal love for the mother. Yet, after three months of such isolation, the subjects make amazingly rapid and successful social adjustment to peers. When released to join cage mates at three months of age, they can still learn socialization because they have not yet reached the age at which most peer interactions normally begin.

If, after three months (from birth) of normal social living, monkeys are subjected to six months of total isolation, relatively little social impairment occurs, although those socially isolated in the last half of the first year tend to be unusually aggressive. This conforms to other data indicating that aggression begins to develop during the last quarter of the first year. Total social isolation during the first half year of life or longer, however, leaves monkeys almost completely unable to interact socially with age mates, impairs heterosexuality in all males and in most females, and eradicates all signs of normal maternal affection in females (which have to be impregnated by artificial means). These effects seem to be permanent; when such total isolates reach three or four years of age, they still show no positive interchange with either age mates or with monkeys half their age. Yet, despite this enormous social–sexual loss, there is no learning deficit as measured in the laboratory by standard learning test procedures, which have been designed for humans, not for monkeys. Even in monkeys totally isolated for nine to 12 months after birth, a procedure that permanently destroys sociability with age mates as well as all forms of sexuality and social communication, no "intellectual loss" has been detected.

Social psychotherapy among monkeys

When there is improvement in behaviour following the harmful effects of social isolation, it seems to come about by a process of social psychotherapy, in which some monkeys act as therapists and others as patients. It has been observed, for example, that the infants of isolate mothers continue to seek maternal attention and, unless actually killed by the mother, are surprisingly successful in forming maternal attachments. Moreover, after three to four months the brutal mothers, even though they still appear to ignore the infants, may begin to behave as if they enjoy the bodily contact with the infants and their sucking for milk. Surprisingly, most of these motherless-mother monkeys who subsequently have a second or third infant come to behave as normal or nearly normal mothers. Apparently, their first infants have psychiatrically rehabilitated them.

Because the first observations of this kind were more or less accidental, experimental psychotherapy was undertaken with a group of monkeys previously kept in total social isolation for six months (a period that had been found to be totally destructive socially if the isolates were then paired with age mates). In this experiment the previously isolated monkeys were intermixed with socially normal monkeys half the age of the former. The assumption was that the younger monkeys would be no threat to the isolates and could eventually re-educate them by initiating social contacts and play. The experiment was successful; gradually the isolate monkeys became progressively less fearful and more socially sophisticated, until they usually could not be distinguished from the normal monkeys.

Personality development. Mild deprivations among natural groupings (extended families) of rhesus monkeys in captivity yields more sophisticated and subtle information about monkey personality than can the more drastic techniques just described. By such means, monkeys can be used as substitutes for human social situations. In some experiments nothing more than an early and a very brief separation of the individual is employed, causing only mild behavioral disturbances. Even these can produce long-lasting personality defects which can be studied with a precision that is impossible with more violent techniques and, of course, is out of the question with humans.

The role of rhesus monkey mothers

Infant rhesus monkeys appear to learn to recognize their mother in the first week or two of life; perceptual and manipulative abilities continue to improve until there is some self-feeding at about three weeks and independent locomotion by about six weeks. The rhesus infants first leave their mothers very briefly when they are about one week old, after which each period of time spent away from the mother increases rapidly from the second to about the 20th week and again just before the end of the first year. During most of its first year the infant spends about 20 percent of the daylight hours awake and attached to the mother's nipple, yet only a small proportion of this time is actually spent sucking. Thus, the nipple serves as a way by which the infant can attach itself to its active mother; it is equally important as a pacifier. Such contact comfort supplied by the mother has been found to be the most important factor in the normal social development of the infant. Moreover, it appears that the increasing independence of the infant with age is the result of a change in the mother's behaviour rather than the infant's. The mother is not merely a guardian and passive provider of the infant's needs; she also plays an active part in promoting its independence. The infant spends more time away from the mother, investigating and exploring the environment, and playing with its peers. Because this play consists primarily of the approach–withdrawal and rough-and-tumble type, the younger infants gradually learn to avoid the playful attacks of their peers. Sex differences appear quite early; males not only initiate more rough-and-tumble play but also exhibit more social threat (*e.g.*, branch-shaking) and mounting behaviour from an early age. Under natural conditions mother and infant live in the complex social structure of a troop of monkeys. Other members in the troop display considerable interest in the infant: looking at, touching, grooming, or holding it; playing with it or behaving aggressively or sexually toward it. The mothers usually resent such attentions and protect their young either by restricting their movements, removing them, or threatening other monkeys.

The attentions of social companions may have a marked effect on the mother–infant relationship, as has been shown by a comparison between mother–infant pairs living in a social environment with mother–infant pairs living in social isolation. In the latter, the infants spent more time away from their mothers because the mothers of such pairs are more permissive. Indeed, in the isolate animals the infants were primarily responsible for maintaining proximity because, in the absence of peer playmates, the mother was the only social companion.

Monkey troop leadership

Effects of social organization. The social organization of the monkey troop also plays a predominant and basic role in controlling what the young animals learn and how they learn it. Females, for example, tend to remain in the centre of the group whereas, in the latter half of their second year, the males move toward the periphery. While the males are doing so, the mother–daughter relation continues, and the dependency on the mother acquired by the daughter in her infancy at the centre of the group seems to be preserved in adult life. As for the males, it is not always the most physically powerful who become leaders of a troop. In fact, a male monkey's chances of becoming a leader are vastly greater if he is born to a female in the troop's organization who is closely associated with the existing leader, even though the leader may not be the father. Because males, including the leader, share in caring for the infants, the children of favoured females, therefore, enjoy a privileged relationship with the leader that the children of other females do not have—a situation curiously reminiscent of patrician families in human societies.

In all monkey and ape societies that have been observed in the wild, clear differences in behaviour have been found from one group to another, even when genetic factors and the general environment are about the same. When Japanese investigators, for example, provided different groups with food, the habit of washing potatoes before eating them spread quite rapidly through one group but not through others. Such habits as swimming, readiness to eat candies, and the use of salt for seasoning provided further examples of new behaviour acquired by the same group that washed its potatoes. Moreover, it was also found that these habits were started by two three-year-old juveniles, after which the habits became popular with their brothers, sisters, and playmates. But, except for some mothers who learned the new habits from their own juveniles, the adults in general resisted adopting any of them. Evidently, it is through the younger ones that new behaviours become widely propagated, eventually being handed down from parent to infant. In fact, both monkey and ape societies show a kind of primitive culture that is acquired by individuals within the society and that constitutes a basic part of the process by which each member of the society develops its own distinctive personality.

Learning of sign language by chimpanzees

Development of "language." The observations described above make it clear that much information passes from one member to another in a troop of monkeys. It is also fairly clear, however, that vocalizations are not the major part of this process; rather, it seems that the chief mode of communication is by facial expression, gesture, posture, and bodily attitude. A number of attempts to teach chimpanzees to communicate information by the use of words have been complete failures. On the other hand, success has been achieved in teaching chimpanzees how to communicate by using a gesture language that is widely used by deaf people to communicate with each other. Two investigators made spectacular progress in understanding the communication of chimpanzees after five years of training a young female named Washoe to use many elements of the gesture language. A great advantage of this technique is that the animal's achievements can be compared directly with those of normal children in learning English and with those of deaf children of the same ages in learning this particular sign language.

After four years of experiment, Washoe had learned to use correctly more than 40 different signs for nouns (*e.g.*, bird, clothes, hammer). In addition, there were four signs of appeal (hurry, please, etc.), five for location (in, up, down, etc.), 24 indicating action (come, go, kiss, etc.), ten for agents or specific people (me, Roger, etc.), and ten for attributes (red, funny, sorry, etc.). Particularly interesting was her ability to use the pronoun "you" appropriately — that is, for any companion — and in combination with a wide variety of signs for actions and attributes. She began to use pronouns and proper names in the third year of teaching and later produced such sign language sentences as "you Roger Washoe out," "you me go out," and "you me go out hurry."

After four years of study, however, the chimpanzee seemed unable to use signs for such combining words as “and,” "for," “with,” or "to." It is significant, however, that words of this type are noticeably absent in the early sentences of young children. Hence, although Washoe’s achievements are not those of a fully articulate adult human being, her earliest two-sign combinations were comparable to the earliest two-word combinations of children.

Other experiments indicate that the linguistic achievements of chimpanzees may be even greater than the results with Washoe have shown. Instead of being taught gesture language, the animal (Sarah) in these experiments was supplied with movable tokens that represented words and that were treated by Sarah accordingly. Although the signs for words were presented to Sarah just as they were to Washoe, she did not have to learn to make the signs herself. Instead, she merely learned their significance and then moved the appropriate tokens on a board to indicate what she meant. In effect, Sarah learned to "write" sentences.

"Gentling." One other interesting phenomenon can with doubt be placed within the category of "learning." Nevertheless, it is significant in learning situations and, in fact, clearly plays a part in the organization of the monkey and ape societies described above. This phenomenon concerns the "gentling" of rats and its result on subsequent behaviour. When female rats are handled gently during their infancy and early life, they appear at maturity less emotional, more ready to leave their cages, and less fearful of strange stimuli than do other rats. In fact, the animal is able to respond more effectively when confronted with normal situations; in other words, its emotional response to novel but normal stimuli is not as intense as it is in rats raised in the usual manner. When such rats become mothers, the body weight and readiness to explore of their young are different from those of a standard controlled group. These effects appear to be mediated through both the prenatal mother–fetus and the postnatal mother–infant relationship. Not only are the young themselves of a different temperament as the result of the early experiences of their gentled mothers, but, when these young bear babies their patterns of mothering are also different from what they would have been normally. Because the effect of the gentling that their mothers received becomes apparent in their grandchildren, it has been called "the grandmother effect."

It appears, therefore, that experiences with both the inanimate and the social environment may affect subsequent fear, avoidance behaviour, and emotionality by a number of mechanisms. These effects in turn will have further ramifying consequences on nearly all other aspects of behaviour. Hence, when considering the effects of environmental manipulations on the development of behaviour, it is necessary to consider this chain of events.

BIBLIOGRAPHY

Standard textbooks related to the whole field: R.A. HINDE, *Animal Behaviour: A Synthesis of Ethology and Comparative Psychology*, 2nd ed. *(1970);* W.H. THORPE, *Learning and Instinct in Animals*, 2nd ed. *(1963)*.

Types of learning: G. HUMPHREY, *The Nature of Learning in Its Relation to the Living System (1933);* G. HORN and R.A. HINDE (eds.), *Short-Term Changes in Neural Activity and Behaviour (1970). (Associative learning)*: J. KONORSKI, *Conditioned Reflexes and Neuron Organization (1948). (Trial-and-error learning—operant* conditioning*)*: B.F. SKINNER, *The Behavior of Organisms: An Experimental Analysis (1938);* W.K. HONIG (ed.), *Operant Behavior* (1966), see especially the section by B.F. Skinner; C.B. FERSTER and B.F. SKINNER, *Schedules of Reinforcement (1957). (Latent learning)*: J. KONORSKI, “Mechanisms of Learning," in *Physiological Mechanisms in Animal Behaviour (1950);* W.H. THORPE, “The Concepts of Learning and Their Relation to Those of Instinct," *ibid. (Insight and insight learning):* W. KOHLER, *Intelligenzpriifungen an anthropoiden*, 2nd rev. ed. *(1921;* Eng. trans., *The Mentality of Apes, 1925). (Perceptual learning)*: W. SLUCKIN and E.A. SALZEN, “Imprinting and Perceptual Learning," Q. *Jl. Exp. Psychol.*, 13:65–77 *(1961);* W.H. THORPE, *Bird-Song: The Biology of Vocal Communication and Expression in Birds (1961);* R.A. HINDE (ed.), *Bird Vocalizations: Their Relation to Current Problems in Biology and Psychology (1969)*.

Behavioral and physiological aspects: W.K. ESTES *et al.*, *Modern Learning Theory (1954);* E.R. HILGARD and D.G. MARQUIS, *Conditioning and Learning*, 2nd ed. *(1961);* S. KOCH (ed.), *Psychology: A Study of a Science*, vol. *1–2 (1959). (Psychological and behavioural aspects of learning theory):* W.H. THORPE and O.L. ZANGWILL (eds.), *Current Problems in Animal Behaviour (1961). (Neurophysiological mechanisms of learning)*: T.H. BULLOCK, "Simple Systems for the Study of Learning Mechanisms," *Neurosci. Res. Prog. Bull.*, 4:105–233 *(1966);* E.R. KANDEL and W.A. SPENCER, "Cellular Neurophysiological Approaches in the Study of Learning," *Physiol. Rev.*, 48:65–134 *(1968);* G. HORN and R.A. HINDE (*op. cit.*); N.S. SCRIMSHAW and J.E. GORDON (eds.), *Malnutrition, Learning and Behavior (1968);* G.S. BRINDLEY, "Nerve Net Models of Plausible Size That Perform Many Simple Learning Tasks," *Proc. R. Soc.*, Series *B* 174:173–*191 (1969)*.

Evolution of learning ability: W.H. THORPE and D. DAVENPORT (eds.), *Conference on Learning and Associated Phenomena in Invertebrates: Report* (1965); W.G. CORNING and E. RICCIO, "The Planarian Controversy," in W.L. BYRNE (ed.), *Molecular. Approaches to Learning and Memory (1970). (Birds)*: G.V.T. MATTHEWS, *Bird Navigation*, 2nd ed. *(1968);* M.C. MICHENER and C. WALCOTT, “Analysis of Tracks of Single Homing Pigeons," *Proc. XIV Int. Orn. Congr.*, pp. *311–329 (1967);* D.R. GRIFFIN, "The Physiology and Geophysics of Bird Navigation," Q. *Rev. Biol.*, 44:255–276 *(1969). (Mammals):* R.M. YERKES, *Chimpanzees: A Laboratory Colony (1943);* R.A. and B.T. GARDNER, “Teaching Sign Language to a Chimpanzee,” *Science*, *165*:664–672 (1969); D PREMACK, “A Functional Analysis of Language,” *J. Exp. Analysis Behav.*, 14:107–125 (1970).

(W.H.T.)

Learning, Perceptual

Perceptual learning refers to the effects of random experience or practice in modifying the way in which an individual processes sensory data. Some restrict this to mean an activity that enhances ability to extract information from external stimuli, permitting more accurate awareness of the environment. This implies the conventional view that learning is confined to improving performance. Nevertheless, practice or ordinary experience also can decrease perceptual efficiency.

For example, human adults are more likely than children to overestimate the size of an object as its distance from the observer increases. Also, many instances of illusory perceiving may result from specific past experience, as in misinterpreting the face of a stranger to be that of an acquaintance. Considering such examples, the term justifiably may be applied to any relatively enduring perceptual modification attributed to learning.

Theories: discovery versus enrichment

Historically, the perceptual role of learning was a source of controversy. Vigorous denials that perceiving is influenced by learning are found in arguments of early Gestalt psychologists (*e.g.*, Max Wertheimer, 1880–1943, a German), By contract, heavy reliance is placed on learning processes in the writings of the German philosopher and scientist H.L.F. von Helmholtz (1821–94). Today, there is virtually full agreement that perceiving is modified by learning. Disputes now focus on the process of perceptual learning itself. Most theoretical alternatives reflect two underlying themes: discovery and enrichment. The discovery thesis is reflected in Eleanor J. Gibson's view that perceptual learning is a process of discovering how to transform previously overlooked potentials of sensory stimulation into effective information. Enrichment theories depict perceptual learning as enriching sensory experience with specific associations and with rules for its interpretation that derive from past experience. Discovery theories propose that perceptual modification results from learning to respond to new aspects of sensory stimuli, while enrichment theories hold that such modification results from learning to respond differently to the same sensory stimuli.

Direct confrontations of these positions are rare, their advocates tending to differ in their selection of experimental procedures and learning situations. It may be that discovery and enrichment theories are compatible, simply accounting for different forms of perceptual learning.

General acceptance of the perceptual role of learning should not be taken to endorse the claim that perceiving originally depends on learning. Indeed, studies of human newborn and very young infants indicate highly organized and stable perceptual functions. Learning is to be regarded as supplementary to unlearned factors that mediate perceiving.

EFFECTS OF PRACTICE

The most direct examination of perceptual learning is provided by investigating the effects of practice.

Detection of stimuli

In so-called detection tasks the observer is required to detect the presence or absence of a selected stimulus. For example, effects of practice on visual acuity were studied by requiring observers to detect simple orientation (left or right) in a row of leaning letters; *e.g.*, E E E E. Practice tended to lower acuity thresholds, defined as the lowest intensity of illumination at which each observer could detect the orientation. Or, observers were asked to say when they just could see that an approaching pair of parallel bars was double. With practice they continued to report seeing the narrow space between the bars at increasing distances. Such improvements suggest that sensitivity to simple (unidimensional) stimuli is not immutable, being modifiable through practice.

Improvement is not limited to simple variables. In one visual-search procedure, subjects scanned a long list of letters to find a single letter that appeared only once. Search time was reduced by a factor of ten following extensive practice, after which ten different letters could be detected as quickly as a single letter. Practice effects with complex targets also have been studied. In one experiment, two rows of figures were displayed on each trial, one with four simple outlines of geometrical figures, the other containing three complicated figures. Subjects were to guess or detect which one of the simple figures was concealed (embedded) in all three of the complex figures. Again, ability to identify the correct simple figure improved with practice.

Tasks involving absolute judgment require much more of the observer than does simple or complex detection. For example, he may be asked to estimate the diameters of circular targets numerically (*e.g.*, in inches or centimetres). In a similar study, two groups of subjects made absolute judgments of widely varying distances outdoors, both before and after interpolated activity. One group spent the interpolated period estimating a large number of other distances, none the same as in the original series. The other group spent the interval on unrelated paper-and-pencil work. In the first (extra-practice) group, judgments became more accurate and less variable than among the pencil-and-paper workers. Increased precision following practice also has been reported for absolute judgments of odour intensities, and of multidimensional visual (colour) and auditory stimuli. Improvement with practice is observed even when the subject remains uninformed of his accuracy; correcting him seems to confer slight benefit.

Many studies have failed to establish a clear basis for observed improvements in altered perceptual sensitivity or discriminability. For example, better performance on an acuity test may result from adopting a new criterion of visual doubleness or from learning how to use characteristics of blur to infer slant among leaning Es. Such uncertainties cloud the theoretical and practical significance of much available data.

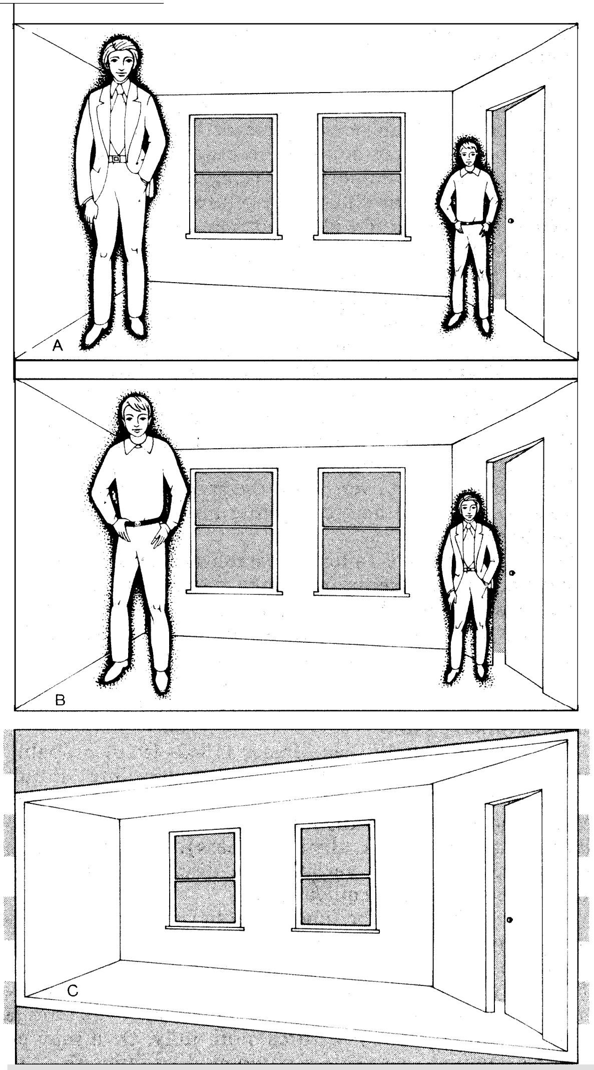
U.S. psychologist William James (1842–1910) probably introduced the notion that practice in labelling stimuli can alter their discriminability. Indeed, sometimes vague visual forms that are distinctively named are easier to discriminate (acquired distinctiveness). If several such stimuli have the same verbal label, discriminability may be reduced (acquired equivalence).

Labelling effects in the laboratory have been discouragingly fragile, however, and factors that favour them are poorly understood. Perhaps labelling affects one's efforts to discover distinguishing characteristics of stimuli. Having him learn distinctive labels may encourage him to analyze sensory features more fully. Or it may be that he begins to perceive a compound stimulus that includes the visual form and its associated label. If labels differ, the presumed compound stimuli are different, and discrimination should be enhanced. These hypotheses express both the discovery and enrichment theses.

EFFECTS OF PERCEPTUAL ASSUMPTIONS

According to one version of the enrichment thesis, exposure to recurrent regularities among stimuli prompts one to assume specific relationships between the environment and his sensory experience. For example, one learns that a continuous sequence of projective transformation (*e.g.*, the circular profile of a dinner plate seems to become eliiptical) is associated with changing positions of the object in view, or that continuous symmetrical expansion of the retinal image is associated with approach. In addition, one presumably learns to make assumptions about what is called reality; *e.g.*, despite alterations in retinal image, one perceives the plate to stay the same size. Psychologists Adelbert Ames, Jr., and Egon Brunswik proposed that one perceives under the strong influence of his learned assumptions and inferences, these providing a context for evaluating sensory data (inputs). In keeping with enrichment theory, Brunswik and Ames contended that sensory stimuli alone inherently lack some of the information needed for mature, adaptive perceiving; enrichment was held necessary to reduce ambiguity.

Much of the evidence for the contention that all perceiving is modified by one's assumptions comes from investigations in which most of the visual, everyday stimuli are eliminated. Often, the subject may view an isolated target in total darkness or look at a motionless display while keeping his head steady. To show that learned assumptions about physical size affect perceived distance, the observer may be asked to judge how far he is from a rectangle of light displayed against total darkness. He is



Perception modified by learned assumptions in a distorted room. The same men change places in A and B in a room modelled after C.

told at one time that the rectangle is a calling card; at another it is called a business envelope. His assumptions about these objects in relation to the size of his retinal image are invoked as prompting him to say that the "envelope" looks more distant than does the "calling card." Dramatic examples of this effect were invented by Ames, including his famous distorted room (see the Figure).

Ames held that perceiving under unusual conditions (*e.g.*, in a dark room) follows the same principles that govern more ordinary experience. The special conditions are said to permit experimental scrutiny of the same processes that are so difficult to examine under ordinary, uncontrolled conditions.

An opposing view is that such perceptual assumptions and inferences operate *only* under specific experimental conditions. It is asserted that only when commonly available sources of information are eliminated is the subject forced to rely on assumptions.

In the tradition of Helmholtz, Ames and Brunswik seemed to liken perceiving to reasoning, although not as a conscious process. They held that perceptual assumptions, once established, are influenced only slightly by logic. Although the floor and ceiling of the distorted room are sloped and all windows are of different size, it projects the same retinal pattern as a normal room; and a naïve subject will report that he sees an ordinary room. But even after he explores the room he remains likely to say it looks rectangular as before, despite his new information. Comparable observations have been reported for a variety of situations. Familiarization or instruction seems to have little effect on long-established perceptual assumptions.

Reasoning and perceiving

INFORMATION DISCREPANCY

Striking examples of perceptual learning are observed when one receives sensory data that contradict earlier experiences. For example, spectacles containing a wedge prism will bend light rays to displace images on the retina. An object thus will be seen as if it were somewhere other than its ordinarily perceived position. The subject's initial attempts to touch the target will be misdirected, and there is a discrepancy between its location as seen and as felt. A right-angle prism will tilt the visual scene to any desired degree, altering the customary direction in which retinal images move. Usually, images of stationary objects move parallel to the direction of head movement; now their motion is at an angle to the head's path.

However, if an observer wears such eyeglasses for an extended period, objects no longer seem displaced, nor does the scene continue to appear tilted. The observer has adapted to the prismatic distortions and comes to perceive the environment as he did pre-experimentally. Similarly adaptation to the perceptual aftereffects rapidly occurs after the prism is removed in such experiments.

Adaptation

Adaptation may be interpreted as perceptual learning that results from exposure to discrepancy. People who wear prism spectacles during active, self-initiated movement tend to show a greater degree of adaptation than do those who sit still or who are moved passively. Apparently conditions that heighten exposure to discrepancies facilitate adaptation. It seems likely that adaptation reflects a learning process during which the perceiver re-evaluates one or more sources of sensory information to reduce his experience of discrepancy. For example, information generated by receptors that respond to tension in skeletal muscles may be re-evaluated to resolve a discrepancy between felt and seen position.

It often is suggested that adaptation to prism eyeglasses may involve the same processes that serve perceptual development in infants. Indeed, some conditions that experimentally facilitate adaptation to prism distortion also seem necessary for everyday perceptual development (*e.g.*, active, self-initiated movement). In work reported by Richard Held (Scientific American, November 1965), actively moving kittens developed visually guided movements normally. When each of these was yoked to a littermate that was pulled passively over the same path, the passive partner failed to develop normal perceptual function. Yet both kittens apparently received identical visual stimuli.

The effects of learning on perceiving are varied. Most of these involve learning to respond to new stimuli or to make new responses to old stimuli. In one case, learning consists of differentiating previously neglected stimulus characteristics; in the other, it is a matter of re-evaluating stimuli and learning to respond to them differently.

BIBLIOGRAPHY. Descriptive reviews of the literature on perceptual learning have appeared at intervals in the *Annual Review of Psychology*, particularly the articles by J.D. DREVER (1960); E.J. GIBSON (1963); and J.F. WOHLWILL (1966). A more extensive review and evaluation of the diverse phenomena and theoretical formulations may be found in W. EPSTEIN, *Varieties of Perceptual Learning* (1967). A major publication is E.J. GIBSON, *Principles of Perceptual Learning and Development* (1969), an examination of the field from the viewpoint of the discovery theory of perceptual learning. This work includes an effort to show the relevance of principles of perceptual learning for a developmental theory of perception.

(W.E.)

Learning, Psychomotor

Human psychomotor skills are organized patterns of muscular activities guided by changing signals from the environment. Driving a car and eye–hand coordination tasks such as drilling a tooth, throwing a ball, typing, operating a lathe, and playing a trombone are behavioral examples. Also called sensorimotor and perceptual-motor skills, they are studied as special topics in the experimental psychology of human learning and performance. In research concerning psychomotor skills, particular attention is given to the learning of coordinated activity of the arms, hands, fingers, and feet; the role of verbal processes (see CONCEPT FORMATION) is not emphasized.

Whenever a person handles knife and fork, signs his name, or taps his foot to music, he is engaging in psychomotor behaviour; and these homely actions are no less perceptual and motor than the sophisticated behaviour of flying an airplane, running a printing press, executing a backhand stroke in tennis, or turning out a piece of ceramics. Some sensorimotor skills rest on eye–foot coordinations (*e.g.*, kicking a ball) or hand–foot coordinations (*e.g.,* playing an organ); others call for complex integrations of many parts of the body (*e.g.*, circus acrobatics, gymnastic routines). Arts and crafts, as well as scientific and technical abilities, require the skillful performance of motor acts under the guidance of sensory processes monitored by the brain.

Factors in performance

Although psychomotor skills are widely distributed—*e.g.,* in military, athletic, musical, and industrial settings—such complex situations typically do not lend themselves to rigorous experimental research. Most scientists have found it more analytically useful to study psychomotor learning under controlled laboratory conditions. Measures of proficiency obtained in the laboratory reflect increasing accuracy and decreasing variability in a learner's performance as training progresses. If there is sufficient genetic aptitude, a person's mastery of a skill depends on his motivation to improve, on his receiving continuous information or sensory feedback about the adequacy of his performance during training, and on such factors as the rewarding effects of corrections made during successive practice periods. Skills are susceptible to inhibitory influences. The full extent of gains in proficiency often is masked by temporary losses and emerges only later, without additional practice sessions.

Psychomotor habits are mediated primarily by the sensory and motor cortex of the brain and by the neural fibres (commissures) that connect the two cerebral hemispheres. According to the majority of theoreticians, learning proceeds (habit strength develops) as a mathematical function of the amount or duration of rewarded (reinforced) practice (see LEARNING THEORIES). The effects of associative and motivational factors are believed to combine mathematically by multiplying one another, while inhibitory and oscillation (variability) factors are thought to have subtractive effects. Despite theoretical and empirical progress, much remains to be discovered about the learning and performance of psychomotor skills, especially about the interrelationships among training variables, feedback contingencies, and human-factor variables.

LABORATORY RESEARCH IN PSYCHOMOTOR LEARNING

The pursuitmeter and the mathometer

Devices and tasks. Hundreds of electrical and mechanical instruments have been developed for research in psychomotor learning, but those commonly used number less than two dozen. In operating a device called a complex coordinator, the learner is instructed to make prompt, synchronized adjustments of handstick and foot-bar controls to match different combinations of stimulus lights. Another device, a discrimination reaction timer, requires that one of several toggle switches be snapped rapidly in response to designated distinctive spatial patterns of coloured signal lamps. In performing on a manual lever, a blindfolded subject must learn how far to move the handle on the basis of numerical information provided by the experimenter. With a so-called mirror tracer, a six-pointed star pattern is followed with an electrical stylus as accurately and quickly as possible, the learner being guided visually only by a mirror image. The operator of an instrument called a multidimensional pursuitmeter is required to scan four dials and to keep their indicators steady by making corrections with four controls of the type found in an airplane cockpit. On a rotary pursuitmeter the trainee's task is to hold a flexible stylus in continuous electrical contact with a small, circular metal target set into a revolving turntable.

Also employed in such research is a selective mathometer, on which the subject's problem is to discover, with cues provided by a signal lamp, which of 19 pushbuttons should be pressed in response to each of a series of distinctive images projected on a screen. While using a star discrimeter, a person receives information about his errors through earphones; his task is to learn to selectively position one lever among six radial slots in accordance with signals from differently coloured stimulus lights. A trainee on a two-hand coordinator has to manipulate two lathe crank handles synchronously to maintain contact with a target disk as it moves through an irregular course.

Measurements. The tasks required by the above devices produce a substantial range of psychomotor difficulty. The elements of skilled behaviour are expressed as numerical scores; *e.g.,* correct response and error percentages, amplitude and speed of movement, hand or foot pressures exerted, time on target, reaction time, rate of response, and indices of time-sharing activity. Most of the behaviour thus recorded lends itself readily to mathematical treatment. Laboratory devices for studying psychomotor learning characteristically exhibit high reliability (*i.e.*, intra-task consistency) and yield scores of useful validity (extra-task correlation) in predicting such behaviour as performance in factory work and the operation of motor vehicles and aircraft. In other words, it would appear that perceptual-motor devices reliably measure what they are designed to measure, and they also tap a significant proportion of the abilities required in real-life situations (see also MEASUREMENT, PSYCHOLOGICAL). When properly maintained and used under standardized conditions, and when the resulting measurements are treated by statistical methods, the above devices are prime choices for many applied and basic research programs.

PHENOMENA OF PSYCHOMOTOR LEARNING

Acquisition. Speed and accuracy in the majority of psychomotor tasks studied are typically acquired very rapidly during the early stages of reinforced practice, the average rate of gain tending to drop off as the number of trials or training time increases (Figure 1). Curves based on such measures as reaction time or errors reflect the learner's improvement by a series of decreasing scores, giving an inverted picture of Figure 1. Tracking scores from the two sexes are seen in Figure 1. Other devices have yielded more complicated functions—*e.g.*, S-shaped curves for complex multiple-choice problems on the selective mathometer (Figure 2). Most acquisition curves (whether the dependent response variable measured is positively or negatively related to the independent practice variable) obey a law of diminishing returns as high levels of skill are approached. Data such as those from tracking and multiple-choice tasks can be explained by rational mathematical equations derived from theoretical models (see formulas and captions in Figures 1 and 2). Between them, these two equations describe psychomotor acquisition curves from a wide variety of learning situa-

From Clyde E. Noble, "Acquisition of Pursuit Tracking Skill Under Extended Training as a Joint Function of Sex and Initial Ability," *Journal of Experimental Psychology*, vol. 86, pp. 360–373 (1970); Copyright (1970) by the American Psychological Association, and reproduced by permission

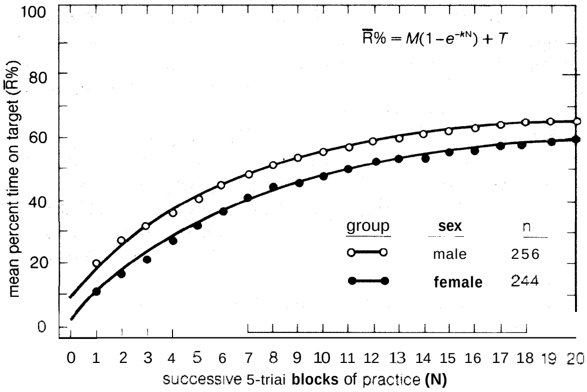


Figure 1: Rotary pursuit acquisition curves.
Average percent of response time on target ($\bar{R}\%$) while using a rotary pursuitmeter is plotted as a mathematical function of the number of successive 5-trial blocks of practice (N), with sex as the distinguishing factor, for 500 subjects. Both curves are predicted by the exponential equation shown in the figure, where T, k, and *M* represent origin, rate, and limit of theoretical response time, respectively.

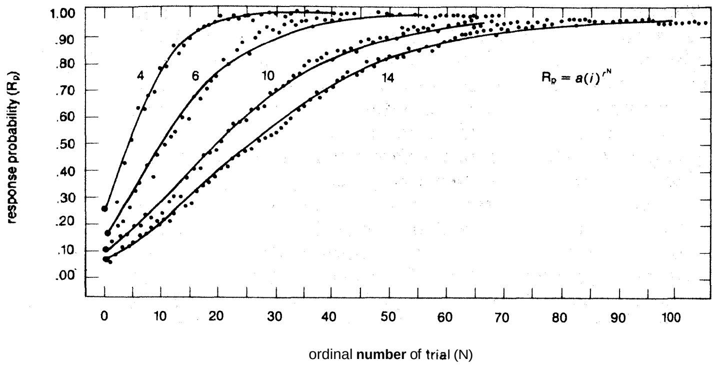


Figure 2: Selective mathometer acquisition curves. Response probability (R_p), or relative frequency of correct multiple choices, in a selective mathometer learning experiment is plotted as a mathematical function of the number of successive practice trials (N), with task length (4,6,10, or 14 units) as the experimental condition, for 192 male subjects. All four curves are predicted by the double-exponential equation shown in the figure, where I,r, and a represent origin, rate, and limit of theoretical response probability, respectively.

From Ciyde E. Noble, "The Length–Difficulty Relationship in Compound Trial-and-Error Learning" Journal of *Experimental Psychology.* vol. 54 (1957), pp. 246258 (1957); Copyright (1957) by the American Psychological Association, and reproduced by permission

tions and of trainees with less than a 2 percent average error of prediction. Contrary to lay opinion, stepwise plateaus of proficiency are seldom seen, not even in learning Morse code. The "natural plateau" is a phantom.

Similarity and performance

Generalization and transfer. The occurrence of the phenomenon of generalization is seen in the tendency of laboratory subjects (conditioned to respond to a particular stimulus—*e.g.*, a light) to respond as well to similar stimuli beyond the original conditions of training. As differences along a physical continuum (*e.g.*, brightness) between the stimuli used in training and those encountered on test trials increase, the effects of generalization decrease until there may be no transfer from one situation to another. Alternatively, the more the two situations have in common, the greater is the amount of predictable transfer. Generalization (or transfer) may be based on temporal patterns of stimuli (*e.g.*, rhythms), spatial cues (*e.g.*, triangularity), or other physical characteristics (see TRANSFER OF TRAINING).

The measured effects of prior training on the performance of a subsequent task define the transfer of psychomotor learning. Although similar, the latter task usually differs measurably from that originally practiced. A common example is the ability required of many automobile drivers to change easily from, say, a three-speed transmission with a horizontal gear lever on the steering wheel to a four-speed mechanism with a vertical floor-mounted gearshift. In laboratory tasks the amount and direction of transfer effects are accurately predicted. In practical skills, transfer is more likely between tennis and badminton than between swimming and football, between comet and trumpet than between piano and tuba. Similarity is not the only correlate of transfer, however, and empirical studies of these effects must take account of such factors as the amount of practice and the sequence of events in previous training.

Transfer effects may be positive, negative, or zero; *i.e.*, learning one task may facilitate, hinder, or have no observable influence upon performance of the next task. Flight simulators are designed to maximize the amount of positive transfer, often by ensuring high levels of behavioral similarity. Negative transfer effects (*e.g.*, reaching for the floor to shift gears when the lever is on the steering wheel) appear occasionally but tend to be rare, transitory, and easily overcome. Since transfer necessarily involves retention, the best schedules minimize forgetting by the inclusion of short time intervals between training and transfer.

The degree and amount of transfer are contingent upon such factors as number of common elements or principles, stimulus and response similarity, amount of predifferentiation training, the variety of learning-to-learn experiences, part–whole relationships, differences in intertask complexity, use of mnemonic aids, and the extent of proactive or retroactive interference, Transfer equations usually assume that the basic indices of performance for experimental and control groups will increase with practice, that the possible measures range from negative 100 percent through zero to positive 100 percent, and that the groups have been equated in aptitude or initial ability to learn before the experimental treatments are begun. Retroactive interference designs typically employ a sequence of original learning, interpolated learning, and relearning.

Effects of rest intervals

Retention. Learning is to acquisition as memory is to retention. Psychomotor retention scores indicate the percentage or degree of originally learned skill that is remembered or recalled as a function of elapsed time. Alterations of motor memory are reflected by changes in means, variances, and correlations between test results. In contrast to verbal behaviour, which is notoriously susceptible to forgetting through interference within a matter of seconds, mean scores for tracking and coordination skills recorded over periods ranging from two days to two years diminish scarcely at all. Yet, when intervals of three minutes to six weeks are interpolated between discrete responses on a manual lever device, performance remains stable for about two days and then becomes inconsistent; variabilities increase and correlations decrease as the subjects mis-recall more and more of their original skill. In the light of this evidence, motor memory may be viewed as a phenomenon of persistence, while forgetting is a case of inconsistence.

One hypothesis advanced to account for the greater retentivity of psychomotor behaviour, as compared to that of newly acquired verbal behaviour, is that nonverbal striped-muscle responses are more often overlearned and are less susceptible to proactive interference (*i.e.*, competition arising from things learned in the past). Distinctions between immediate, short-term, and long-term memory are also less prominent in studies of motor learning, possibly because of the devotion of skills specialists to efficient practice and feedback methods that ensure permanent storage of habits in the brain. This is not to say that motor skills are unforgettable; studies of short-term memory suggest that psychomotor forgetting can be swift indeed. Regardless of theoretical differences, however, psychologists generally agree that psychomotor behaviour is best remembered (and least forgotten) when overlearning is high, interference is low, reinforcing feedback is optimal, and interpolated activities are unrelated to the task being learned. Time is less important in the degradation of memory than are the events that fill the time (see also MEMORY: RETENTION AND FORGETTING).

Reminiscence. The phenomenon of reminiscence is a gain in performance without practice. Thus, when subjects performing trial after trial without rest (massed practice) are given a short break, perhaps midway through training, scores on the very next trial will show a significant improvement when compared with those of a massed group given no break. Reminiscence effects are most prominent in tasks demanding continuous attending and responding; they are least often observed with discrete-responding apparatus. The theoretical importance of this concept derives from its role in testing a hypothesis of reactive inhibition that asserts that a decremental process cumulates in the organism as a positive function of responding to stimulation and a negative function of resting time. That the phenomenon of reminiscence also manifests bilateral transfer of skill (*e.g.*, from the left to the right hand) suggests that the locus of the decrement is in the central nervous system rather than in the peripheral effector organs. Indeed, merely watching another subject practicing on a rotary pursuitmeter has an inhibitory effect on a person's performance; yet the cause is neither boredom nor fatigue.

Warm-up. Athletes and musicians often report that they get "cold" during a layoff (even for a rest period of a mere five minutes); when practice is resumed, the decrement in performance requires a warm-up before it is overcome. Similarly, on a rotary pursuitmeter, it is necessary to regain the optimal posture, grip the stylus correctly, begin the coordinated movements of eyes and hand, and recapture the proper whole-body rhythm. Warm-up produces a further gain in proficiency following the initial reminiscence effect. Mean scores continue to rise for several trials, reach a peak at the level found for distributed practice, and then fall more gradually until they merge with the curve for massed practice. When the duration of rest is extended, the amount of warm-up decrement first increases rapidly and then decreases; similar findings obtain for a succession of work and rest periods. Investigators who have tried to substitute warm-up activities other than actual pursuitmeter practice to offset or reduce the magnitude of the decrement have not been successful. At least for continuous psychomotor tasks of this sort, the need for proper, task-specific warm-up appears to be an intrinsic requirement of efficient performance. Wherever reminiscence goes, warm-up seems to follow; yet the converse does not always hold. The connection between warm-up and forgetting is uncertain.

Warm-up as a practice factor

Refractory period and anticipation. When required to make quick, discrete responses to two stimuli separated in time by one-half second or less, an operator's reaction time (latency) for executing the second response is typically longer than that of his first response. This difference in reaction time is called the psychological refractory period. At one time, it was thought possible that sensory feedback from the first response might stack up in the nerve centres to make the system refractory for a brief time, thereby delaying the processing of the second stimulus. Research findings that erroneous reactions could be corrected within one-tenth second would seem to negate the hypothesis. An alternative suggestion is that corrective movements are facilitated by feedback from the incorrect ones, and controlled observations appear to confirm that error-correcting responses have shorter latencies than those that are either correct or erroneous. Apparently, a false movement can be stopped on the basis of internal cues more promptly than on that of external stimuli.

Expectancy is a collateral factor with which researchers have had to reckon; *i.e.*, a subject may learn to accommodate himself to expect a delay between the first and second stimulus and thus be relatively unprepared should the second arrive earlier than usual. Further, people learn to be more expectant for particular kinds of stimuli than for others. When a person is uncertain about whether regularly occurring stimuli will be auditory or visual, or when their spatial direction is uncertain, performance is significantly degraded. This would suggest the possibility of divided attention; indeed, when pairs of stimuli are made perfectly predictable as to time and type, no impairment of response is observed.

Learning to divide attention

If a subject can acquire suitable expectancies via training and experience, then he can improve the skill of dividing his attention and, within physiological limits, simultaneously handle an increased range of stimuli without loss of proficiency. Results from extended practice on a task requiring successive choice and dual reaction indicate that. with learning, people can reduce the psychological refractory period. The ability to develop anticipatory responses to regularly occurring stimulus cues is well established. A military gunner scanning a distant fixed target for azimuth and elevation, for example, is engaging in a preview of receptor anticipation to maximize his score. An operatic tenor who rehearses covertly the opening notes of his cadenza while the orchestra finishes the introduction is employing perceptual anticipation to optimize his rendition. Anticipatory timing is learned, and reinforcing feedback is necessary.

FACTORS AFFECTING PSYCHOMOTOR SKILL

Amount of practice. It has been noted above (Figure 1) that the practice of sensorimotor tasks usually produces changes in scores that reflect diminishing returns. A major influence in learning generally, repetition is the most powerful experimental variable known in psychomotor-skills research. But practice alone does not make perfect; psychological feedback is also necessary. The consensus among theoreticians is that feedback must be relevant and reinforcing to effect permanent increments of habit strength. Once developed, habit never dies; it does not even fade away.

The effects of feedback and four other important performance variables (*i.e.*, task complexity, work distribution, motive-incentive conditions, and environmental factors) remain to be summarized.

Psychological feedback. Ranking prominently among experimental variables are so-called feedback contingencies (aftereffects, knowledge of results) that may be controlled by the experimenter so as to occur concurrently with or soon after a subject's response. A learner appears to improve by knowing the discrepancy between a response he has made and the response required of him; but, in experimental practice, the investigator manipulates behaviour by transforming functions of error. Since transformations are usually numerical or spatial, sensory returns from one's action may be informative, motivating, or reinforcing. Response-produced stimulation is intrinsic to most skeletal–muscular circuits; the neural consequences of bodily movement are fed back into the central nervous system to serve the organism's regulatory and adaptive functions. When this normal feedback is interrupted or delayed, psychomotor skill is often seriously degraded. Experimentally delayed auditory feedback of a subject's oral reading produces stuttering and other speech problems; delayed visual feedback in simulated automobile steering is a greater hazard under emergency conditions than is the driver's reaction time.

Generalizations about psychomotor learning

Laboratory investigations have supported the following generalizations about psychomotor learning: (1) without some kind of relevant feedback, there is no acquisition of skill; (2) progressive gains in proficiency occur in the presence of relevant feedback; (3) performance is disrupted when relevant feedback is withdrawn; (4) delayed feedback in continuous (but not discrete) tasks is typically decremental; (5) augmented or supplementary feedback usually results in increments; (6) the higher the relative frequency of reinforcing feedback, the greater is the facilitation of skill; and (7) the more specific the feedback (*e.g.*, in designating location, direction, amount), the better is the performance.

Experiments with a manual lever device, for example, suggest that when feedback is introduced and withdrawn at four stages of practice, the effect on error scores is profound. Knowledge of results given early and late has effects similar enough to reject any hypothesis that learning arises merely from repetition. These experiments indicate that practice makes perfect only if reinforced; the result of unreinforced practice is extinction of the correct response and a proliferation of errors. Studies employing a complex mirror-tracking apparatus have clarified the role of reinforcing feedback. Targeting performance was facilitated by presenting distinctive supplementary visual feedback cues previously associated with aversive (electrical shock) and nonaversive consequences. Moreover, the amount of facilitation grew curvilinearly with the number of cue conditioning trials. Work on human incentive learning thus demonstrates that the rate of gain in psychomotor proficiency can be regulated by stimuli that have been accompanied by positive or negative aftereffects. Persistence of the acquired reinforcing effects, considered with their cumulative quantitative properties, enhances the attractiveness of theoretical interpretations that emphasize continuity and reinforcement as contrasted with theories based on discontinuity and contiguity alone. Clark Hull's system (1943) is the classic model.

Human incentive learning

Task complexity. The complexity of discrete psychomotor tasks may be specified either as the number of response sequences a subject can make or as some measure of a subject's uncertainty about choices among stimuli. Still other factors that have been investigated as instances of complexity include variations in the number of possible responses at each choice point, different lengths of series, and regular versus unpredictable stimulus sequences.

Experimental procedures involving an increase of complexity produce more errors, require more trials to reach proficiency, and result in longer latencies per trial. Difficulty in psychomotor learning, therefore, generally increases with the complexity of the task to be mastered. An example of this phenomenon appears in Figure 2. Subjects exhibit continually altered probabilities of response during training sessions, and an average person with enough practice on a discrete sensorimotor task can learn to perceive, select, and react as fast to ten stimuli as he can to two. Apparently, it is not the number of choices among stimuli as much as it is the number of choices among responses that slows up a subject's processing activities and complicates his decision problems. Indeed, by limiting response alternatives (*e.g.*, circumscribing the physical range of a trainee's movements or providing supplementary auditory and visual indicators of error), a training device can facilitate the acquisition or transfer of skill.

Work distribution. Hazardous though generalizations about work and rest in psychomotor learning may be, a few guiding principles are notable: (1) massed practice is usually superior to distributed practice for simple discrete-trial tasks; (2) distributed practice is usually superior for complex continuous-action tasks; (3) short practice sessions are generally superior to long practice sessions; (4) long rest periods are generally superior to short rest periods, although forgetting must be counteracted; (5) for continuous-tracking tasks practiced under constant work sessions and variable rest periods, the final proficiency level grows curvilinearly as the intertrial interval is lengthened; (6) gains in proficiency under distributed practice, or with interpolated rest periods during massed practice, are usually in terms of performance rather than of learning; (7) losses in proficiency under massed practice, or with increased work load, usually pertain to inhibitory rather than motivational decrements; (8) under certain conditions (*e.g.*, "cramming" for examinations) it may be most efficient to mass practice as long as adequate rest can be obtained before criterion performance is demanded; (9) reminiscence increments and warm-up decrements are intimately related to schedules of work and rest; (10) decrement is not the same as fatigue.

Quite apart from the practical question of the optimal management of training programs (*e.g.*, in coaching oarsmen in racing shells), the aversive inhibitory consequences of sustained action that are recognized as subjective fatigue and behavioral decrement are clearly adaptive. By a reflex negative-feedback mechanism, inhibitory impulses may prevent an organism from working itself to exhaustion. With few exceptions, the presumption in favour of spaced practice can safely be taken out of the psychomotor-skills laboratory and applied in the gymnasium, lake, and playing field. Research on the skills involved in, for example, archery, badminton, basketball, golf, javelin throwing, juggling, marksmanship, rowing, and tennis supports the notion of distributing training by means of short workouts and frequent breaks (see also FATIGUE).

Distributed practice applications

Motive-incentive conditions. Motivational processes are states of the organism that serve to activate reaction tendencies. Such states are classified as primary (innate) or secondary (acquired, learned) motivation. Though common physiological needs (*e.g.*, for food, water, avoidance of pain) may evoke psychological drives (*e.g.*, hunger, thirst, pain), the concepts of need and drive are not perfectly correlated. Some needs (*e.g.*, oxygen demand) seem to have no specific behavioral drive, and for some drives, clear-cut biological needs remain to be identified (*e.g.*, curiosity). Despite this apparent discrepancy, there is a theoretical consensus that psychological drive arouses the body to action, energizes its latent responses, and supports its behaviour over time. Most theorists believe that motivation (drive) and learning (habit) interact (in a multiplicative — drive times habit equals action — manner) in generating response. In other words, to produce action both are theoretically indispensable, but neither is sufficient alone. A person is not likely to perform a skill if he does not want to and cannot do so if he does not know what to do. The multiplicative theory implies that the same level of psychomotor proficiency may arise from quite different combinations of learning and motivation. Moreover, the organism's temporary drive state seems clearly to affect the adequacy of reinforcing feedback (*e.g.*, offers of monetary reward do little to arouse one who is already trying his level best). While these theoretical interpretations often apply well to laboratory animals, their application to human acquisition of skill is complicated because incentive learning in man can become very abstract.

Physiological explanations of human behaviour that depend on the concept of primary motives (derived from research with rats and dogs) run into difficulties in view of the fact that primary motivation and reward do not appear to be critical in most studies of human skill acquisition. Thus, instead of giving food pellets (as to a rat), an experimenter delivers praise to a human subject; rather than receiving feedback by electric shock, the human can be guided by a needle moving on a dial or a buzzer signalling an error. At any rate, despite efforts to distinguish such motivational factors as general drives from selective incentives, attempts to demonstrate significant motivational effects in human psychomotor learning have met with only modest success. Among exceptions to the above are a few studies with standard apparatus (*e.g.*, the complex coordinator) and with special devices that have indicated that such incentives as money, verbal threats, electric shock, exhortations, and social competition may be relevant. Significant effects frequently fail to appear in experiments, and findings are often contradictory, so it has been suggested that the intrinsic challenge of the gadgetry, coupled with the subjects' already high pre-experimental motivation, leaves human volunteers unaffected by such weak laboratory manipulations of motive-incentive conditions as the foregoing (see also MOTIVATION).

Contradictory findings about motivation

Environmental factors. Many practical skills must be executed outside the laboratory under unfavourable conditions of temperature, humidity, illumination, and motion. It is generally found that below the limiting levels of extreme stress, such conditions affect psychomotor performance to a greater extent than they affect psychomotor learning. Representative findings have included the following: (1) isolation and sensory deprivation cause dramatic reductions in vigilance and monitoring skills within an hour; (2) environmental temperatures above or below 70° ± 5° F tend to lower scores on tracking apparatus but do not impair learning; (3) lack of oxygen slows reaction time, especially when the atmosphere corresponds to altitudes of 20,000 feet or higher; (4) accelerations of the body in a centrifuge or rotating platform disrupt postural coordination and produce systematic shifts in the perception of the vertical; (5) although such people as acrobats, dancers, pilots, and skaters can adapt well to high accelerations, even they lose equilibrium if deprived of the customary visual frame of reference; (6) rather mild centrifugal effects of slow, constant rotation may induce acute motion sickness and associated degradation of psychomotor proficiency in normal persons; (7) while some controlled work–rest schedules of crews during confinement in a small cabin upset daily sleep rhythms and lead to decrements in watchkeeping, memory, and procedural skills, a schedule of four-hours-on versus four-hours-off duty can be maintained for several months without significant impairment; (8) faulty identifications of visual displays on an eye–hand matching task have been produced in volunteer subjects exposed to controlled infectious diseases (*e.g.*, respiratory tularemia, phlebotomus fever, viral encephalitis).

Other environmental stress variables found to exert negative influences are vibration, low illumination, high atmospheric pressure, noise, glare, toxic gases, ionization, and subgravity. Certain drugs have positive effects on psychomotor performance (*e.g.*, amphetamines, magnesium pemoline, methyl caffeine, pipradrol); some have deleterious effects (*e.g.*, alcohol, barbiturates, diphenhydramine hydrochloride, lysergic acid, meprobamate, phenothiazines, scopolamine, tetrahydrocannabinol, tripelennamine); and others are either neutral or have inconsistent effects (*e.g.*, caffeine, nicotine).

INDIVIDUAL AND GROUP DIFFERENCES

Statistical indices of psychomotor ability (*e.g.*, means, variances, correlations) not only differ among individuals but may also serve to distinguish from each other groups of persons classified by such traits as age, sex, race, personality, and intelligence. Efforts to specify the relative contributions of heredity and environment to these individual and group differences are treated in GENETICS, HUMAN and HUMAN BEHAVIOUR, INNATE FACTORS IN. Comparative psychological studies of monozygotic (identical, one-egg) and dizygotic (fraternal, two-egg) twins have indicated that high coefficients of heritability—measured as a ratio of genotypic to phenotypic variance—exist for perceptual, spatial, and motor abilities.

Types of changes with age

Age. The most pervasive differences in human performance on psychomotor apparatus are associated with chronological age, and scores obtained from nearly all the devices mentioned above are sensitive to age differences. Researchers generally report a rapid increase in psychomotor proficiency from about the age of five years to the end of the second decade, followed by a few years of relative stability and then by a slow, almost linear decrease as the ninth decade is approached. For simple hand or foot reactions, complex discrimination-reaction time, and coordinated automobile steering, the peak of skill is attained between ages 15 and 20 on the average, and then performance at age 70 declines to about the level of age 10. This is a two-stage process: first, a developmental phase (*e.g.*, through maturation), followed by the more gradual deterioration of aging. Common athletic skills (*e.g.*, balancing, catching, gripping, jumping, reaching, running, and throwing) also improve through childhood, and it is well known that most athletes reach their prime before the end of the third decade. Olympic events requiring great muscular strength or stamina (*e.g.*, swimming) are dominated by athletes in their teens and 20s, whereas practitioners of more refined technical skills (*e.g.*, gymnastics) tend to be older. Self-paced, leisurely sports (*e.g.*, golf) are favoured over opponent-paced, combative activities (*e.g.*, tennis) as the aging process continues. Hereditary potentialities require several years to become established. It is probable that the genetic factors that underlie growth rates, and the age sequence in which different kinds of behaviour first appear, affect learning as well as performance.

Sex. Although the assessment of sexual differences in perceptual and reactive abilities is complicated by a number of factors (*e.g.*, age, race, and personality), girls and women tend to be more proficient than boys and men in such psychomotor skills as finger dexterity and inverted-alphabet printing. On the other hand, males generally do better than females at pursuit tracking, repetitive tapping, maze learning, and reaction-time tasks. On rotary pursuitmeter tests, women are not only less accurate but more variable than men of the same age and race (Figure 1). Although males appear to be superior to females in aptitude and capacity, these advantages disappear when subgroups are carefully matched for initial ability. In contrast, speed scores on discrimination-reaction tests reveal clearly diverging trends for college men and women trained intensively for several days (960 trials). This seems to be a genuine sex difference rather than an element of measurement or selection. Though both groups were equated for intelligence and had similar error scores, females began to suffer cumulative impairment on the fourth day of practice, whereas males kept improving. Sizable average differences in reaction latency as well as in movement time are characteristic of the sexes on other tasks.

Athletic skills

Whereas girls tend to attain their maximum proficiency in speeded tasks earlier in life than boys do, males continue to gain over a longer period and maintain their superiority over females for about half a century of the lifespan. After puberty, boys excel at most athletic skills demanding stamina and strength (*e.g.*, jumping, running, throwing). Thus, female Olympic swimming and track-and-field records are inferior to those of males and are achieved by girls who are noticeably younger than male champions in the same events. Sex has also been implicated in experiments employing complex coordinators, mirror tracers, and selective mathometers, with boys and men typically surpassing girls and women. Not all psychomotor differences associated with sex are intrinsically biological; unequal opportunities, distinctive social learning, role playing, and other culturally conditioned influences undoubtedly modulate the learning and execution of skills by males and females.

Race. All mankind is of one species. Zoologically, human races are all mutually interfertile subspecies—*i.e.*, breeding populations that differ in the relative frequencies of one or more genes (see ANTHROPOLOGY; HUMAN POPULATIONS; RACES OF MANKIND). Although the Variety of possible traits is practically limitless, random mating is not the case. Because of historical inbreeding tendencies, it is statistically improbable that any two human races have the same means and variances for all psychological traits. Not surprisingly, therefore, significant differences in psychomotor behaviour are foupd among ethnic groups throughout the world.

In one classic set of data (1904) on form-board skill (fitting nine geometric forms into correct holes), the average time in seconds for completing the task varied for different races—*e.g.*, seven African Pygmies (82.20), 12 Philippine Negritos (63.30), 55 American Indians and Eskimos (34.24), and 74 U.S. whites (27.80). Average error scores fell in the same order and, consistent with a genetic hypothesis, hybrid groups were appropriately ranked in between. These small samples were, however, not necessarily representative; *i.e.*, no effort was made to equate for differences in cultural values and psychomotor experiences found in different societies. Furthermore, since these were average differences, no conclusions about individual persons were warranted. It is interesting that the rankings were found to change for other psychomotor tasks; *e.g.*, on a test of tapping and aiming skill, Eskimos surpassed all others, followed by Filipinos, who were in turn trailed by Caucasians. A more recent study (1967) discovered that a sample of Mongoloid Chamorro people from Saipan and another of Indians from the U.S. exceeded white norms on a test of maze-tracing ability. In both studies, people of mixed races tended to make intermediate scores, a fact consistent with contemporary research in behavioral genetics and physical anthropology.

Absence of a superior race

No particular race is found to be uniformly superior in all psychomotor aptitudes and capacities, but environmental causes seem to be inadequate to explain differences in rate of acquisition and final level of performance on standard apparatus. For some tasks (*e.g.*, on the rotary pursuitmeter) psychologists report that the degree of initial hereditary determination seems to reach 90 percent, thus leaving little room for sociological variables to operate. Teams conducting research with infant tests have noted that Congoid babies in both Africa and the United States are more precocious in sensorimotor maturation than are Caucasoid babies in Africa, the United States, or Europe, and that this precocity lasts about three years, after which whites outperform the blacks. By adolescence, Negro subjects score significantly less well than whites on complex coordinator, rotary pursuitmeter, discrimination reaction, selective mathometer, and two-hand coordinator tasks. On the other hand, Chinese- and Japanese-American infants lag behind Caucasians in motor development but perform better than both whites and blacks on certain eye–hand coordination and dexterity tasks at later ages.

Many research workers believe that social, economic, educational, and attitudinal variables that might unequally influence minority groups are of little consequence in the psychomotor field. They point out, for instance, that Chinese, Jewish, Negro, and Puerto Rican children, when tested by members of their own cultures, show distinctive patterns of basic perceptual and motor abilities, and their particular skill profiles are unaffected by differences in socio-economic level. Malnutrition (*e.g.*, protein deficiency) is not believed to be a plausible explanation unless there has been severe deprivation during the perinatal period. According to much of the literature, blacks generally do better than whites on chemical taste tests, rhythmic discrimination, visual acuity, colour perception, and resistance to special optical illusions; Mongoloids show better taste sensitivity and less colour blindness than Caucasoids; and on certain physical-fitness tests male Afro-American athletes are not only superior to their white countrymen but their relative proficiency is inversely correlated with the degree of Caucasoid admixture.

Genetic behavioral differences among human populations—just as those of morphology—appear to be the rule rather than the exception. Tarahumara Indians far surpass other races in endurance at long-distance running contests; Andeans and Tibetans are superbly adapted to working at high altitudes; Eskimos excel on psychomotor tasks performed under low-temperature stress. It is plausible that the inherited factors underlying behavioral aptitudes and capacities have evolved from different selective pressures in different ecological niches. As is true for age and sex, however, hereditary and environmental variables are complexly intertwined in racial studies. Nevertheless, genetic determinants seem to be far more powerful in the etiology of original psychomotor aptitudes. It does not follow that learnability is weak. Quantitative experiments demonstate that heritabilities can be systematically altered by controlled practice; this is a theoretical discovery of broad implications for practical training programs. At the same time, it would appear that the hereditary control of several psychomotor abilities tends to be less pronounced at the end of training than at the beginning.

IQ correlations

Other factors. A number of other personal characteristics have been found to be of significance in psychomotor behaviour. For instance: (1) speed scores in reaction-time tasks are positively correlated with body temperature in adults, one of the many indices of variation within the individual; (2) psychotics show longer reaction times and poorer tracking scores than do people of normal personality; (3) right-handed operators are favoured on the rotary pursuitmeter, while left-handed persons tend to do better on the complex coordinator; (4) left-handed people are more variable in finger-dexterity and paper-cutting skills and also are more prone to show signs of ambidexterity; (5) intelligence quotients (IQ) are weakly related to physical strength or endurance yet are strongly associated with performance in such psychomotor activities as running the 35-yard dash, balancing on one foot, discrimination reaction, rotary pursuit, and selective mathometry—these correlations are especially high when based on groups that comprise a full range of IQs (from retardates to college students); (6) typically one's body build (somatotype) is associated with his athletic skills—the best fencers, oarsmen, and basketball players, for example, tend to be tall and lean (ectomorphic); top swimmers, divers, and pole-vaulters are likely to be broad-shouldered and slim-hipped (mesomorphic); champion wrestlers, shot putters, and weight lifters are apt to be thick-trunked and short-limbed (endomorphic). While these genetically determined somatotypes do not guarantee athletic prowess, they definitely do favour success in certain sports rather than others. Similar considerations apply to vocal and instrumental musical aptitudes wherein unique combinations of such anatomical structures as lips, teeth, larynx, tongue, eyes, ears, hands, and arms can facilitate the attainment of virtuoso skill.

In short, psychomotor abilities and learning underlie some of the most fundamental human activities, contributing to the full spectrum of work, play, creativity, love, and the very survival of individual and species.

BIBLIOGRAPHY. A.L. IRION, "A Brief History of Research on the Acquisition of Skill," in E.A. BILODEAU (ed.), *Acquisition of Skill* (1966), an excellent historical survey; R.S. WOODWORTH and H. SCHLOSBERG, *Experimental Psychology*, rev. ed. (1954), one of the best standard reference works; C.E. NOBLE, "S-O-R and the Psychology of Human Learning," *Psychol. Rep.*, 18:923–943 (1966), a brief introduction to human learning; A.W. MELTON (ed.), *Apparatus Tests* (1947), a classic volume on psychomotor devices used in aviation psychology. Periodic reviews of the psychomotor field include: E.A. and I.McD. BILODEAU, "Motor-Skills Learning," *A. Rev. Psychol.* 12:243–280 (1961); J.A. ADAMS, "Motor Skills," *ibid.*, 15:181–202 (1964); and C.E. NOBLE, "The Learning of Psychomotor Skills," *ibid.*, 19:203–250 (1968). C.L. HULL, *Principles of Behavior* (1943); and K.W. SPENCE, *Behavior Theory and Conditioning* (1956), are two of the most influential books on general behaviour theory from the reinforcement viewpoint. C.E. NOBLE, "A Theory of Psychomotor Skill: Derivation and Data," *Psychonomic Sci.* 21:344 (1970), makes a specific application to psychomotor learning. Two volumes presenting detailed information-processing analyses of skill are P.M. FITTS and M.I. POSNER, *Human Performance* (1967); and A.T. WELFORD, *Fundamentals of Skill* (1968). E.A. BILODEAU (ed.), *Principles of Skill Acquisition* (1969), provides an eclectic, simplified treatment of current topics by several authors. R.N. SINGER, *Motor Learning and Human Performance* (1968), is oriented mainly toward athletic proficiency and physical education. Articles of specialized interest, as indicated by their titles, are J.A. ADAMS, "Response Feedback and Learning," *Psychol. Bull.*, 70:486–504 (1968); C.E. NOBLE, "Acquisition of Pursuit Tracking Skill Under Extended Training As a Joint Function of Sex and Initial Ability," J. *Exp. Psychol.*, 86:360–373 (1970); and R.B. PAYNE, "Functional Properties of Supplementary Feedback Stimuli," J. *Motor Behnv.*, 2:37–43 (1970).

(C.E.N.)

Learning Theories

A common goal in defining any psychological concept is a statement that corresponds to common usage. Acceptance of that aim, however, entails some peril. It implicitly assumes that common language categorizes in scientifically meaningful ways; that the word learning, for example, corresponds to a definite psychological process. However. there appears to be good reason to doubt the validity of this assumption. The phenomena of learning are so varied and diverse that their inclusion in a single category may not be warranted.

Recognizing this danger (and the corollary that no definition of learning is likely to be totally satisfactory) a definition proposed in 1961 by G.A. Kimble may be considered representative: Learning is a relatively permanent change in a behavioral potentiality that occurs as a result of reinforced practice. Although the definition is useful, it still leaves problems.

The definition may be helpful by indicating that the change need not be an improvement; addictions and prejudices are learned as well as high-level skills and useful knowledge.

The phrase relatively permanent serves to exclude temporary behavioral changes that may depend on such factors as fatigue, the effects of drugs, or alterations in motives.

The word potentiality covers effects that do not appear at once; one might learn about tourniquets by reading a first-aid manual and put the information to use later.

To say that learning occurs as a result of practice excludes the effects of physiological development, aging, and brain damage.

Reinforcement as aid to learning

The stipulation that practice must be reinforced serves to distinguish learning from the opposed loss of unreinforced habits. Reinforcement objectively refers to any condition—often reward or punishment—that may promote learning.

However, the definition raises difficulties. How permanent is relatively permanent? Suppose one looks up an address, writes it on an envelope, but five minutes later has to look it up again to be sure it is correct. Does this qualify as relatively permanent? While commonly accepted as learning, it seems to violate the definition.

What exactly is the result that occurs with practice? Is it a change in the nervous system? Is it a matter of providing stimuli that can evoke responses they previously would not? Does it mean developing associations, gaining insights, or gaining new perspective?

Such questions serve to distinguish Kimble's descriptive definition from theoretical attempts to define learning by identifying the nature of its underlying process. These may be neurophysiological, perceptual, or associationistic; they begin to delineate theoretical issues and to identify the bases for and manifestations of learning.

The range of phenomena called learning. Even the simplest animals display such primitive forms of adaptive activity as habituation, the elimination of practiced responses. For example, a paramecium can learn to escape from a narrow glass tube to get to food. Learning in this case consists of the elimination (habituation) of unnecessary movements. Habituation also has been demonstrated for mammals in which control normally exercised by higher (brain) centres has been impaired by severing the spinal cord. For example, repeated application of electric shock to the paw of a cat so treated leads to habituation of the reflex withdrawal reaction. Whether single-celled animals or cats that function only through the spinal cord are capable of higher forms of learning is a matter of controversy. Sporadic reports that conditioned responses may be possible among such animals have been sharply debated.

At higher evolutionary levels the range of phenomena called learning is more extensive. Many mammalian species display the following varieties of learning.

Classical conditioning. This is the form of learning studied by Ivan Petrovich Pavlov (1849–1936). Some neutral stimulus. such as a bell, is presented just before delivery of some effective stimulus (say. food or acid placed in the mouth of a dog). A response such as salivation, originally evoked only by the effective stimulus, eventually appears when the initially neutral stimulus is presented. The response is said to have become conditioned. Classical conditioning seems easiest to establish for involuntary reactions mediated by the autonomic nervous system.

Conditioning of involuntary reactions

Instrumental conditioning. This indicates learning to obtain reward or to avoid punishment. Laboratory examples of such conditioning among small mammals or birds are common. Rats or pigeons may be taught to press levers for food; they also learn to avoid or terminate electric shock.

Chaining. In the form of learning called chaining the subject is required to make a series of responses in a definite order. For example, a sequence of correct turns in a maze is to be mastered, or a list of words is to be learned in specific sequence.

Acquisition of skill. Within limits, laboratory animals can be taught to regulate the force with which they press a lever or to control the speed at which they run down an alley. Such skills are learned when a reward is made contingent on quantitatively constrained performance. Among human learners complex, precise skills (*e.g.*, tying shoelaces) are routine.

Discrimination learning. In discrimination learning the subject is reinforced to respond only to selected sensory characteristics of stimuli. Discriminations that can be established in this way may be quite subtle. Pigeons, for example, can learn to discriminate differences in colours that are indistinguishable to human beings without the use of special devices.

Concept formation. An organism is said to have learned a concept when it responds uniquely to all objects or events in a given logical class as distinct from other classes. Even geese can master such concepts as roundness and triangularity; after training, they can respond appropriately to round or triangular figures they have never seen before.

Principle learning. A subject may be shown sets of three figures (say, two round and one triangular; next, two square and one round, and so on). With proper rewards, the subject may learn to distinguish any "odd" member of any set from those that are similar. Animals as low in the evolutionary scale as the pigeon can master the principle of this so-called oddity problem.

The oddity problem

Problem solving. Examples of human problem solving are familiar: finding the roots of a quadratic equation, solving a mechanical puzzle, and navigating by the stars. Among other animals, chimpanzees have been observed to solve problems requiring toolmaking.

This list only samples from the remarkable array of animal activities categorized as learning. Beginning with habituation, they range from the simple adjustments of single-celled animals up to the highest intellectual accomplishments of mankind. It would be wonderful indeed if a single theory of learning were enough to account for all this diversity. So far, however, no theory of learning adequately covers more than a small fraction of these phenomena.

THE STATE OF LEARNING THEORIES

Yet, at the start of the 20th century, vast psychological systems, such as behaviourism and Gestalt psychology, indeed were offered as explanations of learning (and of much wider ranges of behaviour as well). And as late as the 1940s, comprehensive theories of learning were still believed to be reasonably near at hand. But during the next three decades it grew clear that such theories are tenable only for very limited sets of data. By the 1970s, learning theory seemed to consist of a set of hypotheses of limited applicability.

Important earlier theorists. Beginning in the 1930s a number of general theories were advanced in attempts to organize most or all of the psychology of learning. The most influential of the contributing theorists are noted below.

E.R. Guthrie (1886–1959) wrote that learning requires only that a response be made in a changing situation. Any response was held to be linked specifically to the situation in which it was learned. Guthrie argued that learning is complete in one trial, that the most recent response in a situation is the one that is learned, and that responses (rather than perceptions or psychological states) provide the raw materials for the learning process.

For E.C. Tolman (1886–1959) the essence of learning was the acquisition by the organism of a set of what he called Sign-Gestalt-Expectations. These referred to propositions said to be made by the learner that his own specific response to given signs (or stimuli) would result in such and such circumstances later on. Tolman seemed to be saying that what the learner acquires is a specific knowledge of "what leads to what." In brief, his theory was that the learner develops expectations based on experience and that learning depends entirely on successions of events. Although less vocal on the point than others, Tolman implied that learning was a gradual process.

The theory offered by Clark L. Hull (1884–1952), over the period between 1929 and his death, was the most detailed and complex of the great theories of learning. The basic concept for Hull was "habit strength," which was said to develop as a function of practice. Habits were depicted as stimulus-response connections based on reward. According to Hull, responses (rather than perceptions or expectancies) participate in habit formation, the process is gradual, and reward is an essential condition.

Comparison of these theories yields major questions for empirical investigation. Is learning continuous or discontinuous; is it a gradual or sudden (one-trial) process? Is learning a matter of establishing stimulus–response (S–R) connections or does it depend on the learner's understanding of perceptual relationships? Is reward necessary for learning?

Are theories of learning necessary? Such major investigators of learning as B.F. Skinner and J.A. McGeoch maintained in the 1930s and 1940s that preoccupation with theory was misguided. For them the approach simply was to discover the conditions that produce and control learned behaviour. Beyond this, their interests diverged. Skinner studied instrumental conditioning (operant conditioning, as he called it) among rats; McGeoch specialized in human rote memory. Although study of rote verbal learning had become heavily theoretical by the 1970s, Skinner and his associates stuck to their empirical guns, guiding a variety of programs for the practical control of behaviour. Teaching machines and computer-aided instruction, behaviour modification (*e.g.*, the use of tokens to reward desired behaviour among psychiatric patients), and planned utopian societies (Walden II) all found scientific origins in Skinner's rejection of theory in favour of direct efforts to produce results.

Intervening variables and hypothetical constructs. Learning is a concept and not a thing, and the activity called learning is inferred only through behavioral symptoms. The distinction implicit here between behaviour and in-

Watson's reflex theory

ferred process is one of Tolman's major contributions and serves to reconcile influential views that might seem completely at odds. Classical behaviourism, as developed by John B. Watson (1878–1958), rejected every mentalistic account and sought to limit analysis to such physiological mechanisms as reflexes. Watson argued that these are objective in a way that so-called thoughts, hopes, expectancies, and images cannot be. The Opposing view holds that experiential (introspective) activity (exactly what Watson sought to dismiss) does require discussion.

Tolman called himself a behaviourist and ostensibly was bound by Watson's insistence on objectivity. But he also was interested in thinking, expectancy, and consciousness. Tolman found his solution to this problem of incompatible theories after his association with the Vienna Circle of Logical Positivists, whose deterministic teachings he brought to the attention of U.S. psychologists about 1920. He maintained that learning is inexorably produced (determined) by such independent (directly manipulable) variables as the organism's previous training and physiological condition and by the response the environment requires. According to Tolman, the development of learning is revealed through the changing probability that given behaviour (the dependent variable) will result. He held that learning itself is not directly observable; it is an intervening variable, one that is inferred as a connecting process between antecedent (independent) variables and consequent (dependent) behaviour.

An attractive possibility is that intervening variables may have discoverable physiological bases. Psychologists Paul E. Meehl and Kenneth MacCorquodale proposed a distinction between the abstractions advocated by some and the physiological mechanisms sought by others. Meehl and MacCorquodale recommended using the term intervening variable for the abstraction and hypothetical construct for the physiological foundation. To illustrate: Hull treated habit strength as an intervening variable, defining it as an abstract mathematical function of the number of times a given response is rewarded. By contrast, Edward L. Thorndike (1874–1949) handled learning as a hypothetical construct, positing a physiological mechanism: improved conduction of nerve impulses.

Intervening variables and hypothetical constructs need not be incompatible; Thorndike's hypothetical neural process could empirically be found to be the mechanism through which Hull's abstraction operates.

Miniature theories. With growing realization of the complexity of learning, the grand theories of Guthrie, Hull, and Tolman generally have been abandoned except as historic landmarks. Hope for any impending, comprehensive theory was almost dead in the 1970s. More modest miniature theories remain, many likely to be of temporary value. An account of their major themes and issues, however, should have more enduring interest.

MAJOR THEMES AND ISSUES

Association. A dominant ancient theme in theories of learning has been that of association. Although the concept was accepted by Aristotle, it was brought into the developing psychology of learning by British empiricist philosophers (Locke, Berkeley, Hume, the Mills, and Hartley) during the 17th, 18th, and 19th centuries. Popular acceptability of the notion of association was related to progress in the physical sciences. The physical universe had been shown to consist of a limited number of chemical elements that can combine in innumerable ways. By analogy, a science of "mental chemistry" seemed appealing. The theorized elements in this new "science" were called ideas, said to be based on what were named sensations. The synthesizing principles by which these posited ideas combined in conscious experience were expressed as so-called laws of association. It was suggested that such conditions as temporal and spatial contiguity, repetition, similarity, and vividness favoured the formation of associations, and each was called a law of association. Thus, there were "laws" of repetition, of similarity, and so on.

At the end of the 19th century the notion of association was widely accepted among psychologists. German psychologist Wilhelm Wundt (1832–1920) took a position nearly identical with that of the British empiricist philosophers. Also in Germany, Hermann Ebbinghaus (1850–1909) began to study rote learning of lists of nonsense verbal items (*e.g.,* XOQ, ZUN, ZIB). He maintained that the association of each word with every succeeding word was the primary mechanism in learning these lists. Pavlov in Russia offered temporary associative connections in the nervous system as a hypothetical basis for conditioned reflexes.

Rote learning

These European influences coalesced in North America. Wundt's notions were introduced there when a student of his from England, Edward Bradford Titchener (1867–1927), came to teach at Cornell University in Ithaca, New York. Ebbinghaus' method and theory became standard in Canadian and U.S. studies of verbal learning; Watson and other behaviourists applied Pavlov's conceptions to their learning experiments. Experimental psychology in the Western Hemisphere came to be dominated by what seemed to be a search for laws of association.

What is associated? Investigators asked whether associations are formed between observable stimuli and responses (S–R) or between subjective sensory impressions (S–S). One group that included Hull, Guthrie, and Thorndike took the relatively objective S–R position, while Tolman and others favoured the more introspective, perceptual S–S approach. For a time S–R theorists held popularity; behavioral responses are readily observable evidence of learning, and many included them in the associative process itself.

But the reduction of learning to mere external stimuli and overt responses raised discordant theoretical objections that the inner activities of the organism were being ignored. S–R theories failed to account for a host of learned phenomena. For example, people could be trained to say they heard sounds even when such auditory stimuli were absent. They said they dreamed about what they had learned, too; yet there need be no immediate external stimulus, nor does the dreamer always make the responses he dreams about.

Physiological psychologists and biologists found ways of delivering electrical stimulation directly to the brain; this eliminated the sensory stimuli and vocal or motor responses on which S–R theories hinge. Direct neural stimulation was found to be an adequate signal and the electrical response of the brain itself proved susceptible to conditioning. At this level of the nervous system, distinctions between stimulus and response mean less than at the periphery, and the S–S versus S–R controversy is no longer such a burning issue.

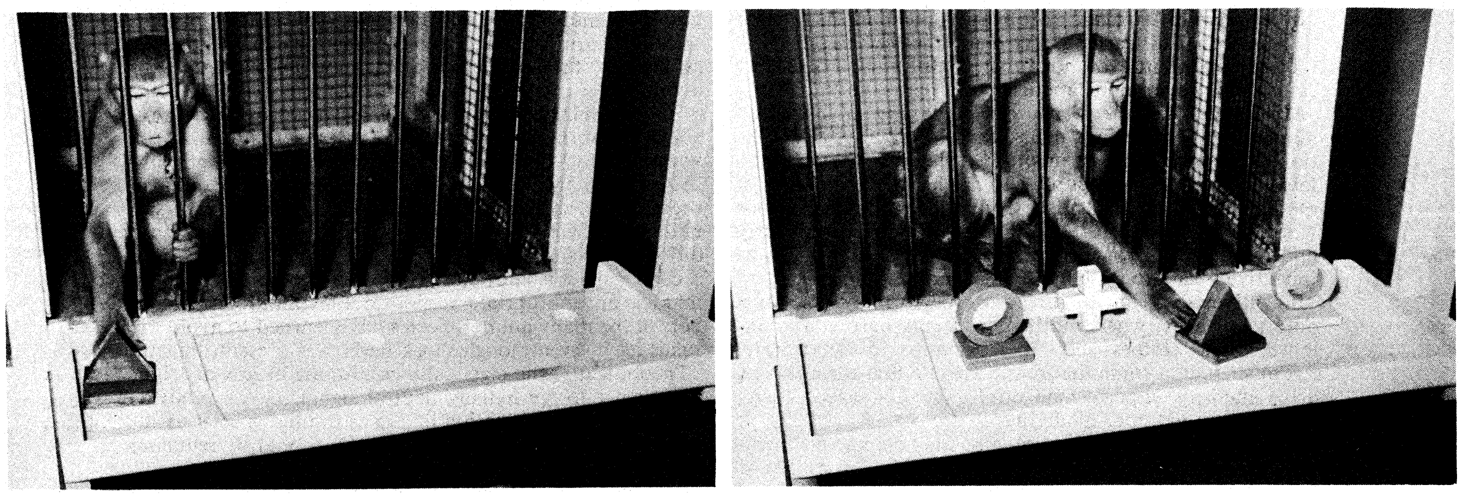
Direction of association. Classical conditioning dependably has been shown to proceed only forward in time. Bell must precede food if a conditioned reaction is to be established. If it had any effect, the reverse procedure (food before bell) would be called backward conditioning; but at most it only inhibits other reactions. There seems to be a relatively brief optimal interval in classical conditioning at which associations are most easily made. For quick reflexes such as the eyeblink, this interval is about one-half second; longer or shorter intervals are less effective. For slower reactions such as salivation the interval is longer, perhaps two seconds or so.

Forward and backward association; context association

In learning verbal associations the situation appears to be quite different. When one learns the Russian–English forward association da–"yes," he also learns the English–Russian backward association "yes"–da. Moreover, timing is much less critical than in classical conditioning. Verbal pairs are learned with almost equal ease whether presented simultaneously or separated by several seconds.

In what is called context association, the general environment may begin to elicit a response that is being conditioned to a specific stimulus. Thus, a dog may salivate simply on being brought into the experimental room—before any bell rings. Verbal associations also can be weakened by changes in the general situation.

Repetition. A major theoretical issue concerns whether associations grow in strength with exercise or whether they are fully established all at once. Evidence is that learning usually proceeds gradually; even when a problem is solved insightfully, practice with similar tasks tends



Learning-to-learn experiment by Harry *F.* Harlow.
The monkey has been trained to expect food after lifting an object. In the first picture, the monkey lifts the triangle after having been shown a sample triangular form. In the second picture, taken several minutes later, the monkey selects the triangle from among other forms. The process is repeated with the circle and the cross.

to improve performance. Some (perhaps most) learning theorists have concluded that repetition gradually enhances some underlying process in learning.

Stimulus-sampling, a statistical learning theory

The view that associations develop at full strength in a single trial leads to a typical question. How can the gradual nature of most learning be explained if all-or-nothing is the rule? One possible answer suggested by Guthrie has led to so-called stimulus-sampling theory. The theory assumes that associations indeed are made in just one trial. However, learning *seems* slow, it is said, because the environment (context) in which it occurs is complex and constantly changing. Given a changing environment, the sample of stimuli will differ from trial to trial. Thus, it is reasoned, it should take many trials before a response is associated with a relatively complete set of all possible stimuli.

In this light, the strength (or probability) of a response should increase with practice even if the elementary associative process occurs in a single trial.

These stimulus-sampling notions translate easily into mathematical form; they are an example of statistical learning theory, a more general development in the quantitative treatment of learning.

Reinforcement. Repetition alone does not ensure learning; eventually it produces fatigue and suppresses responses. An additional process called reinforcement has been invoked to account for learning, and heated disputes have centred on its theoretical mechanism.

Objectively reinforcement refers to the use of stimuli that have been found to facilitate learning. Under appropriate conditions, these include praise, food, water, opportunity to explore, sexual stimuli, money, electric shock, and direct brain stimulation.

More theoretically, the term reinforcement expresses various theoretical hunches about some specialized subjective quality all such stimuli might share. Food for a hungry animal is a well-established reinforcer, conceivably through its distinctive appearance and odour. It tends to elicit a set of responses: approaching, chewing, tasting, swallowing; these may produce additional perceptual activities that reduce the drive or desire for food (*e.g.*, by halting stomach contractions that are experienced as hunger pangs). But no single subjective quality imagined by theorists seems invariably effective in reinforcement studies. Perhaps some combination of introspective influences is critical, or it may be that perceptual processes apply differently from one learning situation to another.

Anti-associationistic positions. Not all psychologists have accepted the general validity of association theories; many have suggested that considerations other than association are crucial to learning.

Organization. Major critics of association theory included such Gestalt psychologists as Wolfgang Kohler (1887–1967), who held that learning often entails a perceptual restructuring of environmental relationships. Kohler cited his own studies of insightful learning by a chimpanzee. The animal learned to join two sticks (akin to a jointed fishing pole) as a tool to pull in a banana that was out of arm's reach and of either short stick alone. The ape was described as sitting quietly (as if in thought), and then suddenly fitting the sticks together to rake in the fruit. It was argued that the ability to perceive new ways of relating the sticks to the banana was essential in solving the problem.

Learning as perceptual organization

Similar organizational processes in perceiving can be demonstrated in serial verbal learning. Memorizing the list *thick, wall, it, tea, of, myrrh, seize, knots, trained* should demand some rehearsal. Yet, notice the phonetic resemblance to Shakespeare's famous line from *The Merchant of Venice:* "The quality of mercy is not strained . . . ." With that kind of perceptual organization, learning can become quick and easy.

A powerful argument also was made by psycholinguists who criticized what they took to be the associationistic account of language learning. Even assuming one-trial acquisition, it was held that such individually learned associations could not account for all combinations of words people use; there are simply too many. They suggested that learning a language requires some general organizing structure on which words are hung. Some proponents of this position hold that this structure does not depend on learning, being transmitted genetically from parent to child.

Inhibition. Gestalt interpretations often reject the associationistic hypothesis wholesale. Other theorists endorse the notion of association, but hold it to be less important than is a process of inhibition through which errors in learning are eliminated. Such theorists find support in evidence for the development of learning sets (what is called learning to learn).

For example, a monkey may learn a long series of discriminations; *e.g.*, red versus green, black versus white, round versus square, large versus small, triangle versus ellipse. After solving several hundred such problems, some monkeys learn to master each new one in a single trial, as if insightfully. The animal is said to have learned to learn such discriminations.

Learning to learn: error-factor theory

Evidence clearly shows that the monkey gradually abandons erroneous tendencies as learning proceeds. At first it might be prone to choose stimuli that are red, black, round, large, or triangular. Correct choices do not always correspond to the animal's initial biases, and their suppression (inhibition, extinction) eventually permits single-trial learning. Theoretically, organisms learn to learn by inhibiting erroneous behaviour; thus, Harry F. Harlow, a proponent of this view, called it an error-factor theory.

Motivation in learning. Motivation popularly is thought to be essential to learning. Yet many theorists suggest that motives make little or no direct contribution—that they simply tend to promote practice.

Motivation and performance. Learning was defined above as a change in a behavioral potentiality. Realization of such potential seems to be related to the learner's level of motivation. A pupil who has learned the names of all members of the British Commonwealth of Nations would be expected to recite them with particular energy under some sort of incentive (reward or punishment). The incentive is said to raise his level of motivation.

Incentives do seem to invigorate performance up to a point; however, when motivation seems particularly intense, some studies show performance to deteriorate. From such data some theorists conclude that the effect of drive intensity on performance follows a U-shaped course, first helping and later hindering.

Greatly increased motivation also may change performance qualitatively by introducing new inefficient modes of behaviour. A student may be so tautly driven to do well on an examination that his tension, fear of failure, and his visceral and muscular discomfort interfere with performance.

Motivation and learning. To show that motivation affects performance of what has been learned is not the same as demonstrating its effect on the process of learning itself. This would require that individuals learn under various levels of motivation and be tested under the same incentive levels. (This is to control for the effects of motivation on performance alone.) And, indeed, the best-controlled experiments of this design indicate learning effects to be the same under different levels of motivation.

Varieties of learning. It is debated whether all forms of learning represent the same process. This question applies even to relatively primitive phenomena such as classical and instrumental conditioning.

Classical and instrumental conditioning

In instrumental conditioning reinforcement is contingent on the learner's response; a rat receives food only if it presses the lever. In classical conditioning there is no such contingency; a dog is fed whether or not it salivates. But this is a distinction in experimental procedure. Whether the underlying process of learning is the same for both is quite another question.

Classical conditioning usually has been reported for glandular, autonomically mediated, involuntary responses (*e.g.*, salivation, heart rate). By contrast, voluntary movements of skeletal muscles more typically have been found to be conditionable instrumentally. However, to theorize that classical conditioning is exclusively effective for one class of responses while instrumental conditioning is uniquely applicable to others seems to be a mistake.

Evidence that seems to demolish such theorizing comes from a series of experiments directed by Neal E. Miller at the Rockefeller University in New York City. Rats were immobilized with curare; this drug blocks the junction between muscle and nerve to paralyze the skeletal muscles. However, a curarized individual still can show autonomic, involuntary signs of emotional activity such as a rapidly beating heart.

Electrical stimulation of selected parts of the brain seems to be rewarding; animals behave as if they seek such stimulation and will learn to press a switch for it (voluntary muscle function). Using curarized animals, Miller and others made the rewarding stimulation contingent on such typically involuntary responses as changes in heart rate, blood pressure, contractions of the bowel, and salivation. Their research has shown such instrumental conditioning to be effective for all these responses. The evidence appears to destroy the once-popular hypothesis that involuntary autonomic reactions are subject only to classical conditioning. In this sense the two primitive forms of learning seem to be the same.

Stages of learning. Should the basic process prove to be the same for all varieties of learning, there would still be reason to believe that it operates differently from one stage of practice to another. For example, in coping with painful stimuli (*e.g.*, electric shocks) laboratory animals seem to learn in two successive, distinguishable phases. Apparently they first learn to fear the situation, then to avoid it.

For example, when an animal learns to avoid painful shock (by turning a paddle wheel or by running away), a warning signal can be given; *e.g.*, with a flash of light or a buzzer. The two stages of learning then can be studied separately. The animal first is subjected to pairings of signal and unavoidable shock to establish (by classical conditioning) signs of fear in response to the signal. In the second stage it is allowed to stop the frightening signal by making an appropriate response. Preconditioned members of the many animal species have learned to avoid the signal itself, even though shock never was presented again.

The role of fear

Theoretically, the classically conditioned signs of fright in response to the initially neutral signal have a motivating function. Termination of that stimulus is seen as instrumental—that is, as rewarding the animal by reducing learned experiences of fear.

Classical conditioning. A two-stage process has been suggested even for classical conditioning. One theory is that in the first stage the subject learns that a neutral stimulus (a ringing bell) is to be presented along with another stimulus (food) whether or not it exhibits a reaction (salivation). Conditioning of any reaction is held to constitute the second stage of learning. The skimpy supporting evidence points to the first stage as a prerequisite, suggesting that responses can only be conditioned after the sensory conditions are recognized.

Verbal learning. Theories that interpret verbal learning as a process that develops in stages also have been worked out. In one variety of rote learning the subject is to respond with a specific word whenever another word with which it has been paired is presented. In learning lists that include such paired-associates as *house–girl*, *table–happy*, and parcel-chair, the correct responses would be girl (for house), happy (for table), and chair (for parcel). By convention the first word in each pair is called the stimulus term and the second the response term. Paired-associate learning is theorized to require subprocesses: one to discriminate among stimulus terms, another to select the second terms as the set of responses, and a third to associate or link each response term with its stimulus term. Although these posited phases seem to overlap, there is evidence indicating that the first two (stimulus discrimination and response selection) precede the associative stage.

Remembering and forgetting. Learning, remembering, and forgetting often have been considered separate processes. Yet these distinctions seem to blur in the face of contemporary research and theory.

Transient and enduring memory. Evidence for stages of learning comes from observations of learners over relatively extended series of trials (or comparatively long periods). The empirical data suggest that several alterations in memory function occur even during a single trial. The process that commits information to memory also seems to have several stages.

Immediate, short-term, and long-term memory

Most theorists attribute at least three stages to memory function: immediate, short-term, and long-term. Immediate memory seems to last little more than a second or so. For example, subjects may be asked to remember where specific objects are located within a complex array they have just seen. Their performance shows that considerable information is retained only briefly, rapidly fading unless it is given special attention.

Short-term memory lasts about 15–30 seconds, as after looking up a telephone number. One makes the call, discovers he has forgotten the number (perhaps in the midst of dialing), and has to look it up again. Nevertheless, such short-term retention does make information available long enough to be rehearsed; if the learner repeats it to himself, the number can be transferred to some sort of longer term storage.

Thus, rehearsal seems to facilitate transfer of data from short-term to long-term memory. Once committed to long-term memory, the results of learning tend to endure but can be abruptly abolished when specific parts of the brain are injured or removed; they also are vulnerable to interference from other learning. Nevertheless, conditioned responses may undergo little or no forgetting over periods of months or years. And electrical stimulation of the surgically exposed brain while a person is awake can make him remember experiences long thought forgotten. Recall is reported to be similarly enhanced during hypnosis.

Retrieval. The amount of information one readily can retrieve from what is stored in memory is prodigious. In locating an item in memory, he apparently activates a system that stores a set of related data; then he searches for the item within that system. For example, a person is shown a long, randomly mixed list of words that belong to different categories (*e.g.*, names of animals, plants, professions, tools). When asked to remember as many words as he can, he spontaneously will tend to group them by category; this is called clustering of recall. Thus, names of animals (spread throughout the original list) are likely to be remembered one after the other.

Clustering; tip-of-the-tongue phenomenon

Studies of the familiar tip-of-the-tongue experience yield analogous results. College students who heard definitions (of this sort: a small, open Chinese boat) were asked to supply the right word (in this case it would be sampan). Those who said they might have it somewhere on the tip of the tongue were significantly accurate in guessing the first letter and the number of syllables. Their tendency also to recall words that sounded the same or that had similar meanings is reminiscent of clustering.

Considerable evidence of this kind supports the theory that the process of retrieval first locates stored data in some sort of associative network and then selects an item with specific characteristics.

Forgetting. Whether immediate and short-term data simply decay or are lost through interference is a matter of controversy. However, evidence is clearer that interference affects retention of information in long-term storage. Retention of the word ***happy*** (learned as a paired associate of ***table)*** seems to be subject to the interference of a strong tendency to associate ***table*** with ***chair.*** Thus, the paired associate ***table–happy*** becomes more readily forgotten when followed by ***parcel–chair*** as the very next item in a list; this seems to help ***chair*** reassert its old tendency to be associated with ***table.*** In general, it is found that associations tend to interfere with or to inhibit one another. Interference deriving from earlier (and later) associations is called proactive inhibition (and retroactive inhibition). These two forms of inhibition commonly are accepted as major processes in forgetting, proactive inhibition being assigned greater importance.

Contemporary trends in learning theory. In the early 1930s the distinction between learned and inherited behaviour seemed clearer than it does now. The view that any bit of behaviour either was learned or simply developed without learning seemed straightforward. Studies based on these expectations led investigators to conclude that rat-killing behaviour among cats is learned rather than instinctive, that human fears are all acquired, or that intelligence is completely the result of experience. Learning theorists were saying then that most behaviour is learned and that biological factors are of little or no importance.

Imprinting, a special form of learning

Forty years later this position seemed grossly untenable. The once-implied sharp distinction between learned and inherited behaviour had become badly blurred. For example, it has been found that the young of many animal species automatically will learn to follow the first large, moving, noisy object presented (as if it were their mother). This special form of learning is called imprinting and seems to occur only during a critical early stage of life. Among mallard ducklings imprinting is most feasible about 15 hours after hatching. During this period a duckling will imprint as easily on an old man or on a rubber ball as it will on a mother duck. Is this instinctive or learned behaviour? Manifestly it is both. The instinctive tendency to be imprinted is part of the duckling's biological heritage; while the object on which it is imprinted is a matter of experience. What is significant for learning theory is that the contribution of biology cannot be ignored.

Learning theorists once ruled a number of concepts out of court on the ground that they seemed objectively unclean. Image, cognition, awareness, and volition, all are concepts that were denied acceptance on this basis. They sounded mentalistic, subjective, introspective, and unverifiable. Yet, by the 1970s all of them were being given more serious scientific consideration.

For example, the concept of image in learning has begun to show real viability. It has long been reported that the more meaningful a list of words is, the easier it will be to learn. Degree of meaningfulness for a word may be defined by the objectively observed probability that people quickly can give another word in response. Using such empirical scales of meaningfulness, a reliable and substantial relationship has been found between meaningfulness and ease of learning. However, meaningful words also may evoke vivid images that subjects can describe when asked. When they do evoke such imagery, they seem to be learned and remembered even more easily. Thus, learning theory seems to be enriched when introspective data are used.

A final fault in much learning theory stems from earlier tendencies to use the laws of physics as a model. Theorists once sought general laws of wide applicability that tended to obscure differences among individuals. For example, so complete was Hull's faith in universal "laws" of animal behaviour, that he based his hypothesis about the optimal interval for classical conditioning in humans, other mammals, and birds on the pattern of nerve conduction in the optic nerve of the horseshoe crab. There was little concern even for species differences. Within the same species, individual differences were viewed as a mere nuisance; it was believed that, by studying many subjects and by computing averages, basic laws of learning could be found. However, so-called laws were developed in this way that failed to represent even one individual whose behaviour contributed to the average. More than any other consideration, this has led learning theorists to take a belated look at the importance of individual differences and species differences in learning.

BIBLIOGRAPHY. J.F. HALL, ***The Psychology of Learning*** (1966), provides a comprehensive account of the empirical data that theories of learning are designed to integrate. E.R. HILGARD and G.H. BOWER, ***Theories of Learning,*** 3rd ed. (1966), is the standard reference on this subject. G.A. KIMBLE (ed.), ***Hilgard and Marquis' Conditioning and Learning,*** 2nd ed. (1961), describes theoretical issues as they apply to simple learning. The separate articles reproduced in KIMBLE, ***Foundations of Conditioning and Learning*** (1967), develop some of these arguments in more detail and also present synopses of the theoretical positions of Hull, Guthrie, and Tolman. G.A. KIMBLE and N. GARMEZY, ***Principles of General Psychology,*** 3rd ed. (1968), is a general textbook that covers several theoretical issues and should be particularly useful on the topics of memory and retrieval.

(G.A.K.)

Leather and Hides

Leather is animal hide or skin that has been processed for use by man. Large animals are said to have "hides" (cowhide, buffalo hide) while smaller animals have "skins" (goatskin, sheepskin). In either case, the hide or skin is composed of water and proteins and, unless preserved, decays quickly.

Basic steps in making leather. Leather is made from rawhide in three steps: (1) Removal of undesirable constituents such as hair, flesh, fat, and some interfibrillary matters, leaving a concentrated network of high-protein collagen fibres, greatly softened and interspaced with water. (2) Tanning; *i.e.*, treating the hide with an agent, called tannin, that displaces the water and then combines with and coats the collagen fibres. Tanning increases resistance to heat, hydrolysis (decomposition caused by water), and micro-organisms. (3) Finishing to obtain proper thickness, moisture, lubrication, and aesthetic appeal. Thus, leather is essentially animal skin protein combined with tannins, small amounts of oils, dyes, finishes, and moisture. The relative proportion and distribution of constituents varies with the type of leather.

By suitable choice of rawhide or skin and tanning method, the tanner obtains leather possessing such properties as the nonstretchiness in upholstery, the drape in clothing, and others.

Properties of leather for footwear

The properties that make leather eminently suitable for footwear are: (1) mechanical: resistance to abrasion, flexing, deforming, and stretch; tensile, tear, and burst strength; suitability for cementing, vulcanizing, and molding; (2) wearing comfort: permeability to air and water vapour; storage capacity for perspiration; thermal insulation; suppleness; (3) appearance: colour fastness; resistance to wet and dry rubbing; easy care.

Major types of leather. The many varieties of leathers produced throughout the world find uses in hundreds of applications. The following classification (condensed from Tanners' Council of America, *Dictionary of Leather Terminology*) lists the world's most important types.

Cattle group. This group includes steer (ox), cow, and bull hides, producing leather for boot and shoe soles, heels, inner soles, uppers; harness, saddles, skirting (for saddles), horse collars; travelling bags, suitcases, briefcases, straps; upholstery; fancy goods such as handbags, belts, gloves, garments; industrial leathers for aprons, buffing wheels, textile carders and combers, hydraulic packings and washers, lithographic purposes, machinery belting; footballs and other sporting goods; razor strops; rawhide.

Also included are kipskins (from large calves or undersized or small breeds of cattle) for shoe uppers; fancy leather goods and handbags; gloves and garments; shoe lining.

Calfskins for boot and shoe uppers and linings; gloves and garments; fancy leather goods and handbags fall under this heading; bookbindings; hat sweatbands, rawhide and parchment; military helmets and gas masks; grips for golf clubs; handicrafts.

Sheep and lamb group. Included are wooled skins, haired skins (cabrettas) for shoes, slippers, gloves, coats, hats; leather goods and handbags; aprons; chamois; parchment; piano action; rollers on textile machinery.

Goat and kid group. Primary uses are skins for shoe uppers; fancy leather goods and handbags; gloves and garments.

Equine group. Included in this group are horse, colt, ass, mule, and zebra hides for shoes; gloves and garments; sporting goods; luggage; belts.

Buffalo group. Domestic land and water buffalo provide leather for shoe soles and uppers, buffing wheels, luggage, and handbags.

Pig and hog group. This group provides pig, hog, boar, peccary, and carpincho leathers for fancy leather goods and luggage, gloves, harness saddlery, shoe uppers.

Aquatic group. Animals whose main habitat is water are seal, sea lion, and walrus for luggage, fancy leather goods, and buffing wheels; shark, whale, blackfish, dolphin, and porpoise for fancy leather goods, luggage, and shoe uppers; and alligator and crocodile for shoes, handbags, and luggage.

Miscellaneous. Skins of deer, kangaroo, ostrich, lizard, etc. are also tanned and made into various kinds of leather goods.

HISTORICAL BACKGROUND

Ice Age man used animal skins to protect his body against inclement weather and his feet against thorns and rocks. Preserved specimens of leather dating back to 5000 BC have been found. Evidence exists for the use of leather by ancient Sumerians (6000 BC), Egyptians, Indians, and Chinese, and later by Greeks and Moors.

Early uses of leather. Early uses can be grouped as follows: (1) apparel: trousers, shirts, footwear, gloves, headgear, belts; (2) containers and holders: wine and water casks, bags, pouches, luggage, tents, sheaths for knives, scissors, and swords; (3) transport and traction: chariots, whips, wagon covers, boats, saddles, bridles, harness; (4) recreation: musical instruments, puppets; (5) communication: parchment and vellum for writing and drums for transmitting messages; (6) war needs: slings, bows, quivers, shields, helmets, leg guards, ankle leather, armour, military clothing, powder horns; (7) interior uses: beds, cushions, upholstery, molded and ornamented articles, bookbinding, tapestries, and paintings.

The Industrial Revolution created other uses for leather such as washers, packings, industrial belting, and bellows; and present-day uses represent a continuous and continuing adaptation to changing circumstances.

Early manufacturing techniques. Leathers may have been man's earliest manufactured goods. Scraping the skin till it was more or less free of flesh by flint tools was probably the only operation carried out by primitive man. Sun drying prevented rapid decay but hides became hard. To make them soft and suitable for wear, animal fats and brains were rubbed in. Similarly, salt was applied to preserve hides, Skin, like meat, was also preserved by smoking.

The need for removal of hair, flesh, fat, and other constituents was slowly recognized. The processes of sweating and enzymatic hair removal are based on the early discovery that hair will fall out of a wet hide that has been allowed to rot slightly. Wood ash and lime were used in the Middle Ages. Much later, the cleaned hides were treated with dog, hen, and pigeon dung and drenched in fermented broth. These procedures were later replaced by acid deliming and bating with enzymes. Bating is a treatment of light types of leather to attain a smoother and clearer grain, or greater softness and pliability. In this process, hides are placed in vats containing ammonium sulfate or chloride as the deliming agent and certain enzymes that act selectively to remove proteins and also to improve colour.

Bating

Eskimos may have rubbed fish oils into animal skins to make the furs soft and durable. This may have been the beginning of the use of oils for tanning (a process termed oil tannage), and of chamoising (the art of making skins and hides extremely soft). Japanese white leather and Ethiopian red leather are still produced with rape and safflower oils.

Smoking of hides and skins involves a reaction with aldehydes, several of which today are used for tanning. Oxidation of oil also produces aldehydes. American Indians still make leather by a process that combines oil and smoke curing.

Development of tanning methods. *Vegetable tannage.* Vegetable tannage is quite ancient. The Hebrews used oak bark; the Egyptians, babul pods (400 BC), and the Arabs barks and roots. A wide variety of vegetable tans are used today. Synthetic tannins were introduced in 1911. In the classical method of vegetable tannage the pelts are placed in a pit or vat, in alternating layers with ground vegetable tanbark, pod, leaf, wood, or root. Water is poured on to cover, and the pelts left for 6 to 12 months. Even today, this "contact," or "sandwich," tanning method is employed with modifications. In another method, called "bag," or "bottle," tanning, the pelt is made into a bag, filled with tanning material and water, and tied to a pole. The water leaches from the material inside the bag, and the tannin diffuses osmotically through the pelt, completing the tanning in three to six days. Long before Christ, the Chinese cured skins with mud and alum. Alum, an accidental substitution for salt, led to tawing, the first mineral tannage. There is evidence for the early use of tawed leather in Assyria, Babylonia, Phoenicia, and India, and later by Greeks and Moors.

Two or more processes were sometimes combined; *e.g.*, alum and oil; alum and gallnuts or sumac leaf, producing soft nappa leathers. By the early 11th century AD, three basic tanning processes were used: (1) oil, (2) vegetable, and (3) alum.

Chemical tannage. A significant change came only toward the end of the 19th century. In 1858 the value of basic chrome salt for tanning was discovered. In 1884 a two-bath tannage process was patented; and in 1893, single-bath chrome tannage was introduced. Chrome tannage, today by far the principal tanning method for light leathers, is rapid; the leather produced has hydrothermal (heat and moisture) resistance to 120° C (248° F). Silica, aluminum, iron, and zirconium salts were also tried, but only zirconium and aluminum salts are used today to any extent.

Two-bath tannage

From early days, leather was made more decorative and adaptable by finishing. In the Middle Ages, vegetable dyes and earth pigments provided the various colours. Leather was frizzed (curled or tufted) to make nap or boarded to make Morocco leather. In boarding, the skin is doubled, grain surface inside, and pressure is applied in the required direction either by hand or with a cork-covered board, or by rotating cork-covered cylinders. The natural grain of the leather is enhanced or brought into sharper definition. The leather was also scented with birch tar oil and curried with grease. Today, a wide variety of dyes, oils, and finishing agents are used.

The earliest tanning tool was flint, used to scrape off flesh. Later, hair and flesh were removed by a curved blunt or sharp knife over a wooden beam. Leather was shaved to the required thickness with a sharp knife. Such hand tools were improved constantly. Machines were introduced in the early 19th century: a splitting machine in 1809 and fleshing and dehairing machines in 1840. Today, most operations may be done by machines, though the leather industry still defies automation because of the irregular size, thickness, and quality of hides handled. Empirical processing methods evolved from observation and experience. The scientific study of leather processing started only at the close of the 18th century. With progress in science, newer methods, materials, and machines have made possible a great variety of improved leathers.

MODERN LEATHER MANUFACTURE

Raw materials. Although any hide or skin can be tanned into leather, those most often used come from cattle, buffalo, sheep, goats, and pigs, and sometimes from horses, camels, crocodiles. snakes, and lizards. Skins of animals such as elephant, zebra, leopard, and tiger bring high prices. Animal skins vary in size, shape, texture, thickness, and quality. Quality depends on how and why the animal was reared, whether it died of old age or was slaughtered, and its age, sex, and environment. Damages of many kinds reduce the value of hides and skins. The most common defects encountered are those produced on the live animal by parasites and diseases, old age, and malnutrition; brand marks and barbed wire scratches; and on the dead animal by poor stripping of the skin, handling, and storage. With proper attention, many of these defects can be avoided.

Characteristics of animal skin

Mammalian hides and skins are divided into three layers distinct in structure and origin. These are (1) a thin outer layer of epithelial cells called the epidermis; (2) a thick layer called corium, or dermis; and (3) subcutaneous adipose or flesh layer. In tanning, the epidermis and flesh layers are removed; the corium is tanned into leather. One layer of the corium containing the grain membrane and hyaline layer, together with the arrangement of hair pores, gives a distinct grain surface pattern for each species of animal. The other corium layer is composed of large collagen fibre bundles interwoven at an angle in a three-dimensional network. The fibre structure varies in different parts of the skin (*e.g.*, belly-loose, butt-tight) and from one species to another (sheep-loose, goat-tight). Elastin fibres make leather that stretches. Fatty skins yield spongy leather.

Fresh hide is composed of 60 to 70 percent water by weight, 30 to 35 percent proteins, about 2 percent lipids, 0.5 percent carbohydrates, and 1 percent mineral salts and other substances such as pigments. Of the solid matter, 90 to 95 percent is made up of proteins, both fibrous and nonfibrous. Of the fibrous proteins, collagen accounts for over 85 percent of corium; elastin and reticulin are present in small quantities; hair and wool are made up of keratinous (horn-producing) fibres. The nonfibrous proteins are albumins (soluble in water), globulins (soluble in salt solutions), and mucins, mucoids, or glycoproteins (soluble in dilute alkalies). All soluble proteins are removed before tanning. The fibrous protein chains are held in position by chemical bonds giving a cohesive structure that may be broken down by heat and hydrolysis to form glue, or strengthened by tanning to produce leather. The science of leather manufacture is the science of skin proteins and their interactions with acids, bases, salts, and organic and inorganic tanning agents.

Preparation of hides for tanning. During preparation and shipping, protection against excessive heat, humidity, rain, and pests is essential. Raw hides are cured or preserved to prevent decay that begins within hours of slaughter. Curing consists of dehydration without disturbing skin structure. Common methods are simple air drying, which yields bony flint hides; salting, which includes treating the hide with a saturated solution (brined); rubbing the flesh side with salt (wet salted); further drying (dry salted); and pickling the pelt with acid and salt. Preservatives are also added to improve curing efficiency. Such methods as cold storage (5° C [41° F]), freeze drying, vacuum drying, irradiation, and solvent dehydration are relatively costly.

The purpose of pretanning, or beamhouse, operations is to remove undesirable constituents and to condition the skin for tanning. The first operation in the tannery beamhouse is the soaking of cured hides in water to rehydrate the hide to its original flaccid condition and to remove dirt, salt, and some soluble proteins. Flint hides need special care, as soluble proteins cement the fibres on air drying. Soaking is assisted by elevated temperature, agitation, and addition of wetting agents, protein dispersants, and preservatives. Well-soaked hides are washed, drained, and weighed.

Sweating and fermenting

Dehairing. Wool skins are still dehaired nonenzymatically by the old methods of sweating them in a hot moist room or keeping them overnight in a fermented bath. In the sweating process, the hide is allowed to rot partially; the hairs then fall out easily. Enzymes with specific chemical activity are available, but enzymatic dehairing is not popular because of the expense involved, and loss of commercial use of the hair, which dissolves or disintegrates. Liming, in which soaked hides are treated with lime for one to two weeks to dehair, is still an old standard practice. By addition of such agents as sodium sulfide, dehairing time is reduced to a few hours. For dehairing sheepskins, lime sulfide paste is applied on the flesh side to save the wool, which is pulled in 3 to 12 hours. Liming not only removes hair, fats, and soluble proteins but also swells and conditions fibre structure.

Deliming and bating. A hide with its hair and flesh removed is known as a pelt. Deliming is done partly or fully to remove excess alkali and is accomplished by treating with such mild acids as boric, acetic, lactic or such acid salts as ammonium chloride or sulfate and sodium bisulfite. The material is then bated to remove interfibrillary proteins and produce a clean white pelt. Bating is essential for special types of hides, such as those for gloving and glace kid leathers.

Pickling. Delimed, bated stock is pickled with a mixture of sulfuric acid and salt in a paddle or drum to prepare it for either long storage or mineral tannage. Fatty skins are degreased with detergent and paraffin solvents, washed, and repickled.

Modern tannery methods and machinery. ***Vegetable*** *tannage.* Vegetable tannage is still an important tanning method. Delimed pelts are treated with infusions of vegetable tanstuffs resulting in firm, full, solid leather. Almost every fibrous plant contains tannins. Some examples are: roots (canaigre, 25–30 percent tan); wood (chestnut, 6–8 percent); bark (mimosa or wattle, 38–40 percent; oak, 7-10 percent); leaves (sumac, 25 percent); fruits (myrobalan, 33–36 percent); pods (dividivi, 40 percent); parasitic growth (oak galls, 50–60 percent); rind (pomegranate, 20 percent); seed husk (cashew, 25 percent); and lignosulfonates, by-products of the paper industry.

Tanning material is ground, leached with water, and concentrated into solid or powdered extract. Sometimes it is sulfited and bleached to yield more soluble and light coloured extracts.

Vegetable tanstuffs contain tannins, which are amorphous, astringent, complex mixtures of polyphenols; and nontanning materials, including carbohydrates, salts, and acids, which also contribute to leather quality.

Vegetable tans are hydrolyzable or condensed. Hydrolyzable tannins are rich in acids and salts. They decompose with water into sugar and acids, and yield spongy leather. Condensed tans, most often used by tanners, yield red-coloured leather unless used in combination with hydrolyzable tans.

The classical year-long oak tannage starting with weak liquors in suspenders and with increasingly concentrated liquors in handlers and layers, or sandwiches, is seldom used now. Bag, or bottle, tannage is still practiced; two other modern methods are widely used. In the first, a delimed pelt is kept stationary and tan liquors are circulated in several sets of pits (rocker vats), terminating by hot pitting with concentrated liquors at 40" C (104" F) to get firmer leather; in the second, accelerated tannage is achieved by fully deliming a pelt; pretreatment with synthetic tanning agents, polyphosphates, or buffer salts; and then tanning with highly concentrated liquors or extracts in drums.

Tannage for belting and harness is the same except that loading and degree of tannage are lower, and stuffing is with 15 to 20 percent fat.

Light leathers

Where leathers are to be more flexible and light coloured, delimed skins are pickled, or pretreated with synthetic tannins, and tanned in paddles, or drums, with a blend of tan extracts for three to six hours. In dry or floatless tannage, powdered extract preceded by synthetic tannins is added onto the pelt in a running drum, leaving no effluent.

Synthetic tans. Synthetic tans originated from studies of the effect of formaldehyde on vegetable tanning materials. The first of them, patented in 1911, was based on phenols and cresols. Later, naphthalene was used. Synthetic tans must have a large polymer molecule and must also be soluble in water; they are made by condensing (forming into a polymer) simple phenol molecules and adding a sulfonate group to make the product soluble. The process also works in reverse, with sulfonation first and condensation second.

At present the wide variety of synthetic tannins are of three types: (1) auxiliary tans, strong simple sulfonic acids that are used to bleach chrome-tanned and vegetable-tanned leathers, to produce shrunken grain, and to aid dyeing; (2) combination tans, sulfonic acids of complex phenolic materials used for light tanning and for white leathers; (3) exchange, or replacement, tans, weak acids containing a large number of phenolic groups to substitute vegetable tans and to produce white, light fast leathers, but these are costly and not widely used. The aim at present is to make specialty products that will produce desired qualities in the leather.

Mineral tannages. The basis of mineral tannage is the use of protein carboxyl groups with a basic metallic salt to form rivets, giving high hydrothermal stability (120" C, 250" F) to leather. Chromium salt is widely used. Aluminum, iron, and zirconium salts are also used.

Chrome liquors are prepared by reducing orange-red acid dichromate to blue with such reducing agents as glucose, molasses, or sulfur dioxide. The liquors may be dried into crystal or powder form. Special extracts may also contain masking, wetting, tanning, and other agents.

Chrome tanning involves drumming the pickled belt in pickle or a separate bath with basic chrome liquor. Masking and other agents are added before, during, or after chrome tanning for special effects. Today the trend is toward the use of salt-free, masked, dry powder tannages. When tannage is complete in three to four hours, a small piece of leather is tested by plunging it into boiling water for two to three minutes (boil test). If it has been properly tanned, it should not curl.

In two-bath tannage the pelt is treated with chromic acid in the first bath. In the second bath the chromic acid is reduced with sodium thiosulfate (hypo) and acid, thereby fixing basic chromium sulfate within the pelt. Precipitated colloidal sulfur imparts strength to the leather and gives it a fine grain. Two-bath tannage involves much work and material waste; it has been almost completely displaced by single-bath tannage. Mixed single- and double-bath tannage is also possible.

Use of aluminum salts

Although an old alum tawing method is still used for dressing furs, basic aluminum sulfate masked with citrate or tartrate is more effective. Aluminum salts often are used in combination with formaldehyde to produce furs and white leathers; in pickling and pretannage for chrome; and in pretannage and retannage of vegetable leathers to give soft, full nappa leathers and perspiration-resistant insole leathers.

Recognized in 1907 and patented in 1933, zirconium tannage employs basic zirconium sulfates masked with acetate, citrate, and other agents to produce dense, dry, full white leathers. Zirconium salts are used as pretanning or retanning agents along with chrome and other tannages to produce kid, suede, and other leathers. The high cost and care required limit the use of this method, though using zirconium in conjunction with silica reduces the cost.

Silica's tanning potency, observed in 1862, was applied and described in 1915. Silica hydrolyzes, polymerizes, and precipitates. Silicate solutions gel readily and fix only on the surface. Sodium metasilicate is considered to give better tannage. Although silica is cheap and abundant, silica tannage is not used commercially because it is not complete and uniform, and leathers are not strong.

Miscellaneous tannages. The oldest oil tanning produces chamois leather, which, unlike other leathers, can be repeatedly wetted and dried without detrimental effects. The special properties of chamois leather result from its open structure. Chamois leather is tanned with aldehydes and peroxides, and coated with polymers from oxidation of fish oil used in the tanning process. Long liming, bating, and mechanical pressing open up and split the fibres; stocking and pounding with fish oil and hanging in heated rooms oxidizes the fish oil. Compounds of copper, manganese, and cobalt accelerate oxidation. Excess oil is washed off with washing soda solution or wetting agents. Pretanning with formaldehyde is common. Though cod oil is favoured, sardine, rape, and safflower oils are also used.

During World War II, sulfochlorinated hydrocarbons were substituted for scarce fish oil when producing white chamois leather. The pelt pretreated with formaldehyde is tanned with a mixture of sulfonyl chloride, wetting agent, and sodium carbonate for four to eight hours. The method is rarely used today, because of its high cost.

Commercially accepted aldehyde tannages include formaldehyde, glutaraldehyde, and dialdehyde starch. Bated pelts are tanned with formaldehyde solution for two to four hours and the excess flesh and bacterial matter in the hide is removed by treatment with magnesium acetate or hydrogen peroxide to produce white leathers. Formaldehyde is mainly used in combination with other tannages. Glutaraldehyde tanning gives leather a yellow colour and high resistance to perspiration and washing with soap, making it suitable for superior shoe and glove leathers.

Resin, or polymer, tannage, employing such materials as urea or melamine, produces fuller and tighter leathers. Engineering and economic problems of solvent tanning remain, but interesting results may be obtained; *e.g.*, completion of vegetable tannage in a few minutes, and production of leathers with a new variety of properties. Quinone, a simple chemical, has good tanning potency and gives good tear resistance to leather, but is seldom used.

The first tannage

Each tanning material imparts special characteristics to leather. Desired properties are obtained by combination of two or more tannages consecutively or simultaneously. The first tannage, however, fixes the character of the leather. Chrome leather is retanned with vegetable to give thicker chrome retan uppers. Vegetable tanned leather is retanned with chrome to give semichrome uppers and suede and glove leather. Invariably combination tannage is the current practice for most leathers.

MODERN LEATHER PROCESSING

Dyeing. After excess water and wrinkles are removed from tanned leather, it is shaved or split to uniform thickness and dyed. Leather dyes may be classified as follows: (1) cationic (basic) dyes, used mainly for vegetable or synthetically tanned leathers; (2) anionic (acid and direct) dyes, used on chrome, vegetable, and aluminum tanned leathers; direct dyes give good surface dyeing with no affinity to vegetable leathers; (3) amphoteric dyes, which easily penetrate chrome but not vegetable leathers; (4)

chrome mordant dyes, which give deep and fast shades; used in gloving leathers; (5) metal incorporated premetallized dyes, which are slow to penetrate but give good levelling and colour fastness despite exposure to alkali and washing; (6) reactive dyes, which will not run when dry cleaned: used mainly for suede and glove leathers.

Dyes are applied to leather either by brushing on one side, tray dyeing, drum dyeing, spray dyeing, solvent dyeing, vacuum dyeing, or dyeing through a tank and squeeze rollers. Fancy colouring is accomplished either by dabbing dye with a sponge, sprinkling it on folded skin, spraying, padding through stencils, or tying and dyeing. Levelness, penetration, richness of shade, and fastness to light, dry cleaning, and rubbing are affected by acidity, temperature, type of leather, and dye concentration and special additives.

Lubrication. Unless lubricated, leather dries hard. Dyed leathers are treated with oils and fats for lubrication, softness, strength, and waterproofing. Oils and fats are incorporated individually, or in blends, in natural condition, as emulsion, or in solvents, by hand or in a drum.

Vegetable tanned light leathers are oiled by swabbing groundnut (peanut) oil on the grain surface. Sole leather is drummed with a mixture of vegetable oil, mineral oil, and small amounts of sulfated oil, epsorn salts, and glucose. The oil keeps loose tanning material from rising to the grain during drying, and produces supple, light-coloured leather.

Belting leather is treated by hand or in a heated drum with a mixture of cod oil, tallow, wool grease, stearine, and paraffin wax. Currying is similar, incorporating such mixtures as a hot melt.

Fat-liquoring

Light leathers are fat-liquored in a drum with an oil and water emulsion. The uniform penetration of the leather by fat liquor gives a soft, stretchy, loose-grained leather; if the fat liquor is deposited only on the surface, the leather is resilient and tight. Raw oils mixed with emulsifiers give desired properties of softness and pliability to suede. Raw oils mixed with soaps or sulfated oils are commonly used. Such vegetable oils as castor, palm, and groundnut; animal oils as tallow and neat's-foot; marine oils as cod, sperm, and sardine; and mineral oils and fatty alcohols are sulfated with sulfuric acid at low temperatures. The greater the reaction, the greater the penetration and stability of the product. Sulfated oils are favoured for gloving, suede, and soft leathers. Oils may be introduced into leather along with a solvent and the solvent then evaporated.

Colloidal graphite and sulfur give mechanical slip to leathers to be used for washers and oil seals. Silicones, fatty acids, waxes, chlorides of chromium, aluminum, and other metals are used for shower proofing or waterproofing of leathers. To avoid the yellowing of aging, especially of white leathers, synthetic oils and sulfated coconut oils are used. The use of newer synthetic tannins, impregnants, and finishes, along with vulcanizing and cementing of shoes, makes constant new demands upon leather lubricants.

Drying and finishing. After dyeing and fat-liquoring, leather contains 45 to 60 percent water and is dried to about 14 percent moisture, chemical and physical reactions taking place. When leather dries to a paler shade, loose tannins, dyes, and oils spread uniformly, penetrate deeply, and are fixed firmly. Uneven drying causes the migration of unfixed tannin dye and oil to the surface, resulting in dark stains in loose areas. During drying, leather tends to shrink, particularly when the water content falls below 25 percent and drying temperatures are high. Vegetable tanned leathers shrink less than others.

Drying techniques. Among the most popular drying techniques are: (1) Air drying: skins are hung from hooks or on sticks or are placed on horizontal racks and dried by the natural passage of air or with the help of overhead fans. Vegetable tanned leathers are air dried. The process is cheap and yields slow drying and gives a uniform, light-coloured product, but productivity is low. (2) Tunnel, or cabinet, countercurrent drying: drying tunnels or cabinets with controlled temperature, humidity, and countercurrent air circulation provide efficient drying. Fresh dry warm air meets dry leather and subsequently moves toward wet leathers. Leathers may be hooked to a moving chain or held on moving toggle frames or plates. (3) Drying under tension: leathers are stretched out, nailed or tacked on wooden boards or stretched on a perforated metal frame by toggles or grip clips, and dried by either process (1) or (2). Toggling gives better air circulation on both sides. (4) Paste drying: leathers are pasted by the grain side on both sides of large plates of glass, porcelain, or metal and then travel through a tunnel drier consisting of four to eight zones of controlled temperature and humidity. In three to six hours, the dry leather is stripped off. Pasting units are expensive; thus they are used largely for upper, split, and lining leathers. They yield flat, smooth grain and larger area. (5) Vacuum drying: leather is spread out, grain down, on a smooth polished metal plate heated by a built-in thermostatically controlled heat exchanger, covered by an airtight hood. The air is then removed from under the hood. Drying to 14 percent moisture is complete in ten minutes for vegetable tanned leather at 50° C (125" F) and three minutes for chrome leather at 80" C (180" F). Vacuum driers are costly, but operate independently of weather conditions, and productivity is high. (6) Others: dyed, seasoned, finished leathers are dried by moving through a tunnel heated by infrared heaters. Radio-frequency heating has the potential of removing water without heating the leather itself.

Paste drying

Finishing. Finishing involves mechanical and chemical treatments. The dried tanned leather is first trimmed, conditioned with damp sawdust to obtain uniform moisture content (around 20 percent), and staked by hand over a steel blade or by machine to stretch and soften it.

Some leathers are impregnated with synthetic resins (acrylics, butadiene, and polyurethanes) to improve their properties. Polyurethanes give the best scuff resistance. Impregnated leathers are usually buffed with an abrasive paper to remove surface blemishes, and to lift nap fibres in suede leather.

Leather is further finished by coating the grain surface. A good coating adheres well to the leather surface, is elastic, and resists abrasion, cracking, peeling, rubbing, light, heat, cold, and moisture. A wide variety of colours, pigment, resins, lacquers, and waxes are used. In unpigmented seasoning, a simple glazing finish or seasoning may contain egg albumin, water, and glycerin; shellac, beeswax, and casein made soluble in mild alkali will also give a glossy surface; pigmented finishes may be aqueous or nonaqueous and usually contain pigment (6–50 percent), protein or resin binder to bind and disperse pigment (10–20 percent), plasticizer (5–20 percent) to soften, dyestuff (0.5–6 percent), a preservative (0.5–2 percent), and water or another solvent. Binder-free pigment finishes are concentrated pigment colours dispersed with a wetting agent. Pigments may be earth pigments, such as iron oxide; or manufactured inorganic pigments, such as lithopone, organic pigments, lakes, and black carbon. Binders may be proteins, such as casein and shellac (used mainly for glaze finish), or synthetic resins for matte finish. The pigment finishes cover the surface defects well; aniline finish is used for leathers that need a natural look. This finish, a casein-free organic pigment in emulsion, is applied with very little loading; lacquers used as finishes are based on nitrocellulose in solvent or in water emulsion, vinyls, urethanes, and alkyds. All are produced in clear, aniline, and pigmented form and applied where gloss and resistance to rubbing are required, as in upholstery, grain garments, and Morocco leathers. Urethanes, instead of partly boiled linseed oil, pigment, and driers, are specially used for patent leather.

The application of finishes

Finishes are applied by hand brush, pad, or spray (with or without air power); or by a seasoning machine, using mechanical padding on a conveyor system; or by means of a flow coater in which the leather passes through a curtain of finish. Automatic spraying is also used.

Water resistance is obtained by treating leather with silicones and waxes.

Glaze finishes may require further treatment by a glazing machine. Matte or resin finished leather is usually plated by passing it between polished, heated cylinders. Sometimes it is embossed with artificial designs by a hydraulic press. Certain types of leathers are boarded by rolling the leather upon itself by hand or machine.

Sole leathers are bleached with oxalic acid or synthetic tannin, seasoned, and rolled under pressure. Belting leathers are stuffed and stretched. Shoe upper leathers with corrected grain are impregnated, buffed, pigment finished, and top dressed with one or two coats of lacquer emulsion. Full natural-grain calf or kid leathers are aniline finished or dilute-protein coated after they have been glazed and ironed to get a glossy, smooth, natural grain.

By proper formulation and combination of materials and methods, desired effects can be obtained in finishing of leather to meet the specific end use.

TANNERY BY-PRODUCTS

Rawhide trimmings

Rawhide trimmings are used in the manufacture of edible gelatin; fleshings for glue, oil, manure, and animal fodder; hair for drugget (coarse cloth), carpets, felt, and manure; soak water for growing coconut trees; lime sludge for cement and road making; tanbark for cardboards and chipboards and to make activated carbon; leather trimmings and shavings for leather board and fertilizer. Effluent water may be treated and used in agriculture.

The leather industry is traditionally associated with bad odours and pollution. Tanneries are generally located close to a river to ensure a plentiful water supply for processing and effluent disposal. Tannery effluent includes proteins, hair, salt, lime sludge, sulfides, acids, chrome and other tanning agents, dyes, and oils. Treatment of this effluent involves screening of hair, precipitation settling and removal of chrome, vegetable tans, and proteins, and filtration and aeration of the effluent. Salt is the most difficult of all effluent ingredients to remove. Sulfides are decomposed by manganese oxidation. Several effluent treatment systems have been proposed, one being the so-called Dutch ditch raceway system, which degrades sludge by biological treatment. Effluent treatment adds a penny per square foot to the cost of leather. Another approach to the pollution problem is to reduce the volume of water and chemicals used and to centralize all tannery wet work up to pickling near the slaughterhouses.

ECONOMICS AND SCOPE OF LEATHER INDUSTRY

Leather is an international commodity and a valuable traditional industry. The total value of the world's leather and allied industries is estimated to be more than $2,000,000,000; the international leather trade is around $1,000,000,000 per year. The industry provides employment to a large number of people. Because a tannery can use either a great deal of labour and low capital investment or a great deal of capital and low labour supply, depending upon the processes used, this industry can be established in both developing and advanced countries. Leather is also a strategic material for defense. Leather's unique properties, rising world population and incomes, new uses for leather, rapid changes in fashion, and more intensive use of resources in developing countries have contributed to an ever-increasing demand for leather. The use of synthetics as leather substitutes, however, is growing. Man-made products in the advanced countries were supplanting 50 percent of leather goods, 70–80 percent of sole leather, and 2–15 percent of the upper leather in footwear in the early 1970s.

World hide production

The world produces around 6,500,000 metric tons of hides and about 1,300,000 metric tons of skins annually. The largest producers of hides are the Soviet Union, the United States, and India; goatskins, India and China; sheepskins, the Soviet Union, New Zealand, and Australia.

The world leather trade is largely concentrated in advanced countries (mostly in the Organization for Economic Co-operation and Development group) that trade among themselves. The major exporters of finished leathers are India, Italy, Japan, West Germany, and the United Kingdom. Italy is the largest exporter of footwear, followed by Japan.

Though the developing nations of Asia, Africa, and Latin America raise nearly 60 percent of the world's cattle, three-quarters of the goats, and more than 40 percent of the sheep, they do not realize the full value of the hides produced because slaughter is inefficient and poorly organized. Developing nations export mostly rawhides and skins and some leathers; they import finished goods.

Tanning is still a craft industry on the cottage level in the developing nations. Modern technology, however, promises to allow these nations fuller use of their animal resources through the establishment of efficient slaughterhouses and large-scale manufacturing units. The development of tanning industries will enable these nations to meet their own internal demands for leather and become exporters of leather and finished leather goods. In the more advanced countries, the scarcity and high cost of labour and the reluctance of workers to perform such operations as dehairing and fleshing, are requiring the purchase of more and more leather, rather than rawhides and skins, from the developing nations. Thus the leather industry may undergo a significant shift from the advanced countries to the developing nations.

BIBLIOGRAPHY. H.R. PROCTOR, ***A Textbook of Tanning*** (1885), a work by an early pioneer in leather science; J.A. WILSON, ***Modern Practice in Leather Manufacture*** (1941), a compendium of information that is still of value to the tanner, with glossary and bibliography; J.W. WATERER, ***Leather in Life, Art and Industry*** (1946), a history of the development of leather making and leather craft in Britain, with glossary and selective bibliography; F. O'FLAHERTY, W.T. RODDY, and R.M. LOLLAR (eds.), ***The Chemistry and Technology of Leather,*** 4 vol. (1956–65), a comprehensive monograph dealing with the theory and practice of leather manufacture, with an extensive bibliography; K.H. GUSTAVSON, ***The Chemistry of Tanning Processes*** (1956), a detailed discussion of the chemical bases of the different types of tannages used in leather manufacture, with an extensive bibliography; D. WOODROFFE, ***Leather Manufacture,*** 2nd rev. ed. (1950), ***Leather Dressing, Dyeing and Finishing*** (1953), two textbook-level works; T.C. THORSTENSEN, ***Practical Leather Technology*** (1969), a general introduction to the principles of leather technology, including information on economic factors and uses for leather; C. SINGER ***et al.*** (eds.), ***A History of Technology,*** 5 vol. (1954–58), traces the development of the different stages in leather making and the uses for leather over the ages.

(Y.N.)

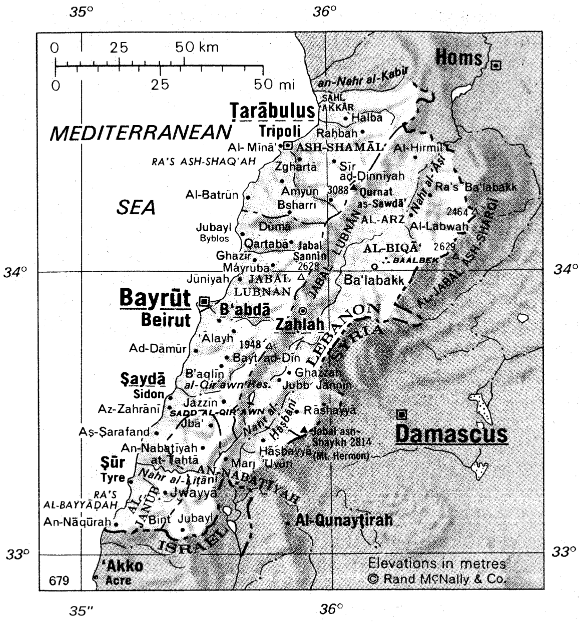
Lebanon

Boundaries

The Republic of Lebanon, a predominantly mountainous country of great scenic beauty, is an Arab republic situated on the eastern shore of the Mediterranean Sea. Consisting of a narrow strip of land about 135 miles (215 kilometres) long from north to south and 20 to 55 miles (30 to 90 kilometres) wide from east to west, it is bounded to the north and east by Syria and to the south by Israel. With an area of 3,950 square miles (10,230 square kilometres), Lebanon is one of the world's smaller sovereign states. The capital is Beirut. In 1978 the country's population was estimated by the United Nations at 3,152,000; the Civil War of 1974–76 led to the emigration of about half a million people, but by 1978 about half of these had returned.

Though Lebanon, particularly its coastal region, was the site of some of the oldest human settlements in the world—the Phoenician ports of Tyre (modern Ṣūr), Sidon (Ṣaydā), and Byblos (Jubayl) were dominant centres of trade and culture in the 3rd millennium BC—it was not until 1920 that the contemporary state came into being. In that year, France, which administered Lebanon as a League of Nations mandate, established the state of Greater Lebanon, which gave the country its present frontiers. Lebanon then became a republic in 1926 and achieved independence in 1943. It is a charter member of the Arab League.

As an Arab republic, Lebanon shares many of the cultural characteristics of the Arab world, yet it has evolved a distinct mode of life and some special socioeconomic and political attributes that differentiate it from many of its Arab neighbours. Its rugged, mountainous terrain has served throughout history as an asylum for diverse religious and ethnic groups and political dissidents.



LEBANON

This, along with the gradual movement of people in the Mediterranean area, has generated its uniquely cosmopolitan character and its receptivity to change. Lebanon is the second most densely populated country in the area. It has the highest rate of literacy. Until 1975 it had the highest standard of living, and—with the exception of the oil-rich states of the Persian Gulf—the highest per capita income. Although its prosperity is unevenly distributed, having bypassed large segments of its population, wealth and privilege appear to be evenly distributed among its middle-income group. Notwithstanding its meagre natural resources, Lebanon managed to become a busy commercial and cultural centre for the Middle East.

This outward image of vitality and growth nevertheless disguised some serious problems. Not only was Lebanon grappling with internal problems of social and economic organization but also, with its Arab neighbours, it was playing a role in a common struggle and was similarly vulnerable to the political vicissitudes of the Middle East. Its pluralistic communal structure gave legitimacy to the vertical hierarchies that dominate the political life of the country. Horizontal relationships that went beyond religious and sectarian affiliations did not have an interest in the system of government. Demands for a secular state were considered unrealistic. The pressures and the complex issues that arose from the absence of a solution to the Palestinian problem led to the breakdown of the system and to a crisis that remained almost unmanageable.

This article deals with contemporary Lebanon. For historical background, see CALIPHATE, EMPIRE OF THE; OTTOMAN EMPIRE AND TURKEY, HISTORY OF THE; and SYRIA AND PALESTINE, HISTORY OF. See also BEIRUT.

THE LANDSCAPE

The natural environment. ***Relief.*** As in any mountainous region, the physical geography of Lebanon is extremely complex and varied. Landforms, climate, soils, and vegetation undergo some sharp and striking changes within short distances. Four distinct physiographic regions may be distinguished: a narrow coastal plain along the Mediterranean Sea, the Lebanon Mountains (Jabal Lubnān), al-Biqāʿ (Beqaa) valley, and the Anti-Lebanon and Hermon ranges running parallel to the coastal mountains.

Physiographic regions

The coastal plain is narrow and discontinuous, almost disappearing in places. It is formed of river-deposited alluvium and marine sediments, which alternate suddenly with rocky beaches and sandy bays, and is generally fertile. In the far north it expands to form the ʿAkkār Plain.

The snow-capped Lebanon Mountains are the most prominent feature of the country's landscape. The range, rising steeply from the coast, forms a ridge of limestone and sandstone, cut by narrow and deep gorges. It is approximately 100 miles (160 kilometres) long and varies in width from 35 to six miles (55 to 10 kilometres). Its maximum elevation is at Qurnat as-Sawdā' (10,131 feet, or 3,088 metres) in the north, where the renowned cedars of Lebanon grow in the shadow of the peak. The range then gradually slopes to the south, rising again to a second peak, Jabal Ṣannīn, northeast of Beirut. To the south, the range gives way to the hills of Galilee, which are lower. The limestone composition of the mountains provides a relatively poor topsoil. The lower and middle slopes, however, are intensively cultivated, the terraced hills standing as a scenic relic of the ingenious tillers of the past.

Al-Biqāʿ

Al-Biqāʿ valley lies between the Lebanon Mountains in the west and the Anti-Lebanon Mountains in the east; its fertile soils consist of alluvial deposits from the mountains on either side. The valley, approximately 110 miles long and from six to 16 miles wide, is part of the great East African Rift System that stretches southward to Mozambique. In the south. al-Biqāʿ becomes hilly and rugged, blending into the foothills of Mt. Hermon (Jabal ash-Shaykh) to form the upper Jordan Valley.

Much like the Lebanon Mountains, the Anti-Lebanon (al-Jabal ash-Sharqi) starts with a high peak in the north and slopes southward until it is interrupted by the compelling heights of Mt. Hermon (9,232 feet).

Drainage. Lebanese rivers, though numerous, are mostly winter torrents, draining the western slopes of the Lebanon Mountains. If not fed by large springs they are only seasonal. The only exception is the Līṭānī (90 miles long), which rises near the famed ruins of Baalbek and flows southward in al-Biqāʿ to empty into the Mediterranean near historic Tyre. The two other important rivers are the Orontes (Nahr al-ʿĀṣī), which also rises in the extreme north of al-Biqāʿ but flows northward through Syria, and the Kabir, which forms the northern boundary between Lebanon and Syria. The Bārid, the Abū 'Ali, the Awwali, and others are only mountain streams.

Climate. There are sharp local contrasts in climatic conditions. Lebanon is included in the Mediterranean climatic region, which extends westward to the Atlantic Ocean. The winter storms formed over the ocean move eastward through the Mediterranean, bringing rain at that season; in summer the Mediterranean receives no rain. The climate of Lebanon is generally subtropical and is characterized by hot, dry summers and mild, humid winters. Mean daily maximum temperatures range from 90° F (32° C) in July, on the coast and in al-Biqāʿ, to 60° F (16° C) on the coast and 50° F (10° C) in al-Biqāʿ in January. Mean minimum temperatures in January are 45° F (7° C) or? the coast and 35° F (2° C) in al-Biqāʿ. At 5,000 feet (1,500 metres), the altitude of the highest settlements, these are reduced by about 15° F (8° C).

Nearly all precipitation falls in winter and averages 30 to 40 inches (750 to 1,000 millimetres) on the coast, rising to more than 60 inches in higher altitudes. Al-Biqāʿ is drier and receives 15 to 25 inches (400 to 600 millimetres). On the higher mountaintops, this precipitation falls as heavy snow that remains until early summer.

Cedar trees

Vegetation and animal life. ***Vegetation.*** Lebanon was heavily forested in ancient and medieval times, and its timber—particularly its famed cedar—was exported for building and shipbuilding. (There are many references to the cedars of Lebanon in the Bible; in I Kings 7:2, King Solomon's palace is called the House of the Forest of Lebanon.) The natural vegetation, however, has been grazed, burned, and cut for so long that little of it is regenerated. What survives is a wild Mediterranean vegetation of brush and low trees, mostly oaks, pines, cypresses, firs, junipers, and carobs.

Animal life. Few large wild animals survive in Lebanon, though bears are occasionally seen in the mountains. Among the smaller animals, deer, wildcats, hedgehogs, squirrels, martens, dormice, and hares are found. Numerous migratory birds from Africa and Europe visit Lebanon. Flamingoes, pelicans, cormorants, ducks, herons, and snipes frequent the marshes; eagles, buzzards, kites, falcons, and hawks inhabit the mountains; and owls, kingfishers, cuckoos, and woodpeckers are common.

THE PEOPLE

Lebanon has a heterogeneous society composed of numerous ethnic, religious, regional, and kinship groups. Primordial attachments and local communalism antedate the creation of the present territorial and political entity and continue to survive with remarkable tenacity. The small size of the country, an effective communications network, and the mobility of the Lebanese, however, are gradually attenuating traditional loyalties.

Ethnic and linguistic groups. Ethnically, the Lebanese compose a mixture in which Phoenician, Greek, Byzantine, crusader, and Arab elements are discernible. Throughout the ages there has been an influx of tribes and families from the Arabian Desert. While the dominant strain in the mountains is related to the Armenians of Asia Minor and the Caucasus, inhabitants of the coastal towns and al-Biqāʿ are descended from inhabitants of the hinterlands of Syria, Palestine, and Arabia.

Arabic is the national language, but French and English are widely spoken, and many educated Lebanese are trilingual. About 5 percent of the population is Armenian-speaking, and Syriac is used in the liturgies of some of the churches of the Maronites (Roman Catholics following an Eastern rite).

Religious groups. Perhaps the most distinctive feature of Lebanon's social structure is its varied religious composition. From the 7th century onward, Lebanon has served as a refuge for persecuted Christian and Muslim sects. The population today is estimated to be 60 percent Muslim and 40 percent Christian; altogether, there are 17 recognized religious sects. Although no official census has been taken since 1932, and although it is now believed that Shi'ite Muslims are the most numerous group, the Maronites have the largest representation in the National Assembly. Among the Christians, Maronites form the largest group (28 percent), followed by Greek Orthodox (10) and Greek Catholics (6). Among the three Muslim denominations, the Shi'ites are followed closely by the Sunnites; the Druzes have remained third, with about 6 percent of the population. There is also a minority of about 7,000 Jews.

Demography. There has been no full census since 1932. In 1978 the population was estimated at 3,152,000. There were about 400,000 Palestinian refugees in the country, and the civil war that was waged after 1974 led to the emigration of about half a million people. Pres. Elias Sarkis in November 1978 stated that, to that date, more than 30,000 Lebanese had been killed and nearly 200,000 mutilated; other estimates put the dead at around 60,000, but this figure includes Syrians and Palestinians.

A youthful population

Because of the country's relatively high fertility (it has a birth rate between 39 and 42 per 1,000) and its rapidly declining mortality rate (9.9 per 1,000), the Lebanese population is extremely youthful. Estimates made in 1978 indicated that 43 percent of the people were 14 years old or less. There was a sex ratio imbalance, with 49 percent female and 51 percent male. Infant mortality (about 14 per 1,000 live births) was much lower than the average rate for Arab countries (127 per 1,000).

One of the most salient demographic features of Lebanon is the uneven distribution of its population. While the country has an overall density of more than 800 persons per square mile (300 per square kilometre), density soars to about 69,000 per square mile (27,000 per square kilometre) in Bayrūt *muḥāfaẓah* (Beirut governorate) but diminishes to 123 in al-Biqāʿ governorate. In other governorates the density is: al-Janūb, 324; ash-Shamāl, 484; Jabal Lubnān, 1,106 (no estimate for an-Nabatiyah, created in 1975).

Before the civil war, the movement of people from rural areas was a major factor in the country's soaring rate of urbanization. About 60.1 percent of the population (1970) lived in urban agglomerations of 10,000 or more, only 39.9 percent in rural areas. Nearly two-thirds of the internal migration had been toward Beirut, which by 1970 had an estimated population of 940,000—86 percent of Lebanon's urban population and 44 percent of the country's total. The civil war led to a substantial return of people to their villages, but accurate figures are not available.

THE NATIONAL ECONOMY

Until 1974, Lebanon sustained a dualistic economy that reconciled traditional and modern elements. The traditional agriculture sector provided primary employment for approximately 20 percent of the labour force but generated only 9 percent of the national income; industry generated 21 percent of income, and services, which employed only 15 percent of the labour force, generated 70 percent. The growth of services was related mainly to international transport and trade and to the position of Beirut as a centre of international banking and tourism.

Economic losses after 1974

Per capita income grew from about $250 U.S. in 1950 to $1,070 in 1974. This figure, however, dropped below $1,000 as a result of the civil war. The war left deep scars on Beirut, several other cities, and many villages. Although only a preliminary inventory of private and public losses attributable to the war has been made, extensive destruction in all sectors, but especially in housing, trade, and public services, is evident, with damages roughly estimated at $2,500,000,000 U.S. Overall, the country's productive capacity was reduced by as much as 40 to 50 percent by the end of 1976, as compared with 1974. The greatest reduction in productive capacity seemed, according to a World Bank report of June 1977, to be in services, followed by industry and agriculture.

Sources of national income. ***Agriculture.*** Arable land is scarce, but the climate and the relatively abundant water supply from springs favour the intensive cultivation of a variety of crops on mountain slopes and in the coastal region. On the irrigated coastal plain, market vegetables, bananas, and citrus are grown. In the foothills, the principal crops are olives, grapes, tobacco, figs, and almonds. At higher altitudes (about 1,500 feet, or 460 metres), peaches, apricots, plums, and cherries are planted, while apples and pears thrive at an altitude of about 3,000 feet. Sugar beets, cereals, and vegetables are the main crops cultivated in al-Biqāʿ. New techniques in agriculture have done much to increase output; Lebanon was formerly an importer of apples, but in 1974 about 160,000 tons were exported. With the completion of the Sadd al-Qir 'awn (Karaoun Dam) across the Līṭānī River in 1965, the amount of irrigated land increased. The government also formulated a "Green Plan" (1964) to restore neglected land and to promote reafforestation. Poultry has become a major source of agricultural income.

Industry. The mineral resources of Lebanon are few. Minor deposits of high-grade iron ore, asphalt, coal, and phosphates are exploited. There are numerous building-stone quarries, and there is some high-quality sand, suitable for glass manufacture, as well as lime.

The industrial sector grew in the 1960s, particularly in quality and diversification of products. Before the war, leading industries were food processing, manufacture of textiles, plywood, and furniture, and printing. Chemicals, pharmaceuticals, and aluminum products were also growing in importance. There were two steel mills, three cement factories, and two oil refineries.

Oil reserves

In April 1975 the Ministry of Industry and Oil announced that commercially workable oil reserves had been discovered in the 'Akkar Plain, western al-Biqāʿ, and the Mayrūbā area near the central coast. Bids for concessions were postponed pending the end of the war. In January 1977, within weeks of the final cease-fire agreement, the government set up a new Higher Oil Committee, putting it in charge of oil imports, exploration agreements, and the country's refineries.

About 70 percent of the country's industry survived the civil war unscathed. Beirut's industrial belt was razed, but some of the country's big complexes were unharmed. Manufacturing in 1978 was back at 60 percent of the still existing capacity, restrained by limited labour mobility, difficulty in acquiring supplies, insufficient working capital, and difficulty in obtaining credit.

Trade and services. Until 1974, trade continued to reflect the mercantile character of the Lebanese economy. Almost anything for sale anywhere could have been bought in Beirut. An enormous propensity to consume generated a huge trade deficit, which was partly covered by "invisible" items such as foreign remittances and transit and tourist services. The trade pattern remained the same during the war, though at a reduced level.

Finance. The monetary unit is the Lebanese pound. The finance sector of Lebanon's economy, including banking and insurance, showed an impressive expansion before the war. Beirut had 72 officially recognized banks (35 Lebanese, 19 foreign, and 18 Lebanese with foreign participation). According to statistics of the International Monetary Fund, monetary reserves of Lebanon continued to rise despite political uncertainties. At the end of 1974, external reserves stood at $1,700,000,000 and, despite the two years of civil war and the extended period of domestic instability following that, reserves continued to rise, reaching $2,000,000,000 at the end of 1977 and $2,260,000,000 at the end of 1978. These amounts included, according to the same source, a considerable portion of gold, which, if valued at the market prices of December 1978, would increase these reserves by approximately $2,000,000,000 more.

Capital inflows

In the late 1970s the balance of payments indicated a surplus. Furthermore, the Lebanese pound appreciated, vis-a-vis the U.S. dollar, after the end of 1977. The strength of the Lebanese pound and of the balance of payments position continued to reflect large capital inflows, mostly from Lebanese living abroad (whose numbers rose considerably during and after the war) and from the high level of liquidity of commercial banks.

Beirut's well-developed seaport and airport and the country's free economic and foreign exchange systems, favourable interest rates, and banking secrecy law (modelled upon that of Switzerland) all contributed to the preeminence of trade and services. Until 1973, tourism was foremost among these services. The country's scenery, its biblical and other historic sites, its hotels, bars, nightclubs, and restaurants, its seaside and mountain resorts, its outdoor sports facilities, and its international cultural festivals had made tourism an all-year-round industry. Tourism was severely damaged by the 1974–76 crisis. As the war progressed, the prosperous hotel district in Beirut became the scene of some of the fiercest fighting, and hotel activity in 1975 dropped by 80 percent below the previous year. The closure of Beirut airport for nine months, the heavy destruction of the port, and the continued political unrest greatly damaged the service industry. The government in its reconstruction policy made plans, ready to be carried out upon the resolution of the security problem, to revive this sector of the economy.

Management of the economy. Lebanon has a liberal economy, characterized by a minimum of government intervention in the activities of private enterprise. There is, however, a government monopoly on the manufacture and sale of tobacco, and some of the railways and public transport systems are now operated by the government.

Though the country had no specific long-term economic development plan, the government spent (between 1971 and 1975), in addition to the regular budget, a sum of $250,000,000 U.S. on development schemes. One of the most important of these was the Līṭānī project, which, when completed, was expected to increase the irrigated area of Lebanon by 50,000 acres (20,000 hectares) and to make possible the production of an additional 626,000,000 kilowatt-hours of electricity per year.

Trade unions. Lebanon is one of the few countries in the Middle East with a comparatively well-developed labour movement. During the war, divisions in many of the trade unions weakened their normal functions, and many of their 60,000 members, distributed among about 170 licensed unions, joined the different warring factions. Many others emigrated. Though the trade unions have secured some tangible gains, such as fringe benefits, collective bargaining contracts, and better working conditions, the movement remains fragmented by regional, religious, and ideological divisions.

TRANSPORTATION

As in antiquity, Lebanon's situation makes it a vital crossroads between East and West.

Roads. A road network of about 2,500 miles (4,000 kilometres) traverses Lebanon, about 300 miles (500 kilometres) being international highways, forming part of major land routes connecting Europe with the Arab countries and the East. There are also about 700 miles (1,100 kilometres) of national highways, 500 miles (800 kilometres) of paved secondary roads, and 1,000 miles of unpaved roads.

Lebanon appears to be a society in perpetual motion; in 1976 there was a ratio of one motor vehicle per 13 persons. There is a low-priced bus and taxi service throughout the country, and regular bus services to Europe and to most Arab countries are available.

Railways. The railway system, which includes lines along the coast and up al-Biqāʿ valley and a cog railway across the Lebanon Mountains, connects with the rail systems of Europe and other Arab countries. Although there is daily passenger service to Aleppo, in Syria, the line is used mainly for long-distance bulk transport.

Port of Beirut

Ports. There are numerous ports along the Lebanese seacoast. Fishing boats and coastal steamers use Tyre, Sidon, and Jubayl (Byblos), all of which have been ports since Phoenician times. Oil tankers load offshore from the pipeline terminals at Tripoli and at az-Zahrānī near Sidon. There is a harbour for yachts and small craft at Jūniyah. The principal cargo and passenger port is that of Beirut, which has a free zone for duty-free handling and storage of transit shipments. About 3,000 ships and some 95,000 passengers utilized Beirut's port facilities in 1974. A major expansion and deepening of the port was completed in 1967, and a large storage silo (for wheat and other grains) was completed in 1970. The docks could hold 20 ships and move 3,500,000 tons of goods per year. Until 1974, traffic in the port of Beirut amounted to 4,000,000 tons annually, of which 1,000 tons was in transit for Arab countries. Port infrastructure remained unaffected by the war, but all superstructures—sheds, warehouses, cargo-handling equipment, workshops, and telephone and power lines—were either destroyed or severely damaged.

Air services. Beirut International Airport was one of the busiest airports in the Middle East before the war. Its runways can take the largest jet airplanes in service, and more than 30 international airlines used Beirut regularly; by 1978 that figure had increased to 40.

ADMINISTRATION AND SOCIAL CONDITIONS

Parliament

The structure of the government. ***The constitutional framework.*** Modern Lebanon is a republic with a parliamentary system of government. Its constitution, promulgated in 1926 during the French mandate and modified by several subsequent amendments, provides for a unicameral Chamber of Deputies (renamed National Assembly in 1979) elected, for a term of four years, by universal adult suffrage (women attained the right to vote and eligibility to run for office in 1953). By the electoral law of 1960, which imposed the principle of the secret ballot, membership of the chamber was increased from 66 to 99. Parliamentary seats are apportioned on a religious basis in the ratio of six Christians to five Muslims, making the total number always a multiple of 11. This sectarian distribution is also observed in all administrative appointments for public office.

The president of the republic is elected by a two-thirds majority of the National Assembly for a term of six years and is eligible for re-election only after the lapse of a further six years. By an unwritten convention, the president must be a Maronite, the premier a Sunni Muslim, and the speaker of the National Assembly a Shiʿite. The president, who is constitutionally invested with the power of chief executive, calls upon a Sunni Muslim to form a Cabinet whose portfolios must reflect the sectarian balance. The Cabinet, to remain in power, requires a vote of confidence from the Assembly. A vote of no confidence, however, is a constitutional right that is rarely exercised in practice. A Cabinet usually falls because of internal dissension or because the president withdraws his support.

Local government. Administratively, Lebanon consists of six *muḥāfaẓāt* (governorates): Bayrūt (Beirut), Jabal Lubnān, ash-Shamāl, al-Janūb, al-Biqāʿ, and an-Nabaṭīah. These are administered by the Ministry of Interior through the *muḥāfiẓ* (governor), who represents the central government in the region. The *muḥāfaẓāt* are further divided into *qaḍā'*s (districts), each of which is presided over by a *qā'im-maqām* (district chief), who, along with the *muḥāfiẓ*, supervises local government in his respective region. Municipalities (communities with at least 500 inhabitants) elect their councils, which in turn elect mayors and vice mayors. Villages and towns (more than 50 and fewer than 500 inhabitants) elect a *mukhṭār* (headman) and a council of elders, who serve on an honorary basis. All officers of local governments serve four-year terms.

Lebanon, Area and Population

	area		population*	
	sq mi	sq km	1961 est.†	1970 est.‡
Governorates (*muḥāfaẓat*)				
al-Biqa'	1,653	4,280	320,967	204,000
al-Janub	364	943	200,818	250,000
ash-Shamāl	765	1,981	506,079	365,000
Bayriit	7	18	298,129	475,000
Jabal Lubnan	753	1,950	622,284	833,000
an-Nabaṭiyah	408	1,058	203,607	...
Total Lebanon%	3,950	10,230	2,151,884	2,126,000‖

*Last complete census was taken in 1932.
†According to population registers. ‡Figures are based on a sample survey made on November 15, 1970, of resident (*de jure*) population; estimate excludes certain categories of persons not covered by the sample, such as Palestinian refugees (who numbered 181,764 in 1971). The total resident population of Lebanon comprised between 2,300,000 and 2,400,000 persons at the end of 1970. §Created in 1975; includes the districts of Nabaṭiyah, Bint Jubayl, Marj ʿUyūn, and Ḥāṣbayya, which were formerly part of al-Janūb. ‖Figures do not add to total given because of rounding.
Source: Official government figures; United Nations.

The political process. The political system in Lebanon remains a curious blend of secular and traditional features. Until 1975, outwardly, the country appeared to support liberal and democratic institutions, yet, in effect, it had hardly any of the political instruments of a civil polity. Its political parties, parliamentary blocs, and pressure groups were so closely identified with parochial, communal, and personal loyalties that they often failed to serve the larger national purpose of the society. A National Pact, a sort of Christian–Muslim entente, sustained the national entity (*al-kiyān*), yet this sense of identity was neither national nor civic. In April 1975 the political process collapsed. The war that had engulfed the Lebanese exposed the vulnerability of the political system. The Syrians in January 1976 introduced suggestions for constitutional amendments that would have led to a more equitable sharing of power among Christians and Muslims, 50:50, rather than the constitutional ratio, which favours the Christian minority. In the light of subsequent developments, these "reforms," which were accepted by the then president, are not considered satisfactory solutions to the political problems facing Lebanon. The legitimate authority maintains the facade of continuity, while the process on which it is based was destroyed by the contending forces in the conflict that has continued to ravage Lebanon.

National Pact

Justice. The system of law and justice is mostly modelled on French concepts. The judiciary consists of courts of the first instance, courts of appeal, courts of cassation, and a Court of Justice that handles cases affecting state security. The Council of State is a court that deals with administrative affairs. In addition, there are religious courts that deal with matters of personal status (such as inheritance, marriage, and property matters) as they pertain to autonomous communities. Despite the country's well-developed legal system and a very high proportion of lawyers, significant numbers of disputes and personal grievances are resolved outside the courts. Justice by feud and vendetta continues.

Armed and security forces. In 1978 it was estimated that the strength of the armed forces amounted to 7,000 in the army, 500 in the air force, and 300 in the navy. Lebanon also had a paramilitary gendarmerie and police force of 5,000. During the civil war, the army practically disintegrated as splinter groups joined the different warring factions. After 1976, security was maintained principally by the 33,000-strong Arab Deterrent Force (ADF) created specifically for the purpose, as a transitional force pending the rebuilding of the army. Reconstruction of the Lebanese armed forces was under way, but the continued existence of private militias continued to prevent a return to full sovereignty by the central government.

Private militias

Social conditions. Lebanon in general has a relatively high standard of living, compared with most other Arab countries. In urban areas, modern health, medical, and sanitation services are easily accessible to all groups.

Education. A well-developed system of education reaches virtually all levels of the population. The estimated literacy rate is 88 percent, the highest in the Middle East. Education was once almost exclusively the responsibility of religious communities or foreign groups, but by the 1970s the number of students in public schools had risen to almost one-third of the total school enrollment.

The compulsory five-year primary school program is followed either by a seven-year secondary program (leading to the official baccalaureate certificate) or by a four-year program preparing students for technical or vocational training. In 1974, some 50,000 students (nearly 22,000 of whom were non-Lebanese) were receiving university education in the country's institutions of higher learning. The number of students in universities decreased during the war, but the prewar level was expected to be regained by 1979. The four major universities are the American University of Beirut, the Universite Saint-Joseph (subsidized by the French government and administered by the Jesuit order), the Lebanese University (Universite Libanaise), and the Beirut Arab University (an affiliate of the University of Alexandria).

Universities

Health and welfare. Public health services are still largely concentrated in the cities, although the government increasingly directs medical aid into rural areas. As in the field of social welfare, nongovernmental voluntary associations—mostly religious, communal, or ethnic—are very active. The Lebanese diet is generally satisfactory, and the high standard of living and the favourable climate have served to reduce the incidence of many diseases that are still common in other Middle Eastern countries.

In proportion to its population, the country also produces the largest number of skilled medical personnel in the Arab Middle East. There is approximately one doctor for 1,290 inhabitants. Hospital facilities are also adequate.

In 1963 the National Social Security Fund was established, covering sickness and maternity insurance, labour accident and occupational disease insurance, family benefits, and termination of service benefits.

Housing. In response to the need for low-cost housing, a Popular Housing Law was enacted in 1965, providing for the rehabilitation of substandard housing. The legislation also stipulated that, instead of investing government funds in low-cost housing, guarantees of mortgage payments would be provided to developers. Some 4,000 "popular houses" were constructed.

A study conducted in 1970 showed that 123,000 homes (23 percent) were without bathrooms and that 8,000 families, excluding the Palestinian refugees, were living in improvised accommodations. ***A Report on Housing*** by the United Nations Economic Commission for Western Asia in July 1977 indicated that there had been no improvement by the start of the war. The economic boom of 1973–74 attracted thousands of villagers to the capital, and the housing shortage worsened considerably. The war drastically increased the problem. Thousands of homes on the front line were destroyed, and entire villages were evacuated and others occupied. The result was chaos in which property rights were violated as a matter of course. The government, in an attempt to remedy the situation, in 1978 set up a Housing Bank with a capital of L.£50,000,000 ($16,800,000 U.S.) to make housing loans.

Housing bank

Wages and cost of living. A minimum wage is set by the Labour Code, and legislation provides for cost-of-living increases. The cost of living increased sharply prior to and during the civil war, mainly because of a substantial rise in the cost of rent, education, food, and petroleum products.

Social and economic division. Lebanese society was able for a long time to give a semblance of relative economic stability and convey an impression of social cohesion. The existence of a large middle-income group, in addition to the political and social legitimacy that kinship ties and religious and communal attachments had, reinforced the false veneer that masked the growing socioeconomic dislocations. The interaction of these factors covered up the growing class polarization, not only between the regions but also around the industrial belt that encircled the capital, Beirut. This rendered the body politic vulnerable. The explosion that ensued was attributable in part to the fact that the system of government was unresponsive to the acute social problems and grievances.

CULTURAL LIFE AND INSTITUTIONS

The cultural milieu. Historically, Lebanon is the heir of a long succession of Mediterranean cultures—Phoenician, Hittite, ancient Egyptian, Greek, Roman, Byzantine, and Arab. Its cultural milieu continues to show clear manifestations of a rich and diverse heritage. As an Arab country, Lebanon shares more than a common language with neighbouring Arab states; it also has a similar cultural heritage and common interests.

In the 19th century, Lebanese linguists were in the vanguard of the Arabic literary awakening. In more recent times, writers of the calibre of Khalil Gibran, Georges Shehade, and Michel Chiha have been widely translated and have reached an international audience.

While originally cultural life in Lebanon was predominantly centred around universities and affiliated institutions, in more recent times there has been an impressive proliferation of cultural activities. Beirut has five museums and a number of libraries, learned societies, and research institutions.

Music festivals

The state of the arts. Lebanon's antiquities and ruins have provided not only inspiration for artists but also magnificent backdrops for annual music festivals, most notably the Baalbek International Festival. International opera, ballet, symphony, and drama companies, of nearly all nationalities, competed to enrich the cultural life of Beirut. Lebanon has produced a number of gifted young artists who have shown a refreshing readiness to experiment with new expressive forms, from Surrealistic painting to Theatre of the Absurd. Several young Lebanese are active in European opera and theatre companies, while others are intent on creating a wider audience for classical Arabic music and theatre.

The cultural awakening encouraged the revival of national folk arts, particularly song, *dabkah* (the national dance), and *zajal* (folk poetry), and the refinement of traditional crafts. Although the Baalbek International Festival was suspended during the civil war, popular theatre flourished in the war-ridden country.

The communications media. In addition to the wide variety of foreign newspapers and magazines that can be found in Beirut, Lebanon has almost 470 registered publications in Arabic, English, French, and Armenian, of which about 40 are dailies.

Impressions suggest strongly that Beirut may be the movie-going capital of the world. Because of the wide choice of films, comfortable surroundings, and low prices, going to the pictures has emerged as the national pastime of the Lebanese. In 1965 UNESCO estimated per capita film attendance to be the 10th highest in the world.

By 1976 there were more than 420,000 television sets, which suggests that more than half of the population is exposed to the programming offered by Beirut's two private companies and by Egyptian and Syrian television.

With the advent of the transistor, Lebanon became virtually saturated with radios, with about 1,600,000 in 1976. The government-run radio station broadcasts Arabic, French, English, and Armenian programs. During the war, there was a proliferation of clandestine radio stations seeking to put forward the news and views of the warring parties. The government is trying to put an end to these radio stations as a manifestation of its commitment to national unity.

PROSPECTS FOR THE FUTURE

Lebanon is groping for a resolution of the conflicts that tore it apart after 1974. After the installation of Pres. Elias Sarkis, the government sought to ensure the restoration of public order with the aid of the Arab Deterrent Force, pending the reformation of the Lebanese national army. Reconstruction projects were suspended, awaiting the solution of the political and security problems.

Plans for constitutional and administrative reforms, however, are contingent on steps being taken toward an overall settlement of the Middle East crises. Despite the gloom that prevails and the suspicions that have developed among many sectors of the population, the prospects for the future cannot with certainty be said to be bleak. There may be sufficient exasperation, and there is a proved reservoir of resilience, to bring about a new order that will restore Lebanon to its historical role as a cultural, intellectual, and artistic centre of the Arab East.

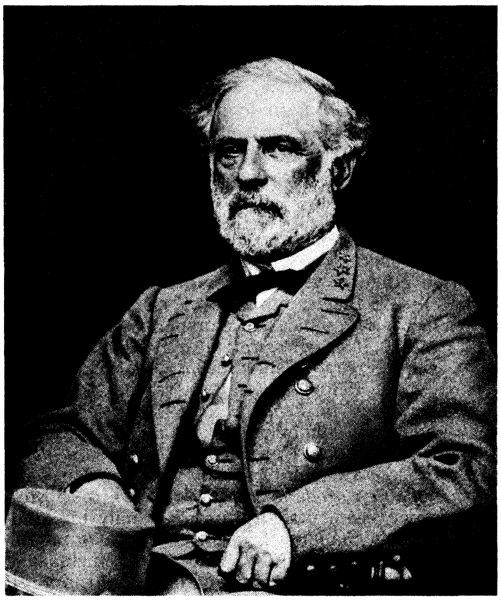
BIBLIOGRAPHY. H. BARAKAT, ***Lebanon in Strife: Student Preludes to the Civil War*** (1977), a study of youth and political orientation prior to the civil war in Lebanon; BEIRUT COLLEGE FOR WOMEN, ***Cultural Resources in Lebanon*** (1969), a collection of 23 lectures by Lebanese scholars, journalists, and men of letters on a wide range of topics; L. BINDER (ed.), ***Politics in Lebanon*** (1966), essays that explore the effects of traditional and cultural elements on political developments and modernization; A.H. FULLER, ***Buarij: Portrait of a Lebanese Muslim Village*** (1961), still a perceptive study even though it is based on field research in the 1930s; G.L. GRASSMUCK and K.S. SALIBI, ***Reformed Administration in Lebanon*** (1964), a descriptive study of government and administrative structure; J. GULICK, ***Social Structure and Culture Change in a Lebanese Village*** (1955), a description of a Greek Orthodox village in northern Lebanon, and ***Tripoli: A Modern Arab City*** (1967), the history, economic and governmental institutions, family life, and social structure of this changing Muslim city; A.H. HOURANI, ***Arabic Thought in the Liberal Age, 1798–1939*** (1962), includes a penetrating analysis of the modern intellectual history of Lebanon, and ***Syria and Lebanon*** (1946; reprinted, 1968), a scholarly political essay by one of the most distinguished historians of the Middle East: M.C. HUDSON, ***The Precarious Republic: Political Modernization in Lebanon*** (1968), an examination of the scope and direction of social mobilization and its implications for political development; INSTITUT INTERNATIONAL DE RECHERCHES ET DE FORMATIONS EN VUE DU DEVELOPPEMENT (IRFED), ***Étude préliminaire sur les besoins et les possibilités de dkveloppement au Liban*** (1959–60), a report outlining the requirements and prospects for a balanced development, the only comprehensive plan for the country; N. KHALAF, ***The Economic Implications of the Size of Nations with Special Reference to Lebanon*** (1971), detailed analysis of the influence of Lebanon's size on its economic structure, stability, and development; S. KHALAF and P. KONGSTAD, ***Hamra of Beirut: A Case of Rapid Urbanization*** (1972), empirical study exploring the ecological transformation and the social structure of a fast-growing urban community; S. KHALAF, ***Prostitution in a Changing Society*** (1965), a sociological survey of legal prostitution in Beirut; L.M.T. MEO, ***Lebanon: Improbable Nation*** (1965), a consideration of the implications of Lebanon's feudal past and religious composition for its political process, though predominantly concerned with the civil disturbances in 1958; I. RABBAT, ***La Formation historique du Liban politique et constitutionnel*** (1973), a scholarly analysis of the Lebanese constitutional system by one of the country's leading jurists; P. RONDOT, *Les Chrktiens d'Orient* (1955), an analysis of the intellectual and political attitudes of Lebanese Christians; E. SALEM, ***Contemporary Lebanon: An Essay on Modernization*** (1971), an analysis of the capability of Lebanon's political and bureaucratic system to generate change and economic progress while preserving its democratic and liberal institutions; K.S. SALIBI, ***Crossroads to Civil War: Lebanon, 1958–1976*** (1976), a reliable account of developments that led to the civil war; Y.A. SAYIGH, ***Entrepreneurs of Lebanon*** (1962), an empirical study of the role of entrepreneurs in the national development of the country; M.W. SULEIMAN, ***Political Parties in Lebanon*** (1967), an analysis of the political parties; J.R. WILLIAMS, ***The Youth of Haouch El Harimi: A Lebanese Village*** (1968), an anthropological monograph containing observations on various aspects of village life.

(S.G.K./C.F.M.)

Lee, Robert E.

As commander of the most successful of the Confederate armies during the United States Civil War, Robert E. Lee became synonymous with the Confederacy's defense. Because he continued to confront the massed Federal forces after four years of war while the rest of the Confederacy collapsed, his surrender at Appomattox is commonly viewed as signifying the end of the Civil War. His subtler fame and more significant contribution came after the war, however, when, as the idol of a defeated people, he served as an example of fortitude and magnanimity during the ruin and dislocations, the anguish and bitterness of the war's long aftermath. In those years, Lee became an enduring symbol to the Southern people of what was best in their heritage.

By courtesy of the Library of Congress, Washington, D.C.



Robert E. Lee, 1865:

Heritage and youth. Robert Edward Lee was born at Stratford, Westmoreland County, Virginia, on January 19, 1807, the fourth child of Colonel Henry Lee and Ann Hill Carter. On both sides, his family had produced many of the dominant figures in the ruling class that had determined Virginia's colonial destiny and contributed leadership to the American Revolution.

Lee's father, "Light Horse" Harry Lee, had been a cavalry leader during the Revolution, a post-Revolution governor of Virginia, and the author of the famous congressional memorial eulogy to his friend, George Washington. Intermarriage with most of Virginia's ruling families was a tradition of both the Lees and 'the Carters, and Lee married a distant cousin, Mary Anne Randolph Custis, the great-granddaughter of George Washington's wife and heiress of several plantation properties.

Yet, with all his aristocratic connections, Lee lacked the advantages of wealth. His colourful father had no aptitude for finance and, dying when Lee was a child, left in straitened circumstances an ailing widow with seven children. Robert, the youngest boy, was the closest of all the children to his mother and was deeply influenced by her strength of character and high moral principles. All reports of his childhood and youth stress that the pinched gentility of his formative years, in such marked contrast to the lordly life on the great plantations of his kinspeople, was another strong influence goading him to excel at whatever task he was assigned.

Military education at West Point

Unable to afford a university education, Lee obtained an appointment to the United States Military Academy at West Point, where his high aspirations and native gifts produced what a fellow cadet, the Confederate general Joseph Johnston, called his natural superiority. Always near the top of his class, he won the appointment to corps adjutant, the highest rank a cadet could attain, and graduated in 1829. With handsome features, a massive head, and superb build, he combined dignity with kindness and sympathy with good humour, to win, as Johnston said, "warm friendship and command high respect."

Early military career. Commissioned into the elite engineering corps, later transferring to the cavalry because of slow advancement in the engineers, he did the best he could at routine assignments and on relatively uninspiring engineering projects. Not until the Mexican War (1846–48), when he was a captain on the staff of General Winfield Scott, did he have the opportunity to demonstrate the brilliance and heroism that prompted General Scott to write that Lee was "the very best soldier I ever saw in the field." In October 1859, while on leave at Arlington to straighten out the entangled affairs of his late father-in-law, he was ordered to suppress the slave insurrection attempted by John Brown at Harpers Ferry, Virginia. Although Lee put down the insurgency in less than an hour, the very fact that it was led by a white man made him aware of the gathering crisis between the North and the South.

Lee was back at his command in Texas when on February 1, 1861, Texas became the seventh Southern state to secede, and with the rest of the U.S. Army forces, he was ordered out of the state. Without a command, he returned to Arlington to wait to see what Virginia would do. On April 18 he was called to Washington and offered command of a new army being formed to force the seceded states back into the Union. Lee, while he opposed secession, also opposed war, and "could take no part in an invasion of the Southern states." Meanwhile, President Lincoln called on Virginia to furnish troops for the invasion. A Virginia convention, which had previously voted 2 to 1 against secession, now voted 2 to 1 against furnishing troops for an invasion and to secede, and Lee resigned from the army in which he had served for 36 years to offer his services to the "defense of [his] native state."

Role in Civil War. As commander in chief of Virginia's forces, Lee saw it as his first task to concentrate troops, armaments and equipment at major points where the invasion might be expected. During this period, Confederate troops joined the Virginia forces and subdued the Federal Army at the first Battle of Bull Run. The attempt at a quick suppression of the Southern states was over and, as Lee was one of the first to realize, a long, all-out war began. Between July 1861 and June 1862, Confederate president Jefferson Davis appointed Lee to several unrewarding positions, the last of which was the trying post of military adviser to the president. Here, however, Lee, working independently of Davis, was able to introduce a coherent strategy into the Confederacy's defense.

Commander of Virginia's forces

During May 1862, General Johnston was leading a heterogeneous collection of Confederate troops back toward Richmond from the east, before the methodical advance of Gen. George B. McClellan's superbly organized, heavily equipped Army of the Potomac. Lee collaborated with the then unknown Stonewall Jackson to concentrate scattered garrisons in Virginia into a striking force in the Shenandoah Valley, where he surprised the Federal forces into retreating and posed a threat to Washington. Jackson's threat from the valley caused Lincoln to withhold from McClellan the large corps of Gen. Irvin McDowell, with whom McClellan planned a pincer movement on Richmond from the east and north. On June 1, Johnston delivered a poorly organized attack on McClellan's forces seven miles east of Richmond in the indecisive Battle of Fair Oaks (Seven Pines). The battle became a turning point for Lee: Johnston was seriously wounded and Lee was at last given field command.

In three weeks he organized Confederate troops into what became the famed Army of Northern Virginia; he tightened command and discipline, improved morale, and convinced the soldiers that headquarters was in full command. McClellan, waiting vainly for McDowell to join the wing of his army on the north side of the Chickahominy River, was moving heavy siege artillery from the east for the subjugation of Richmond when Lee struck. Combin-

Seven Days' Battles

ing with Jackson, who moved in from the valley, Lee defeated Porter's right wing and was on McClellan's supply line to his base on the York River. In a series of hard fights, the Seven Days' Battles (around Richmond), McClellan withdrew his army to the wharves of Berkeley Plantation, where he was aided by the U.S. Navy. Because it was the first major victory for the Confederacy since Bull Run, and because it halted a succession of military reversals, Lee emerged overnight as the people's hero, and his soldiers developed an almost mystical belief in him.

Lee never believed that the Confederate troops had the strength to win in the field; for the next two years his objectives were to keep the enemy as far away as possible from the armament-producing centre of Richmond as well as from the northern part of the state, where farmers were harvesting their crops, and, finally, to inflict defeats of such decisiveness as to weaken the enemy's will to continue the war. To nullify the Federals' superiority in manpower, armaments, and supply, Lee always sought to seize the initiative by destroying the enemy's prearranged plans. Until the spring of 1864, he was successful in keeping the enemy away from Richmond and from the northern part of the state, twice expelling the enemy out of Virginia altogether. He inflicted several severe defeats on the enemy, most strikingly at the Second Battle of Bull Run (Second Manassas), August 29–30, 1862. To shift the fighting out of Virginia, Lee crossed into Maryland, where he hoped for support from Southern sympathizers. But his plans fell into Northern hands, and his forces were nearly destroyed at Antietam (Sharpsburg) on September 17, 1862. He was, however, able to withdraw the remnants across the Potomac, to reorganize his army, and to resume his series of victories at Fredericksburg in December of that year. At Chancellorsville (May 1–4, 1863) he achieved another notable victory, although outnumbered two to one, by splitting up his army and encircling the enemy in one of the most audacious moves in military history.

Gettysburg

But he was producing no more than a stalemate on the Virginia front, while Federal forces won important victories in other parts of the Confederacy, and time was against him. While the Federals always replaced their losses, Lee's army was dwindling in size, suffering an irreplaceable drain in its command—particularly through the loss of Stonewall Jackson, who had been mortally wounded at Chancellorsville—and increasingly acute shortages of food and clothing, which undermined the physical condition of the soldiers. Largely to resupply his troops and to draw the invading armies out of Virginia, Lee once more crossed the Potomac. The first invasion had ended with the Battle of Antietam and the second ended in Lee's repulse at Gettysburg (July 1–3, 1863). There, operating for the first time without Jackson, Lee was failed by three of his top generals in using the discretionary orders that had worked so effectively with Jackson, his "right arm."

Then, in May 1864, Ulysses S. Grant, the newly appointed commanding general of all Union forces, drove at Lee with enormous superiority in numbers, armaments, and cavalry. The horses of the troopers of Confederate general Jeb Stuart were in poor condition, and Stuart was killed early in the campaign. Grant could neither defeat nor outmanoeuvre Lee, however, and the superb army Grant inherited sustained losses of 50,000 men in the May and early June battles of the Wilderness, Spotsylvania Court House, the North Anna, and Cold Harbor. Grant, however, his losses replaced by fresh recruits, had advanced within seven miles of Richmond, while Lee, his soldiers too weakened physically and his officers too inexperienced to attempt countering manoeuvres, had lost the initiative. Lee himself was, moreover, physically declining and frequently incapacitated by illness. When Grant, abandoning his advance on Richmond, moved south of the James River to Petersburg—Richmond's rail connection with the South—Lee could only place his starving tatterdemalions in defensive lines in front of Petersburg and Richmond.

Beginning at Spotsylvania Court House, Lee had nullified Grant's numbers by using his engineering experience to erect fortifications that were in advance of any fieldworks previously seen in warfare. At Petersburg, Lee extended the field fortifications into permanent lines that presaged trench warfare. While Lee's lines enabled him to withstand Grant's siege of the two cities from late June 1864 to April 1, 1865, once his mobile army was reduced to siege conditions, Lee said the end would be "a mere question of time."

Surrender at Appomattox

The time came on Sunday, April 2, when his defensive lines were stretched so thin that the far right broke under massive assaults, and Lee was forced to evacuate Petersburg and at last uncover Richmond. When the survivors of his army pulled out of the trenches an agonizing week of a forlorn retreat began for him; his men fell out from hunger, animals dropped in the traces, and units dissolved under demoralized officers. At Appomattox Court House on April 9, 1865, his way west was blocked and there was nothing left except to bear with dignity the ordeal of surrender, which was made less painful for him by Grant's considerate behaviour.

Postwar years. Lee spent several months recuperating from the physical and mental strain of retreat and surrender, but he never regained his health. He was, moreover, deeply concerned about the future of his seven children, for his wife's Arlington plantation had been confiscated by the United States government and he was without income at the age of 58. Both to earn subsistence for his family and to set an example for his unemployed fellow officers, he accepted the post of president of Washington College (later Washington and Lee University) in Lexington, Virginia.

Lee was a surprisingly progressive educator; by employing his lifelong practices in economy, he placed the institution on a sound basis and awakened in his students—many of whom were veterans of the recent war—the desire to rebuild their state with the goal of good citizenship in a nation that in time would become reunited.

He died on October 12, 1870, in the president's home at Washington College.

Reputation. Although history knows him mostly as "the Rebel General," Lee was a disbeliever in slavery and secession and was devoutly attached to the republic that his father and kinsmen had helped bring into being. He was, moreover, very advanced in his rejection of war as a resolution of political conflicts—a fact that has been almost entirely ignored by posterity. As a U.S. Army colonel in Texas during the secession crises of late 1860, he wrote, "[If] strife and civil war are to take the place of brotherly love and kindness, I shall mourn for my country and for the welfare and progress of mankind."

BIBLIOGRAPHY. DOUGLAS SOUTHALL FREEMAN, *R.E. Lee*, 4 vol. (1934–35), is the definitive biography of Lee; its footnotes provide a rich source of bibliography. Material that has appeared since Freeman's book does not significantly alter the character presented in his work. New material has, however, caused a re-evaluation of Freeman's interpretation of the second day of the battle of Gettysburg. CLIFFORD DOWDEY, ***Lee*** (1965), is a one-volume biography that treats the second day at Gettysburg in the light of the new material now generally available; his ***Death of a Nation*** **(1958)**, is a narrative study of the Gettysburg campaign that also uses the new material on the second day.

CLIFFORD DOWDEY and LOUIS MANARIN (eds.), ***The Wartime Papers of Robert E. Lee*** **(1961)**, draws upon an enormous volume of letters and dispatches not previously published.

Among the most reliable reminiscences of Lee's contemporaries are JOHN WILLIAM JONES, ***Personal Reminiscences, Anecdotes, and Letters of Gen. Robert E. Lee*** (1874) and ***Life and Letters of Robert Edward Lee*** (1906); A.L. LONG, ***Memoirs of Robert E. Lee*** (1886)—Long was a literate member of Lee's early staff and later chief of Second Corps artillery; WALTER H. TAYLOR, ***Four Years with General Lee*** (1877) and ***General Lee: His Campaigns in Virginia, 1861–1865*** (1906)—Taylor was on Lee's staff to the end. There are scores of reminiscences and narratives of the Civil War by participants that include references to Lee. E.P. ALEXANDER, ***Military Memoirs of a Confederate*** **(1907)**, a sound standard work on the campaigns in Virginia, has the best account of Lee's reorganization of the Army of Northern Virginia. There are only two accounts of the last illness and death of Lee: MRS. ROBERT E. LEE, "Letter to Miss Mary Meade,

October 12th, 1870," *Virginia Magazine of History and Biography,* 35:23–26 (1927); and COLONEL WILLIAM PRESTON JOHNSTON, "Death and Funeral of General Lee," contained in Jones's *Personal Reminiscences . . . ,* pp. 44—60.

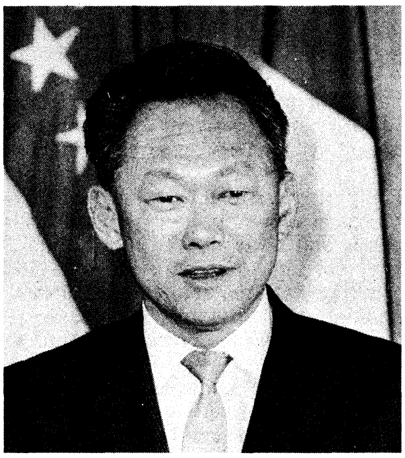
Two good books on Lee by an Englishman are SIR FREDERICK MAURICE, *Robert E. Lee, the Soldier* (1925) and *An Aide-de-Camp of Lee: Being the Papers of Colonel Charles Marshall* (1927). FRANKLIN L. RILEY (ed.), *General Robert E. Lee After Appomattox* (1922), is a collection of sketches, data, and reminiscences of Lee as president of Washington College; not entirely accurate, and lacking much detailed factual material, it is nonetheless valuable for the impressions of Lee by his students.

(C.Do.)

Lee Kuan Yew

Lee Kuan Yew, the first prime minister of the Republic of Singapore, became one of the leading political figures in Southeast Asia, as an anti-colonialist and anti-Communist supporter of parliamentary democracy. The strategic location of his country commanding the Strait of Malacca, between the China Sea and the Indian Ocean, has further served to enhance his position both in Asia and in the Commonwealth of Nations.

Keystone



Lee Kuan Yew.

Lee was born on September 16, 1923, into a wealthy Chinese family that had been established in Singapore since the 19th century. His first language was English; only upon entering politics did he acquire a command of Chinese as well as Malay and Tamil. After attending school in Singapore, Lee, a brilliant pupil, won a scholarship to Raffles College in Singapore, which he attended before going to England to enter the University of Cambridge. There he took a double first (first-class honours in two different subjects) in law and headed the honours list; he also became a Socialist. In 1950 he was admitted to the English bar. In the same year he married a Malayan, Kwa Geok Choo, who had also been a student at Cambridge. Instead of practicing as a barrister, however, Lee returned to Singapore. Appointed legal adviser to the Postal Union, he participated in negotiations to obtain higher wages for postal workers and subsequently did similar work for other trade unions.

Beginnings of Lee's political career

A British crown colony and the site of Britain's principal naval base in the Far East, Singapore was ruled by a governor assisted by a legislative council. Its members consisted primarily of wealthy Chinese businessmen, most of whom were appointed rather than elected. When, in the early 1950s, constitutional reform was in the air in Singapore, as elsewhere in Britain's colonial empire, Lee formed an alliance with two other political newcomers —David Saul Marshall, a lawyer, and Lim Yew Hock, a trade unionist—to challenge the hold of the businessmen on the council. Lee, however, soon broke with his two colleagues to take a more radical stand, becoming secretary general of his own party, the People's Action Party (PAP), which included some Communists, Lee having accepted some Communist support for some years.

In 1955 a new constitution was introduced increasing the number of elected seats on the council to 25 out of a total of 32. In the elections, the Labour Front, founded by Lee's former colleagues, won 13 seats, while the PAP won three—one of which, for a district inhabited by many of the poorest Chinese in Singapore, was won by Lee.

The following year Lee returned to London as a member of a delegation that unsuccessfully sought self-rule for the colony. Unrest followed during which a number of PAP leaders were imprisoned. In 1957 the London negotiations were resumed, again with Lee on the delegation. After agreement was reached on a measure of self-government, Lee won a by-election by an overwhelming majority, and a PAP power struggle ensued. In August Lee was ousted from the secretary generalship by the party's left wing but regained his post in October. The next year (1958) in London he helped negotiate the status of a self-governing state within the Commonwealth for Singapore. Elections were held under the new constitution in May 1959, but, despite the fact that Lee's party won a decisive victory, gaining 43 of the 51 seats, Lee refused to form a government until the British freed the left-wing members of his party who had been imprisoned in 1956. With their release, he then formed a cabinet and introduced a five-year plan that sought to promote reforms (such as the emancipation of women and the expansion of educational services) and industrialization. In 1961 the party's left-wing members broke away from the PAP to form the Barisan Socialis ("Socialist Front"), which shared sympathies with President Sukarno of Indonesia, who for some time had represented a challenge to Lee's leadership. In the following year, a national referendum resulted in an open confrontation between Lee and his left-wing opponents; its results were a victory for Lee.

In 1963 Lee took Singapore into the newly created federation of Malaysia. In elections held in 1964, however, Lee made the mistake of entering his party, 75 percent of whose members were Chinese, in the Malayan national elections. The growing tension between Chinese and Malayans resulted in communal rioting in Singapore itself. In August 1965 Lee was told by his Malaysian colleagues in the federal government that Singapore must leave the federation. Although Lee passionately believed in the multiracialism that the federation represented, Singapore had to secede. It then became a sovereign state with Lee as its first prime minister.

Leader of a new state

Lee's principal aims were to ensure the physical survival of the new state and to retain Singapore's national identity, traditionally marked by interracial tolerance. Surrounded by more powerful neighbours (including China and Indonesia), Lee did not press for the immediate withdrawal of British, Australian, and New Zealand forces from Singapore. Instead, he sought to phase them out slowly and to replace them with a Singapore force locally trained and patterned on the Israeli model. Under his leadership trade and industry prospered, and the new state achieved a high standard of living, with the second highest per capita income in Asia after Japan. At Commonwealth prime ministers' conferences, Lee also played a significant role, expressing support of a multiracial approach to Commonwealth problems in Africa and elsewhere.

Shrewd and resolute in political life, Lee is said to lead a frugal private life, living quietly at home with his wife, two sons, and daughter. Among the anti-colonial leaders of the Commonwealth, he represents a new prototype. Combining an Asian outlook with some Western methods, he has brought his country an efficient administration, a measure of economic prosperity, and an alternative to the tyranny of either racial demagoguery or ideological conformity.

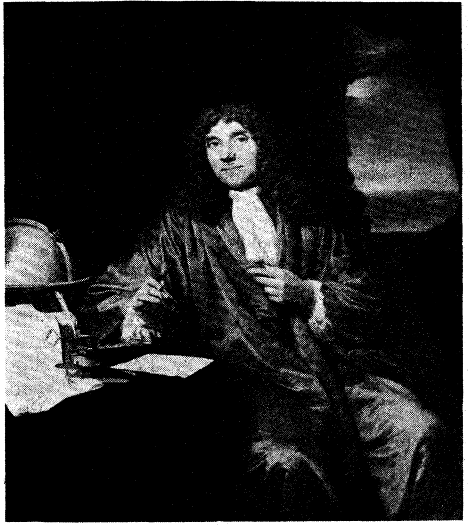
BIBLIOGRAPHY. C. NORTHCOTE PARKINSON, *A Law unto Themselves* (1966), contains a biographical sketch. See also ALEX JOSEY, *Lee Kuan Yew* (1968).

(K.I.)

Leeuwenhoek, Antonie van

One of the most remarkable of the 17th-century microscopists, Antonie van Leeuwenhoek was the first man ever to observe bacteria and protozoa through a lens. Although there were only isolated observations of these organisms during the century following his death, his accurate interpretations of what he saw led to the sciences of bacteriology and protozoology.

By courtesy ot the Rijksmuseum, Amsterdam



Leeuwenhoek, portrait by Jan Verkolje (1650–93). In the Rijksrnuseurn, Amsterdam.

Leeuwenhoek was born at Delft, Holland, on October 24, 1632. Little is known of his early life, but it is certain that he did not have an extensive scientific education. When his stepfather died in 1648, he was sent to Amsterdam to become an apprentice to a linendraper. Returning to Delft when he was 20, he established himself as a draper and haberdasher. In 1660 he obtained a position as chamberlain to the sheriffs of Delft. His income was thus secure and sufficient enough to enable him to devote much of his time to his all-absorbing hobby of grinding lenses.

The "little animalcules"

High-quality single lenses of very short focal length were preferable to the compound microscope, which increased the problem of chromatic aberration. Although Leeuwenhoek's studies lacked the organization of formal scientific research, his powers of careful observation enabled him to make discoveries of fundamental importance. In 1674 he began to observe bacteria and protozoa, his "very little animalcules," which he was able to isolate from different sources, such as rainwater, pond and well water, and the human mouth and intestine; and he calculated their sizes.

In 1677 he described for the first time the spermatozoa from insects, dogs, and man, though Stephen Hamm probably was a codiscoverer. Leeuwenhoek studied the structure of the optic lens, striations in muscles, the mouthparts of insects, the fine structure of plants, and discovered parthenogenesis in aphids. In 1680 he noticed that yeasts consist of minute globular particles. He extended Marcello Malpighi's demonstration in 1660 of the blood capillaries by giving (in 1684) the first accurate description of red blood cells. In his observations on rotifers in 1702, Leeuwenhoek remarked that, "In all falling rain, carried from gutters into water-butts, animalcules are to be found; and that in all kinds of water, standing in the open air, animalcules can turn up. For these animalcules can be carried over by the wind, along with the bits of dust floating in the air."

A friend of Leeuwenhoek put him in touch with the Royal Society of England, to which, from 1673 until 1723, he communicated by means of informal letters most of his discoveries and to which he was elected a fellow in 1680. His discoveries were for the most part made public in the Society's *Philosophical Transactions.* The first representation of bacteria is to be found in a drawing by Leeuwenhoek in that publication in 1683.

His researches on the life histories of various low forms of animal life were in opposition to the doctrine that they could be produced spontaneously or bred from corruption. Thus he showed that the weevils of granaries (in his time commonly supposed to be bred from wheat as well as in it) are really grubs hatched from eggs deposited by winged insects. His letter on the flea, in which he not only described its structure but traced out the whole history of its metamorphosis, beginning with its first emergence from the egg, is of great interest, not so much for the exactness of his observations as for an illustration of his opposition to the spontaneous generation of many lower organisms, such as "this minute and despised creature." Some asserted that the flea was produced from sand, others from dust, etc., but Leeuwenhoek showed it to be "endowed with as great perfection in its kind as any large animal" and proved that it bred in the regular way of winged insects.

Leeuwenhoek also carefully studied the history of the ant and was the first to show that what had been commonly reputed to be ants' eggs were really their pupae, containing the perfect insect nearly ready for emergence, and that the true eggs were much smaller and gave origin to maggots, or larvae. He argued that the sea mussel and other shellfish were not generated out of sand found at the seashore or mud in the beds of rivers at low water but from spawn, by the regular course of generation. He maintained the same to be true of the freshwater mussel, whose embryos he examined so carefully he was able to observe how they were consumed by "animalcules," many of which, according to his description, must have included ciliates in conjugation, flagellates, and the *Vorticella.* Similarly, he investigated the generation of eels, which were at that time supposed to be produced from dew without the ordinary process of generation.

The dramatic nature of his discoveries made him world famous, and he was visited by many notables, including the Queen of England and Peter the Great.

Leeuwenhoek's microscopy

Leeuwenhoek's methods of microscopy, which he kept secret, remain something of a mystery. He ground over 400 lenses during his lifetime, most of which were very small—some no larger than a pinhead—and usually mounted them between two thin brass plates, rivetted together. A large sample of these lenses, bequeathed to the Royal Society, were found to have magnifying powers of between 50 and, at the most, 300 times. In order to observe phenomena as small as bacteria, Leeuwenhoek must have employed some form of oblique illumination, or other technique, for enhancing the effectiveness of the lens, but this method he would not reveal.

Leeuwenhoek continued his work almost to the end of his long life of 90 years. He died at Delft on August 26, 1723.

BIBLIOGRAPHY. Leeuwenhoek's contributions to the *Philosophical Transactions* amounted to 375, and those to the *Memoirs of the Paris Academy of Sciences* to 27. Two collections of his works appeared during his life, one in Dutch (1685–1718) and the other in Latin (1715–22); a selection was translated by S. HOOLE, *The Select Works of A. van Leeuwenhoek* (1798–1807). See also C. DOBELL, *Antony van Leeuwenhoek and His "Little Animals"* (1932); B. COHEN, "On Leeuwenhoek's Method of Seeing Bacteria," *J. Bact.,* 34:343–346 (1937), and *The Leeuwenhoek Letter* (1937).

Legal Education

The rise of legal education

Schools of law are of comparatively recent origin. The ancient Romans had schools of rhetoric that provided training useful to someone planning a career as an advocate, but there was no systematic study of the law. During the 3rd century BC, Tiberius Coruncanius, the first plebeian *pontifex maximus* (chief of the priestly officials) gave public legal instruction, and a class of nonpriests (*jurisprudentes)* who acted as legal consultants grew up. A student, in addition to reading the few lawbooks that were available, might attach himself to a particular *jurisprudens* and learn the law by attending consultations and by discussing points with his master. Over the ensuing centuries a body of legal literature developed, and some *jurisprudentes* set themselves up as regular law teachers.

In the medieval universities of Europe, including England, it was possible to study canon law and Roman law but not the local or customary legal system. The study of national laws at universities is, in most European countries, a development that began in the 18th century; the study of Swedish law at Uppsala dates from the early 17th century.

On the continent of Europe, the transition to the study of national law was facilitated by the fact that modern legal systems grew mostly from Roman law. In England, on the other hand, the national law, known as the common law, was indigenous. In medieval times, education in the common law was provided for legal practitioners by societies of lawyers called Inns of Court through reading and practical exercises. These methods fell into a decline in the late 16th century, mainly because students came to rely on printed books, and after the middle of the 17th century there was virtually no organized education in English law until the introduction of apprenticeship for solicitors in 1729. The famous jurist Sir William Blackstone lectured at Oxford in the 1750s, but university teaching of the common law did not develop significantly until the 19th century. The Council of Legal Education for barristers was established in 1852. In the United States, systematic legal education began with the founding of the Harvatd Law School in 1817.

THE AIMS OF LEGAL EDUCATION

Legal education generally has a number of aims, not all of which are pursued simultaneously. The emphasis placed on various objectives varies from one law school to another and from one teacher to another. One aim is to make the student familiar with legal concepts and institutions. Like most intellectual disciplines the law has its technical concepts, frequently expressed in technical terms. These terms are sometimes special legal words, such as "easement" (a landowner's right over some neighbour's land, such as a right of way), and sometimes ordinary words used in special legal senses, such as "distress" (a right to seize the goods of a tenant who has not paid his rent). All lawyers must become acquainted with the processes of making law, settling disputes, and regulating the legal profession. They must study the structure of government and the organization of courts of law, including the system of appeals and other adjudicating bodies such as social security tribunals in Great Rritain, administrative agencies in the United States, the Conseil d'État in France, and industrial arbitration bodies in Australia. Attention is also paid nowadays to plans of legal aid and advice for clients who are too poor to pay.

Traditional legal education also included the teaching of legal history. Once regarded as an essential part of any educated lawyer's equipment, legal history now receives much less emphasis. Separate courses in the subject are offered in fewer and fewer law schools and, when optional, are not very popular among students. But much legal history is taught in the context of other courses. The corpus of the law is a constantly evolving collection of rules and principles, and many teachers consider it necessary to trace the development of the branch of law they are discussing.

In countries where most parts of the law are codified (as in continental Europe, Ethiopia, Thailand, and Japan, for example) it is not generally necessary to go back beyond the code. On the other hand, in countries with a common-law system (England, most members of the British Commonwealth, and most parts of the United States), in which few branches of law are codified, knowledge of the law depends largely on the study of court decisions and statutes out of which common law evolved. This makes the study of legal history of more practical significance.

The place of the social sciences in legal education

Another aim of legal education is the teaching of law in its social, economic, political, and scientific context. While law schools have never ignored the social context of their subject, the 20th century has seen the development of a more or less scientific approach to social studies. It is now possible for lawyers to seek knowledge from professional psychologists, sociologists, economists, and statisticians, rather than relying solely on their own observations. Some law schools appoint psychiatrists or sociologists to their staffs, while others require or allow their students to take courses outside the law school as part of their work toward a degree. This awareness of the other social studies is thought to be more advanced in the United States than in Great Britain. In France and Italy nonlegal subjects are compulsory for a first degree in law; whereas in Germany and The Netherlands they are not. In Japan, the first year or two of the work toward a law degree is spent in liberal arts training.

The graduating law student is not expected to have studied the whole body of substantive law. He is, however, expected to be familiar with the general principles of the main branches of law. To this end, certain subjects are regarded as basic: constitutional law, governing the major organs of state; the law of contract, governing obligations entered into by agreement; the law of tort (or delicts), governing compensation for personal injury and damage to property, income, or reputation; the law of real (or immovable) property, governing transactions with land; and criminal (or penal) law, governing punishment, deterrence, and prevention of dramatically antisocial acts such as homicide and robbery. Less basic subjects include commercial law, the law of marriage and domestic relations, and tax law. A traditional subject like public international law, which deals with the legal relations between sovereign states, may be regarded as basic in one university and yet not even be offered in another. The chief materials are the same everywhere: codes (where these exist), reports of decided cases, legislation, government and other public reports, institutional books (in civil-law countries), textbooks, and articles in learned periodicals. The aim is not that the student should remember "the law" but that he should understand basic concepts and become sufficiently familiar with a law library to carry out the necessary research on any legal problem that may come his way.

Teaching method and procedure

The extent to which legal education aims to teach method and procedure varies from place to place. Attention is always given to the methods of ascertaining the law from the books but not always to the methods of using this knowledge of the law in various roles, such as legal adviser, advocate, draftsman, or judge. Discussion of these matters tends to be more widespread in universities in the United States and in other countries where the main part of qualification to practice the law is a university degree than it is in countries such as England, where professional examinations are largely divorced from academic study. Much the same is true with regard to the rules and principles of court procedure. Few English universities teach these, leaving them to the bar and solicitor's examinations though the law of evidence (governing what facts may be proved in court, and how) is usually an optional subject; some knowledge of civil and criminal procedure may be picked up incidentally during the study of substantive law. In other countries, courses in procedure are frequently compulsory in university law schools.

The establishment of a professional standard and the communication of professional ethics should perhaps be one of the aims of legal education. This receives more emphasis in countries where lawyers qualify to practice mainly by graduating in law at a university. In Great Britain such matters are imparted by the professional bodies and by personal experience rather than through formal instruction. In Japan, legal ethics is taught as a graduate course.

TEACHING METHODS

Pupillage or apprenticeship

The variety of approaches. Methods of legal education are constantly changing. In England and Wales a barrister (an advocate and consultant) may not practice at all without having undergone six months' pupillage in chambers and may not practice independently until he has been a pupil for a year. Pupillage causes some difficulty, partly because of the cost but mainly because of the shortage of places in chambers. To qualify as a solicitor (general legal adviser dealing with all kinds of legal business out of court and advocate in some of the lower courts) it is necessary to serve as an articled clerk for two or four years, depending on whether the candidate is a university graduate or not. In Scotland and Ireland there are similar requirements, though the arrangements differ in detail. In the United States, qualification to practice normally requires a law degree and the passing of state bar examinations; a handful of states require that the candidate serve an apprenticeship (six months as a law clerk). In continental European countries the requirements vary, some placing more emphasis on apprenticeship and others placing more emphasis on examination. Some newly independent countries that previously had no local facilities for legal education have established professional law schools to give practical training to graduates in lieu of pupillage (*e.g.*, the Nigerian Law School in Lagos).

University law schools tend to differ along national lines in their methods of teaching. In the United States, following pioneer work by Christopher Langdell at Harvard in the latter half of the 19th century, the case method came to prevail, in which the student reads reported cases and other materials collected in a case book and the class answers questions about them instead of listening to a lecture by the teacher. The case-book method has been adopted at some institutions in England and other common-law countries but has found scarcely any adherents elsewhere. Even in the United States, many law schools now use seminars and lectures as well. The case method has the advantage of emphasizing the fundamental feature of the common law—the evolution of principles from decisions in actual cases—and thus of focussing the student's attention on method. It has the disadvantage of being relatively time-consuming in relation to the amount of knowledge of legal principle that can be imparted.

The traditional teaching techniques in English universities have been lectures and tutorials (or seminars). In addition, moot courts have been an integral part of English legal education since the Middle Ages. The student argues a point of law before "judges" as if he were appearing in an appellate court. Organized over the centuries for students and young barristers by the Inns of Court, moots spread to the universities when these undertook the teaching of English law. They are sometimes organized by teachers and sometimes by student societies. Most law schools in the United States also have moot programs.

In continental European countries the backbone of legal education is the formal lecture. Class sizes are sometimes very large compared with those in the United States and Britain, and lectures tend to be magisterial performances. Attendance is frequently voluntary, and those who stay away are sometimes able to secure the text of what they have missed orally. Seminars are given too, particularly in the medium-sized and smaller law schools and for specialized subjects. Similar methods are used in other countries with large numbers of law students—*e.g.*, Japan, India, and Thailand.

The case method and the lecture

Teaching methods are not unrelated to the nature of the legal system. The fact that in common-law systems principles of law are largely derived by a process of inductive reasoning from very numerous decisions of superior courts lay behind the development of the case method. In continental Europe, the fact that law is found mainly in systematic codes is one of the chief reasons for the lecture method, in which the subject can be approached through its philosophical background. And there are other considerations. A high staff-to-student ratio makes more seminar work possible. A desire to expound a body of principle rather than a method of approach to materials calls for formal lectures, and this is reinforced in some of the newer countries where published reports of local court decisions and texts on the local law are as yet scanty.

Study and practice. To some extent, all law courses are out of harmony with legal practice, for a case in real life is not presented neatly by a client to his lawyer as a case, say, of contract or divorce. The case begins as a statement, often jumbled, of facts and problems that cut across pedagogical categories. A story of a road accident may involve the lawyer in considering questions, for example, of the civil responsibility for the cause of the accident; of contract (in relation to insurance); of criminal law (in relation to a traffic offense); and of other branches of law as well. It is therefore important, while making divisions of law for convenience of study and examination, to guard students against the danger of thinking in compartments. Lawyers must also contend in practice with branches of law in which they have received no formal education. Patent law, for example, is absent from most law-school curricula; and new legal structures come into existence during every lawyer's lifetime. His task may be eased if he has been trained in not too parochial a fashion and has learned to look to the experience of other countries. Since World War II many new countries have emerged with new constitutions, and lawyers working with them have had to seek assistance from the law of other countries. Thus, Indian lawyers just after 1950 could get little help from their British constitutional law background but did make considerable use of decisions of the Supreme Court of the United States. A good law school produces a graduate who is not constricted by pedagogy; rather, he is equipped to adapt himself to—and perhaps lead in bringing about—legal changes related to social, economic, and political developments.

The curriculum of the law school must also allow for the great diversity of careers followed by those who have been trained in the law. In most countries, large numbers of persons with a legal training seek a career outside the legal profession, commonly in the civil service, in municipal government service, in legal education, and in commerce and industry. Student requirements and tastes differ; and most law schools, therefore, offer a choice. It is common to prescribe a certain number of compulsory subjects, which are regarded as basic to any law student's education, and leave a freedom of selection as to other subjects, stipulating only the number of courses to be studied. There is not even uniformity from law school to law school within the same country as to which subjects are compulsory, and lists of optional subjects vary spectacularly. In the early 1970s more and more opportunities were being given to study comparative law, foreign legal systems, and transnational laws, such as those of the European Economic Community.

EXAMINATIONS AND QUALIFICATIONS

The process of selecting members of the legal profession begins in the universities and law schools and continues afterward in the form of professional entrance requirements.

School examinations. In the United States, Great Britain, and other common-law countries, students are generally required to pass an examination in each subject. Four or five subjects are studied simultaneously during the academic term, and students must take examinations in all of them at the end of the term.

In some countries more comprehensive examinations are the rule. In France, the law student must pass an examination at the end of each of his four years' study for the licence-en-droit, but that is not the general pattern in continental Europe. In West Germany, the course for the university law degree normally takes five years, with a single comprehensive examination at the end on everything studied during those five years. Students are admitted to the final examination if they produce certificates of satisfactory work in each subject, in a jurisprudence seminar, and in a course on economics and finance. The Netherlands has an intermediate system: the course for a first degree in law lasts five years, with an examination at the end of the second year and another at the end of the fifth.

The method of subject-by-subject examination is less taxing on the memory than the system of comprehensive examination. It may well enable the student to do more detailed work on the problems of each subject. It has the disadvantage of encouraging him to think in terms of separate subjects, whereas the comprehenisve examination leads him to consider legal problems in all their aspects. Being aware of the dangers of compartmentalized thinking, some law schools in the common-law world have introduced into their curricula "general" subjects, such as "common law," in place of separate courses in contract and tort, or they require the student to write papers about issues that relate to several of the subjects studied.

Limitations of formal tests

No formal test is wholly satisfactory as a method of screening potential lawyers. The type used most widely, in which students write answers to questions in an examination hall, has been criticized for placing too much emphasis on memory. This criticism is met to some extent in some universities by allowing candidates to consult books and reference materials during the examination, thus bringing the test a little closer to what a lawyer will do when confronted with a real problem. Another objection is that testing creates a situation of stress, in which a candidate does not necessarily demonstrate how he has benefitted from his legal education, and also that the skill demonstrated in the examination hall is not all the skill required of a lawyer. In particular, the examination does not test capacity for patient research nor the capacity for oral argument, which require theses and oral examinations. Examinations to be done outside orthodox examination halls have also been proposed, as a way of providing a more realistic mixture of law and fact and of relevant and irrelevant matter.

Some universities in Great Britain and the Commonwealth countries require one or more long essays or a short thesis or research paper as part of the work for a first degree in law (as opposed to the more substantial dissertation, or thesis, for a postgraduate law degree). This is commonly written during the final year, with no restriction on the resources employed. Universities in civil-law countries sometimes use this type of examination. Critics have noted that mediocre and poor students tend to get quite good marks in this type of exercise by cribbing from published works. Credit is also sometimes given for articles or notes published by students in law reviews. Such student publishing is commoner in the United States than elsewhere, partly because most U.S. law shcools have their own legal journals and partly because American law students are nearly always graduates. Such student work also enhances prospects of employment, particularly when the student becomes an editor of the law review.

Oral examinations are no longer widely used. British and overseas Commonwealth universities hold oral examinations to resolve doubtful results on written papers or as a prerequisite to the award of first class honours. In the United States oral examinations are rare. Some French universities have abandoned them. In Italy, where a law student has to present a thesis after passing his other examinations, the thesis has to be orally defended before examiners. The German law student, after passing his written examination, has an oral one. In Japan, for professional qualification at the Legal Training and Research Institute (see below), there is an oral examination in each of the compulsory subjects after the written examination has been passed.

National differences in requirements

Qualifications for practice. *Common-law countries.* In England a practicing lawyer must be either a barrister or a solicitor (see above *Teaching methods*). It is not necessary to hold a law degree or any university degree to qualify; in the 1960s just under half of those admitted as solicitors had graduated from a university. Barristers must pass two-part examinations in legal subjects, and graduates may obtain partial or total exemption from the first part, depending on their degrees. A barrister's work also includes practical courses and a period of pupillage administered under the authority of the senate of the Inns of Court. For a solicitor, there are also law examinations in two parts, as well as a (varying) period of articled clerkship under the authority of the Law Society.

In the United States, admission to the bar qualifies one for all types of legal work. The only formal requirements are the passing of state bar examinations after graduating from a law school; in a few states, the law degree alone is sufficient.

In both England and the United States as in many other common-law countries, becoming a judge or magistrate is a promotion (by appointment or election) from the ranks of the bar and there is no special training for the exercise of judicial functions. But in some other common-law countries, especially in Africa and Asia, a newly qualified lawyer may enter the government legal service and find himself appointed in a short time to a junior magistracy. Even in these countries there is generally no special training for the job of adjudicating, though refresher courses for those already serving are organized in some countries (*e.g.*, in Tanzania).

Civil-law countries. Various divisions of the legal profession exist in civil-law countries. In France, for example, a legal practitioner may be an advocate, an *avoué*, a notary, or a judge. Each receives a different training, but all will normally have gone through third- and fourth-year law degree courses. The advocate (roughly corresponding to the English barrister) will also have completed a one-year course at the university to train him for his profession and have been examined for his qualifying certificate by a panel of academic and practicing lawyers. The *avoué* (something of a cross between a junior barrister and a senior solicitor), serves a period of articled clerkship, the length of which depends on his degree, and undergoes a professional examination by practicing lawyers. The notary (who does the noncontentious work performed in England by a solicitor) need not be a graduate, but his period of training (six years in a notary's office) is shorter if he is; he also takes a professional examination. To be a judge, one must have a law degree and pass an entrance examination for the magistracy; this must be followed by 28 months of professional training, both theoretical and practical, at the National Centre for Judicial Studies.

In West Germany, the graduate in law who seeks a legal career must embark upon a period of practical training as a *Referendar*. This is a uniform program involving two and a half years of practical work in the courts, in the office of a lawyer in private practice, in the office of a public prosecutor, in the civil service, and, possibly, in the legal department of a commercial concern. Upon its completion, he must pass a state examination.

A somewhat similar procedure is followed in Japan. Law graduates who seek careers as judges, public procurators, or lawyers in private practice must (with the exception of summary court magistrates and assistant procurators) pass the National Law Examination for entrance to the Legal Training and Research Institute. This is an organ of the Supreme Court. Like his German counterpart, the *Referendar* training to become an *Assessor*, the Japanese student at the institute is paid but, unlike the *Referendar*, is not regarded as a civil servant during this period. The period of training at the Institute lasts two years. The bulk of the work consists of practical exercises and discussions, lectures on legal topics, and visits to institutions (such as prisons) of concern to lawyers. The training is uniform, leads to a single examination, and qualifies the graduate for any branch of legal practice.

LEVELS OF STUDY

Law degrees are first degrees in most countries. The student embarks upon the study of law at a university at about the age of 18. In France, the universities offer a course of two years in duration called the ***capacité-en-droit***, which may be taken by anyone at least 17 years of age, regardless of educational background. High marks in this entitle the candidate to enroll for the *licence-en-droit.*

In the United States, most law schools require the entrant to be a university graduate, and none will admit students without substantial liberal arts college education. Consequently, the U.S. student of law is generally in his early 20s. Other countries with this system include India and Pakistan.

University law schools

University law schools in many countries accept all candidates with a certain level of pre-legal education. One drawback of this system is a substantial failure rate in the examination at the end of the first year. In some countries, candidates are screened. In England, for example. each university imposes a quota on entry to its law school and selects among candidates on the basis, usually, of academic performance. (It is possible, however, to enter the legal profession without a law degree; anyone of good character who reaches a certain level of pre-legal education will be admitted to an Inn of Court or to training for qualification as a solicitor.) In the United States, candidates are selected on the basis of academic performance and their demonstrated aptitude for the study of law. In both the United States and England, entry requirements vary according to the prestige of the law school.

Most countries also provide for higher degrees in law. In common-law countries there is usually a series of steps, ascending through a degree of master of laws to a doctorate or senior doctorate. In civil-law countries it is normal to go straight from a first degree to a doctorate. Masters' degrees are, as a rule, based on advanced examination after courses of instruction, though sometimes they are awarded for research or for a combination of examination and dissertation. Doctorates are awarded for theses expounding the results of original research and senior doctorates for published contributions to scholarship in the subject. In many countries there are also specialized postgraduate diplomas or certificates in particular subjects, such as criminology, representing work in a narrower area than that for a master's degree.

TRENDS IN LEGAL EDUCATION

Expansion of legal studies

Modern legal education is expanding both in quantity and in scope. A big upsurge of the teaching of law has occurred in Africa, where newly independent countries have established universities and professional schools concentrating on local laws and practice. Civil-law universities have departed in some countries from the rigidity of prescribed syllabi to allow a greater range of student choice in selecting subjects. In France, for example, universities were empowered after 1968 to arrange their own curricula, which had previously been a matter for government regulation. In Italy, law students now submit their own study plans for approval. In most countries, more attention is paid than formerly to foreign legal systems and to comparative law. The expansion of law schools allows for greater specialization, larger libraries, new subjects, and more student choices.

In some countries, nonlegal subjects have long been part of the syllabus. In France, courses in economics and history are part of the study for the *licence,* and Japanese universities devote the whole of the first 18–24 months of the law curriculum to liberal arts. In some other countries where law as a first-degree specialization has hitherto comprised only law studies, there has been a tendency to include nonlegal studies, joint courses in law and social sciences, or a more sociological approach to law.

The opportunities for university and professional education in law increased greatly after World War II. Many governments have made provision, or greater provision, for the maintenance of students; and legal education has been opened to a larger cross section of society in many places. Nevertheless, the children of middle class parents continue to predominate in the field.

BIBLIOGRAPHY. For England and Wales, much interesting information may be found in the *Report of the Committee on Legal Education* ("Omrod Report," HMSO, 1971), which also contains, in Appendix D, a survey of legal education in Australia, Canada, France, West Germany, India, Italy, The Netherlands, New Zealand, Nigeria, Scotland, South Africa, Sweden, and the United States. In Great Britain the *Journal of the Society of Public Teachers of Law* (semi-annual), publishes articles on legal education, chiefly but not exclusively relating to the United Kingdom; a similar periodical in the United States is the *Journal of Legal Education* (quarterly). ALBERT J. HARNO, *Legal Education in the United States* (1953), is a survey with history and criticism; JULIUS STONE, *Legal Education and Public Responsibility* (1959), is devoted to professional ethics and standards in the United States. Legal education in parts of Asia is discussed in S.P. KHETARPAL (ed.), *1962 Regional Conference on Legal Education* (1965); HAKARU ABE, "Education of the Legal Profession in Japan," in ARTHUR T. VON MEHREN (ed.), *Law in Japan* (1963); and JAY MURPHY, *Legal Education in a Developing Nation: The Korea Experience* (1967). For Africa, see JACQUES VANDERLINDEN (ed.), *Conference on Legal Education in Africa* (1968).

(L.A.S.)

Legal Ethics

Sources of legal ethics

Legal ethics are the ethical principles that members of the legal profession are expected to observe in the practice of law. They are an outgrowth of the development of the profession itself.

Practitioners of law arose when legal systems became too complex for those affected by them to understand and apply the law. Certain individuals with the required ability mastered the law and offered their skills for hire. No prescribed qualifications existed, and these specialists were not subject to legal controls. The incompetent, unscrupulous, and dishonest charged exorbitant fees, failed to perform as promised, and engaged in delaying and obstructive tactics in the tribunals before which they appeared. Action to prevent such abuses was taken by legislation and by judicial and other governmental measures. The right to practice law came to be limited to those who met prescribed qualifications. Expulsion from practice and criminal penalties were introduced for various types of misconduct.

These measures did more than correct the abuses. They also gave recognition to the social importance of the functions performed by lawyers and identified those who were qualified to perform them. This marked the formal beginning of the legal profession. A consciousness developed within the profession of the need for standards of conduct. This became the core of legal, or professional, ethics.

Prior statutes, court rules, and other government directives remained in force along with the profession's self-imposed ethical standards. Taken together, they constituted the sum total of the restraints placed upon lawyers in regard to their professional conduct. This pattern has continued to the present time.

In many countries professional associations of lawyers have sought to reduce the principles of ethical conduct to written form, but a written code is not essential. Ethical principles may exist by common understanding, as well as in the literature and writings of the profession. This is the case in England. A code, however, makes ethically obligatory principles readily available to the practitioner and hence assures greater observance of them. When such a code does exist, it usually contains both statements of general ethical principles and particular rules governing specific problems of professional ethics. But no code can foresee every ethical problem that may arise in the practice of law. Hence, codes are supplemented by opinions rendered and published by committees of bar associations. In the United States, for example, the most important and influential of these are the so-called formal opinions of the committee of the American Bar Association. While not having the same authoritative status as codes of professional ethics, they nevertheless indicate what is deemed proper professional conduct. In England, the General Council of the Bar serves a similar function, all the more important in the absence of a formal code of professional ethics.

Differences among countries. Principles of legal ethics, whether written or unwritten, not only seek to control the conduct of legal practice but they also reflect the basic assumptions, premises, and methods of the legal system within which the lawyer operates. They reflect, as well, the profession's conception of its own role in the administration of justice. In western European legal systems the role of the practitioner, in both civil and criminal litigation, differs from what it is in Anglo-American legal systems, and this is reflected in the ethics of their legal professions. Practitioners in countries having a Communist form of government are ordinarily salaried employees of the government, and their ethical obligations consequently have a focus different from those in countries where lawyers engage in independent practice or are employed by private firms. In England and France, where the profession is divided into separate branches, the principles of professional ethics reflect the relationships incident to that division.

In western European and Anglo-American countries and others with similar systems of justice, such as Japan and India, in which the lawyer is not an employee of the state but engages in private practice to serve clients who employ him, professional ethics are addressed to two basic aspects of the lawyer's status. On the one hand, he is employed by clients to serve and represent their interests; on the other, he is participating in an important social function—the application of rules of law through advice, trial of cases, preparation of legal documents, and negotiation with others for his clients. Hence, the principles of legal ethics stress that the lawyer's chief interest lies in serving his client and in securing justice—not in promoting his own income. He is an agent of his client but deemed to retain a large measure of independent judgment as to the proper course to pursue. He represents his client's interests but may not engage in tactics that defeat the fair administration of justice. The lawyer is engaged, it is said, in a profession and not in a business.

The legal ethics prevailing in countries where the lawyer practices privately have been directed largely toward the general practitioner serving individual clients. They are not addressed to the ethical problems encountered by such specialists as the house counsel of corporations, government lawyers, lawyers employed by labour unions, law teachers, large metropolitan law firms serving important segments of the industrial, commercial and financial communities, and lawyers engaged in providing legal services for poor clients, supported either by private or by public funds.

Conflicting interests of client and society

Naturally the interests of client and society sometimes conflict, and the principles of legal ethics do not always indicate how these conflicts should be resolved. Should a lawyer cross-examine an adverse witness in such a way as to undermine or destroy his testimony, when the lawyer believes the witness is actually telling the truth? May he invoke rules of evidence to exclude points that would weigh against his case but that he considers to be probably true? May he take advantage of the errors of an unskilled opponent? Should he demand a jury trial for purposes of delay when a jury trial has no advantage for his client? These questions may be answered differently in legal systems that operate on different premises. A system in which a lawyer presents his client's case in the most favourable light permitted by law and in which the court must decide the merits of the case may well produce different answers than those produced in a system that assigns a higher priority to the lawyer's duty to the state to assure proper administration of justice.

Areas of application. ***Conflicting interests.*** A lawyer will at times be faced with the question of whether or not he may represent two or more clients whose interests conflict. Quite aside from his ethical obligation, the legal systems of the world generally forbid a lawyer from representing a client whose interests conflict with those of another, unless both consent. This applies not only to lawyers but also to other fiduciaries, such as agents, brokers, trustees, and guardians.

In Anglo-American legal systems the prohibition has three aspects. First, the attorney is not permitted to concurrently represent two or more clients if, in order to further the interests of one, he must forego advancing the conflicting interests of another. In short, he cannot be both for and against a client. Second, he cannot subsequently accept employment from another for the purpose of undoing what he had earlier been retained to accomplish. Third, he may not accept subsequent employment from another if it involves the use, the appearance of use, or possible use of confidential information received from his former client. Such actions are forbidden by law and by legal ethics.

Obligations to the client

To illustrate, an attorney may not ordinarily prepare an instrument for both buyer and seller in which their respective rights are defined. He may not prepare an instrument or negotiate a settlement for a client and later accept employment from another to defeat that instrument or settlement. He may not represent both a driver and his passenger in recovering damages from another party charged with negligent driving in a collision, since the passenger may have a claim against his own driver as well. He may not represent two or more defendants in a criminal prosecution if their respective defenses are inconsistent or, possibly, even when the case against one is stronger than the case against the other. The same principles apply with respect to interest of the attorney that may detract from the full and faithful representation of his clients. For example, he may not purchase property that he has been retained to acquire for his client, nor may he draw a will in which he is a beneficiary.

These conflict-of-interest prohibitions are not absolute. The client may consent to the representation after full disclosure of the actual or possible conflict. But the client's consent may not suffice if public interest is deemed to be adversely affected, as it would be if a lawyer undertook to secure a divorce for one spouse and also act as attorney for the other. Consent divorces are deemed against public policy.

The practicing lawyer who is also a member of a legislature is confronted with a conflict of interest whenever his clients enlist his support to promote or oppose legislation or to secure favourable decisions from administrative agencies that are dependent on legislative financial support. The problem is an important one in the United States, where members of legislatures frequently maintain private law practices, but it has received insufficient consideration by the U.S. legal profession.

Confidential communications. In Anglo-American countries judicial decisions, legislation, and professional ethics forbid a lawyer to testify about his client's communications to him unless the client consents. Similar provisions are found in such diverse legal systems as those of Japan, Germany, and the Soviet Union. In countries in which the attorney's obligation to protect state interests is given relatively greater emphasis, there may be a duty to disclose information when it is deemed to be to the state's advantage. In Anglo-American law the obligation does not apply when the client seems about to commit a fraud or crime. An attorney also may disclose his client's communication if it is necessary to refute a false charge made against him by the client. A similar obligation is recognized by members of the medical profession, clergymen, accountants, social workers, and journalists.

Dignity of the profession

Advertising and solicitation. It has been a long-standing principle of professional ethics in Anglo-American countries that an attorney must not seek professional employment through advertising or solicitation, direct or indirect. The reasons commonly given have been that seeking employment through these means lowers the tone of the profession, that it leads to extravagant claims by attorneys and to unrealistic expectations on the part of clients, and that it is inconsistent with the personal relationship that should exist between attorney and client. A more basic reason appears to have been the social necessity of restraining the motive of personal gain and of stressing the objective of service. Until 1977 the legal profession in all Anglo-American countries took the position that, with some exceptions, the prohibition must be complete. Many specific rules were laid down by bar associations, covering even such details as the content of a lawyer's business card, letterheads, office sign, and announcements of changes in partnerships and office locations. The rules also governed his relation to lay and professional groups that might serve as feeders to his practice and his participation in the publicity attending the trial of a case in which he served as counsel. The situation changed in the United States in 1977 when the U.S. Supreme Court ruled that lawyers could not be barred from advertising their fees. The American Bar Association subsequently revised its code of ethics to include provisions and guidelines for advertising, and suggested that lawyers limit their advertising to basic information about practice and fees.

Fees. Attorneys are ethically enjoined to keep fees reasonable, neither too high nor too low. Many state and local bar associations in the United States have issued minimum-fee schedules as guidance for their members but do not make it an ethical obligation to observe them.

The profession in the United States has assumed, in principle, the obligation to serve poor clients without compensation. The task, however, has become of such magnitude, especially in view of the expansion of the constitutional right to counsel in criminal cases, that ways of providing paid legal services for the poor have emerged, such as legal-aid societies and public defenders.

Fees that are contingent on the successful outcome of litigation or settlement are widely used in the United States, particularly in automobile-accident and other negligence cases, and they are accepted as ethical by the U.S. legal profession. The fee is usually an agreed percentage of the recovery. The justification given is that the injured poor are enabled to assert their claims in court. But contingent fees give the attorney a financial stake in the outcome of litigation—which is ordinarily frowned upon. An offsetting merit may be that in this type of case, where the outcome is difficult to predict, the lawyer also assumes the risk of losing his fee. Nevertheless, in countries other than the United States, contingent fees are generally not permitted. Nor are they permitted in the United States in criminal and divorce cases or to secure a pardon or in the enactment of legislation. Fees for the collection of money claims are frequently based on a percentage of the amount collected.

To avoid the cost of attorney's fees or to supply legal services to those who cannot pay the required fees, various groups in the United States, such as labour unions, corporations, and service organizations, have hired attorneys to provide their members with legal counsel. This practice has been resisted by the U.S. legal profession on several grounds. These include the ground that such action places an intermediary between the attorney and his client, thus diluting the direct personal relationship; and that the interests of the employing organization may at times be in conflict with those of the client. Some decisions of the United States Supreme Court have rejected these contentions and given constitutional status to the right of a labour union or of an organization representing a minority group to provide legal services to its members. But the full scope of this constitutional right remains to be determined.

Honesty in court

Criminal cases. Both the prosecution and the defense of criminal cases raise special ethical issues. The prosecutor represents the state, and the state's concern is not only in convicting the guilty but also in acquitting the innocent. The state's interest may also dictate that prosecution be foregone in some cases in which a conviction could be obtained, as when a person who could be prosecuted supplies evidence against others and thus aids the state in its case. The decision to prosecute must be governed by such considerations. The prosecutor also has an ethical and, in considerable measure, a legal duty to disclose to the defense any information known to him and unknown to the defense that might exonerate the defendant or mitigate the punishment. He must not employ trial tactics that may lead to unfair convictions, nor should he prosecute merely to enhance his political prospects.

The defense counsel has different concerns. Under Anglo-American law, an accused may compel the state to prove that he is guilty beyond a reasonable doubt. The defense counsel, therefore, becomes ethically obligated to require the state or produce such proof, whether or not the attorney believes his client to be guilty. His client's guilt is for the tribunal to determine. The attorney may not, however, deliberately resort to perjured or other false testimony. Similar principles hold in civil-law countries. When the client, against the attorney's advice, insists on testifying falsely, the ethical course to be pursued has not been fully settled. Some maintain that the attorney should withdraw, if possible, or else merely permit the client to testify without aiding him or asserting the truth of the testimony given.

BIBLIOGRAPHY. On the social nature and function of professional ethics, see R.M. MACIVER "The Social Significance of Professional Ethics," *Ann. Am. Acad. Polit. Soc. Sci.,* 297:118–124 (1955); and EMILE DURKHEIM, *Leçons de sociologie: Physique des moeurs et du droit* (1950; Eng. trans. by CORNELIA BROOKFIELD, *Professional Ethics and Civic Morals,* 1957).

The leading texts on the ethics of the legal profession in the United States are H.S. DRINKER, *Legal Ethics* (1953); and L. RAY PATTERSON and ELLIOTT E. CHEATHAM, *The Profession of Law* (1971). The legal ethics of the Canadian legal profession are dealt with in M.M. ORKIN, *Legal Ethics* (1957). For the English legal profession, see T.G. LUND, *Professional Ethics* (1970); and W.W. BOULTON, *A Guide to Conduct and Etiquette at the Bar of England and Wales,* 4th ed. (1965). Information on legal ethics in other countries may be found in P.G. LEPAULLE, "Law Practice in France," *Columbia Law Review,* 50:945–958 (1950); M. CAPPELLETTI, J.H. MERRYMAN, and J.M. PERILLO, *The Italian Legal System* (1967); T. HATTORI, "The Legal Profession in Japan: Its Historical Development and Present State," in A.T. VON MEHREN (ed.), *Law in Japan* (1963); and K. OHIRA and G.N. STEVENS, "Admission to the Bar, Disbarment and Disqualification of Lawyers in Japan and the United States: A Comparative Study," *Washington Law Review,* 38:22–27 (1963).

On the subject of contingent fees, see F.B. MACKINNON, *Contingent Fees for Legal Services* (1965). Specialization in the U.S. legal profession is discussed in B.F. CHRISTENSEN, *Specialization* (1967). The same author discusses group practice in *Group Legal Services* (1967). His *Lawyers for People of Moderate Means* (1970), covers the subject of fees, specialization, group services, and advertising and solicitation.

(M.E.P.)

Legal Profession

One definition of the legal profession calls it "the vocation based on expertness in the law and its application." This simple definition may be best, despite the fact that in some countries there are several professions and even some occupations (*e.g.,* police service) that require this expertness but may not be considered to be within the "legal profession" at all.

HISTORY

Distinct legal systems emerged relatively early in history, but legal professions of size and importance are relatively modern. It was previously thought that all law in ancient times grew out of situations where law, religion, and morals were fused, and that those who applied it necessarily combined an expertness in all three. In some cases, however, and notably in the early Mesopotamian system recorded in the Code of Hammurabi (18th century BC), the law was already substantially secular and distinct from religion and morals. It is likely, though less capable of proof, that a similar state of affairs existed in 3rd-millennium Egypt and China. The evidence of the Hebrew Bible, the Hindu Vedas and Upanišads, and the Muslim Qur'an points to a mixture of law, religion, and morals at *later* stages. On the other hand, early Greek and Roman records suggest a religious stage *before* the secular law became distinct.

But whether early archaic law was distinctly secular or "mixed," there is not the slightest trace in ancient times of a distinct legal profession in the modern sense. The earliest known legal specialist was the judge, and he was only a part-time specialist. The chief, prince, or king of small societies discharged the judicial function as part of the general role of political leader. As his power spread he delegated the function, though not to legal specialists; in the secular stages of the early systems, legal duties were taken over by royal officials who were "generalists." In the wake of powerful religious or quasi-religious movements, priests or savants often judged or advised the judges, a situation that persisted in Muslim countries and in China until the 20th century AD. It may be suspected that, in some of these cases, specialized legal aid to the ordinary citizen did exist, but at levels of social status below the notice of chroniclers or tomb inscriptions, and perhaps without benefit of official approval.

Classical beginnings of a legal profession. **A** distinct class of legal specialists other than judges first emerged in the Greco-Roman civilization, and as with the law itself, the main contribution was from Rome, in the period from 200 BC to 600 AD. In the early stages of both Greece and Rome, as later among the German tribes who overran the Roman Empire, there was a prejudice against the idea of specialists in law being generally available for fee. The assumption was that the citizen knew the customary law and would apply it in transactions or in litigation personally, with advice from kinsmen. As the law became more complex, men prominent in public life—usually patricians—found it necessary to acquire legal knowledge, and some acquired a reputation as experts. Often they also spent periods serving as magistrates and, in Rome, as priests of the official religion having special powers in matters of family law. Among the German tribes noble experts were allowed to assist in litigation, not in a partisan fashion but as interpreters *(Vorsprecher)* for those unready of speech who wished to present a case. The peculiar system of development of the early Roman law, by annual edict and by the extension of trial formulas, gave the Roman patrician legal expert an influential position; he became the jurisconsult, the first nonofficial lawyer to be regarded with social approbation, but he owed this partly to the fact that he did not attempt to act as an advocate at trial—a function left to the separate class of orators—and was prohibited from receiving fees.

The jurisconsult and his successors

The modern legal professional, earning his living by fee-paid legal services, first became clearly visible in the later Roman Empire, when the fiction that a jurisconsult received only gifts was abandoned and when at the same time the permissible fees were regulated. Changes in the methods of trial and other legal developments caused the jurisconsult to disappear in time. The orator, who now was required to acquire legal training, became the advocate. A subordinate legal agent of the classical system, the procurator, who attended to the formal aspects of litigation, took on added importance because later imperial legal procedure depended largely on written documents drawn by procurators. The jurisconsults had been important as teachers and writers on law; with their decline this function passed to government-conducted law schools at Rome, Constantinople, and Berytus, and their salaried professors. There was also a humbler class of paid legal documentary experts, the ***tabelliones,*** useful in nonlitigious transactions.

Medieval Europe. This late Roman pattern of legal organization profoundly influenced the Europe that arose after the barbarian invasions, from AD 1000 on; and even during the invasions the methods of Roman imperial administration never ceased to operate in some parts of southern France and central Italy. The Christian Church, which became the official Roman imperial church after AD 400, developed its own canon law, courts, and practitioners, and followed the general outline of later Roman legal organization. The church's success among the invaders established its jurisdiction in many matters of family law and inheritance. Hence both the idea of a legal profession and the method of its operation remained visible to offset Germanic and feudal objections to legal representation. After the revival of learning in the 12th century, in particular the revived study of Roman law at Bologna, the influence of the late Roman professional system was greatly strengthened.

Classes of legal practitioners

From then on, every country in continental Europe acquired, by various stages and with local variations, a legal profession in which four main constituents could be observed. Procurators attended to the formal and especially the documentary steps in litigation. (In France after the Revolution, the *procureurs* were renamed *avoués.*) Advocates, usually university graduates in Romanic learning, gave direct advice to clients and to procurators and presented oral arguments in court. Among a miscellany of legal scribes, the notaries acquired importance because, in addition to being drafting experts, they provided officially recognized document authentication and archives for internal as well as international purposes. University teachers of law took over the main task of explaining and adapting the mixture of Roman law and Germanic custom that produced the modern laws of the major European countries, and continued to dominate in the scholarly interpretation of the law even after the 19th-century codifications. The relative importance of these classes varied enormously from place to place and from century to century. At times the teaching doctors almost supplanted the advocates; in some courts the procurators swallowed up the advocates, and in others the converse occurred; only the notaries survived with little change.

England after the Conquest. England after the Norman conquest also was influenced by Roman example, and the clerics who staffed the Norman and Plantagenet monarchies and who provided the earliest of their judges enabled the notion of a legal profession, and especially of litigious representation, to be accepted. Only in the ecclesiastical and admiralty courts, however, did procurators (proctors) and doctors of the civil and canon laws become established as practitioners. The native "common law" was developed instead by a specialized legal university, the Inns of Court in London; there, through lectures and apprentice training, men acquired admission to practice before the royal courts. More particularly, they could become serjeants—the most dignified of the advocates, from whom alone, after about 1300, the royal judges were appointed. Various agents for litigation resembling procurators also became known. The "attorneys," authorized by legislation, at first shared the life of the Inns with the "apprentices" in advocacy, who themselves, in time, acquired the title of barristers. Indeed there were cases of men working as both barristers and attorneys. When in the 16th century the court of chancery was established as the dispenser of "equity," the appropriate agent for litigation was called a solicitor, but the common law serjeants and barristers secured the right of advocacy in that court. It was not until the 17th century that the attorneys and solicitors were expelled from the Inns and the division between advocate and attorney became rigid, and not until the 18th century that the barristers accepted a rule that they would function only on the engagement of an attorney—not directly for the client. Other types of legal agents also developed in England, but in the 19th century all the nonbarristers were brought under the one name, solicitor. The order of serjeants was wound up, leaving only barristers, of whom the most senior could be made "Queen's Counsel."

The English division

In its final development the English profession thus bore a resemblance to the European—particularly to that of northern France, where the *parlements* (courts) had a corporate life and apprentice training not unlike that of the Inns, and the Ordre des Avocats acquired a prestige and autonomy resembling that of the English bar. But there were four great differences between England and the Continent. No distinct class of great university teachers and commentators on the national law developed in England, and nonstatutory development of the law took place chiefly through precedent based on the reported judgments of the courts. The continental monarchies also developed a system of career judicial office, in which the young university licentiate went straight into government service, whereas in England appointment of judges from the senior practicing profession remained the settled practice. In addition, the division between barristers and solicitors ultimately became much more rigid in England than did the division between the advocate and procurator in Europe, and Europe never adopted an equivalent of the English practice requiring a barrister to be employed by a solicitor; procurator and advocate were separately and directly employed by the client. England never developed the profession of notary for internal purposes, so that the whole burden of transactional work fell on those who are now the solicitors, with legal advice from the bar.

The worldwide legal profession. The main patterns both of law and of legal practice were exported by Europe and England to their overseas colonies and possessions, and most of the noncolonial countries of the rest of the world imitated one or the other system. Thus the Romano-Germanic (frequently called civil law) practices became the norm for Scandinavia, Scotland, Latin America, most of the Muslim countries of the Middle East, for French-speaking areas, Portuguese and Spanish Africa, and for Japan, Thailand, and the former French parts of Southeast Asia. They have also influenced Communist practice. The English system provided the model for English-speaking North America, most former English colonies in Africa including South Africa, most of the Indian subcontinent, Malaysia, Australia, and New Zealand. The original model has undergone considerable modification by both the countries of export and the countries of influence. In particular, the specialization of procurator-advocate and solicitor-barrister has tended to be replaced by a "fused" profession of legal practitioners qualified to perform both functions and usually doing so. (Such a fusion also occurred in 19th-century Germany, producing a "Rechtsanwalt" corresponding to the U.S. attorney-at-law.) In Latin America, the fused profession is general, except that the specialized notary survives.

COMPOSITION AND CHARACTERISTICS OF THE PROFESSION

The legal profession has always had an ambiguous social position. Leading lawyers have usually been socially prominent and respected—the sections of the profession so favoured varying with the general structure of the law in the particular community. The family status of early Roman jurisconsults may have been more important than their legal expertise in securing such a position, but by the time of the Principate it was their legal eminence that made them respected. The English serjeants lived magnificently, especially in Elizabethan times, and the French Order of Advocates was established (14th century AD) by feudal aristocrats in circumstances reminiscent of early Rome—including an insistence on receiving gifts rather than fees. The early Italian doctors of the civil and canon law (12th–15th centuries) were revered throughout Europe. In England and the countries influenced by its system, the highest prestige gradually came to be concentrated on the judges, rather than on the order of serjeants, of which they were members; and the judges remain the only legal class in the liberal-capitalist common law countries of today to enjoy something like charisma. In the Romano-Germanic systems, it is the notaries and the advocates who have come to be most trusted or admired, the judiciary being more closely identified with the civil service.

The legal profession and social conflict

Yet along with this high repute, sustained over two millennia, lawyers have also been among the most hated and distrusted elements in society. In a few cases this has been the consequence of a powerful ideology in the society which was hostile to the whole idea of law, China being the most important example. Confucian teaching (6th century BC) was against the use of civil law as a major means of social control, and this influence remained powerful there and in Japan until the 20th century. In Communist Russia, the early leaders (1917–22) imagined that law and lawyers were peculiar to "exploiting" societies and would soon wither away in classless Communism, and this belief was revived in the early days of Chinese Communism (1947–55). Further experience persuaded the Communist governments that there was room for "socialist legality" and for lawyers to serve it, but a degree of mistrust remains and the repute of the legal expert is lower than that of the political and technological expert.

Most lawyers are conservative, because the law itself is predominantly intended to satisfy expectations arising from an inherited pattern of behaviour; in a particular social setting, this tends to identify the lawyer with the established and successful classes and to make him seem an enemy to oppressed classes or "new men." Individual lawyers have, nevertheless, often been on the side of rebels: Robespierre and Lenin were both lawyers. But the dominant attitude of the legal profession is for moderation. Thus many lawyers took the British side in the American Revolution, and even among the lawyers who took the other side, the predominant influence was against any attempt to turn the political revolution into a social revolution.

Along with these varying ideological and political reasons for popular distrust, and even more deep-seated, are the inherent difficulties associated with law and with some of the legal functions. Most men would like law to be so certain that its application in any case is equally certain, and so simple that any person of sense can see how it applies. In a discipline sharing the imperfection and complexity of society itself, no such situation is attainable, and the lawyers are blamed for the basic difficulty of their craft—which in some instances they worsen needlessly themselves by multiplying obscurities, contradictions, and complexities. The legal function likely to be most distrusted by the average person, though it also produces some of the law's heroes, is litigious advocacy, particularly in the criminal law. Plato and Aristotle condemned the advocate as one who was paid to make the better cause appear the worse, or endeavoured by sophisticated tricks of argument to establish as true what any man of common sense could see was false. The feeling against advocacy in the criminal law was so strong that, at least in the case of the more serious kinds of crime, a right to representation by a trained advocate was nowhere generally recognized until the 18th century AD.

Governments and the members of organized legal professions have from the infancy of the craft endeavoured to meet the basic problem of representation by a basic rule of professional ethics—that the dominant duty of the advocates is not to his client but to truth and the law. Since the later Roman Empire, advocates have been required to take oaths to this effect, and lawyers are often technically classed as "officers of court." The duty of the advocate is to fight for the rights of his client, but only up to the point where an honourable man could fairly put the case on his own behalf. He must not identify with his client's possible willingness to tell untruths, or to misrepresent the relevant law.

THE FUNCTIONS SERVED BY THE LEGAL PROFESSION

Private practice. Client-directed lawyers often are called counsellors, but in the original sense of that word—giving advice as to how the law stands—this is rarely an independent function; it is an inseparable part of the other functions. In his client-directed activities the lawyer is concerned with how the law affects specific circumstances, which can for convenience be divided into two main types: transactional and litigious.

In the transactional type, the lawyer is concerned with the validity or legal efficacy of a transaction, independent of any immediate concern with the outcome of litigation. Such activities comprise the largest area of professional activity, whether considered from the point of view of the number of lawyers involved, or of the time they have to spend on the task, or of the number of clients affected. If the events constituting the transaction in question happened before the lawyer is consulted, he can only advise on their legal significance and perhaps suggest methods of overcoming legal deficiencies in what has been done. If future conduct is involved, he is better placed to plan his client's course of conduct so as to secure the required end in the most economical fashion that the law permits. Transactions may concern words and acts, but characteristically they require the drafting of documents. In the Romano-Germanic systems these often require notarization. Typical activities falling in this category today are: transferring interests in property (especially if, as in the case of land, more than a mere expression of intention is required to satisfy formal legal requirements) ; transmitting property on death; settling property within a family; making an agreement (especially if a commercial agreement of some complexity and duration is involved); incorporating or winding up a corporate entity; varying the terms on which a corporate entity is conducted (classes of shares, managerial rights, distribution of profits, etc.); and adjusting the ownership and control of property and income so as to comply with the requirements of taxation laws and minimize their impact on the property and income in question, or so as to ensure the proper management of the assets and distribution of the proceeds among beneficiaries (estate planning), or both. In the Romano-Germanic systems many of these functions are discharged by notaries, and in the English and similar divided systems by solicitors, though in difficult situations the opinions of advocates or barristers may be obtained. In the fused professions of North America, some firms of attorneys specialize in business of this type and avoid, so far as they can, the litigious function. A frequent accompaniment of such work is the handling of clients' money, and in many contemporary legal systems it has become common for statutory regulation to require rigid accounting and the provision of a guarantee system to protect such funds.

Typical legal functions

The litigious function is subdivided into three main stages. First is the preparation of the case—interviewing the client and witnesses, investigating the circumstances in the light of the leads provided by the client, and attending to the formal requirements of the procedure in question, which may involve writs, summonses, filing of statements of claim or defense, gathering information through interrogatories and discovery of documents between the parties, and preparing for trial. Second is the trial proper, in which the facts and law are established and argued before the judge and a decision is made. Third is the execution of the judgment—obtaining payment of damages, delivery of property, or performance of obligation in civil cases; payment of fine or imprisonment, etc. in criminal matters. Similar stages arise on appeal. In the divided professions, the sharing of these functions is intricate and varies between one system and another. The advocate or barrister is especially responsible for the second stage, the day in court, but he may advise upon or draft many of the documents used in other stages. If incidental disputes concerning procedure have to be litigated, he is likely to conduct the proceedings; and if the procedure includes a pretrial conference, he is likely to represent the client. Otherwise the first and third stages are primarily the province of procurator or solicitor.

Public-directed practice. As noted above, the highest function served by the legal specialist today is that of judging, but it can be regarded as a separate profession. Where judges are elected, as they are in many instances in the U.S., they may return to practice at their term's end, and practitioners may consider judicial office as merely an interlude in their careers as attorneys. In England, senior barristers often accept positions as temporary judges—recorders, or chairmen of quarter sessions—while continuing in the main their professional career, though for a majority this is an apprentice stage before life appointment to a judicial post. In the Romano-Germanic systems, the judges traditionally have been considered a separate profession within the civil service, and in many respects subject to civil service rather than law-profession control. This, however, may be changing in some countries. Thus in West Germany and in Japan (whose modern legal profession and judiciary were modelled on the French and German systems), there is a contemporary move to extricate the judiciary from a primary identification with government service and associate them actively with the private practicing lawyers. As yet, however, appointment of judges from the advocates, though legally possible, is unusual.

Governmental roles of the legal profession

Governments have always themselves required a staff of legal specialists, and the scope for such employment today is enormous. There is usually a senior political officer—minister for justice, attorney general, solicitor general—who by convention needs to be a lawyer, and a government department concerned mainly with the legal problems of the government as client (in the English-derived systems, usually the department of the attorney general). Increasingly, however, it is found that the great departments of state need their own legal subbranch. In Communist countries, lawyers tend to work mainly for government and for collectivized industrial and farm organizations.

One of the oldest and still most difficult of the governmental legal functions is that of prosecutor in criminal cases. Prosecution is often in part carried on by private persons acting through private lawyers, but the trend is to concentrate the function in government legal officers. In most Commonwealth countries, the Crown or public prosecutor is a specialized officer under the general control of the attorney general. England has an independent "director of public prosecutions," concerned only with the most serious types of crime, but most prosecutions are conducted by private barristers briefed by him or by police. In the U.S. this function has come to be mainly local, and prosecutors are elected for short terms, the commonest title being district attorney. In the Romano-Germanic systems, prosecuting is a career service but often organized in substantial independence from other forms of government legal service.

The prosecuting function is particularly delicate because criminal prosecution can be used as an instrument of oppression and political persecution, even where conviction is not obtained, and because in most systems prosecutors are expected to act with a degree of fairness and restraint not necessarily expected of the parties to civil litigation. Many Romano-Germanic systems employ officers who keep a general supervision over the working of the courts and especially their criminal jurisdiction. This is the office of the "procurator general," or "officer of justice," and a similar service exists in most Communist countries.

Another branch of government, the legislature, usually requires legal assistance. For one thing the presiding officers apply complex standing orders with legal advice, such as that given by the parliamentarian at Washington and the standing counsel to the speaker at Westminster. Legislation also needs to be expressed in language readily applicable by judges and lawyers and to be framed in harmony with the existing body of law. This requires the service of parliamentary draftsmen who are expert lawyers. A further specialized branch of advisory activity associated with legislation has recently become prominent—the law reform commission or committee.

Teaching and scholarship. A prominent role for the legal profession since Roman times has been concerned with teaching and scholarship in the law. Until the 18th century, teaching of the English common law was vested exclusively in the Inns of Court, and a good deal of European teaching for professional practice—particularly in the case of notaries and procurators—was also professionally organized. Even university law teaching in Europe was rarely aloof from legal practice; there was usually a fruitful interchange between practitioner and teacher, exemplified in such great figures as the French 18th-century teacher, advocate, and judge Robert Joseph Pothier, whose commentaries provided the foundation for the Napoleonic Code of Civil Law. Much law teaching in the new university law schools that sprang up in the U.S., the United Kingdom, and the Commonwealth in the 19th and 20th centuries was carried on part-time by attorneys, barristers, and judges, and some still is. Blackstone, the first Vinerian professor of English law at Oxford, came from the bar and became a judge. Law teaching has today tended to become a distinct, full-time profession, usually carried on at a university. Teachers and practitioners contribute to an enormous professional literature in all countries, with students' texts, practical manuals, theoretical monographs, and a periodic literature whose bulk is coming to be almost as big a problem as the enormous bulk of reported judicial decisions that are consulted for guidance and precedent.

THE AUTONOMY OF LAW AND LAWYERS

At least since the classical Greeks, a recurring political theme has been the need for a government of laws rather than of men. Actually, however, as the legal philosopher Julius Stone has said, society of necessity has a government of laws and yet of men, and the demand for *legal* autonomy is often seen in practice as a demand for freedom of the *lawyers* from political dictation or influence. The main issue has been the independence of the judiciary, and democracies have been particularly assiduous in cultivating both a spirit and traditions that respect judicial independence. The details of their government structure or constitutional guarantees tend in that direction, offering obstacles to the ready dismissal of judges, charging their salaries on consolidated revenue, and prohibiting the vesting of judicial functions other than in duly constituted courts of law.

Political context of the judiciary

The special position of the judiciary in constitutional states is usually considered to be an aspect of the division of powers, but it should also be considered in its relation to the structure of the legal profession. Since later Roman times, admission to the practice of law and the regulation of the practicing profession has been habitually vested in the judiciary. The duty of the advocate to speak fearlessly for his client has often required courage in the face of political threats directed against the advocate and also against the court before which he appeared, so that judicial courage was also required. The legal profession as a whole is then seen defending "the rule of law" against the political regime.

The issue of judicial independence may sometimes, however, be seen in the context of tension between the judges and the advocates. In the Romano-Germanic systems, the judges often are subject to a strong corporate discipline within their own craft, and differences can occur between them and the body of advocates, and also between them and the university teacher-commentators. These differences may relate to questions of legal ethics, especially the limits of advocate identification with client, or to questions of legal doctrine; the judges are then apt to be considered as representing "the state" and the advocates and teachers the autonomy of the law. In the English-derived systems, judges are much less subject to corporate discipline, and disputes with the bar are more likely to arise with individual judges and to be highly personal. Even in stable countries, where the rule of law and the independence of judiciary and profession are respected, there is a less dramatic tension between the standards and tone of the lawyers on the one hand and the political administration on the other. For the lawyers, policy is largely concealed in the propositions that constitute the normative system, and legal reasoning usually involves definitions and processes of inference from the body of such propositions themselves rather than directly from the policies which the norms subserve. There have often been revolts against such "logic" within the legal profession itself, especially in the 20th century, but it still remains the commonest method of thinking among lawyers and it is doubtful whether one can speak of a "rule of law" at all unless a good deal of legal reasoning is conceptual in style. Politicians and administrators, on the other hand, are more likely to reason directly from policies and purposes and from the considerations relevant to their attainment. The classic illustration of the tension is that between the police officer, confident that he has the guilty man and intent only on putting him in jail, and the lawyers and judge, who insist on the need for "conviction according to law," which may involve the application of rules of evidence seeming artificial and even absurd to the police officer. In rigid consitutional systems where there is judicial review of legislation, politicians may be affronted at the way in which political issues are transformed by the lawyers into legal issues. In many modern countries there has been a tendency to remove certain kinds of dispute both from the courts and from the lawyers, and to vest their determination in administrative bodies before which lawyers are denied standing, so as to escape what has been regarded as the blight of legal reasoning; as often there have been reactions in favour of restoring the rule of "law" and the lawyers. In such disputes, it is often difficult to distinguish between lawyer attitudes that reflect the necessary features of a rule of law from those that merely reflect the temporary self-interest of particular lawyers or their clients.

ORGANIZATION AND DISCIPLINE OF THE PROFESSION

Since about 1800, most countries have brought their legal professions under systems of statutory control with three main principles: admission to practice automatically and compulsorily makes the lawyer a member of an appropriate professional association; those associations are given substantial powers in relation to legal education, admission to practice, and the disciplining of the profession, but subject to overriding powers vested in the courts and/or (especially in the Romano-Germanic systems) in government legal departments; the practice of law for reward is prohibited—generally or as to particular functions—to persons not admitted under the system. In the U.S., about half of the states have such a system, there called that of the "integrated bar"; in the other states, bar associations are voluntary and have few controlling powers. England has retained the traditional Inns of Court (in whose management the judges play a leading role) for barristers, but solicitors are subject to a statutory system as above. In some countries (*e.g.*, France) professional organization is regionalized to correspond with judicial organization, and in some federal countries (*e.g.*, the **U.S.**, Canada, and Australia) professional control is vested in the states; such situations create the problem of a national organization which is generally a voluntary federation of regional bodies and therefore lacking in compulsive authority. The American Bar Association, established 1878, is a leading example. In other federal countries (*e.g.*, West Germany, India) the central government has created national law associations responding to the system of control. The law associations, apart from the functions already mentioned, help their members to understand and apply professional ethics, and they develop canons of ethics to cover new problems. They are often active in policing the prohibition of legal practice by unqualified persons, which tends to bring them into dispute with other professions—*e.g.*, tax accountants and land salesmen—whose members wish to perform legal functions in relation to their tasks and often have considerable knowledge of the relevant law.

Profes-
sional
associa-
tions:
protection
and control

Where the profession is divided, it is today usually possible to transfer from one branch to another, though sometimes after delay or subject to additional training. In many of the Romano-Germanic systems, however, professional mobility is severely restricted by another factor—numerical limits on the numbers admitted to a branch of the profession. There are usually limits to the permitted numbers of procurators and notaries, and in some cases, notably the highest French courts, advocate and procurator functions have been combined in relation to a particular jurisdiction, and a limit has been placed on numbers; otherwise the number of advocates is generally not restricted. In the restricted cases a person admitted to practice can actually work in the profession only as an employee of an existing practitioner or after buying out such a practitioner.

The practical ability to enter or carry on in the profession can also be much influenced by the varying national rules as to legal partnerships. They are prohibited for English barristers and for most divided bars derived from that system, and among most of the Romano-Germanic specialized advocates and notaries; France now permits *avocats* to have partnerships, but they are rare. Incorporation of legal practitioners is almost universally prohibited. These restrictions result from the emphasis on personal responsibility of the individual lawyer to his client, to the court, and to the ethical system. In the fused professions, however, partnership is usually permitted, and in the U.S. some very large firms have developed, particularly among the "corporation lawyers"—referring to their clients—of New York. But even in the **U.S.** single-person practices and small partnerships are commonest.

THE FUTURE OF THE LEGAL PROFESSION

Law faces problems that are common to most professions, such as the consequences of rapidly growing size and complexity of society and of knowledge, and the impact of technological change; thus lawyers will have to adjust to the use of computers to store and retrieve information about the proliferating body of statutes and judicial decisions, and more and more frequently must use scientific and social expertise in the solution of specifically legal problems.

In many countries, changes in the demand for legal services are also compelling changes in organization. Generally, lawyers have in the past been readily available only to those with above-average means. Today there is a demand for making such services available in principle to all citizens. In the **U.S.** the particular case of legal aid to persons accused of crime became urgent because of the Supreme Court's interpretation of the Constitution as making such aid necessary, if requested, to the validity of convictions. There is also a demand for wider legal aid in civil matters. Britain and many Commonwealth countries have established systems for the provision of aid in both criminal and civil matters, which is jointly administered by government bodies and legal associations; the policy is to retain the principle of a private legal profession although financial support is provided by government. In other countries, chief reliance may be placed on legal aid wholly provided by government; this creates the problem of the independence of the aid, especially where government is also a party in a case. In the Communist countries, a situation where the lawyers' responsibility to the government or to the collectivity was very heavily stressed is

Legal aid

gradually giving way to one in which it is recognized that the individual requires some independent aid; in the U.S.S.R. this is difficult because the "cooperatives" of advocates available to individuals are heavily regulated, but less difficult in other European Communist countries where the lawyer cooperatives have greater freedom.

The need for legal specialization also creates problems. Early in the 20th century it seemed likely that the barrister-solicitor and procurator-advocate division would soon disappear entirely, as it has done in many countries, but later in the century it appeared that surviving cases of the division (notably England and Scotland) would continue; in some Australian states there has been reversion to a divided profession after earlier fusion measures. This testifies to the technical advantages of separating the functions of advocacy and specialist advice, but there are rigidities in the English situation that multiply costs; nor does the notary-procurator-advocate division altogether fit the contemporary needs of the Romanist systems where it survives. The very large partnerships that arise where there are fused professions allow some specialization to occur within the firm. In a regime of small general practitioners, specialization is more difficult because the individual attorney wishes to keep his client for all purposes; this difficulty can be mitigated by such devices as the "lawyer referral" services now operated by many U.S. bar associations, though these grew up to provide guidance on reliability rather than on specialized ability.

With the decline of apprentice law training and growth of full-time law teaching, there is a risk of discontinuity between teaching and practice. This is being bridged by the development of practical training and experience courses in law schools, such as legal-aid clinics and drafting seminars, and by teachers spending periods in government or private practice; but the cure is itself creating problems by placing excessive stress on the teaching of technique. The lawyer is caught between cross fires of demand; some would like him to be the chief moral conscience and applied sociologist of society, while others think he would do better as a specialized technician in his craft who leaves "generalizing" to others. In the open societies, lawyers will make their own diverse choices.

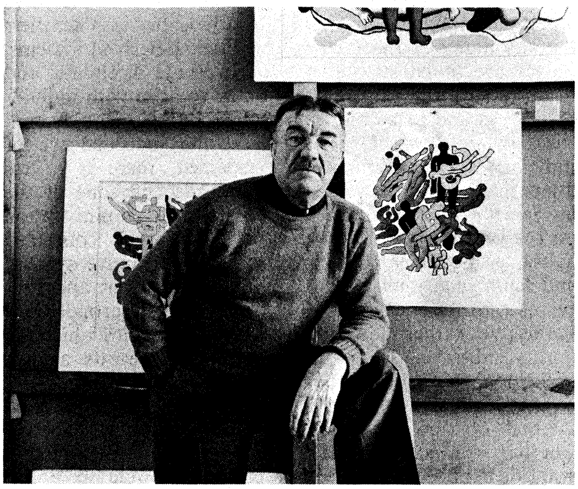
BIBLIOGRAPHY. J.H. WIGMORE, *Panorama of the World's Legal Systems,* 3 vol. (1928), still the best general survey; D. BODDE and C. MORRIS, *Law in Imperial China* (1967), a comprehensive account closely based on original sources; R.J. BONNER and T.C. BURGESS, *Lawyers and Litigants in Ancient Athens: Genesis of the Legal Profession* (1927); F. SCHULZ, *History of Roman Legal Science* (1946), one of the few books on Roman law concentrating on the role of lawyers; *Cambridge Mediaeval History,* vol. 5, c. 21 by H.D. HAZELTINE (1926), a classic account of Roman and Canon Law in the Middle Ages; A. WEISSLER, *Geschichte der Rechtsanwaltschaft* (1967), the only comprehensive account of the legal profession for Germany and France; H.J. COHEN, *History of the English Bar and Attornatus to 1450* (1929), a classic account of the history of the legal profession in the Middle Ages; M. BIRKS, *Gentlemen of the Law* (1960), a definitive work on the English solicitor; R.E. MEGARRY, *Lawyer and Litigant in England* (1962), and B. ABEL-SMITH and R. STEVENS, *Lawyers and the Courts: A Sociological Study of the English Legal System, 1750–1965* (1967), two opposing views of the legal profession in England; Q. JOHNSTONE and D. HOPSON, *Lawyers and Their Work* (1967), a balanced comparative survey of the U.S. and English professions and their future development; M. CAPPELLETTI, J.H. MERRYMAN, and J.M. PERILLO, *The Italian Legal System,* ch. 3 (1967); J.N. HAZARD, I. SHAPIRO, and P.B. MAGGS, *The Soviet Legal System* (1969); A.T. VON MEHREN (ed.), *Law in Japan* (1963); *International Lawyer* (quarterly), and the *Journal of the International Commission of Jurists* (quarterly), the only periodicals that regularly carry articles and notes on professional organization throughout the world.

(G.S.)

LCger, Fernand

Few 20th-century artists have accepted tbe Industrial Revolution and a possible era of the common man with as much enthusiasm as the French painter Fernand Léger displayed during a long and, although qualitatively uneven, remarkably consistent career. Beginning in the years before World War I with a personal variety of Cubism, he developed a vigorous style that combined forms reminiscent of machinery with elements of geometrical abstraction, bright, poster-like zones of colour, and a hint of the popular lithographs of around 1900. His favourite subject matter included acrobats, cyclists, and construction workers. Since his death in 1955 his reputation, although somewhat that of a period figure, has grown.

© Arnold Newman



Léger, photograph by Arnold Newman, 1941.

Léger was born on February 4, 1881, at Argentan, a small town in Normandy. His father and his grandfather were peasants engaged in livestock raising. After completing his education in local secondary schools, he served a two-year apprenticeship in an architect's office at Caen and then, in 1900, went to work in Paris, first as an architectural draftsman and later as a retoucher in a photographer's studio. In 1903 he enrolled in the Paris Ecole des Arts Décoratifs and, although failing to get into the Ecole des Beaux-Arts, began to study under two of its professors as an unofficial pupil. A large retrospective of the work of Paul Ctzanne at the Paris Salon d'Automne of 1907 influenced him profoundly, as it did many other young painters who were looking at the time for a solid way forward from the form-dissolving dazzle of the Impressionists and the undisciplined colour of the recent Fauvist movement. "All of us," Léger said later, "had Cézanne as a point of departure." In 1908, the year Cubism began, he rented a studio at La Ruche (The Beehive), an artists' settlement on the edge of Montpamasse, and there he soon found himself in the centre of several avant-garde tendencies. Eventually, he got to know the painters Robert Delaunay, Marc Chagall, and Chaim Soutine; the sculptors Jacques Lipchitz, Henri Laurens, and Alexander Archipenko; and the poets Guillaume Apollinaire, Max Jacob, Blaise Cendrars, and Pierre Reverdy. Through the poets, in particular, there was a connection with the Cubist movement, the early centre of which was Montmartre, where Pablo Picasso and Georges Braque had their studios.

Léger had been painting in a style that mixed Impressionism with Fauvism; now, under the pressure of his new environment, he evolved rapidly. In 1909 he produced "The Seamstress" ("La Couseuse"; private collection, Paris), in which he reduced his colours to a combination of blue gray and buff and the human body to a construction of slabs and cylinders that resembled a robot. That same year he began "Nude Figures in a Wood," in which the geometrical volumes composing the figures are broken into large fragments. By 1913 he was painting, in brighter colours, the series of dynamic, sometimes completely abstract studies he called "Contrast of Forms" (Philadelphia Museum of Art, among other places); here his style, aptly nicknamed "tubism," was intended to illustrate his theory that the way to achieve a maximum of pictorial effect was to multiply contrasts of colour, contrasts of curved and straight lines, and contrasts of solids with each other or with flat planes. In 1914, in a lecture

Development of "tubism"

entitled "Les Révélations picturales actuelles" ("Contemporary Achievements in Painting"), he added to this aesthetic basis of his art an affirmation of his faith in modern life and popular culture:

> The breaks with the past that have occurred in our visual world are innumerable. . . . The advertisement hoarding, which brutally cuts across a landscape in obedience to the dictates of modern commerce, is one of the things that have aroused most fury among men of so-called good taste. And yet that yellow or red bill-board, shouting in a timid landscape, is the finest of possible reasons for the new painting; it knocks head over heels the whole sentimental and literary conception of art . . . .

During World War I, in which he fought as a sapper in the front lines, he acquired a strengthened sense of reality and a renewed interest in cylinders. "Without transition," he remembered, "I found myself at the level of the entire French people . . . . At the same time I was dazzled by the breech of a 75 [artillery piece] in full sunlight, by the magic of the light on the bare metal . . . . Total revolution, as man and as painter." Gassed at Verdun, he was hospitalized for a long period and was finally released from the army in 1917. That year he completed "Soldiers Playing at Cards," which he regarded as "the first picture in which I deliberately took my subject from our own epoch." By 1919 he was in what has been called his mechanical period, which was marked by a fascination for motors, gears, bearings, furnaces, railway crossings, and factory interiors. In the mid-1920s he was associated with the French formalist movement called Purism, which had been launched by the painter Amédée Ozenfant and the painter-architect Charles-Édouard Jeanneret (better known as Le Corbusier). But from then on, Léger's art was essentially figurative, and the only significant change in his style was a tendency, begun during World War II, to separate his bands of colour from his drawing and to leave them abstract.

Work in mediums other than painting

Léger took an interest in many arts besides painting. He designed sets for ballets and motion pictures, and in 1926 he conceived, directed, and produced ***Le Ballet mécanique*** ("The Mechanical Ballet"), a purely non-narrative film with photography by Man Ray and Dudley Murphy and music by the U.S. composer George Antheil. He was concerned all his life about the relationship of colour to public buildings, and he was able to realize some of his ideas in the mosaic facade of Notre-Dame-de-Toute-Grâce at Assy, in southeastern France (1949); in a mosaic for the crypt of the American memorial at Bastogne (1950); in a mural for the United Nations building in New York; and in several projects for stained-glass windows, such as those for Sacré-Coeur of Audincourt, France (1951). His desire to bring his art closer to the life of ordinary people may have been one of his reasons for deciding, in 1945, to join the French Communist Party—although in fact he never practiced the Social Realism that was then favoured by Communist leaders.

During the last years of his life his major paintings were "Les Constructeurs" and "The Great Parade." A large number of studies and variations can be linked to both pictures. Leger died on August 17, 1955, at Gif-sur-Yvette, in the suburbs of Paris. At Biot, in southern France, there is a museum devoted to his work.

MAJOR WORKS

PAINTINGS: "Nude Figures in a Wood" (1909–10; Rijksmuseum Kröller-Müller, Otterlo, The Netherlands); "Soldiers Playing at Cards" (1917; Rijksmuseum Kröller-Müller, Otterlo); "Acrobats in the Circus" (1918; Kunstmuseum, Basel, Switzerland); "The City" (1919; Philadelphia Museum of Art); "The Mechanic" (1920; Galerie Louis Carré, Paris); "Three Women" ("Le Grand Déjeuner," 1921; Museum of Modern Art, New York); "The Great Tug" (1923; Musée Fernand Léger, Biot, France); "Woman with Book" ("La Femme au livre," 1923; Nelson A. Rockefeller Collection, New York); "Composition" (1925; Solomon R. Guggenheim Museum, New York); "Three Musicians" (1925–44; Wright Ludington Collection, Santa Barbara, California); "Three Faces" (1926; private collection, New York); "The Vase" (1927; private collection, Chicago); "Still Life with Lamps" (1928; private collection, Chicago); "Divers on a Yellow Background" (1941; Art Institute of Chicago); "The Black Trellis" (1943–44; Solomon R. Guggenheim Museum, New York); "Leisure" ("Les Loisirs," 1944–49; Musée National d'Art Moderne, Paris); "The Great Julie" (1945; Museum of Modern Art, New York); "Les Constructeurs" ("The Builders," 1950; Musée Fernand Léger, Biot, France); "The Great Parade" (1954; Solomon R. Guggenheim Museum, New York).

MURALS: designs for the Swedish Ballet (1921–22; Paris); mosaics (1949; Notre-Dame-de-Toute-Grâce, Plateau d'Assy, France); memorial (1950; Bastogne, France); stained glass (1951; Sacré-Coeur, Audincourt, France); murals (1952; U.N. General Assembly auditorium, New York); stained glass (1954; church, Courfaivre, Switzerland).

BIBLIOGRAPHY. Excellent biographical and critical studies are PIERRE DESCARGUES, *Fernand Léger* (1955); and ROBERT DELEVOY, *Léger* (1962). DOUGLAS COOPER, *Fernand Leger et le nouvel espace* (1949), is a monograph with an extensive bibliography by HANNAH MULLER of the Museum of Modern Art, New York. The artist's principal writings have been collected in ***Fonctions de la peinture*** (1965).

(R.McMu.)

Leibniz, Gottfried Wilhelm

Gottfried Wilhelm Leibniz, whose universal genius has influenced such fields as logic, mathematics, mechanics, geology, jurisprudence, history, linguistics, and theology, dominated the intellectual life of Germany in the late 17th and early 18th centuries, influenced the German Idealists Johann Gottlieb Fichte and G.W.F. Hegel, reappeared as an intellectual force at the beginning of the 20th century, and became again in the 1970s, particularly through his scientific thought, one of the most relevant philosophers. His theory of monads—ultimate, simple, unextended, spiritual substances that are the basis of all composites—was the characteristic element of his Rationalist metaphysical system.

By courtesy of the Herzog Anton Ulrich-Museum, Braunschweig

Leibniz, oil painting by A. Scheits, 1695. In the Herzog Anton Ulrich-Museum, Braunschweig.

Early life and education. Leibniz was born into a pious Lutheran family in Leipzig on July 1, 1646, near the end of the Thirty Years' War, which had laid Germany in ruins. As a child, he was educated in the Nicolai School but was largely self-taught in the library of his father, who had died in 1652. At Easter time in 1661, he entered the University of Leipzig as a law student; there he came into contact with the thought of men who had revolutionized science and philosophy—men such as Galileo, Francis Bacon, Thomas Hobbes, and René Descartes. Leibniz dreamed of reconciling—a verb that he did not hesitate to use time and again throughout his career—these modem thinkers with the Aristotle of the Scholastics. His baccalaureate thesis, ***De Principio Individui*** ("On the Principle of the Individual"), which appeared in May 1663, was inspired partly by Lutheran nominalism (the theory that universals have no reality but are mere names) and emphasized the existential value of the individual, who is not to be explained either by matter alone or by form alone but rather by his whole

Contact with new schools of science and philosophy being (entitate *tota*). This notion was the first germ of the future "monad." In 1666, he wrote De Arte *Combinatoria* ("On the Art of Combination"), in which he formulated a model that is the theoretical ancestor of some modern computers: all reasoning, all discovery, verbal or not, is reducible to an ordered combination of elements, such as numbers, words, sounds, or colours.

After completing his legal studies in 1666, Leibniz applied for the degree of doctor of law. He was refused because of his age and consequently left his native city forever. At Altdorf—the university town of the free city of Numberg—his dissertation De Casibus Perplexis ("On Perplexing Cases") procured him the doctor's degree at once, as well as the immediate offer of a professor's chair, which, however, he declined. During his stay in Nurnberg, he met Johann Christian, Freiherr von Boyneburg, one of the most distinguished German statesmen of the day. Boyneburg took him into his service and introduced him to the court of the prince elector, the archbishop of Mainz, Johann Philipp von Schonborn, where he was concerned with questions of law and politics.

King Louis XIV of France was a growing threat to the German Holy Roman Empire. To ward off this danger and divert the King's interests elsewhere, the Archbishop hoped to propose to Louis a project for an expedition into Egypt; because he was using religion as a pretext, he expressed the hope that the project would promote the reunion of the church. Leibniz, with a view toward this reunion, worked on the Demonstrationes Catholicae. His research led him to situate the soul in a point—this was new progress toward the monad—and to develop the principle of sufficient reason (nothing occurs without a reason). His meditations on the difficult theory of the point were related to problems encountered in optics, space, and movement; they were published in 1671 under the general title Hypothesis *Physica* Nova ("New Physical Hypothesis"). He asserted that movement depends, as in the theory of the German astronomer Johannes Kepler, on the action of a spirit (God).

Period in Paris. In 1672 the Elector sent the young jurist on a mission to Paris, where he arrived at the end of March. In September, Leibniz met with Antoine Arnauld, a Jansenist theologian (Jansenism was a nonorthodox Roman Catholic movement that spawned a rigoristic form of morality) known for his writings against the Jesuits. Leibniz sought Arnauld's help for the reunion of the church. He was soon left without protectors by the deaths of Freiherr von Boyneburg in December 1672 and of the Elector of Mainz in February 1673; he was now, however, free to pursue his scientific studies. In search of financial support, he constructed a calculating machine—a perfected version of an earlier machine developed by Blaise Pascal, a French scientist and writer—and presented it to the Royal Society during his first journey to London, from January to March 1673.

His circle of friends was also constantly expanding. While in London, he met many of the noted scientists and mathematicians, including Robert Boyle, John Pell, and possibly John Collins, a friend of the physicist Sir Isaac Newton. In Paris, he had already met Christiaan Huygens, a Dutch mathematician and astronomer who had first helped him with mathematics. Arnauld introduced him to many important Jansenists, particularly Étienne Périer, a nephew of Pascal who, in the summer of 1674, entrusted Leibniz with the unpublished papers of Pascal. During the first part of 1675, he spent some time with Nicolas Malebranche, another Cartesian philosopher and geometer, and in September 1675, he befriended Ehrenfried Walter von Tschirnhaus, a German mathematician.

Foundations of calculus and the dynamic theory of motion

Late in 1675 Leibniz laid the foundations of both integral and differential calculus. With this discovery, he ceased to consider time and space as substances—another step closer to monadology. He began to develop the notion that the concepts of extension and motion contained an element of the imaginary, so that the basic laws of motion could not be discovered merely from a study of their nature. Nevertheless, he continued to hold that extension and motion could provide a means for explaining and predicting the course of phenomena. Thus, contrary to Descartes, Leibniz held that it would not be contradictory to posit that this world is a well-related dream. If visible movement depends on the imaginary element found in the concept of extension, it can no longer be defined by simple local movement; it must be the result of a force. In criticizing the Cartesian formulation of the laws of motion, known as mechanics, Leibniz became, in 1676, the founder of a new formulation, known as dynamics, which substituted kinetic energy for the conservation of movement. At the same time, beginning with the principle that light follows the path of least resistance, he believed that he could demonstrate the ordering of nature toward a final goal or cause.

The Hanoverian period. Leibniz continued his work in Paris, but he was still without an income-producing position. By October 1676, however, he had accepted a position in the employment of John Frederick, the duke of Braunschweig-Liineburg (Brunswick-Liineburg; including the duchies of Zelle and Hanover), with whom he had been corresponding for some time; against his wishes he was forced to leave the French capital for Hanover by way of London and the Netherlands. In London he visited Collins, who allowed him to see some unpublished papers of James Gregory, a Scottish mathematician, and of Newton. In the Netherlands, he met the naturalists Jan Swammerdam and Antonie von Leeuwenhoek; he also stopped at The Hague, where he had long conversations with the Jewish Rationalist Benedict de Spinoza. He finally arrived in Hanover in mid-December.

Service under John Frederick. John Frederick, a convert to Catholicism from Lutheranism in 1651, had become duke of Hanover in 1665. He appointed Leibniz librarian, but, beginning in February 1677, Leibniz solicited the post of councillor, which he was finally granted in 1678. It should be noted that, unlike Descartes, Leibniz had no fortune, nor did he belong to a religious order as did Malebranche; and unlike Huygens he had not been granted a pension: among the great philosophers of his time, he was the only one who had to earn a living. As a result, he was always a jack-of-all-trades to royalty. In 1677 he was involved in simultaneous discussions about the reunion of the church with Bishop Cristóbal Rojas de Spínola, an envoy of the Emperor, and with Nicolaus Steno, a prelate who had been a scientist specializing in geology. In February 1679, he started a correspondence with Jacques Bénigne Bossuet, a French Roman Catholic bishop and renowned orator, that continued until 1702.

Jack-of-all-trades to royalty

For the negotiations conducted before the treaties of Nijmegen in 1678 and 1679, which made Louis XIV the arbitrator of Europe, Leibniz tried in vain to obtain the rank of ambassador for the members of the delegation from Hanover. Trying to make himself useful in all ways, Leibniz proposed that he inspect cloisters, that education be made more practical, that academies be founded; he worked on hydraulic presses, windmills, lamps, submarines, clocks, and a wide variety of mechanical devices; he devised a means of perfecting carriages and experimented with phosphorus (which Hennig Brand, a German alchemist, had just discovered). He also developed a water pump run by windmills, which ameliorated the exploitation of the mines of the Harz Mountains, and he worked in these mines as an engineer frequently from 1680 to 1685. Leibniz is considered to be among the creators of geology because of the observations he compiled there, including the hypothesis that the earth was at first molten; this hypothesis, from which Georges-Louis Leclerc, comte de Buffon, a French naturalist, later drew his inspiration, was presented in Protogeae, which was not published until 1749. These many occupations did not stop his work in mathematics: In March 1679 he perfected the binary system of numeration (*i.e.*, using two as a base), and at the end of the same year he proposed the basis for analysis situs, now known as general topology, a branch of mathematics that deals with selected properties of collections of related physical or abstract elements. He was also working on his dynamics and his philosophy, which was becoming increasingly anti-Cartesian. At this point, Duke John Frederick died on January 7, 1680, and his brother, Ernest Augustus I, succeeded him.

Service under Ernest Augustus. France was growing more intolerant at home—from 1680 to 1682 there were harsh persecutions of the Protestants that paved the way for the revocation of the Edict of Nantes on October 18, 1685—and increasingly menacing on its frontiers, for as early as 1681, despite the reigning peace, Louis XIV took Strasbourg and laid claim to ten cities in Alsace. France was thus becoming a real danger to the empire, which had already been shaken on the east by a Hungarian revolt and by the advance of the Turks, who had been stopped only by the victory of John III Sobieski, king of Poland, at the siege of Vienna in 1683. Leibniz served both his prince and the empire as a patriot. He suggested to his prince a means of increasing the production of linen and proposed a process for the desalinization of water; he recommended classifying the archives and in 1682 suggested publishing the periodical *Acta Eruditorum*, which was to become part of the *Journal des Savunts*. For the empire, in 1681, he made an analysis of the current stale of affairs; in 1683 he wrote, in both French and Latin, a violent pamphlet against Louis XIV, entitled *Mars Christianissimus* ("The Most Christian War God"); in the same year he expressed his thoughts on the war with Hungary, in the form of notes; and in 1684 he stated precisely the *Raisons touchant la guerre ou l'accommodement avec la France* ("Reasons concerning War or Settlement with France").

Development of Rationalistic system

During this same period Leibniz continued to perfect his metaphysical system through research into the notion of a universal cause of all being, attempting to arrive at a starting point that would reduce reasoning to an algebra of thought. He also continued his developments in mathematics; in 1681 he was concerned with the proportion between a circle and a circumscribed square and, in 1684, with the resistance of solids. In the latter year he published *Nova Methodus pro Maximis et Minimis* ("New Method for the Greatest and the Least"), which was an exposition of his differential calculus. Newton had also discovered calculus as early as 1665, but he had not published his findings; he had only hinted at them to Gregory and Collins. Leibniz' article, which appeared in *Acta,* and the discovery that he had seen some of Newton's writings opened up the question of the priority of the invention of calculus, one of the most famous disputes of the 18th century.

Leibniz' noted *Meditationes de Cognitione, Veritate et Ideis (Reflections on Knowledge, Truth, and Ideas)* appeared at this time and defined his theory of knowledge: things are not seen in God—as Malebranche suggested—but rather there is an analogy, a strict relation, between God's ideas and man's, an identity between God's logic and man's. In February 1686, Leibniz wrote his *Discours de métaphysique (Discourse on Metaphysics).* In the March publication of *Acta,* he disclosed his dynamics in a piece entitled *Brevis Demonstratio Erroris Memborabilis Cartesii et Aliorum Circa Legem Naturae* ("Brief Demonstration of the Memorable Error of Descartes and Others about the Law of Nature"). A further development of Leibniz' views, revealed in a text written in 1686 but long unpublished, was his generalization concerning propositions that in every true affirmative proposition, whether necessary or contingent, the predicate is contained in the notion of the subject. Contingent propositions (stating what can be or not be), which are the foundation of liberty, although they do not appear to be identified with the subject, nevertheless tend toward such an identity in the same way that a curve approaches but never reaches its asymptote. It can be said that, at this time, with the exception of the word monad (which did not appear until 1695), his philosophy of monadology was defined. In January 1687 he began corresponding with Pierre Bayle, the French philosopher and encyclopaedist who edited the influential journal *Nouvelles de la république des lettres*; ("News of the Republic of Letters"); in his letters, Leibniz established his independence from the Cartesians. This correspondence anticipated the *Essais de théodicée sur la bontk de Dieu, la libertk de l'homme et l'origine du mal* (1710; Eng. trans., *Theodicy,* 1952), his only large book on philosophy published in his lifetime.

In 1685 Leibniz was named historian for the House of Brunswick and, on this occasion, *Hofrat* ("court adviser"). His job was to prove, by means of genealogy, that the princely house had its origins in the House of Este, an Italian princely family, which would allow Hanover to lay claim to a ninth electorate. In search of these documents, Leibniz began travelling in November 1687. Going by way of southern Germany, he arrived in Austria, where he learned that Louis XIV had once again declared a state of war; in Vienna, he was well received by the Emperor; he then went to Italy. Everywhere he went, he met scientists and continued his scholarly work. The *Acta* published, in February 1689, his essays on the movement of celestial bodies and, in April 1689, on the duration of things. In November of that year he read Newton's *Principia* for the first time. He returned to Hanover in mid-July 1690. His efforts had not been in vain. In October 1692 Ernest Augustus obtained the electoral investiture.

Work on a history of the Earth

Until the end of his life, Leibniz continued his duties as historian. He did not, however, restrict himself to a genealogy of the House of Brunswick; he enlarged his goal to a history of the Earth, which included such matters as geological events and descriptions of fossils. He searched by way of monuments and linguistics for the origins and migrations of peoples; then for the birth and progress of the sciences, ethics, and politics; and, finally, for the elements of a *historia sacra.* In this project of a universal history, Leibniz never lost sight of the fact that everything interlocks. Even though he did not succeed in writing this history, his effort was influential because he devised new combinations of old ideas and invented totally new ones.

In 1691 Leibniz was named librarian at Wolfenbüttel. He renewed his correspondence with Bossuet and propagated his system and his discoveries by means of articles in scientific journals. In 1695 he explained a portion of his dynamic theory of motion in the *Système nouveau* ("New System"), which treated the relationship of substances and the pre-established harmony between the soul and the body: God does not need to bring about man's action by means of his thoughts, as Malebranche asserted, or to wind some sort of watch in order to reconcile the two; rather, the Supreme Watchmaker has so exactly matched body and soul that they correspond—they give meaning to each other—from the beginning.

In 1697, *De Rerum Originatione (On the Ultimate Origin of Things)* tried to prove that the ultimate origin of things can be none other than God. In 1698, *De Ipsa Natura* ("On Nature Itself") explained the internal activity of nature in terms of Leibniz' theory of dynamics.

All of these writings opposed Cartesianism, which was judged to be damaging to faith. Plans for the creation of German academies followed in rapid succession. With the help of his friend the electress Sophia Charlotte, daughter of Ernest Augustus and soon to become the first queen of Prussia (January 1701), the German Academy of Sciences in Berlin was founded on July 11, 1700; the academy, however, had no financial support until the time of Frederick II the Great, who became king of Prussia in 1740.

Service under George Louis. On January *23,* 1698, Ernest Augustus died, and his son, George Louis, succeeded him. Leibniz found himself confronted with an uneducated, boorish prince, a reveller who kept him in the background. Leibniz took advantage of every pretext to leave Hanover; he was constantly on the move; his only comfort lay in his friendship with Sophia Charlotte and her mother, Princess Sophia. Once again, he set to work on the reunion of the church: in Berlin, it was a question of uniting the Lutherans and the Calvinists; in Paris, he had to subdue Bossuet's opposition; in Vienna (to which Leibniz returned in 1700) he enlisted the support of the Emperor, which carried great weight; in England, it was the Anglicans who needed convincing.

Work toward the reunion of the church

The death in England of William, duke of Gloucester, in 1700 made George Louis, great-grandson of James I, a possible heir to the throne. It fell to Leibniz, jurist and historian, to develop his arguments concerning the rights of the House of Braunschweig-Liineburg with respect to this succession. He communicated with many important Englishmen: John Toland, a Deist who accompanied an ambassador sent to Hanover in June 1702; Gilbert Burnet, bishop of Salisbury, head of the Anglican Church; the poet and essayist Joseph Addison; and Lady Damaris Masham, at whose house John Locke, the Empiricist, died in 1704. Since 1696, Leibniz had been annotating Locke's *Essay Concerning Human Understanding;* Locke's death, however, dissuaded him from publishing these notes.

The *Théodicée* and the *Monadologia*

The War of the Spanish Succession began in March 1701 and did not come to a close until September 1714, with the Treaty of Baden. Leibniz followed its episodes as a patriot hostile to Louis XIV. His fame as a philosopher and scientist had by this time spread all over Europe; he was named a foreign member by the Academy of Sciences of Paris in 1700 and was in correspondence with most of the important European scholars of the day. If he was publishing little at this point, it was because he was writing *Théodicée,* which was published in 1710. In this work he set down his ideas on divine justice: metaphysical evil is inevitable because it is nothing other than the finite nature inherent in every creature; each creature is defined by his place in creation and by everything that he is fated to experience in virtue of his complete identity; therefore, each creature acts according to his own nature without the least material hindrance from others, with whom he is in tune in a universal harmony; each creature is autonomous and, if it is endowed with reason, free. Evil, moral or physical, is in no way positive; it is a lack that, like dissonance in music, increases the beauty of the whole. Finally, he held that God chose the best of all possible worlds.

Leibniz was impressed with the qualities of the Russian tsar Peter the Great, and in October 1711 the ruler received him for the first time. Following this, he stayed in Vienna until September 1714; and during this time the Emperor promoted him to the post of *Reichhofrat* ("adviser to the empire") and gave him the title of *Freiherr* ("baron"). About this time he wrote the *Principes de la nature et de la Grâce fondés en raison,* which inaugurated a kind of pre-established harmony between these two orders. Further, in 1714 he wrote the *Monadologia,* which synthesized the philosophy of the *Théodicée.*

In August 1714, the death of Queen Anne brought George Lewis to the English throne under the name of George I. Returning to Hanover, where he was virtually placed under house arrest, Leibniz set to work once again on the *Anrzales Imperii Occidentis Brunsvicenses* (1843–46; "Braunschweig [Brunswick] Annals of the Western Empire"). He also maintained a lengthy correspondence with Samuel Clarke, a close associate of Newton, about the concepts of space and time. At Bad-Pyrmont, he met with Peter the Great for the last time in June 1716. From that point on, he suffered greatly from gout and was confined to his bed. He died on November 14, 1716.

Leibniz was a man of medium height with a stoop, broad-shouldered but bandy-legged, as capable of thinking for several days sitting in the same chair as of travelling the roads of Europe summer and winter. He was an indefatigable worker whose manuscripts have not yet all been published, a universal letter writer (he had more than 600 correspondents), a patriot and cosmopolitan, a great scientist, and one of the most powerful spirits of Western civilization.

MAJOR WORKS

METAPHYSICS, EPISTEMOLOGY, AND THEOLOGY: *Disputatio Metaphysica de Principio lndividui* (1663); *Confessio Naturae Contra Atheistas* (1669); *Defensio Trinitatis per Nova Reperta Logica* (1669, printed 1841); "Dissertatio Praeliminaris" for an edition of *Marii Nizolii, . . . de Veris Principiis . . . Contra Pseudophilosophos Libri IV* (1670); *Meditationes de Cognitione, Veritate, et Ideis* (1684); *Discours de me'taphysique* (written 1686, printed 1846; *Discourse on Metaphysics,* 1902; new version, 1953 and 1961); *Brevis Demonstratio Erroris Memorabilis Cartesii et Gliorum Circa Legem Naturae Secundum Quam Volunt a Deo eandem Semper Quantitatem Motus Conservari* (1686); *Systema Theologicum* (written 1686, printed 1819; *A System of Theology,* 1850); *Système nouveau de la nature et de la communication des substances, aussi bien que de I'union qu'il y a entre l'âme et le corps* (1695); *Animadversiones in Partem Generalem Principiorum Cartesianorum* (drafts written 1697, printed 1844 and 1880); *De lpsa Natura, sive de Vi Insita, Actionibusque Creaturarum* (1698); *Conside'rations sur la doctrine d'un esprit universe1 unique* (written 1702 but not printed until 1839); *Nouveaux essais sur I'entendement humain* (written 1700–05, printed 1765; *New Essays Concerning Human Understanding,* 1898; new edition 1916); *Essais de théodicée sur la bonté de Dieu, la liberté de l'homme et l'origine du mal,* 2 vol. (1710; *Theodicy: Essays on the Goodness of God, the Freedom of Man, and the Origin of Evil,* 1952); *Causa Dei Asserta per Justitiam Ejus, cum Caeteris Ejus Perfectionibus Cunctisque Actionibus Conciliatam* (1710); *Principes de la nature et de la grâce fonde's en raison* (written 1714, printed 1718); *Principia Philosophiae, More Geometrico Demonstrata* (written 1714, printed 1720–21 and generally known as *Monadologia* or *La Monadologie; The Monadology,* 1898, 1925; new version 1930); *Entretien de Philarète et d'Ariste* (written *c.* 1715, printed 1720); *A Collection of Papers, Which Passed Between the Late Learned Mr. Leibnitz and Dr. Clarke, in the Years 1715 and 1716,* namely, the Leibniz–Samuel Clarke correspondence, with Leibniz' text in the original French, 2 parts (1717; new edition, *The Leibniz–Clarke Correspondence,* 1956).

LOGIC AND MATHEMATICS: *Dissertatio de Arte Combinatoria* (1666, with unauthorized republication as *Ars Combizatoria,* 1690); *Nova Methodus pro Maximis et Minimis, Itemque Tangentibus, Quae nec Fractas, nec Irrationales Quantitates Moratum, et Singulare pro Illis Calculi Genus* (1684); *De Geometria Recondita et Analysi Indivisibilium et Infinitorum* (1686); *Tentamen Anagogicum: Essai anagogique dans la recherche des causes* (*c.* 1690–95, printed 1890).

PHYSICS AND PRACTICAL SCIENCES: *Hypothesis Physica Nova* (1671); *Protogaea, sive de Prima Facie Telluris et Antiquissimae Historiae Vestigiis in Ipsis Naturae Monumentis Dissertatio* (written 1691 and published 1749); *Specimen Dynamicum* (1695); *Relatio . . . de Antidysenterico Americano* (1696); edition of *Miscellanea Berolinensia ad Incrementum Scientiarum* (1710); *Remarques sur le discours de Mr. H. S. . . . , touchant la manière de gouverner les horloges à pendule et les monfres à spirale* (1714).

LAW AND CONTEMPORARY POLITICS: *Specimen Quaestionum Philosophicarum ex Jure Collectarum* (1664); *Disputatio Juridic de Conditionibus* (1665); *Disputatio Inauguralis de Casibus Perplexis in Jure* (1666); *Nova Methodus Discendae Docendaeque Jurisprudentine* (1667); *Specimen Demonstationum Politicarum pro Eligendo Rege Polonorum* (1669); *Bedenken, welchergestalt Securitas Publica lnterna et Externa und Status praesens jetzigen Umständen nach im Reich auf festen Fuss zu stellen* (1670); *De Expeditione Aegyptiaca Regi Franciae Proponenda Justa Dissertatio* (1672; *A Summary Account of Leibnitz's Memoir, Addressed to Lewis the Fourteenth,* 1803), with synopsis *Consilium Aegyptiacum* (also 1672, both finally printed 1864); *Politische Betrachtung iiber gegenwaertigen Krieges-Zustand, zwischen Frankreich und Ober- und Nieder Teutsch-Land* (1674); *De Jure Suprematus ac Legationis Principum Germaniae,* published under the pseudonym Caesarinus Fuerstenerius (1677, enlarged edition 1678); *Entretien de Philarète et d'Eugène sur la question du temps, agite'e à Nimwègue, touchant le droit d'ambassade des électeurs et princes de l'Empire* (1678); *Mars Christianissimus* (1683); *De la tolerance des religions: de M. de Leibniz et réponses de M. de Pellisson* (1692); edition of *Codex Juris Gentium Diplomaticus* (1693; with appendix *Mantissa Codicis . . . ,* 1700); *Novissima Sinica Historiam Nostri Temporis Illustratura* (1697; 2nd ed., 1699); *Manifeste contenant les droits de Charles III, roi d'Espagne, et les justes motifs de son expe'dition* (1703); *Bericht iiber die Reunionssache an Clemens XI* (1710, printed 1903).

HISTORICAL RESEARCH: *Lettre sur la connexion des maisons de Brunsvic et d'Este* (1695); edition of J. Burchard, *Specimen Historiae Arcanae, Anecdotae de Vita Alexandri VI Papae* (1696 and 1697); edition of Albericus Trium Fontium (Aubry de Trois-Fontaines), *Chronicon* (1698); collection of *Accessiones Historicae, Quibus Utilia Superiorum Temporum Historiis Illustrandis Scripta Monumentaque Nondum Hactenus Edita inque Iis Scriptores Diu Desiderafi Continentur* 4 part (1698 and 2 vol. 1700); edition of *Scriptores rerum Brunsvicensium,* 3 vol. 1707–11); *Dissertatio de Numis Gratiani* (1709); *De Origine Francorum Disquisitio* (1715); collection of material for *Anrzales Imperii Occidentis Brunsvicenses* printed 1843–46).

PHILOLOGY AND GENERAL SYNTHESIS: *Dissertatio de Variis Linguis* (1715); *Collectanea Etymologica*, 2 pt. (1717); materials for a *Scientia Generalis* and for a related *Characteristica* drafted from *c.* 1670 onward (printed as a collection 1890).

BIBLIOGRAPHY

Collected editions of Leibniz' works: Editions, far from complete, of Leibniz' collected works were undertaken by R.E. RASPE (1765); by L. DUTENS, 6 vol. (1768); by G.H. PERTZ *et al.*, 13 vol. (1843–63); by A. FOUCHER DE CAREIL, 7 vol. out of 20 planned (1859–75); by O. KLOPP, 11 vol. (1864–84); by C.J. GERHARDT, 14 vol. (1850–63, 1875–90). The Berlin Academy, after plans concerted with the French Academy from 1900, began a definitive edition in 1923, and it is still in progress. Selections in English translation from Leibniz' philosophical writings appeared in 1890, in 1934, and in 1956; from his mathematical manuscripts in 1920; and from his logical papers in 1966.

Biographical and critical studies: G.G. GUHRAUER, *Gottfried Wilhelm, Freiherr v. Leibniz: Eine Biographie*, 2 vol. and suppl. (1842–46; with partial Eng. trans. by J.D. MACKIE, *Life of G.W. von Leibniz*, 1845), remains the fundamental work. WILHELM TOTOK and CARL HAASE, *Leibniz: Sein Leben, sein Werken, seine Welt* (1966), completes and corrects Guhrauer in several respects. KURT MULLER and GISELA KRONERT, *Leben und Werk von Gottfried Wilhelm Leibniz: Eine Chronik* (1969), gives references to the fundamental documentation. YVON BELAVAL, *Leibniz: Initiation à sa philosophie*, 3rd ed. (1969), places the philosophical development of Leibniz in the context of the history of his time. Still of value is the great work of BERTRAND RUSSELL, *A Critical Exposition of the Philosophy of Leibniz* (1900).

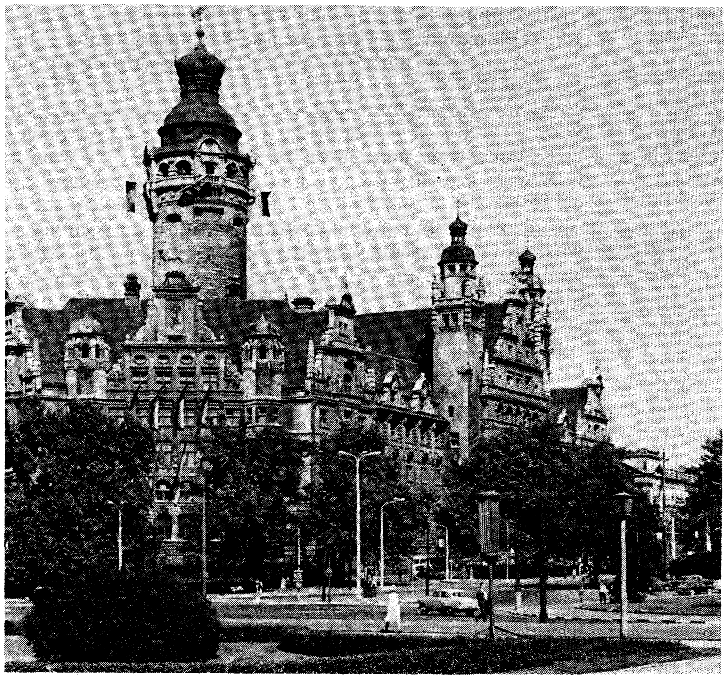
(Y.B.)

Leipzig

The character of the city

The home of about 566,600 persons in the mid-1970s, Leipzig is second only to East Berlin, the capital, among the cities of the German Democratic Republic. It lies in the south of the country, at the heart of the low-lying, fertile Leipzig Basin with its deposits of lignite (brown coal). Although encircled by a pleasant green belt of parks and gardens (the Schrebergärten), the city is a major industrial centre and transport junction and the core of the Halle–Leipzig metropolitan agglomeration. The fur-dressing and book-manufacturing industries of Leipzig are well known, but its contemporary industrial structure also includes emphasis on heavy constructional engineering and the electrical, textile, clothing, chemical, and machine tool industries. The annual Leipzig Fair is probably the most important forum for international East–West trade, and the city is also an intellectual centre, with a university (renamed Karl-Marx-Universität Leipzig in 1953) dating back to 1409. A quarter of what was known throughout Europe as an exceptionally beautiful city was destroyed in the last years of World War II, but Leipzig has since been rebuilt in the modern style. In addition to the restored historic buildings—notably the Altes Rathaus (Old Town Hall), the Alte Handelsborse (Old Commercial Exchange), the old residential and market squares, Auerbach's Cellar, and the 13th-century Church of St. Thomas (Thomaskirche)—the skyline of the modern city includes the new university tower and new hotels and commercial and residential buildings. (For related information see GERMAN DEMOCRATIC REPUBLIC.)

History. Located in the middle of a basin covered thickly with the windblown deposits known as loess, Leipzig developed from a small community on the periphery of the area of German settlement. It enters recorded history in AD 1015 as the fortified town of Urbs Libzi and was granted municipal status by 1170. Its position on important trade routes soon promoted a flourishing foreign trade. Its two annual markets, at Easter and at Michaelmas, were raised in 1497 to the rank of imperial fairs. Additional economic privileges enabled Leipzig to become the foremost German commercial centre by around 1700, a development that, in turn, promoted the growth of the network of roads converging on the town. Leipzig's focal geographical situation had another, less fortunate consequence: several important battles were fought in or near it. These included two at the village of Breitenfeld (now a suburb) in 1631 and 1642 and one at Lützen in 1632, during the Thirty Years' War; and in particular the Battle of Leipzig (or Battle of the Nations) in October 1813, in the Napoleonic Wars.



The Neues Rathaus (New Town Hall) in Leipzig, built 1899–1905 on the site of the Pleissenburg citadel in the southwestern corner of the inner town. The round tower of the former citadel is seen at upper left.

Karl Droste—Bavaria Verlag

The town's enviable economic status stimulated a notable cultural life, based particularly on the early development of the printing industry but also including the musical efflorescence associated with Johann Sebastian Bach. Trade continued to be the most important economic activity in the town (which had a population of about 70,000 by mid-19th century), the main commodities traded being books, furs, yarns, and textiles. As early as 1839 the first German overland railway was opened between Leipzig and Dresden, and the concomitant growth of banks provided capital for the growing textile and metallurgical industries. By the end of the 19th century, the contemporary pattern of heavy constructional engineering and machine tool enterprises had been established.

The development of industry also resulted in Leipzig's becoming a major centre of the German labour movement. After the devastation of World War II, the restoration and reconstruction of the city and its political and social institutions were carried out under the Communist policies of the German Democratic Republic. With renewed international attention being paid to the Leipzig Fair and other exhibitions and congresses held in the city, Leipzig continues to play an important role among European cities.

Location and boundaries. Leipzig covers 54 square miles (141 square kilometres) at the confluence of the Weisse Elster, Pleisse, and Parthe rivers; the city centre is 387 feet (118 metres) above sea level. The surrounding countryside is a plain, the original natural forest cover of which since the Middle Ages has been replaced by intensive farming on the predominantly brown soils. Within the city limits, the woodlands—Ratsholz, Nonne, Rosental, and Stadtforst—along the riverbanks have been partly converted to parks and fulfill an important recreational function; Leipzig's surroundings otherwise have few scenic attractions. To the north, the city boundary reaches out to the Halle–Dresden *Autobahn* (expressway); the Berlin–Munich *Autobahn* runs to the west; to the south, the city is edging toward the opencast lignite mines of Bohlen and Espenhain; and fertile farmlands lie to the east.

Demographic structure

The people. As early as the 17th century, Leipzig was the home of 20,000 persons. The population reached 41,000 by 1841 and 107,000 in 1871, when Leipzig became officially a city. The growth was accompanied by a corresponding increase in the urban area, yet at its peak, with 717,000 in 1931, Leipzig was one of Germany's most densely populated cities. Having suffered severely in World War II, Leipzig had only 608,000 inhabitants in 1946; and as an indirect result of the war (in terms of unfavourable age and sex distribution) the population was still decreasing slightly in the mid-1970s, when it numbered about 567,000—and this in spite of an in-migration of labour. More than 40 percent of the resident population is economically inactive because of extreme youth or age: about 20 percent are children and the other 20 percent retired. There are about 125 women for each 100 men. Of the 342,000 persons gainfully employed early in the 1970s, more than 125,000 worked in industry; some 30,000 were commuters.

Economic life. Leipzig generates about 5 percent of the entire industrial gross national product of the German Democratic Republic and half the industrial output of its district. There are almost 800 enterprises, about 30 of them very large, and half the work force is employed in metalworking. Products range from heavy construction equipment to computers, and every third book published in the German Democratic Republic is manufactured in the city.

Economic life is greatly influenced by the Leipzig Fair, which each spring attracts some 600,000 visitors (including scientists and businessmen) from all over the world. The fair has more than 420,000 square yards (350,000 square metres) of floor space. The international fur auction is also important.

Administration. Leipzig is an important political and administrative centre: it is the seat of administration of the district and of the municipal and rural *Kreise* (equivalent to the U.S. county) bearing its name. Early in the 1970s there were nearly 600 municipal district council members and 200 town councillors, representing a wide range of occupational backgrounds and a multiplicity of political and mass organizations. In addition, nearly 70 other permanent commissions were active in public affairs. Citizens were also active as lay judges and members of parents' advisory councils.

Transport services

Services. Leipzig's main railway station is one of the largest terminals and one of the most important passenger stations in central Europe. About 125,000 travellers pass through it each day, and their number rises by about 50,000 when the fair is in session. Leipzig is the focus of many railway lines and major roads, and there are two small airports, Leipzig–Mockau and Leipzig–Schkeuditz. Internal transportation is provided by streetcars (carrying 270,000 passengers daily), trolley buses (11,000), motor buses (33,000), and private and state-owned taxis. Parks and gardens cover, in all, an area of 612 acres (248 hectares).

The 25 or so hospitals provided about 128 beds for each 10,000 inhabitants by the early 1970s. There were also large numbers of ancillary health centres and nearly 30 homes for older persons. Schools included almost 200 kindergartens and more than 80 general education institutions, three-quarters of them general technical secondary schools. There were also 30 or more vocational training schools.

Cultural life. Leipzig is a major European intellectual and cultural centre. The university—attended by Goethe and Wagner, among others—has been joined by a number of other institutions of higher education. The academies of dramatic art, musical history, graphic arts, and bookmaking are well-known internationally; and the physical culture academy is largely responsible for East Germany's important showing in international sport. The Deutsche Bucherei (German Library) was set up as a national library and since 1913 has been collecting the entire range of German literary work. The Comenius-Bucherei (Comenius Library), founded in 1871, has become Europe's largest library specializing in education. The university library and the Stadt- und Bezirksbibliothek Leipzig (Leipzig City and District Library) and the Stadtarchiv (City Archives) are also of major significance.

Historical sites and museums (including the Museum für Geschichte der Stadt Leipzig, or historical museum, in the Altes Rathaus) are visited by well over 1,000,000 people each year. Musical traditions are carried on by the Thomaner Choir, the Gewandhaus Orchestra (Leipziger Gewandhausorchester), and the Radio Symphony Orchestra. Cultural centres and a fine opera house (1963) also ensure that cultural life flourishes. Finally, Leipzig is one of Europe's great convention cities, and economic, scientific, and sports congresses attract more than 2,000,000 people each year. Sports are particularly popular, and the central stadium seats more than 100,000 spectators.

BIBLIOGRAPHY. Accurate and detailed statistics on the history, economy, culture, and social life of the city and district of Leipzig may be found in the two annual *Statistische Jahrbiicher.* Detailed individual articles are published in the scientific periodicals *Geographische Berichte, Sachsische Heimatblätter,* and also in the *Wissenschaftliche Zeitschrift* of Karl-Marx-Universitat Leipzig. A standard comprehensive publication is the volume commemorating the decennial of the Geographical Society of the German Democratic Republic, *Das Leipziger Land* (1964), which contains physical, economic, and geographical studies. Another notable publication is WALTER FELLMANN and KARL CZOK, *Leipzig* (1971), a city guide, in German, that contains important information on the landscape, history, economy, and culture.

(H.-J.K./R.E.H.M.)

Lena River

The Lena (meaning in Yakut "Big River") is, in terms of water volume, the second largest river in the Soviet Union; in terms of length it is one of the largest rivers in the world, flowing 2,730 miles (4,400 kilometres) from its source in a small lake in the mountains west of Lake Baikal, in Central Asia, to the mouth of its delta on the Arctic Laptev Sea. The area of the river basin is about 961,000 square miles (2,490,000 square kilometres).

Physiography. The Lena has three main sections, each of about 900 miles (1,450 kilometres): the upper section from the source to the Vitim tributary; the middle course from the Vitim to the mouth of the Aldan; and the lower section from the Aldan to the Laptev Sea.

In the section from the source to the Vitim, the Lena flows in a deep-cut valley, the rocky and steep slopes of which are raised above the river, often up to 1,000 feet (300 metres). These slopes are formed on the right bank by the ledges of the Northern-Baikal Upland. The width of the river valley varies from one to six miles, but here and there it narrows in ravines to only 700 feet. The best known ravine, named Pyany Byk (Drunken Bull), is situated 147 miles below Kirensk.

In the first 110 miles from its source the Lena has a great number of shallow, rocky shoals, which occur as far as the Kirenga tributary. Below the mouth of the Kirenga the depth of the riverbed in the pools increases to 30 feet, and the rate of the flow decreases with a decrease in gradient. In the middle course, from the mouth of the Vitim to the Aldan, the Lena becomes a large, deep river. The water supply increases, especially after the junction with the Olyokma, and the width of the river reaches one mile. From the mouth of the Vitim to the Olyokma, the river skirts the Patom Plateau, on the right bank, forming an enormous bend; the width of the valley increases in places to 20 miles. Its slopes are gentle and green with forests, and along them run well-marked terraces formed by rivers. In this section of the valley there occurs an extensive water meadow, in which small lakes are scattered.

Scenic attractions

Below the Olyokma, the character of the valley changes sharply. In a stretch of about 400 miles, from Olyokma to Pokrovsk (60 miles above Yakutsk), the Lena flows along the bottom of a narrow valley with sheer, broken slopes. The enormous rocks sometimes resemble the ruins of a castle, or columns, or the figures of people and animals; and the area is a favourite place for tourists and rock climbers. In this section the Lena receives its largest tributaries. In addition to the Vitim and the Aldan, these include the Great Patom and the Olyokma, flowing in on the right, and the Nyuya, on the left. Below the mouth of the Aldan the width of the valley of the Lena extends 12 to 16 miles, and the width of the water meadow reaches four to nine miles. Here the Lena enters into the borders of the Central Yakut Plain. The water meadow abounds with lakes, often marsh-ridden, and the riverbed divides, forming many islands and branches. The depth is from 50 to 70 feet, but there are many shallow sections with sandbanks.

In the lower section between the island of Zholdongo and the beginning of the delta the Lena Valley becomes narrow, its width being about one mile. The islands of the delta are low-lying, and covered with peat bogs; some fossilized ice may be found.

Climate. The climatic features of the Lena Basin are determined by its location well inside the Asian continent. In winter a powerful anticyclone is formed, the spur of which occupies all of Eastern Siberia. Because of the anticyclone the winter is notable for its clear skies and prevailing calm. The frosts reach —76" to —94" F (—60" to —70" C), the average monthly air temperature in January being —22" to —40" F. In July averages range between 50° and 68" F (10" to 20° C). Due to its remoteness from the sea, the amount of precipitation in the basin is slight. Only in the southern mountains does the yearly total reach 24–28 inches; in most of the basin it is 8–16 inches and in the delta four inches. Between 70 and 80 percent of the precipitation falls in the summer in the form of steady rain. In winter, on the average, not more than two inches of snow fall, so that the snow cover is very slight.

The very cold temperatures result in large accumulations of ice in the form of icy knolls. These are formed of groundwaters that accumulate between the layers of soil frozen permanently over many years and layers of seasonally frozen soil. Sometimes the icy knolls disintegrate with considerable force, scattering ice blocks. The riverbeds and floodlands are also covered with permanent ice in some places.

Vegetation and animal life. The vegetation of the Lena Valley reflects the features of the natural zones across which the river flows. The main part of the river basin is covered with taiga (swampy northern coniferous forest); in the lower reaches are found tundra and scattered forest. Spruce, cedar, birch, and poplar predominate in the moister regions. In the central valley occur some expanses of steppe, a rare occurrence above latitude 60° N. Characteristic of the water-meadow area are peat bogs and swamps and thickets of willow, alder, and dwarfish birches. In the lower valley course appear tundra plants such as mosses and lichens, partridge grasses, whitlow grass, arctic poppy, and cotton grass.

The plankton of the Lena is rather poor and restricted in variety. More than 100 species of animals are found. Commercially important fish include sturgeon, salmon, roach, dace, and perch. The main concentration of these varieties is in the mouth region, where in the summer there is comparatively warm water.

Hydrography. More than 95 percent of the Lena's water is obtained from melting snow and from rain; 1 to 2 percent of the yearly flow comes from subsoil waters. Typical of the Lena Basin are high floods, especially flash floods, in summer, and very little flow in winter. Complete cessation of flow may occur with the freezing of the river through to the bottom. The average volume of the Lena at the mouth is 21,300 cubic yards (16,300 cubic metres) per second; the greatest volume is 260,000 cubic yards per second, and the minimum is 478 cubic yards per second. The total yearly volume approaches 100 cubic miles. During the high-water period the water level rises by an average of 30 to 50 feet, and in the lower course reaches 60 feet.

The river's volume

The highest temperature of the water in the upper course of the river reaches 66° F (19" C) and in the lower course about 57" F (14" C). The river is free of ice in the south for five to six months and in the north four to five months. The breakup of ice in the spring causes significant damage to the shores: the ice floes grind the rocks, pull trees out by their roots, and carry away large sections of the banks.

The Lena carries out into the sea about 12,000,000 tons of suspended alluvium and 41,000,000 tons of dissolved matter a year. The proportion of suspended alluvium in the water is small: even in floodwaters it does not exceed 50 to 60 grams per cubic metre. The mineralization of the water in the lower Lena during low water is from 80 to 100 grams per cubic metre, and in floodwaters 160 to 500 grams per cubic metre.

Human activity. ***Exploration.*** The first exploration of the Lena was conducted by Russians at the beginning of the 17th century. In 1631 a fortress and a settlement was founded at Ust-Kut. The first scientific research was conducted by the Great Northern Expedition in 1733–43. Cartography was begun in 1910. In 1912, in the icebreakers "Taymir" and "Vaygach," the delta was surveyed and mapped. Further surveying was conducted in the interwar period, and is currently being conducted by the Yakut branch of the Academy of Science of the USSR and by other government bodies.

Settlement. Among the peoples living in the Lena Basin, mainly on the banks of the river and its tributaries, are Russians, Evenks (Tungus) Yakuts, and Yukaghirs. There are many industrial and cultural centres, collective farms, and state farms. In the territory of the basin is found the Yakut Autonomous Soviet Socialist Republic, with the capital of Yakutsk.

Economic activity. The Lena is navigable by small boats in sections below Kachug and from Ust-Kut by larger vessels. The largest wharves are Bulun, Zhigansk, Yakutsk, Vitim, Kirensk, Osetrovo, Zhigalovo, and Kachug; Osetrovo is an important river port, equipped with modern machinery. The largest navigable tributaries are the Kirenga, Vitim, Olyokma, Aldan, and Vilyuy. Wood products, furs, gold, mica, industrial products, and food are the main cargoes.

Within the borders of the Central Yakut Plain various agricultural crops—barley, oats, wheat, potatoes, cucumbers, and others—are grown. The large meadows and pastures present successful livestock raising. Future prospects for the economic development of the Lena Basin are extremely good. The basin is rich in useful fossil minerals, and gold and coal occur in exploitable quantities. In 1955, in western Yakutia, rich deposits of diamonds were discovered in the Vilyuy Basin. To the east of the diamond-bearing region are found vast deposits of natural gas; near Olyokminsk are found salt beds, and to the south of Yakutsk, iron ore and coking coal.

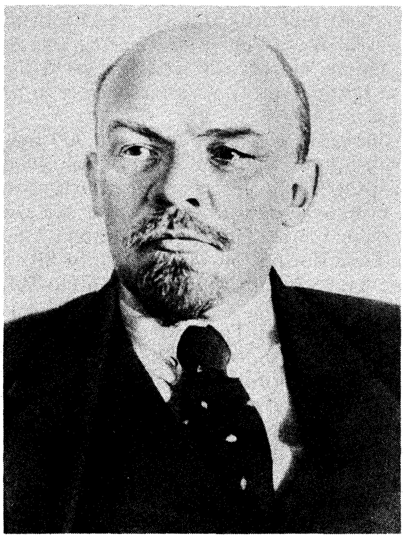
The hydroenergy potential of the Lena and its tributaries is about 40,000,000 kilowatts, and hydroelectric stations are planned at Lensk, Yakutsk, and Nizhne-Lensky. (I.V.P.)

Lenin

Militant Marxist, founder of the Russian Communist Party (Bolsheviks), inspirer and leader of the Bolshevik Revolution, Lenin was the architect, builder, **and first** head of the Soviet state. He was the founder of the Comintern (Communist International) and posthumous source of "Leninism," the doctrine codified and conjoined with Marx's works by Lenin's successors to form Marxism–Leninism, the Communist world view. If the Bolshevik Revolution is—as some people have called it—the most significant political event of the 20th century, then Lenin must for good or ill be regarded as this century's most significant political leader. Not only in the Soviet Union but even among many non-Communist scholars, he is regarded as both the greatest revolutionary leader and revolutionary statesman in history, as well as the greatest revolutionary thinker since Marx.

EARLY LIFE

The making of a revolutionary. Yet it is difficult to find anything in his childhood that might have turned



Lenin, 1918.
Tass—Sovfoto

him onto the path of a professional revolutionary. Vladimir Ilich Ulyanov was born in Simbirsk (renamed Ulyanovsk) on April 22 (April 10, old style), 1870. (He adopted the pseudonym Lenin in 1901 during his clandestine party work after exile in Siberia.) He was the third of six children born into a close-knit, happy family of highly educated and cultured parents. His mother was the daughter of a physician, while his father, though the son of a serf, became a schoolteacher and rose to the position of inspector of schools. Lenin, intellectually gifted, physically strong, and reared in a warm, loving home, early displayed a voracious passion for learning. He was graduated from high school ranking first in his class. He distinguished himself in Latin and Greek and seemed destined for the life of a classical scholar. When he was 16, nothing in Lenin indicated a future rebel, still less a professional revolutionary--except, perhaps, his turn to atheism. But, despite an ideal upbringing, all five of the Ulyanov children who reached maturity joined the revolutionary movement. This was not uncommon in tsarist Russia, where even the highly educated and cultured intelligentsia were denied elementary civil and political rights.

Execution of Lenin's brother

In adolescence Lenin suffered two blows that unquestionably impelled him to take the path of revolution. First, his father was threatened shortly before his untimely death with premature retirement by a reactionary government grown fearful of the spread of public education. Second, in 1887 his beloved eldest brother, Aleksandr, a student at the University of St. Petersburg (now Leningrad State University), was hanged for conspiring with a revolutionary terrorist group that plotted to assassinate Emperor Alexander III. Suddenly, at age 17, Lenin became the male head of the family, which was now stigmatized as having reared a "state criminal."

Fortunately his mother's pension and inheritance kept the family in comfortable circumstances, despite the frequent imprisonment or exile of her children. Moreover, Lenin's high school principal (the father of Aleksandr Kerensky, who was later to lead the Provisional government deposed by Lenin in November [October, O.S.] 1917) did not turn his back on the "criminal's" family. He courageously wrote a character reference that smoothed Lenin's admission to a university.

In autumn 1887 Lenin enrolled in the faculty of law of the imperial Kazan University (now Kazan [V.I. Lenin] State University), but within three months he was expelled, accused of participating in an illegal student assembly. He was arrested and banished from Kazan to his grandfather's estate in the village of Kokushkino, where his older sister Anna had already been ordered to reside by the police. In the autumn of 1888, the authorities permitted him to return to Kazan but denied him readmission to the university. During his enforced idleness, he met exiled revolutionaries of the older generation and avidly read revolutionary political literature, especially Marx's *Das Kapital.* He became a Marxist in January 1889.

Formation of a revolutionary party. In May 1889 the Ulyanov family moved to Samara (now Kuybyshev). After much petitioning, Lenin was granted permission to take his law examinations. In November 1891 he passed his examinations, taking a first in all subjects, and was graduated with a first-class degree. After the police finally waived their political objections, Lenin was admitted to the bar and practiced law in Samara in 1892–93, his clients being mainly poor peasants and artisans. In his experience at law he acquired an intense loathing for the class bias of the legal system and a lifelong revulsion for lawyers, even those who claimed to be Social-Democrats.

Exile in Siberia

But law was an extremely useful cover for a revolutionary activist. He moved to St. Petersburg (now Leningrad) in August 1893 and, while working as a public defender, made contact with revolutionary Marxists. In 1895 his comrades sent him abroad to make contact with Russian exiles in western Europe, especially with Russia's most commanding Marxist thinker, Georgy Plekhanov. Upon his return in 1895, Lenin and other Marxists, including L. Martov, the future leader of the Mensheviks, succeeded in unifying the Marxist groups of the capital in an organization known as The Union for the Struggle for the Liberation of the Working Class. The Union issued leaflets and proclamations on the workers' behalf, supported workers' strikes, and infiltrated workers' education classes to impart the rudiments of Marxism. In December 1895, the leaders of the Union were arrested. Lenin was jailed for 15 months, then exiled to Shushenskoye, in Siberia, for a term of three years. He was joined there in exile by his fiancée, Nadezhda Krupskaya, a Union member, whom he had met in the capital. They were married in Siberia, and she became Lenin's indispensable secretary and comrade. In exile they conducted clandestine party correspondence and collaborated (legally) on a Russian translation of Sidney and Beatrice Webb's *Industrial Democracy.*

Upon completing his term of exile in January 1900, Lenin went abroad and was joined later by Krupskaya in Munich. His first major task abroad was to join Plekhanov, Martov, and three other editors in bringing out the newspaper *Iskra* ("The Spark"), which they hoped would unify the Russian Marxist groups scattered throughout Russia and western Europe into a cohesive Social-Democratic party.

Up to the point that Lenin joined the *Iskra,* his writings had focussed on three problems: first, he had written a number of leaflets that aimed to shake the workers' traditional veneration of the tsar by showing them that their harsh life was caused, in part, by the support tsarism rendered the capitalists; second, he attacked those self-styled Marxists who urged Social-Democrats and workers to concentrate on wage and hour issues, leaving the political struggle for the present to the bourgeoisie; third, and ultimately most important, he addressed himself to the peasant question.

The principal obstacle to the acceptance of Marxism by many of the intelligentsia was their adherence to the widespread belief of the Populists (Russian pre-Marxist radicals) that Marxism was inapplicable to peasant Russia, in which a proletariat was almost nonexistent. Russia, they held, was immune to capitalism, owing to joint ownership of peasant land by the village commune. This view had been first attacked by Plekhanov in the 1880s. Plekhanov had argued that Russia had already entered the capitalist stage, as evidenced by the rapid growth of industry. Despite the denials of the Populists, the man of the future in Russia was the proletarian, not the peasant. Applying the Marxist scheme of social development to Russia, Plekhanov had come to the view that the revolution in Russia must pass through two discrete stages: first, a bourgeois revolution that would establish a democratic republic and full-blown capitalism; and second, a proletarian revolution after mature capitalism had generated a numerous proletariat that had attained a high level of political organization, socialist consciousness, and culture, enabling them to usher in full Socialism.

The peasant question

It was this view that Lenin adhered to after he read Plekhanov's work in the late 1880s. But almost immediately, Lenin went a step beyond his former mentor, especially in the peasant question. In an attack on the Populists published in 1894, Lenin charged that, even if they realized their fondest dream and divided all the land among the peasant communes, the result would not be Socialism but capitalism spawned by a free market in agricultural produce. The "Socialism" of the Populists would favour the growth of small-scale capitalism; hence the Populists were not Socialists but "petty bourgeois democrats." Outside of Marxism, which aimed ultimately to abolish the market system as well as the private ownership of the means of production, Lenin concluded, there could be no Socialism.

Even while in exile in Siberia, Lenin had begun research on his investigation of the peasant question, which culminated in his magisterial ***Development of Capitalism in Russia*** (published legally in 1899). In this work he argued that capitalism was rapidly destroying the peasant commune. The peasantry. for the Populists a homogeneous social class, was in actuality rapidly stratifying into a well-off rural bourgeoisie, a middling peasantry, and an impoverished rural "proletariat and semi-proletariat." In this last group, which comprised half the peasant population, Lenin found an ally for the extremely small industrial proletariat in Russia.

Iskra's success in recruiting Russian intellectuals to Marxism led Lenin and his comrades to believe that the time was ripe to found a revolutionary Marxist party that would weld together all the disparate Marxist groups at home and abroad. An abortive First Congress, held in 1898 in Minsk, had failed to achieve this objective, for most of the delegates were arrested shortly after the congress. The organizing committee of the Second Congress decided to convene the congress in Brussels in 1903, but police pressure forced it to transfer to London.

The congress sessions wore on for nearly three weeks, for no point appeared too trivial to debate. But the main issues quickly became plain: eligibility for membership and the character of party discipline; but, above all, the key issue focussed on the relation between the party and the proletariat, for whom the party claimed to speak.

In his *What Is To Be Done?* (1902), Lenin totally rejected the view that the proletariat was being driven spontaneously to revolutionary Socialism by capitalism and that the party should merely coordinate the struggle of the proletariat's diverse sections on a national and international scale. Capitalism, he contended, predisposed the workers to the acceptance of Socialism but did not spontaneously make them conscious Socialists. The proletariat by its own efforts in the everyday struggle against the capitalist could achieve "trade-union consciousness." But the proletariat could not by its own efforts grasp that it could win complete emancipation only by overthrowing capitalism and building Socialism, unless the party from without infused it with Socialist consciousness.

The "vanguard of the proletariat"

In his ***What Is To Be Done?*** and in his other works on party organization, Lenin created one of his most momentous political innovations, his theory of the party as the "vanguard of the proletariat." He saw the vanguard as a highly disciplined, centralized party that worked unremittingly to suffuse the proletariat with Socialist consciousness and served as mentor, leader, and guide, constantly showing the proletariat where its true class interests lie.

At the Second Congress the ***Iskra*** group split, and Lenin found himself in a minority on this very issue. Nevertheless, he continued to develop his view of "the party of a new type," which must be guided by "democratic centralism," or absolute party discipline. The party must be a highly centralized body organized around a small, ideologically homogeneous, hardened core of experienced professional revolutionaries, who were elected to the central committee by the party congress and who led a ramified hierarchy of lower party organizations that enjoyed the support and sympathy of the proletariat and all groups opposed to tsarism. "Give us an organization of revolutionaries," Lenin exclaimed, "and we will overturn Russia!"

Lenin spared no effort to build just this kind of party over the next 20 years, despite fierce attacks on his conception by some of his closest comrades of the ***Iskra*** days, Plekhanov, Martov, and Leon Trotsky. They charged that his scheme of party organization and discipline tended toward "Jacobinism," suppression of free intraparty discussion, a dictatorship ***over*** the proletariat, not ***of*** the proletariat, and, finally, establishment of a one-man dictatorship.

The Bolsheviks and the Mensheviks

Lenin found himself in the minority in the early sessions of the Second Congress of what was now proclaimed to be the Russian Social-Democratic Workers' Party (RSDWP). But a walkout by a disgruntled group of Jewish Social-Democrats, the Bund, left Lenin with a slight majority, Consequently, the members of Lenin's adventitious majority were called Bolsheviks (majoritarians), and Martov's group were dubbed Mensheviks (minoritarians). The two groups fought each other ceaselessly within the same RSDWP and professed the same program until 1912, when Lenin made the split final at the Prague Conference of the Bolshevik Party.

CHALLENGES OF THE REVOLUTION OF 1905 AND WORLD WAR I

The differences between Lenin and the Mensheviks became sharper in the Revolution of 1905 and its aftermath, when Lenin moved to a distinctly original view on two issues: class alignments in the revolution and the character of the post-revolutionary regime.

The outbreak of the revolution, in January 1905, found Lenin abroad in Switzerland, and he did not return to Russia until November. Immediately Lenin set down a novel strategy. Both wings of the RSDWP, Bolshevik and Menshevik, adhered to Plekhanov's view of the revolution in two stages: first, a bourgeois revolution; second, a proletarian revolution (see above). But the Mensheviks argued that the bourgeois revolution must be led by the bourgeoisie, with whom the proletariat must ally itself in order to make the democratic revolution. This would bring the liberal bourgeoisie to full power, whereupon the RSDWP would act as the party of opposition. Lenin defiantly rejected this kind of alliance and post-revolutionary regime. Hitherto he had spoken of the need for the proletariat to win "hegemony" in the democratic revolution. Now he flatly declared that the proletariat was the driving force of the revolution and that its only reliable ally was the peasantry. The bourgeoisie he branded as hopelessly counterrevolutionary and too cowardly to make its own revolution. Thus, unlike the Mensheviks, Lenin henceforth banked on an alliance that would establish a "revolutionary democratic dictatorship of the proletariat and the peasantry."

Nor would the revolution necessarily stop at the first stage, the bourgeois revolution. If the Russian revolution should inspire the western European proletariat to make the Socialist revolution, for which industrial Europe was ripe, the Russian revolution might well pass over directly to the second stage, the Socialist revolution. Then, the Russian proletariat, supported by the rural proletariat: and semi-proletariat at home and assisted by the triumphant industrial proletariat of the West, which had established its "dictatorship of the proletariat," could cut short the life-span of Russian capitalism.

Years of discouragement

After the defeat of the Revolution of 1905, the issue between Lenin and the Mensheviks was more clearly drawn than ever, despite efforts at reunion. But, forced again into exile from 1907 to 1917, Lenin found serious challenges to his policies not only from the Mensheviks but within his own faction as well. The combination of repression and modest reform effected by the tsarist regime led to a decline of party membership. Disillusionment and despair in the chances of successful revolution swept the dwindled party ranks, rent by controversies over tactics and philosophy. Attempts to unite the Bolshevik and Menshevik factions came to naught, all breaking on Lenin's intransigent insistence that his conditions for reunification be adopted. As one Menshevik opponent described Lenin: "There is no other man who is absorbed by the revolution twenty-four hours a day, who has no other thoughts but the thought of revolution, and who even when he sleeps, dreams of nothing but revolution." Placing revolution above party unity, Lenin would accept no unity compromise if he thought it might delay, not accelerate, revolution.

Desperately fighting to maintain the cohesion of the Bolsheviks against internal differences and the Mensheviks' growing strength at home, Lenin convened the Bolshevik Party Conference at Prague, in 1912, which split the RSDWP forever. Lenin proclaimed that the Bolsheviks were the RSDWP and that the Mensheviks were schismatics. Thereafter, each faction maintained its separate central committee, party apparatus, and press.

When war broke out, in August 1914, Socialist parties throughout Europe rallied behind their governments despite the resolutions of prewar congresses of the Second International obliging them to resist or even overthrow their respective governments if they plunged their countries into an imperialist war.

Denunciation of pro-war Socialists

After Lenin recovered from his initial disbelief in this "betrayal" of the International, he proclaimed a policy whose audacity stunned his own Bolshevik comrades. He denounced the pro-war Socialists as "social-chauvinists" who had betrayed the international working-class cause by support of a war that was imperialist on both sides. He pronounced the Second International as dead and appealed for the creation of a new, Third International composed of genuinely revolutionary Socialist parties. More immediately, revolutionary Socialists must work to "transform the imperialist war into civil war." The real enemy of the worker was not the worker in the opposite trench but the capitalist at home. Workers and soldiers should therefore turn their guns on their rulers and destroy the system that had plunged them into imperialist carnage.

Lenin's policy found few advocates in Russia or elsewhere in the first months of the war. Indeed, in the first flush of patriotic fervour, not a few Bolsheviks supported the war effort. Lenin and his closest comrades were left an isolated band swimming against the current.

Lenin succeeded in reaching neutral Switzerland in September 1914, there joining a small group of anti-war Bolshevik and Menshevik *émigrés.* The war virtually cut them off from all contact with Russia and with like-minded Socialists in other countries. Nevertheless, in 1915 and 1916, anti-war Socialists in various countries managed to hold two anti-war conferences in Zimmerwald and Kienthal, Switzerland. Lenin failed at both meetings to persuade his comrades to adopt his slogan: "transform the imperialist war into civil war!" They adopted instead the more moderate formula: "An immediate peace without annexations or indemnities and the right of the peoples to self-determination." Lenin consequently found his party a minority within the group of anti-war Socialists, who, in turn, constituted a small minority of the international Socialist movement compared with the pro-war Socialists.

Lenin's theory of imperialism

Undaunted, Lenin continued to hammer home his views on the war, confident that eventually he would win decisive support. In his ***Imperialism, the Highest Stage of Capitalism*** **(1917),** he set out to explain, first, the real causes of the war; second, why Socialists had abandoned internationalism for patriotism and supported the war; and third, why revolution alone could bring about a just, democratic peace.

War erupted, he wrote, because of the insatiable, expansionist character of imperialism, itself a product of monopoly finance capitalism. At the end of the 19th century, a handful of banks had come to dominate the advanced countries, which, by 1914, had in their respective empires brought the rest of the world under their direct or indirect controls. Amassing vast quantities of "surplus" capital, the giant banks found they could garner superprofits on investments in colonies and semi-colonies, and this intensified the race for empire among the great powers. By 1914, dissatisfied with the way the world had been shared out, rival coalitions of imperialists launched the war to bring about a redivision of the world at the expense of the other coalition. The war was therefore imperialist in its origins and aims and deserved the condemnation of genuine Socialists.

Socialist Party and trade-union leaders had rallied to support their respective imperialist governments because they represented the "labour aristocracy," the better paid workers who received a small share of the colonial "superprofits" the imperialists proffered them. "Bribed" by the imperialists, the "labour aristocracy" took the side of their paymasters in the imperialist war and betrayed the most exploited workers at home and the super-exploited in the colonies. The imperialists, Lenin contended, driven by an annexationist dynamic, could not conclude a just, lasting peace. Future wars were inevitable so long as imperialism existed; imperialism was inevitable so long as capitalism existed; only the overthrow of capitalism everywhere could end the imperialist war and prevent such wars in the future. First published in Russia in 1917, *Imperialism* to this day provides the instrument that Communists everywhere employ to evaluate major trends in the non-Communist world.

LEADERSHIP IN THE RUSSIAN REVOLUTION

By 1917 it seemed to Lenin that the war would never end and that the prospect of revolution was rapidly receding. But in the week of March 8–15, the starving, freezing, war-weary workers and soldiers of Petrograd (until 1914, St. Petersburg) succeeded in deposing the Tsar. Lenin and his closest lieutenants hastened home after the German authorities agreed to permit their passage through Germany to neutral Sweden. Berlin hoped that the return of anti-war Socialists to Russia would undermine the Russian war effort.

The Provisional Government

First return to Petrograd. Lenin arrived in Petrograd on April 16, 1917, one month after the Tsar had been forced to abdicate. Out of the revolution was born the Provisional Government, formed by a group of leaders of the bourgeois liberal parties. This government's accession to power was made possible only by the assent of the Petrograd Soviet, a council of workers' deputies elected in the factories of the capital. Similar soviets of workers' deputies sprang up in all the major cities and towns throughout the country, as did soviets of soldiers' deputies and of peasants' deputies. Although the Petrograd Soviet had been the sole political power recognized by the revolutionary workers and soldiers in March 1917, its leaders had hastily turned full power over to the Provisional Government. The Petrograd Soviet was headed by a majority composed of Menshevik and Socialist Revolutionary (SR), or peasant party, leaders who regarded the March (February, O.S.) Revolution as bourgeois; hence, they believed that the new regime should be headed by leaders of the bourgeois parties.

On his return to Russia, Lenin electrified his own comrades, most of whom accepted the authority of the Provisional Government. Lenin called this government, despite its democratic pretensions, thoroughly imperialist and undeserving of support by Socialists. It was incapable of satisfying the most profound desires of the workers, soldiers, and peasants for immediate place and division of landed estates among the peasants.

Only a soviet government—that is, direct rule by workers, soldiers, and peasants--could fulfill these demands. Therefore, he raised the battle cry, "All power to the Soviets!"—although the Bolsheviks still constituted a minority within the soviets and despite the manifest unwillingness of the Menshevik–SR majority to exercise such power. This introduced what Lenin called the period of "dual power." Under the leadership of "opportunist" Socialists, the soviets, the real power, had relinquished power to the Provisional Government, the nominal power in the land. The Bolsheviks, Lenin exhorted, must persuade the workers, peasants, and soldiers, temporarily deceived by the "opportunists," to retrieve state power for the soviets from the Provisional Government. This would constitute a second revolution. But, so long as the government did not suppress the revolutionary parties, this revolution could be achieved peacefully, since the Provisional Government existed only by the sufferance of the soviets.

Initially, Lenin's fellow Bolsheviks thought that he was temporarily disoriented by the complexity of the situation; moderate Socialists thought him mad. It required several weeks of sedulous persuasion by Lenin before he won the Bolshevik Party Central Committee to his view. The April Party Conference endorsed his program: the party must withhold support from the Provisional Government and win a majority in the soviets in favour of soviet power. A soviet government, once established, should begin immediate negotiations for a general peace on all fronts. The soviets should forthwith confiscate landlords' estates without compensation, nationalize all land, and divide it among the peasants. And the government should establish tight controls over privately owned industry to the benefit of labour.

Rise of the Bolsheviks

From March to September 1917, the Bolsheviks remained a minority in the soviets. By autumn, however, the Provisional Government (since July headed by the moderate Socialist Aleksandr Kerensky, who was supported by the moderate Socialist leadership of the soviets) had lost popular support. Increasing war-weariness and the breakdown of the economy overtaxed the patience of the workers, peasants, and soldiers, who demanded immediate and fundamental change. Lenin capitalized on the growing disillusionment of the people with Kerensky's ability and willingness to complete the revolution. Kerensky, in turn, claimed that only a freely elected constituent assembly would have the power to decide Russia's political future—but that must await the return of order. Meanwhile, Lenin and the party demanded peace, land, and bread—immediately, without further delay. The Bolshevik line won increasing support among the workers, soldiers, and peasants. By September they voted in a Bolshevik majority in the Petrograd Soviet and in the soviets of the major cities and towns throughout the country.

Decision to seize power. Lenin, who had gone underground in July after he had been accused as a "German agent" by Kerensky's government, now decided that the time was ripe to seize power. The party must immediately begin preparations for an armed uprising to depose the Provisional Government and transfer state power to the soviets, now headed by a Bolshevik majority.

Lenin's decision to establish soviet power derived from his belief that the proletarian revolution must smash the existing state machinery and introduce a "dictatorship of the proletariat"; that is, direct rule by the armed workers and peasants which would eventually "wither away" into a non-coercive, classless, stateless, Communist society. He expounded this view most trenchantly in his brochure *The State and Revolution,* written while he was still in hiding. The brochure, though never completed and often dismissed as Lenin's most "Utopian" work, nevertheless served as Lenin's doctrinal springboard to power.

Until 1917 all revolutionary Socialists rightly believed, Lenin wrote, that a parliamentary republic could serve a Socialist system as well as a capitalist. But the Russian Revolution had brought forth something new, the soviets. Created by workers, soldiers, and peasants and excluding the propertied classes, the soviets infinitely surpassed the most democratic of parliaments in democracy, because parliaments everywhere virtually excluded workers and peasants. The choice before Russia in early September 1917, as Lenin saw it, was either a soviet republic—a dictatorship of the propertyless majority--or a parliamentary republic—as he saw it, a dictatorship of the propertied minority.

Lenin therefore raised the slogan, "All power to the Soviets!", even though he had willingly conceded in the spring of 1917 that the Provisional Government was the "freest of all the belligerent countries." To Lenin, however, the Provisional Government was merely a "dictatorship of the bourgeoisie" that kept Russia in the imperialist war. What is more, it had turned openly counterrevolutionary in the month of July when it accused the Bolshevik leaders of treason.

From late September, Lenin, a fugitive in Finland, sent a stream of articles and letters to Petrograd feverishly exhorting the Party Central Committee to organize an armed uprising without delay. The opportune moment might be lost. But for nearly a month Lenin's forceful urgings from afar were unsuccessful. As in April, Lenin again found himself in the party minority. He resorted to a desperate stratagem.

Around October 20, Lenin, in disguise and at considerable personal risk, slipped into Petrograd and attended a secret meeting of the Bolshevik Central Committee held on the evening of October 23. Not until after a heated ten-hour debate did he finally win a majority in favour of preparing an armed take-over. Now steps to enlist the support of soldiers and sailors and to train the Red Guards, the Bolshevik-led workers' militia, for an armed take-over proceeded openly under the guise of self-defense of the Petrograd Soviet. But preparations moved haltingly, because serious opposition to the fateful decision persisted in the Central Committee. Enthusiastically in accord with Lenin on the timeliness of an armed uprising, Trotsky led its preparation from his strategic position as newly elected chairman of the Petrograd Soviet. Lenin, now hiding in Petrograd and fearful of further procrastination, desperately pressed the Central Committee to fix an early date for the uprising. On the evening of November 6, he wrote a letter to the members of the Central Committee exhorting them to proceed that very evening to arrest the members of the Provisional Government. To delay would be "fatal." The Second All-Russian Congress of Soviets, scheduled to convene the next evening, should be placed before a *fait accompli.*

Overthrow of the Provisional Government

On November 7 and 8, the Bolshevik-led Red Guards and revolutionary soldiers and sailors, meeting only slight resistance, deposed the Provisional Government and proclaimed that state power had passed into the hands of the Soviets. By this time the Bolsheviks, with their allies among the Left SR's (dissidents who broke with the pro-Kerensky SR leaders), constituted an absolute majority of the Second All-Russian Congress of Soviets. The delegates therefore voted overwhelmingly to accept full power and elected Lenin as chairman of the Council of People's Commissars, the new Soviet Government, and approved his Peace Decree and Land Decree. Overnight, Lenin had vaulted from his hideout as a fugitive to head the Revolutionary government of the largest country in the world. Since his youth he had spent his life building a party that would win such a victory, and now at the age of 47 he and his party had triumphed. "It makes one's head spin," he confessed. But power neither intoxicated nor frightened Lenin; it cleared his head. Soberly, he steered the Soviet government toward the consolidation of its power and negotiations for peace.

Saving the Revolution. In both spheres, Lenin was plagued by breaks within the ranks of Bolshevik leaders. He reluctantly agreed with the right-wingers that it would be desirable to include the Menshevik and Right SR parties in a coalition government—but on Lenin's terms. They must above all accept the soviet form of government, not a parliamentary one; they refused. Only the Left SR's agreed, and several were included in the Soviet government. Likewise, when the freely elected Constituent Assembly met in January 1918, the Mensheviks and Right SR majority flatly rejected sovietism. Lenin without hesitation ordered the dispersal of the Constituent Assembly.

The Allies refused to recognize the Soviet government; consequently it entered alone into peace negotiations with the Central Powers (Germany and her allies Austro-Hungary and Turkey) at the town of Brest-Litovsk. They imposed ruinous conditions that would strip away from Soviet Russia the western tier of non-Russian nations of the old Russian Empire. Left Communists fanatically opposed acceptance and preached a revolutionary war, even if it imperilled the Soviet government. Lenin insisted that the terms, however ruinous and humiliating, must be accepted or he would resign from the government. He sensed that peace was the deepest yearning of the people; in any case, the shattered army could not raise effective resistance to the invader. Finally, in March 1918, after a still larger part had been carved out of old Russia by the enemy, Lenin succeeded in winning the Central Committee's acceptance of the Treaty of Brest-Litovsk. At last Russia was at peace.

Treaty of Brest-Litovsk

But Brest-Litovsk only intensified the determination of counterrevolutionary forces and the Allies who supported them to bring about the overthrow of the Soviet government. That determination hardened when, in 1918, Lenin's government repudiated repayments of all foreign loans obtained by the tsarist and Provisional governments and nationalized foreign properties in Russia without compensation. From 1918 to 1920 Russia was torn by a Civil War, which cost millions of lives and untold destruction. One of the earliest victims was Lenin himself. In August 1918 an assassin fired two bullets into Lenin as he left a factory in which he had just delivered a speech. Because of his robust constitution, he recovered rapidly.

The Soviet government faced tremendous odds. The anti-Soviet forces, or Whites, headed mainly by former tsarist generals and admirals, fought desperately to overthrow the Red regime. Moreover, the Whites were lavishly supplied by the Allies with matériel, money, and support troops that secured White bases. Yet, the Whites failed.

It was largely because of Lenin's inspired leadership that the Soviet government managed to survive against such military odds. He caused the formation and guided the strategy of the Workers' and Peasants' Red Army, commanded by Trotsky. Although the economy had collapsed, he managed to mobilize sufficient resources to sustain the Red Army and the industrial worker. But above all it was his political leadership that saved the day for the Soviets. By proclaiming the right of the peoples to self-determination, including the right to secession, he won the active sympathy, or at least the benevolent neutrality, of the non-Russian nationalities within Russia, because the Whites did not recognize that right. Indeed, his perceptive, skillful policy on the national question enabled Soviet Russia to avoid total disintegration and to remain a huge multinational state. By making the industrial workers the new privileged class, favoured in the distribution of rations, housing, and political power, he retained the loyalty of the proletariat. His championing of the peasants' demand that they take all the land from the gentry, church, and crown without compensation won over the peasants, without whose support the government could not survive.

Because of the breakdown of the economy, however, Lenin adopted a policy toward the peasant that threatened to destroy the Soviet government. Lacking funds or goods to exchange against grain needed to feed the Red Army and the towns, Lenin instituted a system of requisitioning grain surpluses without compensation. Many peasants resisted—at least until they experienced White "liberation." On the territories that the Whites won, they restored landed property to the previous owners and savagely punished the peasants who had dared seize the land. Despite the peasants' detestation of the Soviet's grain requisitioning, the peasants, when forced to choose between Reds and Whites, chose the Reds.

The New Economic Policy

After the defeat of the Whites, the peasants no longer had to make that choice. They now totally refused to surrender their grain to the government. Threatened by mass peasant rebellion, Lenin called a retreat. In March 1921 the government introduced the New Economic Policy, which ended the system of grain requisitioning and permitted the peasant to sell his harvest on an open market. This constituted a partial retreat to capitalism.

From the moment Lenin came to power, his abiding aims in international relations were twofold: to prevent the formation of an imperialist united front against Soviet Russia; but, even more important, to stimulate proletarian revolutions abroad.

In his first aim he largely succeeded. In 1924, shortly after his death, Soviet Russia had won de jure recognition of all the major world powers except the United States. But his greater hope of the formation of a world republic of soviets failed to materialize, and Soviet Russia was left isolated in hostile capitalist encirclement.

Formation of the Third International. To break this encirclement, he had called on revolutionaries to form Communist parties that would emulate the example of the Bolshevik Revolution in all countries. Dramatizing his break with the reformist Second International, in 1918 he had changed the name of the RSDWP to the Russian Communist Party (Bolsheviks), and in March 1919 he founded the Communist, or Third, International. This International accepted the affiliation only of parties that accepted its decisions as binding, imposed iron discipline, and made a clean break with the Second International. In sum, Lenin now held up the Russian Communist Party, the only party that had made a successful revolution, as the model for Communist parties in all countries. One result of this policy was to engender a split in the world labour movement between the adherents of the two internationals.

The Communist International scored its greatest success in the colonial world. By championing the rights of the peoples in the colonies and semi-colonies to self-determination and independence, the International won considerable sympathy for Communism. Lenin's policy in this question still reverberates through the world today. And it offers another example of Lenin's unique ability to find allies where revolutionaries had not found them before. By taking the side of the national liberation movements, Lenin could claim that the overwhelming majority of the world's population, then living under imperialist rule, as well as the European proletariat, were the natural allies of the Bolshevik Revolution.

Thus Lenin's revolutionary genius was not confined to his ability to divide his enemies; more important was his skill in finding allies and friends for the exiguous proletariat of Russia. First, he won the Russian peasants to the side of the proletariat. Second, while he did not win the workers to make successful Communist revolutions in the West, they did compel their governments to curtail armed intervention against the Bolshevik Revolution. Third, while the Asian revolutions barely stirred in his lifetime, they did strengthen the Soviet Communists in the belief that they were not alone in a hostile world.

By 1921 Lenin's government had crushed all opposition parties on the grounds that they had opposed or failed to support sufficiently the Soviet cause in the Civil War. Now that peace had come, Lenin believed that their opposition was more dangerous than ever, since the peasantry and even a large section of the working class had become disaffected with the Soviet regime. To repress opponents of Bolshevism, Lenin demanded the harshest measures, including "show" trials and frequent resort to the death penalty. Moreover, he insisted on even tighter control over dissent within the party. Lenin's insistence on merciless destruction of the opposition to the Bolshevik dictatorship subsequently led many observers to conclude that Lenin, though personally opposed to one-man rule, nevertheless unwittingly cleared the way for the rise of Joseph Stalin's dictatorship.

Suppression of dissent

By 1922 Lenin had become keenly aware that degeneration of the Soviet system and party was the greatest danger to the cause of Socialism in Russia. He found the party and Soviet state apparatus hopelessly entangled in red tape and incompetence. Even the agency headed by Stalin that was responsible for streamlining administration was, in fact, less efficient than the rest of the government. The Soviets of Workers' and Peasants' Deputies had been drained of all power, which had flowed to the centre. Most disturbing was the Great Russian chauvinism that leading Bolsheviks manifested toward the non-Russian nationalities in the reorganization of the state in which Stalin was playing a key role. Moreover, in April 1922 Stalin won appointment as general secretary of the party, in which post he was rapidly concentrating immense power in his hands. Soviet Russia in Lenin's last years could not have been more remote from the picture of Socialism he had portrayed in ***State and***

Revolution. Lenin strained every nerve to reverse these trends, which he regarded as antithetical to Socialism, and to replace Stalin.

Illness and death. In the spring of 1922, .however, Lenin fell seriously ill. In April his doctors extracted from his neck one of the bullets he had received from the assassin's gun in August 1918. He recovered rapidly from the operation, but a month later he fell ill, partially paralyzed and unable to speak. In June he made a partial recovery and threw himself into the formation of the Union of Soviet Socialist Republics, the federal system of reorganization he favoured against Stalin's unitary scheme. But in December he was again incapacitated by semi-paralysis. Although no longer the active leader of the state and party, he did muster the strength to dictate several prescient articles and what is called his political "Testament," dictated to his secretary between December 23, 1922, and January 4, 1923, in which he expressed a great fear for the stability of the party under the leadership of disparate, forceful personalities such as Stalin and Trotsky. On March 10, 1923, another stroke deprived him of speech. His political activity came to an end. He suffered yet another stroke on the morning of January 21, 1924, and died that evening at Gorky, near Moscow.

The last year of Lenin's political life, when he fought to eradicate abuses of his Socialist ideals and the corruption of power, may well have been his greatest. Whether the history of the Soviet Union would have been fundamentally different had he survived beyond his 54th birthday, no one can say with certainty.

MAJOR WORKS

BOOKS: ***Chto Takoye "Druzya Naroda," kak oni voyuyut protiv Sotsial-Demokratov?*** (hectographic editions from 1894, printed from 1920; ***What the "Friends of the People" Are, and How They Fight the Social-Democrats,*** 1946); ***Razvitiye kapitalizma v Rossi*** (1899, augmented ed. 1908; ***The Development of Capitalism in Russia,*** 1956); ***Chto delat?*** (1902; ***What Is To Be Done?,*** 1929, 1933, 1950,1963); ***Shag vperyod, dva shaga nazad*** (1904; ***One Step Forward, Two Steps Back,*** 1941); ***Dve taktiki Sotsial-Demokraty v demokraticheskoy revolyutsi*** (1905; ***Two Tactics of Social-Democracy in the Democratic Revolution,*** 1935, 1947); ***Agrarny vopros i "Kritiki Marksa"*** (1908, first chapters already published in periodicals from 1901); ***Agrarnaya programma Sotsial-Demokratyi v pervoy russkoy revolyutsl 1905–1907 godov*** (1908, 1917; ***The Agrarian Programme of Social-Democracy in the First Russian Revolution,*** 1954); ***Imperializm, kak noveyshy etap kapitalizma*** (1917, later retitled ***Imperializm, kak vysshaya stadiya kapitalizma; Imperialism: The Latest Stage in the Development of Capitalism,*** 1924; and ***Imperialism, the Highest Stage of Capitalism,*** 1933, 1939, 1947); ***Gosudarstvo i revolyutsiya*** (1917; ***The State and Revolution,*** 1919); ***Proletarskaya revolyutsiya i renegat Kautsky*** (1918; ***The Proletarian Revolution and Kautsky the Renegade,*** 1920); ***Detskaya bolezn "levizny" v kommunizme*** (1920; ***"Left Wing" Communism: An Infantile Disorder,*** 1920, 1934, 1961).

MARXIST PERIODICALS EDITED OR CONTROLLED BY LENIN: ***Iskra,*** 51 numbers (1900–03); ***Zarya,*** 4 numbers (1901–02); ***Vperyod,*** weekly, 18 numbers (1905); ***Novaya Zhizn,*** daily, 28 numbers (1905); ***Proletary,*** weekly, 26 numbers (1905); ***Volna,*** daily, 25 numbers (1906); ***Vperyod,*** daily, 17 numbers (1906); ***Ekho,*** daily, 14 numbers (1906); ***Proletary,*** 50 numbers (1906–09); ***Sotsial-Demokrat,*** 58 numbers (1908–17, wholly Leninist from 1911); ***Zvezda,*** at first weekly, then more frequent, 69 numbers (1910–12); ***Rabochaya Gazeta,*** 9 numbers (1910–12); ***Prosveshcheniye,*** monthly, 31 numbers (1911–14 and 1 final number, 1917); ***Pravda,*** daily (1912–14 and from 1917 with numerous changes of name 1913–14 and July-October 1917); ***Kommunist,*** 1 double number (1915); ***Sbornik Sotsial-Demokrafa,*** 2 numbers (1916).

JOURNALISM AND PARTY THESES: "O zadachakh proletariata v dannoy revolyutsi" ***(The April Theses,*** 1951), "O dvoevlasti," and "Uroki revolyutsi" ***(Lessons of the Revolution,*** 1918), all in ***Pravda,*** 1917; "Ocheredniye zadachi Sovietskoy vlasti," ***Pravda*** (1918; ***The Soviets at Work,*** 1918; and ***The Immediate Tasks of the Soviet Government,*** 1951).

BIBLIOGRAPHY

Editions: The fifth and most complete collection of Lenin's works was published in Moscow, 55 vol. (1958–65, with index, 1966; reference volume in two parts, 1969–70). The ***Collected Works,*** 45 vol. (1960–70), is an English translation of the 4th Russian edition of Lenin's works, enriched by editorial notes from the 5th edition. ***Selected Works,*** 3 vol. (1967), includes most of the works mentioned in this article and many more.

Biographical and critical studies: LOUIS FISCHER, ***The Life of Lenin*** (1964), a major, detailed work, enlivened by the author's own reminiscences of Moscow in Lenin's time; MOSHE LEWIN, ***Le Dernier Combat de Lénine*** (1967; Eng. trans., ***Lenin's Last Struggle,*** 1968), a brief but incisive account of Lenin's effort to mobilize party forces to curb Stalin and remove him from power; ALFRED MEYER, ***Leninism*** (1957), the best analysis of Lenin's political philosophy; DAVID SHUB, ***Lenin,*** rev. ed. (1966), readable and informative; ADAM B. ULAM, ***The Bolsheviks: The Intellectual and Political History of the Triumph of Communism in Russia*** (1965), a learned political biography of Lenin; BERTRAM D. WOLFE, ***Three Who Made a Revolution: A Biographical History,*** 4th rev. ed. (1964), a pioneering work (biographies of Lenin, Trotsky, and Stalin) that demolishes many myths; NADEZHDA KRUPSKAYA, ***Reminiscences of Lenin*** (1970; Eng. trans. of the 2nd Russian ed., 1968), reticent, impersonal recollections by Lenin's widow; LEON TROTSKY, ***Lenin: Notes for a Biographer*** (1971; Eng. trans. of the 2nd Russian ed., 1924), an appreciation of Lenin of the ***Iskra*** period and 1917–18, the periods of Trotsky's closest collaboration with Lenin; NIKOLAI VALENTINOV, ***Encounters with Lenin*** (1968; orig. pub. in Russian, 1953), revealing observations on Lenin's personality by a former associate; DIETRICH GEYER, ***Lenin in der Russischen Sozialdemokratie*** (1962), a scholarly study of Lenin and the origins of the Bolshevik–Menshevik split.

(Al.Re.)

Leningrad

The second largest city of the Soviet Union and one of the world's major cities, Leningrad has played a vital role in Russian history—for two centuries as capital of the Russian Empire—and it maintains today outstanding importance as an industrial and cultural centre and as a seaport. Founded by Peter I the Great in 1703 as St. Petersburg, it was renamed Petrograd in 1914 and, finally, Leningrad in 1924. The city is particularly renowned as the scene of the February and October revolutions in 1917, as the besieged and fiercely defended city of World War II, and, architecturally, as one of the most splendid and harmonious cities of Europe.

Leningrad is sited on the delta of the Neva River, where it debouches into the head of the Gulf of Finland. The city spreads across nearly 100 islands of the delta and across adjacent parts of the mainland floodplain. The very low and originally marshy site has made the city subject to recurrent flooding, especially at the time of the spring thaw. Exceptionally severe inundations occurred in 1777, 1824, and 1924, the last two being the highest on record and flooding most of the city. A number of canals have been cut to assist drainage. These, together with the many natural channels (now embanked with stone), make Leningrad a city of waterways and bridges and have earned it the nickname of "Venice of the North."

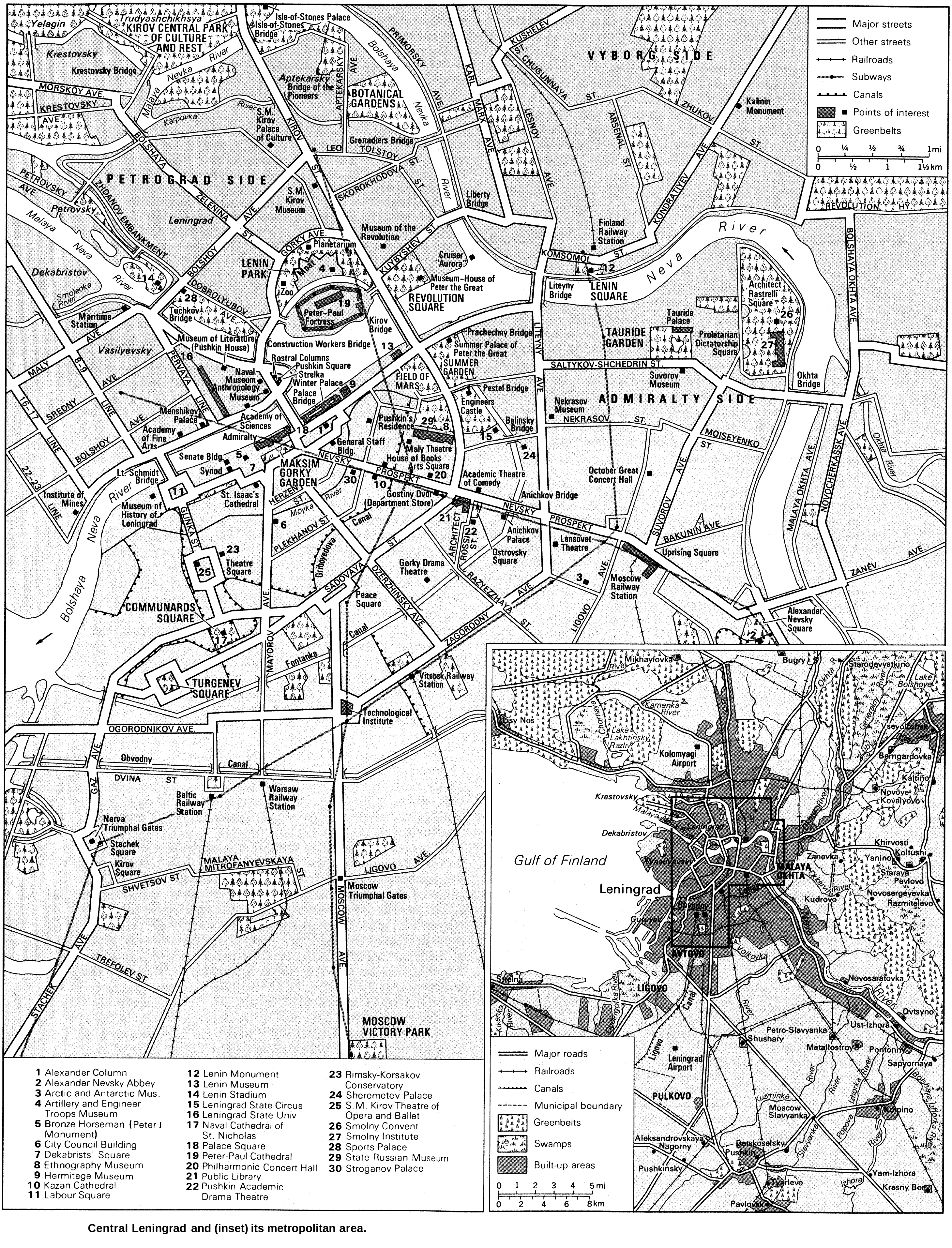
Floods

The climate is modified continental, with marked maritime influences. Winters are cold, with a February average temperature of 18" F (−8° C), more keenly felt because of the moist sea air. Snow cover lasts on average 132 days. The Neva begins to freeze normally about mid-November, and the ice is solid by the start of December; breakup begins in mid-April and is completed by the end of the month. Icebreakers prolong the navigation season. Summers are only moderately warm, with a July average temperature of 63.9" F (17.7" C). The mean annual precipitation is 23 inches (585 millimetres), the maximum coming in summer. The northerly location (latitude 60° N passes through Leningrad's northern suburbs) gives long winter nights, while the half-light "white nights" of early summer are a renowned feature of the city.

Greater Leningrad covers an area of 521 square miles (1,350 square kilometres), which includes a number of satellite towns and urban districts under the city administration. Its population in 1975 was 4,311,000, of whom 3,853,000 lived within Leningrad itself.

HISTORY

Foundation and early growth. Settlement of the region around the head of the Gulf of Finland by Russians began in the 8th or 9th century AD. Known then as Izhor-



Central Leningrad and (inset) its metropolitan area.

skaya Zemlya, or more commonly as Ingermanland, or Ingria, the region came under the control of Novgorod, but for long it remained very thinly populated and underdeveloped. In the 15th century the area passed, with Novgorod, into the possession of the grand princes of Moscow. Sweden annexed Ingria in 1617 and established fortresses along the Neva. During the Great Northern War of 1700–21, Peter I the Great, seeking a maritime outlet to the west for Russia, constructed a fleet on the Svir River (which connects Lakes Onega and Ladoga) and sailing across Lake Ladoga launched an attack on the fortress of Noteburg (later Shlisselburg, now Petrokrepost), where the Neva flows out of Ladoga. In 1703 Noteburg fell to Peter, and shortly afterward he captured the Swedish fortress of Nienshants on the lower Neva, thus gaining control of the delta and the head of the gulf.

Peter–Paul Fortress

On May 16 (May 27, new style), 1703, shortly after the fall of Nienshants, Peter himself laid the foundation stones of a new fortress in the delta, the Peter–Paul Fortress on Zayachy Island near the right bank of the Neva. This date is taken as the founding date of Sankt-Peterburg, or St. Petersburg. The massive fortification, the city's first structure, initially had earthen walls, but these were soon replaced by stone walls 40 feet high and 12 feet thick, with 300 cannon mounted on the bastions. In the spring of the following year Peter established the fortress of Kronshlot, later Kronshtadt, on Kotlin Island in the Gulf of Finland, to protect the approaches to the delta. At the same time he founded the Admiralty shipyard on the riverbank opposite the Peter–Paul Fortress; in 1706 its first warship was launched. Around the fortress and shipyard Peter began the building of a new city to serve as his "window on Europe." Just upstream of the Peter–Paul Fortress, the first small, single-storied house, built for Peter himself in the early days of the city's construction, is still preserved as a museum.

Although the first dwellings were single-storied and made of wood (and later, as a precaution against fire, of clay on wooden frames), it was not long before stone buildings were erected. The first stone palace, still preserved, was completed in 1714 for Prince Aleksandr Danilovich Menshikov, first governor of the city. From the start the city was planned as an imposing capital, on a regular street pattern, with broad avenues radiating out from the Admiralty and spacious squares. Architects, craftsmen, and artisans were brought from all over Russia and from many foreign countries to construct and embellish the new town. In 1712 the capital of Russia was transferred there from Moscow, although it was not until 1721 that Sweden, in the Treaty of Nystad, formally ceded sovereignty of the area to Russia. Members of the nobility and merchant class were compelled to move to the new capital and to build houses for themselves. Government buildings and private palaces and houses arose swiftly; among the earliest buildings were the Merchants' Exchange (now the Naval Museum), Customs House (now the Museum of Literature), and marine hospital, together with the Summer Palace. Canals for drainage were cut through the marshy left bank of the Admiralty Side. The first floating bridge over the Neva was constructed in 1727, and soon more than 370 bridges, some of them drawbridges, had been built across the many canals and river channels. The watercourses themselves were controlled with embankments of gray and pink granite.

A harbour was constructed, and Peter took measures to curtail traffic through Arkhangelsk on the White Sea, previously Russia's major port. In consequence, as early as 1726 St. Petersburg was handling 90 percent of Russia's foreign trade. In 1703 Peter had begun work on the Vyshnevolotsky Canal in the Valdai Hills to link the upper Volga to the Msta and so, by way of the Volkhov, Lake Ladoga, and the Neva, to St. Petersburg. The canal was opened to traffic in 1709, giving the capital a direct water link to central Russia and all the Volga Basin. Industry soon began to develop. The original and flourishing Admiralty shipyard was joined by enterprises to supply its needs and those of the growing fleet—a foundry to produce cannon, a gunpowder factory, and a tar works. Merchantmen as well as warships were built, and before the end of the 18th century papermaking, printing, and food, clothing, and footwear industries had been established; as early as the 1740s a factory was set up to make china. By 1765 the population numbered 150,300 and by the end of the century it had reached 220,200, of whom more than a third were in the armed forces or the administration.

The rise to splendour. The growing city displayed a remarkable richness of architecture and harmony of style. Initially the style was one of simple but elegant restraint, seen in the cathedral of the Peter–Paul Fortress and in the Summer Palace. In the mid-18th century an indelible stamp was put on the city's appearance by the architects Bartolomeo Rastrelli, S.I. Chevakinsky, and V.P. Stasov, working in the style known as Russian Baroque, which combined clearcut, even austere lines with richness of decoration and use of colour. To this period belong the Winter Palace, the Smolny Convent, and the Vorontsov and Stroganov palaces, among others; outside the city were built the summer palaces of Peterhof on the gulf shores and of Tsarskoye Selo. After a transitional period, dominated by the architecture of Jean-Baptiste de La Motte and Aleksandr Kokorinov, toward the end of the 18th century a pure Classical style emerged under the architects Giacomo Quarenghi, Carlo Rossi, A.N. Voronikhin, and others. The Kazan and St. Isaac's cathedrals, the Smolny Institute (the most famous girls' school of tsarist Russia), the new Admiralty, the Senate, and the Yelagin Palace and Mikhaylovsky palace (now the State Russian Museum) are notable representatives of the many splendid buildings of this period.

Russian Baroque architecture

The cultural life of the capital developed and flourished. In 1773 the Institute of Mines, one of the first higher technical colleges in Russia, was established. The University of St. Petersburg (now Leningrad State University) was founded in 1819. Many of the most celebrated names in Russia in the spheres of learning, science, and the arts are closely associated with the city: M.V. Lomonosov, D.I. Mendeleyev, Ivan Pavlov, Aleksandr Pushkin, Fyodor Dostoyevsky, and Leo Tolstoy among many others. As early as 1738 the first ballet school in Russia was opened in St. Petersburg; in the 19th century, under the guidance of Marius Petipa, the Russian ballet rose to worldwide pre-eminence and renown and produced such dancers as Vaslav Nijinsky, Tamara Karsavina, and Anna Pavlova. In 1862 the first conservatory of music in Russia opened its doors, and there the premieres of works by Tchaikovsky, Rimsky-Korsakov, Rachmaninoff, and other composers were performed. Over all, as focus and patron of the city's cultural life, stood the imperial court; its ostentatious splendour and wealth were legendary throughout Europe. Its presence fostered the growth of jewelry making and other luxury trades.

The road to revolution. The imperial magnificence, centred on the tsarist autocracy, was in sharp contrast with the other side of St. Petersburg's development, the growth of its industrial proletariat. The 19th century witnessed very great growth of industry in the city, accelerated by improvements in communications and extension of trade. Navigation was opened on the Mariinsky (1810) and Tikhvin (1811) canal systems, which linked the capital with the Volga Basin and replaced the old and inadequate Vyshnevolotsky system. In 1813 the first Russian steamship was built in St. Petersburg and in 1837 the first Russian railway was constructed from the city to the Summer Palace in Tsarskoye Selo (now Pushkin). Five years later work started on the railway to Moscow, opened to traffic in 1851. A line to Warsaw followed in 1861–62, followed by others that made the city a focus of rail routes. In particular, the cotton textile and metalworking industries flourished, the former using imported raw materials. By the 1840s more than 60 percent of Russian imports were entering by way of St. Petersburg, but the increasing size of vessels necessitated transshipment at Kronshtadt until 1885, when a channel was dredged to give access to the port. City growth and industrialization were greatly stimulated by the emancipa-

New lines of communication

tion of serfs in 1861, which allowed far greater mobility of labour. From 539,400 inhabitants in 1864 the population rose to 1,439,000 in 1900, largely by migration from rural areas; in 1910 only 32 percent of the population had been born in the city. By 1914 the total had risen to 2,100,000.

Politicized work force

With the development of metalworking and engineering as the primary industries, there arose a skilled labour force, increasingly alert politically. Moreover, the factory workers, who numbered nearly 250,000 in 1914, tended to be concentrated in plants of far larger size than was usual in Russia; the Putilov (now Kirov) armaments works alone employed about 13,000. It was thus easier for revolutionaries to spread their ideas and for workers' groups to organize themselves than it was elsewhere in Russia. At the same time the growth of the city was characterized by a very belated and slow development of public transport, making it necessary for workers to live close to their place of work, in conditions of exceptionally high population density (more than 180,000 per square mile in the centre), appalling overcrowding, squalor, and lack of sanitation. Throughout the period before 1917 the city administration was lacking in efficiency and often in funds, and the provision of all public services--even a water supply—was very inadequate. Outbreaks of serious epidemics were commonplace, and St. Petersburg was notorious as one of the most unhealthful capitals of Europe. For many of the inhabitants the conditions of life were extremely bad.

The first serious revolutionary outbreak in St. Petersburg came on December 14 (December 26, N.S.), 1825, the so-called Dekabrist (Decembrist) insurrection, organized largely by liberal aristocrats and army officers seeking a liberal constitution and an end to serfdom. It was ruthlessly suppressed. During the rest of the century, workers' revolutionary activity by many different groups and general unrest steadily increased, with ever more frequent strikes and outbreaks of violence. These culminated in the general strike of January 1905, when some 150,000 workers took part. On Bloody Sunday, January 9 (22, N.S.), a mass march to the Winter Palace, bearing a petition to the Tsar, was met by troops, who opened fire; more than 1,000 people were killed. The situation developed into revolution, spreading throughout Russia. Although it was again crushed, underground revolutionary activity continued.

The outbreak of World War I in 1914 saw an upsurge of patriotic fervour focussed on the Tsar. The Germanic form of the city's name was changed to its Russian version, Petrograd. The military disasters of the war and the worsening economic situation, however, revived and intensified discontent. Transport inefficiencies led to severe shortages of food and other supplies in Petrograd. On February 26 (March 11, N.S.), 1917, disorders broke out as a result of food shortage.' Authority was slow to act and lost all control. The city was in chaos and the Petrograd Soviet of Workers' and Soldiers' Deputies was formed on February 27 (March 12, N.S.). The Tsar was prevented from entering the city, and on March 2 (15, N.S.) he abdicated. A provisional government was set up, eventually under the premiership of Aleksandr Kerensky.

Return of Lenin

On April 3 (16, N.S.) Lenin returned to Petrograd from Switzerland and set about organizing the overthrow of the provisional government. A Bolshevik rising in July was suppressed, but on October 25 (November 7, N.S.) Communist-led workers and sailors stormed the Winter Palace, deposing the provisional government and establishing Lenin and the Communist Party in power. The battle cruiser "Aurora" of the Baltic Fleet, which bombarded the palace, remains berthed in the Neva as a museum.

The Russian Revolution of 1917, which changed the course of history, was in large measure the work of the Petrograd proletariat and the sailors from Kronshtadt. In January 1918 a Constituent Assembly met in Petrograd, but the Bolsheviks, who had won only 168 seats out of 703, dispersed it by force and consolidated their authority. In March 1921 the sailors of Kronshtadt rose in revolt against Lenin, but after a night attack on the fortress across the frozen gulf, the rising was put down with great severity.

The Soviet period. In March 1918 the capital of the young Soviet state was moved back to Moscow. The years of the civil war after the Revolution had a disastrous effect on the city's economy. Industry came very nearly to a standstill. The population fell sharply to 722,000 in 1920, a mere third of the pre-Revolutionary size. Many died of hunger. In 1924, following Lenin's death, the city was renamed Leningrad. With the ending of the civil war, recovery began. When in 1928 the era of five-year plans began, much of the initial burden of developing the national economy fell on Leningrad and its established industrial plant and work force, especially in the provision of power equipment and machinery. This stimulated further growth, and by 1939 the city was responsible for 11 percent of all Soviet industrial output and its population had passed the 3,000,000 mark.

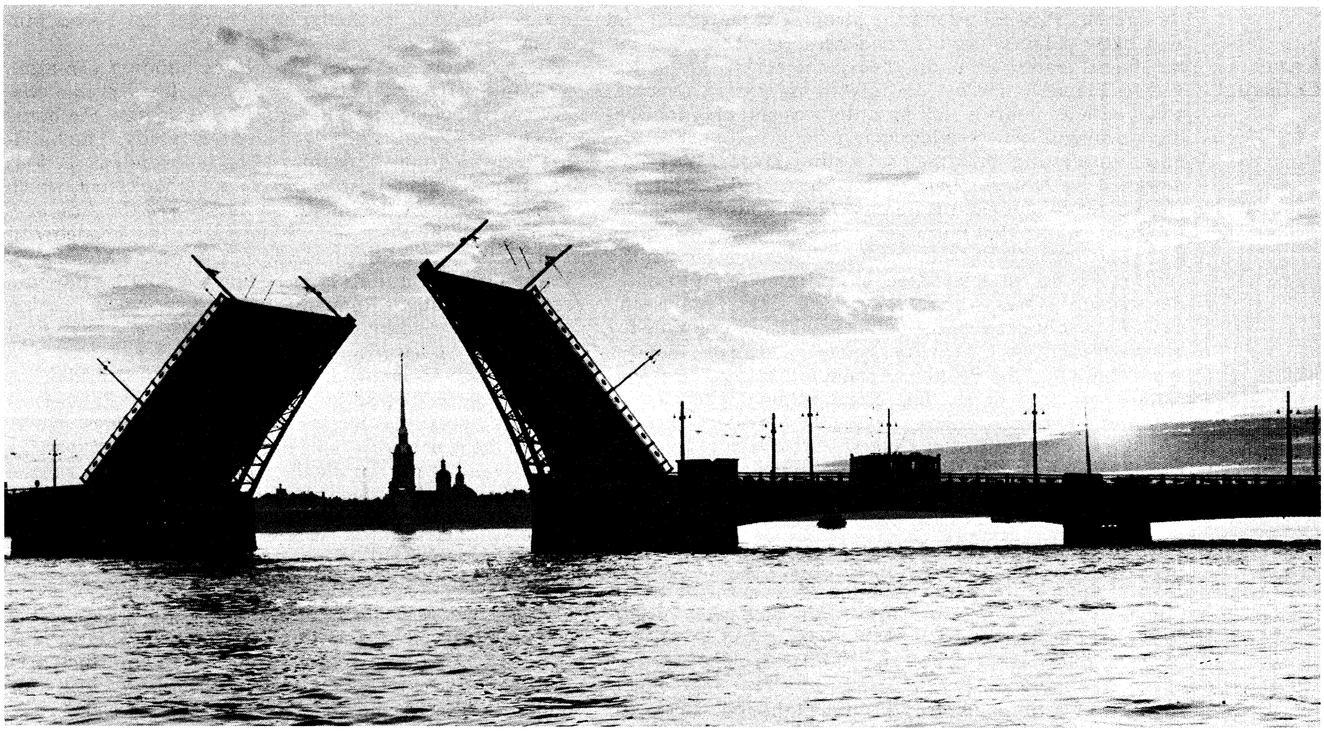
Then once again Leningrad was struck by a period of loss and destruction. It was one of the initial targets of the German invasion in 1941; by September of that year German troops were on the outskirts of the city and had cut off communication with the rest of the Soviet Union, while Finnish troops advanced across the Karelian Isthmus from the north. Many of the inhabitants and nearly three-quarters of the industrial plant were evacuated eastward ahead of the German advance. The remainder of the population and the garrison began what has become known as the 900-day siege; in fact it lasted 872 days, from September 8, 1941, to January 27, 1944. Leningrad put up a desperate and courageous resistance in the face of many assaults, constant artillery and air bombardment, and appalling suffering from shortages of supplies. An estimated one million people died, a very high proportion from scurvy and starvation. In particular the exceptionally bitter winter of 1941–42, when temperatures fell to —40" F (—40" C), was one of extreme hardship and loss of life. The only route for supplies was the "road of life" across the ice of Lake Ladoga; later an oil pipeline and electric cables were laid on the lake bed. The strict blockade was broken in January 1943, but it was another year before the Germans had been driven back from the outskirts. Enormous damage had been caused by the bombardment, and before retreating the Germans deliberately destroyed the palaces at Petrodvorets and Pushkin. So great was the city's loss that 15 years later, in the census of 1959, its population, 2,899,955, had still not regained prewar size. For its role in the war Leningrad was subsequently honoured with the title of Hero City, a special defense medal, and the Order of Lenin. To these were later added a second Order of Lenin, Gold Star, and Order of the October Revolution.

The 900-day siege

In Leningrad, the first postwar Five-Year Plan was devoted to reconstruction of the city. By 1950 industry was once more on its feet, and it has continued to expand and develop. The architectural heritage of the past was rebuilt and restored with meticulous care. By the census of 1970 the population had reached 3,512,974 in the city itself and 3,949,501 including the suburban towns within Greater Leningrad. From the late 1950s onward a program of housing construction added extensive new suburbs of apartment buildings, arranged in "micro-regions," or neighbourhood units, with populations usually of 3,000 to 8,000 and with local services of shops, cinemas, polyclinics, and schools. The apartment buildings, usually of five to nine stories in the 1960s, from the 1970s tended to be higher, with 12 to 20 or more stories. Great use has been made of standardized prefabrication in housing construction. Since 1966 development has been coordinated in a master plan, covering a 25-year period.

THE CONTEMPORARY CITY

Greater Leningrad, the city itself with its satellites, forms a large horseshoe around the head of the Gulf of Finland and including the island of Kotlin in the gulf. On the north it stretches westward along the shores for nearly 50 miles to include Zelenogorsk. This northern extension is an area of dormitory towns, resorts, sanatoriums, and holiday centres set among extensive conif-



Drawbridge rising over the Neva River, Leningrad. In the background on the right embankment is the Peter-Paul Cathedral (1712–33).

Novosti Press Agency—E. Ettinger

erous forests and fringed by fine beaches and sand dunes. Here many of the more important and influential citizens have summer cottages, or dachas. There are also many holiday camps for children. On the southern side of the gulf the metropolitan limits extend westward to include Petrodvorets (formerly Peterhof) and the towns of Pushkin (formerly Tsarskoye Selo) and Pavlovsk. Eastward, Greater Leningrad stretches up the Neva River to Ivanovskoye.

The central city. Leningrad is divided by distributaries of the Neva River into four sections. The Admiralty Side lies along the left (south) bank of the Neva itself. Between the two major arms of the Neva, the Bolshaya (Great) Neva and the Malaya (Little) Neva, is Vasilyevsky Island. The Malaya Neva and the distributary known as the Bolshaya Nevka enclose a group of islands known as the Petrograd Side, while east of the Bolshaya Navka lies the Vyborg Side.

The Admiralty

The Admiralty Side. The southern side of the Neva, known as the Admiralty Side, centres on the Admiralty. This, the nucleus of Peter's original city, was reconstructed in 1806–23 by A.D. Zakharov, as a development of the earlier building of I.K. Korobov, constructed in 1727–38, but retaining the layout of the original. Its elegant spire, topped by a weather vane in the form of a ship, is one of the principal landmarks of the city. The building today houses a naval college.

Just to the north lies the great Palace Square, the city's oldest. The 600-ton granite monolith of the Alexander Column (1830–34), tallest of its kind in the world, thrusts up for 155.8 feet (47.5 metres) near the centre of the square, so finely set that its base is not fastened.

Between the square and the river rises the huge and massive rectangle of the Winter Palace, the former principal residence of the tsars. The present structure, the fifth to be built, was the Baroque masterpiece of Rastrelli, and was built in 1754–62. Both the exterior and the interior of the palace were dazzlingly luxurious. In 1837 the building was destroyed by fire, and only the adjoining Hermitage survived; the Winter Palace was recreated in 1839 almost exactly according to Rastrelli's plans. The appearance of the palace is striking: the background is green, the columns white, and the stucco moldings golden, and 176 sculptured figures line the roof. The whole complex now houses the State Hermitage Museum, a treasure-house of fine art of worldwide significance that originated in 1764 as the private collection of the empress Catherine II.

Opposite the Winter Palace, the great crescent of Rossi's General Staff building (1811–29) dominates the square. The two wings of the building are joined by a huge triumphal arch, topped by heroic figures and crowned by a chariot containing a figure representing Glory, expressing the Russian victory in the war of 1812.

Dekabrists Square

On the western (downstream) side of the Admiralty stretches the expanse called Senate Square when that body moved to the city in 1763 and now called Dekabrists Square in commemoration of the revolt in 1825. The buildings of the former Senate and Synod (now housing archives) dominate the western side of the square, their decorated facades dating from the 1830s and representing the last great work of Rossi. They are separated by an arch looking across to the centre of the square where the equestrian statue of Peter I the Great, known as the "Bronze Horseman" and created in 1782 by Étienne Falconet, looks out to the Neva. Beside the Senate and Synod buildings stand two columns (1845) bearing statues of the goddess of glory; next to them rises the Classical front of the Horse Guards Riding School (1807), reminiscent of a Greek temple, and beyond, dominating the south side of St. Isaac's Square, is the cathedral of the same name, an outstanding monument of late Classical Russian architecture built by A. Montferrand (1818–58). St. Isaac's is one of the largest domed buildings in the world, and its golden cupola, 71.60 feet (21.83 metres) in diameter, soars to 331 feet (101 metres) and is visible all over Leningrad. Majestic and massive, the cathedral can hold 14,000 people. It is now a museum.

From the Admiralty and its surrounding squares radiate three great avenues, of which the most important and best known is the Nevsky Prospekt. One of the world's great thoroughfares, the Nevsky Prospekt cuts eastward across the peninsula formed by the northward loop of the Neva to the vicinity of the Alexander Nevsky Abbey, crossing the smaller Moyka River and Fontanka Canal. The Anichkov Bridge across the latter is graced by four groups of sculptured horses. The street has a special beauty; the architecture is majestic, the buildings graceful and finely proportioned, the construction complex. On

Kazan Cathedral

the Nevsky Prospekt stand the Stroganov, Shuvalov, and Anichkov palaces, private residences, and churches, but the finest feature is undoubtedly the Kazan Cathedral (1801–11), designed by Voronikhin in Russian Classical style, with an interior rich in sculptures and paintings behind a magnificent semicircular frontal colonnade. Another interesting building is Gostiny Dvor (1761–85), designed by Jean-Baptiste de La Motte and rebuilt 100 years later by Alexandre Benois. This building forms an irregular square that has a perimeter of half a mile or so and that opens onto four streets; formerly a mercantile centre, it is now a department store. In addition to other department stores, the Nevsky Prospekt contains the city's largest bookshop, the Saltykov-Shchedrin Public Library (located on Ostrovsky Square), and many theatres —most notably the Pushkin Academic Drama Theatre— restaurants, and cafes. The street is always noisy and humming with activity.

At the eastern end of the Nevsky Prospekt, Alexander Nevsky Square fronts the main entrance to the abbey of the same name and its surrounding gardens. Beyond the square's main entrance lie, on the left and right, respectively, monuments and sculptures of the 18th-century Lazarus Cemetery (where Lomonosov and many of the city's architects are buried) and the 19th-century Tikhvin Cemetery (containing the graves of such writers and composers as Dostoyevsky, Mussorgsky, and Tchaikovsky). Behind rise the spires and cupolas of the Cathedral of the Annunciation (1720, designed by Trezzini), which is now a museum, and Holy Trinity Cathedral (1778–90, designed by Ivan Starov).

Through the Admiralty Side and intersecting the radial avenues cut the canals that so characterize the city. The most important, in concentric order from the Admiralty, are the Moyka River and the Griboyedova, Fontanka, and Obvodny (Bypass) canals. Downstream of the northern entrance of the Fontanka into the Neva lies the Field of Mars, one of the most beautiful open spaces in Leningrad. Begun under Peter (when it was known as the Field of Amusement), it was intended for popular festivities and fireworks. A favourite haunt of the 18th-century nobility under the name Tsarina's Meadow, it derives its present name from a monument erected in 1801 and portraying the great Russian military leader A.V. Suvorov (buried in the Cathedral of the Annunciation) as the god of war. In the 19th century the square was used for military parades and exercises. The fallen of the 1917 February Revolution and defenders of the city during the civil war and foreign military intervention (1918–20) were buried here. Since 19.57 an eternal flame has commemorated the defenders of the Revolution.

Summer Palace

Just to the east lies the distinctive Summer Garden. Founded on an island in 1704, it has well laid out parks and gardens that contained more than 250 magnificent statues and busts, mostly the work of Venetian masters, by the end of the 18th century. The Summer Palace, Peter's first in the city, erected 1710–14 in early Russian Baroque style and designed by Trezzini, stands in the northeastern portion. The whole area suffered greatly in the great flood of 1777. The Neva embankment is fronted by an iron grille (1784) that is at the same time majestically monumental and light and delicate in design.

At the extreme eastern side of the central city, within the sharp bend of the Neva itself, lies the Smolny complex of buildings; these include the former convent, with the beautiful five-domed cathedral, built by Rastrelli in 1748–64, and the fine Classical building of the Smolny Institute constructed by Quarenghi in 1806–08. The institute was used as Lenin's headquarters during the events of 1917.

The Strelka. Opposite the Admiralty and Winter Palace is the eastern tip of Vasilyevsky Island, where the remarkable architectural complex known as the Strelka (literally, "Pointer") faces the bifurcation of the Neva. Behind the two great Rostral Columns, decorated by carved ships' prows, and across Pushkin Square, the point rises majestically to the Merchants' Exchange (Thomas de Thomon, 1805–lo), the city's finest example of early 19th-century style and reminiscent of a classical temple in appearance. The Exchange now houses the Naval Museum.

Farther back, the Twelve Colleges building (Trezzini, 1722–42), originally intended to house the supreme governmental bodies of Peter I the Great, is now the home of Leningrad A.A. Zhdanov State University. The building is divided into 12 identical but independent sections and runs at right angles to the Neva embankment, which is fronted here by the impressive facades of the Academy of Sciences, the Menshikov Palace, and the Academy of Fine Arts. On the far, or northern, side of the Merchants' Exchange, the Customs House (now the literary museum and the Institute of Russian Literature known as Pushkin House), designed by Lucchini (1829–32), is a crisp final link in the Classical complex of the Strelka.

Peter–Paul Cathedral

The Petrograd Side. Upstream of the bifurcation of the Neva is the Petrograd Side, where the great Peter–Paul Fortress faces the Strelka across the Malaya Neva. Above the squat horizontal lines of the fortress' massive stone walls soars the slender, arrow-like spire of the Peter–Paul Cathedral, a golden landmark for the city that reaches 400 feet (122 metres). The cathedral was built in 1712–33 by Domenico Trezzini, and all the tsars and tsarinas of Russia from the time of Peter the Great (except for Peter II and Nicholas II) are buried in it. Trezzini also designed St. Peter's (Petrovsky) Gate (1718) as the eastern entrance to the fortress. The Neva Gate, designed by N.A. Lvov, dates from 1787. From the early 19th century the fortress was used as a prison, chiefly for political prisoners and including a number of Dekabrists. Today it is a historical museum. At noon each day a signal cannon is fired from its battlements.

Just to the east of the Peter–Paul Fortress, at the beginning of the distributary known as the Bolshaya Nevka River, the cruiser "Aurora," which fired on the Winter Palace in 1917, is permanently moored as a museum and training vessel for the Naval College.

The Vyborg Side. Farther upstream and east of the Bolshaya Nevka is the Vyborg Side. It contains the Finland Railway Station, in front of which stands a statue of Lenin commemorating his return to the city in 1917.

The outer city. In general terms this may be defined as the area north of the Bolshaya Nevka and south of the Obvodny Canal. Buildings of historic or architectural interest are relatively few. Extensive industrial and residential suburbs spread north and south, those closest to the centre mostly of late 19th- or early 20th-century date. These parts, like the central city, are characterized by very high population densities, between 26,000 and 73,000 persons per square mile, although rehousing since the late 1950s has reduced the previously exceptionally high figures of the centre. The city now extends well to the north and south of the original delta site. The outer suburbs of later development include extensive open areas, reducing overall population densities to between 5,000 and 21,000 per square mile. The use of large apartment buildings with two- or three-room flats means that densities within the built-up area remain very high. It is in these outer parts, especially on the eastern and southern sides of the present built-up area, that future expansion is envisaged under the master plan of 1966. Other parts of the periphery are designated as green belt for amenity purposes, and urban development is prohibited. Scattered through the outlying sections of Greater Leningrad are the satellite towns and dormitories.

Petrodvorets. The communities around Leningrad are also rich in historic monuments. The most famous is Petrodvorets (Peterhof before 1944), whose unique garden-park setting, stretching in terraces rising above the Gulf of Finland, contains representative works from two centuries of Russian architectural and park styles. The Great Palace, former residence of Peter I, stands at the edge of the second terrace, its bright yellow walls contrasting with white stucco decorations and the gilt domes of its lateral wings. Built in the Baroque style (1714–28), it was reconstructed and expanded by Rastrelli in 1747–52. On the north, the building commands a view of the Grand Cascade, a grandiose structure including a grotto, 64 fountains, and two cascading staircases, which lead to an enormous semicircular basin that originally contained a giant statue of Samson wrestling with a lion that symbolizes the military glory of Russia.

This magnificent vista is all the more remarkable when it is remembered that much of it is a post-World War II restoration that at the beginning of the 1960s was still under way. The Nazis carried off the Samson statue and many other statues and left only the charred walls of the palace itself.

Pushkin. The town of Pushkin (before 1917 Tsarskoye Selo; called Detskoye Selo in 1918–37) arose in the beginning of the 18th century as one of the tsarist residencies. The palace of Catherine I (1717–23; enlarged by A.V. Kvasov and S.I. Chevakinsky, 1743–48; rebuilt by V.V. Rastrelli, 1752–57) is notable for its dimensions, the beauty and majesty of its form, and the wealth of its sculptural decoration. The golden suite of splendid halls (including the Amber Room) exemplifies Russian Baroque at its peak. The community also is the site of the Chinese Village (1782–96) in Alexander Park and the gallery (1780–90) named after the architect Charles Cameron, the terraces of which contain more than 50 busts of figures from ancient Greek and Roman history. The Lycte, a school for the offspring of the nobility, had the great Pushkin as a student. The town suffered severe destruction during the German onslaught but has been much restored to its former glory as a museum town.

The palace of Catherine

Pavlovsk. The town of Pavlovsk, a southern suburb, is the site of a remarkable late 18th- and early 19th-century palace-park ensemble in the Classical style created as a country residence for Tsar Paul I. The central Great Palace (1782–86; Cameron) is crowned by a dome supported by 64 columns. It, too, was severely damaged by the Nazis and has been restored. Some 60,000 of its destroyed trees have been replaced.

Gatchina. Another southern suburb, Gatchina, is noted for the palace built in 1766–72 by Rinaldi for Count Orlov, a favourite of Catherine II. Gatchina Park was created at the same time. Its monuments, sculptures, and gardens, like those of all Leningrad, are under state supervision that provides for protection and restoration.

ADMINISTRATIONAND THE URBAN ECONOMY

Since 1931 Leningrad has been designated a "city of republican subordination"—that is, its City Council (Soviet) is directly subordinate to the government of the Russian Soviet Federated Socialist Republic. The council consists of 640 members, elected for two-year terms. The city itself is divided into 19 administrative wards. In addition to these, there are five towns in the outer part of Greater Leningrad with status equivalent to that of the administrative wards—Kolpino, Petrodvorets, Pushkin, Sestroretsk, and the island of Kronshtadt. Two other towns, Zelenogorsk and Pavlovsk, and 17 urban districts are subordinate to the administrative ward councils. Leningrad is also the administrative centre of Leningrad *oblast* (region), which has an area of 33,200 square miles (85,900 square kilometres) and a population of 1,495,000 (1975 estimate), excluding the city. As a residue of the city's former status as capital, certain organizations still maintain their national headquarters in Leningrad, among them the all-union geographical, chemical, and medical societies.

Industry. Leningrad is second only to Moscow in the Soviet Union as a centre of industry, and more than half of its working population is employed in factories and the building trade. More than 4,000 enterprises produce about 6 percent of all Soviet industrial production. The entire Soviet Union forms the hinterland for the distribution of Leningrad's manufactures.

Engineering industries

Foremost position is taken by engineering, which accounts for more than half of the city's industrial output. Special emphasis is placed on those branches of engineering that require a skilled labour force and relatively small quantities of metal. Even so, Leningrad uses about a tenth of all steel made in the Soviet Union. Some steel, of specialized high-grade type, is made locally in the suburb of Kolpino, but the bulk of Leningrad's requirements comes from Cherepovets, 400 miles to the east on the Rybinsk Reservoir of the Volga. Leningrad's original industry, shipbuilding, is still of prime importance and the largest of its kind in the Soviet Union; it produces icebreakers (some atomic-powered), tankers, timber carriers, and fishing vessels. Other sectors of heavy engineering make armaments and rolling stock. Of outstanding national significance is the manufacture of electrical and power machinery, such as steam, hydraulic, and gas turbines, armatures, and generators. Precision engineering includes tools and instruments, refrigerators, radios, television apparatus, and other electric and electronic goods. Industrial machinery is produced for automated factories and for the knitwear and footwear industries.

Second to engineering is the group of chemical industries. These make a wide range of items, such as superphosphate fertilizers (using apatite from the Kola Peninsula), tires and other synthetic rubber goods, plastics and plastic goods, artificial and synthetic fibres, paints, and pharmaceutical preparations. Consumer goods industries also comprise a very wide range, with the city itself as their principal market. These include the production of cotton and woollen textiles, clothing, footwear, tobacco products, beer, and foodstuffs. Leningrad is the largest centre of the printing industry in the Soviet Union.

Electrical power for these industries comes from peat-burning generating stations near the city and from hydroelectric plants on the Volkhov, Svir, and Vuoksa rivers. Natural gas is piped to Leningrad from the southern regions of European Russia.

Electricity

Transport. Leningrad has a major function as the largest port of the Soviet Union. The main harbour, in the southwest of the city on the shores of the Gulf of Finland, is protected by breakwaters and is reached by a dredged channel from Kronshtadt. Imports to Leningrad include metal pipes, factory equipment, chemicals, sugar, cotton, and fruit, while machinery, timber, coal, potassium salts, and pyrites form the bulk of exports. Passenger liners maintain regular summer services through the Baltic and to Great Britain. Smaller seagoing ships have access by way of the Neva (all the bridges over which can open) to Lake Ladoga and thence throughout the inland waterway system of European Russia. From Lake Ladoga the Svir River, Lake Onega, and the White Sea Canal (Belomor Kanal) lead to the White Sea and thence by the Northern Sea Route all along the Arctic coast of the Soviet Union. From Lake Onega the Volga–Baltic Waterway system leads to the Volga Basin and Caspian Sea and, via the Volga–Don Canal, to the Black and Azov seas.

The city is a focus of rail routes, with 10 trunk lines radiating out to Helsinki and Warsaw as well as to Moscow and other major cities of the Soviet Union. There are five principal passenger rail termini—the Moscow, Vitebsk, Warsaw, Baltic, and Finland stations. A network of suburban electric services connects the outer parts of Leningrad and its satellite towns. Internal city traffic is carried by a subway system (begun in 1946 and opened in 1955) and a well-developed network of bus, streetcar, and trolleybus lines. Altogether some 6,000,000 people are transported daily, two-thuds of them on the buses and streetcars.

HEALTH. EDUCATION. AND CULTURE

Leningrad is fully equipped with the health services of a modern city. There are some 500 polyclinics providing medical and dental care and maternity and nursing services for the residential "micro-regions" and suburbs. The higher level of medical care is provided by 154 general and specialized hospitals.

The city is one of the most important centres for education and scientific research in the Soviet Union; no less than 15 percent of the total employed population is engaged in education, the arts, and the sciences. Heading the educational establishments is the Leningrad A.A. Zhdanov State University, founded in 1819; its main building is Trezzini's Twelve Colleges on the Bolshaya Neva opposite the Admiralty. No less renowned and even older are the Academy of Fine Arts, founded in 1757, the

Institute of Mines (1773), and the Military Medical Academy (1798). Altogether Leningrad has some 40 higher educational establishments, taking over a quarter of a million students from all parts of the Soviet Union. There are about 740 general schools and a further 86 specialist and technical secondary schools.

Library of the Academy of Sciences

A focus for research is the library of the Academy of Sciences of the U.S.S.R., which remained in Leningrad when the academy's headquarters moved back to Moscow after the Revolution. More than 40,000 scientific research workers are employed in approximately 400 research establishments. These include the Botanical, Forestry, and Zoological institutes of the Academy of Sciences, the Pulkovo Observatory, and the Geological Institute. The city is the principal Soviet centre for Arctic research, notably at the Arctic and Antarctic Research Institute and the Institute for the Study of Permafrost. In the 1970s plans were going forward for an academic satellite suburb for Leningrad, on the lines of that outside Novosibirsk. Eventually the university and other research and educational establishments will be located there.

There is a large number of libraries in the city, headed by the Saltykov-Shchedrin Public Library on the Nevsky Prospekt, established in 1795; its 14,000,000 volumes and many valuable manuscripts make it second only to Moscow's Lenin Library in the Soviet Union. Another important specialized collection is in the Pushkin House literary museum on Vasilyevsky Island. Leningrad has 15 theatres. The Kirov State Academic Theatre of Opera and Ballet, formerly the Mariinsky Theatre, has long enjoyed an international reputation, and its resident company is frequently on tour abroad. Other important theatres are the Maly (Little), Gorky, Pushkin, and Musical Comedy theatres. The largest of several concert halls is the October Great Concert Hall, which seats 4,000 people. Apart from the Hermitage, of world rank, there are 47 museums, notably the State Russian Museum founded in 1895, which specializes in Russian painting. Many books and journals and 125 newspapers are published in the city. There are three radio stations and a large television centre, transmitting on three channels. Leningrad is a significant centre of the motion-picture industry, with three studios. Provision of sports facilities is good; the Kirov Stadium, largest in the city, has seating for 80,000. Other forms of open-air recreation are catered for by the large zoo, the oldest botanical gardens in the Soviet Union, and numerous parks and gardens.

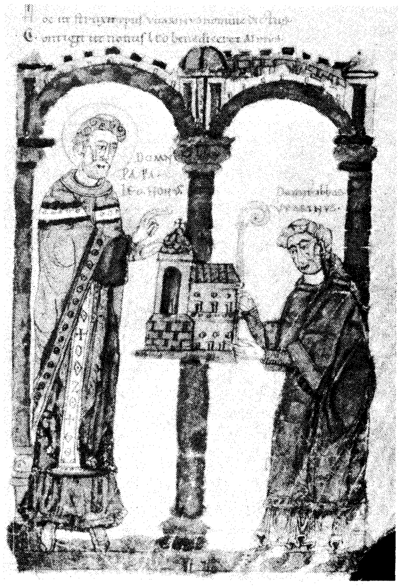
Museums

(Y.M.D./R.A.F.)

BIBLIOGRAPHY. Л.С. ШАУМЯАН (ed.), *Ленинград; энциклопедический справочник* (1957), a survey of the history, economics, public education, cultural and educational institutions, architecture and construction, public health service, science, literature, and art; С.М. СЕРПОКРЫЛ (comp.), *Достопримечательности Ленинграда* (1967), an account of the basic architectural aggregations, squares, embankments, avenues, and suburban parks; М.П. ВЯТКИН (ed.), *Очерки истории Ленинграда*, 5 vol. (1955–67), an analysis of the development of the city and its construction with an account of the stages of its revolutionary development; В.А. КАМЕНСКИЙ, *Город смотрит в завтра: Генеральный план разеитип Ленинграда* (1968), on development prospects; В.Е. МУШТУКОВ, *Ленинград: Путеводитель* (1970), descriptions of the architectural buildings and monuments of the city; М.Д. ФИЛОНОВ and И.Д. КОЗЛОВ (eds.), *Ленинград аа 50 nem; статистический сборник* (1967), data concerning the development of the whole city following the October Revolution; Е.Я. ЗАЗЕРСКИЙ and X.X. ICAPHMOB, *Наш город [Ленинград. Справочник]* (1968), economic and geographical characteristics of the city and its structure.

General descriptive works in English are: NIGEL GOSLING, *Leningrad: History, Art, Architecture* (1965); KAREL NEUBERT, *Portrait of Leningrad* (1966); PAVEL KANN, *Leningrad* (1962), a guide; and A. KAGANOVICH, *Splendours of Leningrad* (1969), the last two translated from the Russian. On politics and government, see DAVID T. CATTELL, *Leningrad: A Case Study of Soviet Urban Government* (1968). On the siege of Leningrad, see HARRISON SALISBURY, *The 900 Days* (1969).

Leo IX, Saint and Pope



Leo IX (left) blesses Abbot Warinus in Metz; codex, 11th century. In the Burgerbibliothek, Bern, Switzerland (Cod. 292, f. 73).

By courtesy of the Burgerbibliothek, Bern. Switzerland

Leo IX, pope from 1049 to 1054, inaugurated the period of the great 11th-century reform of the Roman Catholic Church and began to make the papacy a focal point of western Europe. During his pontificate occurred the schism between the Eastern and Western churches (1054).

Bruno of Egisheim (in Alsace) was born into an aristocratic family on June 21, 1002. He was educated at Toul, where he first became canon and then was consecrated bishop on September 9, 1027, at the early age of 25. Dynamic, purposeful, and zealous in the cause of reform, he began to raise the moral standards of important monasteries in his diocese, as well as those of the secular diocesan clergy, by holding frequent meetings and by constant exhortations. In accordance with prevailing practice, he was appointed pope at the age of 47 by the emperor Henry III. He insisted, however, upon being elected by the people and clergy of Rome, an action that implicitly indicated his opposition to the firmly entrenched lay intervention, especially by the emperors, in purely ecclesiastical matters. After having obtained approval by the Romans, he was enthroned as pope on February 12, 1049.

Attack on concubinage, simony, and lay investiture

Leo IX's aim was the eradication of what he saw as the chief evils of the time—that is, concubinage (clerical marriage), simony (buying and selling of ecclesiastical offices), and lay investiture (conferment of an ecclesiastical office by a lay ruler). In order to achieve these ends it was necessary for the Roman Church itself to be made the centre of Christian society and life. Leo therefore called to Rome men whom he had known in his capacity as bishop of Toul. They not only were aware of the pressing need for reform but were also first-class scholars and administrators as well as men who realized the difficulties with which they were to be confronted. Among them were Humbert of Moyenmoutier, Frederick of Lorraine (later Pope Stephen IX), and Hugh of Remiremont, all of whom became cardinals. A notable monk at Cluny, Hildebrand, also obeyed the call to Rome, where he was destined to play a historic role as Pope Gregory VII, becoming the consummator of the reform initiated by Leo. These men and their assistants infused new blood into the Roman Church. Leo also entertained regular contact with other leading churchmen, such as Peter Damian and Abbot Hugh of Cluny, who by virtue of their reputations exercised great influence upon their immediate surroundings and thus prepared the way for the acceptance of measures to reform Christian society.

These men succeeded in transforming the papacy from a local Roman institution into an international power. This farsighted and able group was determined to make papal ideology a social reality. The pivotal point in this ideology was the primatial position of the pope as so-called successor of St. Peter—an ecclesiastical expression for papal monarchy. The organizational apparatus of the papacy experienced a great expansion at this time, notably the chancery, which became the nerve centre of the papacy in which the universally valid and applicable law and the instructions and mandates to distant ecclesiastical officers were drafted.

Although the immediate effect of these legal measures was not immediately conspicuous, they nevertheless prepared the ground for their subsequent successful implementation. During Leo IX's pontificate the cardinals became more and more prominent as the most intimate counsellors of the pope, and within a few years they were to form the body known as the Sacred College of Cardinals.

The validity of priestly ordinations administered by simoniac bishops proved a serious problem, because most theologians held that simony prostituted the sacrament of ordination. Leo IX ordered a number of simoniacally ordained priests to be "reordained." This order called forth a great spate of controversial literature but the problem was not solved until several decades later. A synod that was under Leo's presidency condemned as heretical in 1050 the views propounded by Berengar of Tours (died 1088) on the Eucharist (that the bread and wine only symbolically became the body and blood of Christ).

Leo IX was intent on making the primatial position of the pope real by his own physical presence outside Rome. To this end he held more than a dozen councils in Italy, France, Germany, and Sicily, which re-enacted the decrees of earlier councils and popes and initiated practical measures to eliminate the worst excesses from which Christian society suffered. The personal attendance of St. Peter's successor and his chairmanship of these councils were factors that powerfully contributed to the accelerated ascendency of the papacy. The frequent journeyings enabled the pope to establish direct contact with the higher and lower clergy as well as with leading secular personages.

The most significant event of Leo IX's pontificate—the actual break with the Eastern Church—resulted, at least partially, from an ill-fated military involvement.

After their settlement in Sicily in the second decade of the 11th century, the Normans presented considerable dangers to the existence of the papal state. In their marauding expeditions they plundered and devastated many churches and monasteries. In conjunction with Emperor Henry III, Leo resolved to undertake a military campaign against the Normans; but Henry withdrew and, with a weak and inexperienced army under his command, Leo had to face the Normans alone. They inflicted a crushing defeat upon the papal army, and on June 18, 1053, they took the Pope prisoner. He was nevertheless allowed to maintain contact with the outside world and to receive visitors. After nine months he was released.

The schism of 1054

The Norman venture, however, brought the papacy into conflict with the Eastern Church centred in Constantinople, which, since the 8th century, had exercised jurisdiction over large areas of southern Italy and Sicily. The forcefully enunciated papal theme of primacy in Leo's pontificate complicated the relations between Rome and Constantinople still further because the patriarch of Constantinople, Michael I Cerularius, considered this sheer provocation.

He closed the Latin (Western) churches in Constantinople and raised serious dogmatic charges against the Roman Church, notably in connection with the Eucharist. Cardinal Humbert attacked the Patriarch in a vitriolic and passionate manner by arguing the case for Roman primacy and also quoting extensively from the forged "Donation of Constantine" (allegedly bestowing sovereignty in the West on the papacy). A legation-under Humbert's leadership left for Constantinople in April 1054 to reach a settlement, but despite several meetings between Patriarch, Emperor, and legates, no concrete results emerged. On July 16, 1054, in the full view of the congregation, Humbert put the papal bull of excommunication—already prepared before the legation left Rome—on the altar of the church of Hagia Sophia in Constantinople. Thereupon the Patriarch excommunicated the legation and its supporters. This marked the final breach between Rome and Constantinople. This schism was to last, with short interruptions, until the modern age. Whether the excommunication of Michael I Cerularius was valid, because Leo had been dead for three months (he died suddenly on April 19, 1054), is merely a technical problem. The Roman legates were legates of the papacy and the bull of excommunication had been a measure of the reigning pontiff. In any case, the excommunication merely formalized in a dramatic and spectacular manner a state of affairs that had long existed. Although this occurred after the death of Leo IX, the outbreak of the formal schism correctly belongs to his pontificate, which in several ways therefore marked a caesura in the history of the papacy in medieval times.

BIBLIOGRAPHY. LUCIEN SITTLER and PAUL STINTZI (eds.), ***Saint Léon IX, le pape alsacien*** (1950), a competent and readable account; HORACE K. MANN, ***The Lives of the Popes in the Early Middle Ages***, vol. 6, ***The Popes of the Gregorian Renaissance***, pp. 19–182 (1910), the fullest account in English, but rather uncritical and in many respects outdated, though largely based on original evidence; ***Pour le neuvième cente'naire de la mort de saint Ldon IX*** (1954), a memorial volume commemorating the ninth centenary of Leo's death and concerning a number of aspects of his pontificate; OWEN J. BLUM in the ***New Catholic Encyclopedia***, vol. 8, pp. 642–643 (1967); V. GRUMEL, "Les Préliminaires du schisme de Michel Cérulaire ou la question romaine avant 1054," ***Revue des études byzantines***, 10:5–23 (1952), an examination of the historical background of the schism; STEVEN RUNCIMAN, ***The Eastern Schism: A Study of the Papacy and the Eastern Churches During the XIth and XIIth Centuries*** (1955), a general outline relative to the outbreak of the schism; DONALD M. NICOL, "Byzantium and the Papacy in the 11th century," ***Journal of Ecclesiastical History***, 13:1–20 (1962), a very competent survey of the subject; F. DVORNIK, ***Byzance et la primaute' romaine*** (1964; Eng. trans., ***Byzantium and the Roman Primacy***, 1966), a fresh and original assessment of the primatial question based on the author's own research.

(W.U.)

Leo X, Pope

Pope Leo X, one of the most extravagant of the Renaissance popes, presided over the dissolution of the unified Western Church by refusing to accept with due seriousness the questions and actions of the founder of the Protestant Reformation, Martin Luther (1483–1546). By his lavish expenditures in making Rome a centre of European culture, he depleted the papal treasury that had been built up by his predecessor, Julius II (1445–1513), but by his political policies he made the papacy a significant power in Europe in the 16th century.

Born Giovanni de' Medici on December 11, 1475, in the city of Florence, he was the second son of Lorenzo the Magnificent, the ruler of the Florentine republic, and by

Leo X, contemporary medallion. In the coin collection of the Vatican Library.

Leonard von Matt—EB Inc.

custom thus destined for a religious life. At the early age of eight young Giovanni received the tonsure—a ceremony involving the cutting of hair from the head, thus indicating the change of status from lay to clerical—and five years later became the cardinal-deacon of Sta. Maria in Dominica. He received the finest education available in Europe at the court of his father under several tutors, including the philosopher Pico della Mirandola (1463–94). From 1489 to 1491 the youthful ecclesiastic studied theology and canon law at the University of Pisa. In 1492 he became a member of the Sacred College of Cardinals and attempted to take up residence in Rome. The death of his father later in the same year, however, brought him back to Florence where he lived with his older brother Piero.

The election of Pope Alexander VI took him back to Rome for the conclave (assembly of cardinals to elect the pope); otherwise he lived in Florence until he was exiled in November 1494 with the other members of the Medici family on the charge of their betraying the republic. For the next six years Cardinal de' Medici travelled throughout northern Europe. In 1500 he returned to Italy and settled in Rome. Upon the death of his brother Piero, Giovanni became the head of the Medici family. During 1503 he took part in the conclaves that elected first Pius III (September) and then Julius II (October). Named papal legate to Bologna and Romagna in 1511, he supervised the re-establishment of Medici control of Florence the following year; and although his younger brother Giuliano actually held the first place in the Florentine Republic, it was the Cardinal who ruled.

Election to the papacy

With the death of Julius II on February 21, 1513, the Sacred College of Cardinals was summoned to elect a successor. The conclave met on March 4, and with minimal deliberation the cardinals, who desired a peace-loving successor to the warlike Julius, elected Cardinal de' Medici on March 11. Taking the title of Leo X the pontiff-elect was ordained a priest on March 15 and consecrated bishop of Rome on the 17th. Two days later the papal coronation took place.

At 37 years of age the new Pope was the personification of the Renaissance ideals. Having spent his youth at the court of Lorenzo de' Medici he had acquired the mannerisms and tastes of one of the most brilliant societies of Europe and posed a sharp contrast to the soldier-pope whom he succeeded. He fit extremely well into the atmosphere of calm and quiet of which Rome was desirous after ten years of Julius II. Leo was lavish in his spending not only of the church's money but also of his own. Under his patronage Rome again became the cultural centre of the Western world. The construction of St. Peter's Basilica—initiated under Julius II—was accelerated, the holdings of the Vatican Library were greatly increased, and the arts flourished. Even the piety of the papacy was revised to some extent after the low reputation it had reached under the Borgia popes (Calixtus III and Alexander VI).

The fifth Lateran Council occupied the new Pope during the first five years of his pontificate. Called by Julius II two years before his death, the council was designed to nullify the efforts of nine rebellious cardinals who had called for a council to meet at Pisa in order to revive the conciliar movement—intending to place ecclesiastical authority in councils, rather than the papacy—of the previous century. Although "Pisa II" collapsed when first the German emperor (Maximilian I, 1459–1519) and then the French king (Louis XII, 1462–1515) withdrew their support, the Lateran Council opened in 1512. Leo X, who inherited the council before it was a year old, was little inclined to preside over the sweeping reforms that the church so desperately needed on the eve of the Protestant Reformation. Poorly attended, and dominated by Italian bishops, the council debated the principal issues of the day; but there was neither direction nor encouragement from the pontiff, nor the urgency and necessity that would spur on the Council of Trent some 40 years later. Thus the council was dissolved (March 16, 1517), without significant action, on the very eve of Martin Luther's posting of his 95 theses.

Political, economic, and religious issues

Leo X was not only the head of the Christian Church, he was the temporal ruler of the Papal States and head of the Medici family that ruled the Florentine Republic. To exert his influence in Italy he resorted to the common practice of nepotism (granting offices or benefits to relatives, regardless of merit). To the influential archbishopric of Florence he appointed his cousin, and future pope (Clement VII), Giulio de' Medici. He also named his younger brother Giuliano and his nephew Lorenzo to be Roman patricians. Giuliano's premature death in 1516 brought an end to the Pope's plan to create a central Italian kingdom for him. On July 1, 1517, following and as a result of an attempt upon his life earlier in the same year, Leo named 31 new cardinals, thus prejudicing the College of Cardinals with papal supporters. One cardinal (Alfonso Petrucci) was strangled in prison, and others were imprisoned and executed when they were implicated in the attempted assassination.

In his struggle to dominate Italy, Leo X was confronted by the awesome power of Spain and the determination of the French kings. Louis XII of France marched into Italy in 1513 to make good his claims to Milan and Naples. Reluctantly Leo formed the League of Mechlin, in which Spain provided the major military strength. The French were defeated at Novara, and Louis renounced his claims and withdrew his army. The peace was short-lived. The ascent of Francis I (1494–1547) in 1515 to the throne of France led to the renewal of the war. Although Leo renewed the coalition of Spain, the Empire, and England, Francis won the Battle of Marignano (September 14, 1515). The Pope made peace with the French king and then followed it up with the Concordat of Bologna. Promulgated in the form of a papal bull (Primitiva) on August 18, 1516, the concordat regulated church-state relations in France for the next 275 years. The French kings were given the power to nominate bishops, abbots, and priors, though the popes did retain the right to nominate candidates to fill vacant benefices in curia and certain other benefices. Though the pope always had the power to veto a nomination of the king's, in practice the lay monarch's choice was tantamount to an appointment. This control over the church in France on the part of the kings explains, in part, why the monarchy showed little interest in Protestantism during the 16th century.

The death of the Holy Roman emperor, Maximilian I, in 1519 brought Leo further into the political arena. The Habsburg candidate, Charles (1500–58) of Spain, had succeeded his maternal grandparents Ferdinand II of Aragon (1452–1516) and Isabella I of Castile (1451–1504) in 1516 and now sought to follow his paternal grandfather, Maximilian, to the powerful German throne. Both Francis I and Frederick the Wise (1463–1525) of Saxony, however, immediately put forward their candidacy. Leo—fearing that if the empire were joined to either France or Spain, Italy would come under the power of the victor—threw his support in favour of Frederick. The election of Charles I of Spain as Charles V of the Holy Roman Empire led to war between France and Spain, and although Leo would have preferred to remain neutral he cast his lot with the new emperor when Francis again invaded Italy.

The ever-pressing financial undertakings of the papacy kept Leo X in constant need of new means of raising revenue. The wars with France, his lavish support of the arts, the construction of St. Peter's, and a projected crusade against the Turks all contributed to the financial needs of the papacy. One important source of revenue had long been the dispensing of indulgences (remission of the temporal penalty for sins) for money. During the reign of Julius II, indulgences had been authorized for financial contributions for the construction of the great basilica in Rome, St. Peter's. Leo, who was very much interested in continuing this work, reaffirmed the indulgence shortly after his ascent. Nevertheless, because of its unpopularity in northern Europe, based primarily on economic reasons, it was not until early in 1517 that a Dominican friar actually began to preach the indulgence in the archdioceses of Mainz and Magdeburg (Germany). In response to this preaching, Martin Luther posted his 95 theses on the door of the Wittenberg Castle Church.

By the following year (1518) Luther's ideas had reached Rome, and Leo ordered the head of the Augustinian order, of which Luther was a member, to discipline the troublesome monk. When this failed the Pope tried to work through Frederick of Saxony, but again to no avail. During the course of the Leipzig debates in the summer of 1519, the Catholic theologian Johann Eck (1486–1543) manoeuvred Luther into publicly professing heresy on three accounts. On June 15, 1520, Leo issued a document condemning Luther of heresy on 41 accounts and ordered him to submit to the authority of Rome within 60 days or suffer excommunication. Luther, who by this time had gained the support of influential figures in Germany, openly defied the Pope. Thus, Leo was left no alternative but to issue a papal bull (Decet Romanum *Pontificem*) of excommunication on January 3, 1521.

Leo X had not taken the Lutheran movement with the seriousness that history later indicated was warranted. But the church, after all, had withstood the teachings of an English reformer, John Wycliffe (c. 1330–84), a Bohemian reformer, Jan Hus (c. 1372/73–1415), and was at the time weathering those of the Renaissance humanist Erasmus (c. 1466–1536). Leo viewed Luther as merely another heretic whose teachings would lead some of the faithful astray but that, as had happened in the past, the true religion would triumph in time. Furthermore, the political climate of Germany was not favourable for strong action against Luther and his supporters. Thus, the new movement spread rapidly in the north with relatively little interference from the papacy. On December 1, 1521, Leo X died suddenly, leaving behind him political turmoil in Italy and religious turmoil spreading across northern Europe.

BIBLIOGRAPHY. WILLIAM ROSCOE, ***The Life and Pontificate of Leo the Tenth,*** 2 vol. (1853), the standard biography of Leo X (well-documented and contains a good index); LUDWIG VON PASTOR, ***The History of the Popes, from the Close of the Middle Ages,*** 3rd ed., vol. 7–8 (1950), the best source on the life and times of Leo X (well-documented and well-indexed); E.P. RODOCANACHI, ***Le Pontificat de Léon X, 1513–1521*** (1930), contains an excellent bibliography; G.K. BROWN, ***Italy and the Reformation to 1550*** (1933), a study of Pope Leo as a principal figure of the Reformation.

(J.G.G.)

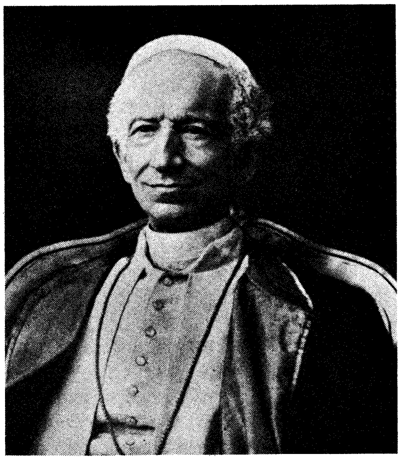
Leo XIII, Pope

Leo XIII, elected pope when he was almost 68 years old, was head of the Roman Catholic Church from 1878 to 1903. Although he kept to many of the paths trodden by his predecessors, he also brought a new spirit to the papacy, manifested in more conciliatory positions toward civil governments, by care taken that the Roman Catholic Church not be opposed to the needs of the times, and by a new conception of how the church should attempt to exercise its influence in society.

Early life and career

Vincenzo Gioacchino Pecci, born March 2, 1810, in Carpineto Romano, in central Italy, was the sixth child of a family of the lower nobility. After his early education in Viterbo and Rome, he completed his studies at the Accademia dei Nobili Ecclesiastici (Academy of Noble Ecclesiastics) in Rome. In 1837 he was ordained a priest and entered the diplomatic service of the Papal States. His superiors immediately appreciated his qualities: flexibility and lucidity and great energy, despite his delicate health. Thus promotions came quickly; he was made delegate (the equivalent of provincial governor) of Benevento in 1838 and was transferred in 1841 to the more important delegation of Perugia. In January 1843 he was appointed nuncio (a papal legate of the highest rank, permanently attached to a civil government) to Brussels and shortly after was consecrated an archbishop.

The Bettmann Archive



Leo XIII, 1878.

Pecci's stay in Belgium, lasting only three years, was an important stage in the life of the future pope. He discovered how Catholics in a modern constitutional government could profit from the parliamentary system and from freedom of the press. But the Belgian nunciature halted the young prelate's career, which had begun so auspiciously. Pecci showed initiative and independence in several delicate situations, but he was severely criticized at the time; and King Leopold I, considering him less docile than his predecessor, soon demanded his recall.

He was then named, early in 1846, bishop of Perugia, a small diocese to which he was confined for 32 years, despite his having been made a cardinal in 1853. He suffered from this obscurity and made many attempts to win Rome's favour, but in vain: his harsh judgment of the opposition in the Papal States to the Roman Revolution of 1848, and his concern to avoid useless conflicts with the Italian authorities after the annexation of Umbria in 1860, made Rome suspect him––quite wrongly— of liberal sympathies and of tepidity with respect to temporal powers.

A weaker personality would undoubtedly have been dulled and embittered by this prolonged period of disfavour, but for Pecci these years of retreat were extremely fertile. He zealously applied himself to the systematic reorganization of his diocese and to the spiritual and intellectual improvement of his clergy. He also had available a great deal of leisure time in which to read and meditate. He occupied himself with the renewal of Christian philosophy and studied particularly the writings of St. Thomas Aquinas, the 13th-century Scholastic philosopher, to whom he had been introduced by his brother Giuseppe, a Jesuit seminary professor. He was also led to reconsider the problem of the relations between the church and modern society and became increasingly convinced of the mistake committed by ecclesiastical authorities in taking a fearful, negative attitude toward the aspirations of the times. The fruits of this silent maturation were revealed to his surprised contemporaries in his pastoral letters of 1877 and 1878, which attracted attention even beyond Italy's borders. He also received notice when, in 1877, he was named camerlengo, the office of chief administrator of the church in the event that the pope dies.

Election to the papacy

At the death of Pius IX in February 1878, his name was mentioned frequently among those of the principal *papabili,* those considered possible successors to the papacy. His candidacy was strongly supported by most of the non-Italian cardinals, who were impressed by the self-control and energy with which he acquitted his duties as camerlengo and who noted that one who had been away for so long from Rome would be less compromised by the decisions of the preceding pope. Cardinal Pecci was elected on February 20, 1878, on the third ballot. He announced that he would take the name Leo in memory of Leo XII, whom he had always admired for his interest in education, for his conciliatory attitude toward temporal governments, and for his desire to create links with Christians who had separated from the Roman Catholic Church. The age of the new pope and his delicate health caused speculation that his pontificate would be brief. But, in fact, he directed the church for a quarter of a century.

The pontificate of Leo XIII's predecessor, Pius IX, had been long and controversial. From shortly after the beginning of his reign, Pius IX had been a strong, conservative authoritarian, both in his governing of the church and in his opposition to the new Italian government that annexed the Papal States. Although the pontificate of Leo XIII had a new spirit, the new pope was as intractable as his predecessor on the principle of the temporal sovereignty of the pope and continued to consider the traditional doctrine of the Christian state as an ideal. He reacted as strongly as had Pius IX against Freemasonry (a secret society that both popes viewed as opposed to Christianity) and secular liberalism. In church administration he continued to accentuate the centralization of authority in the papacy rather than in the national churches and reinforced the power of the nuncios. In addition, Leo XIII followed Pius IX in encouraging the devotion to the Sacred Heart of Jesus and to Mary. He renewed the condemnations of Rationalism—the theory that reason is the primary source of knowledge and of spiritual truth—and pursued with fresh vigour the re-establishment of the philosophy of St. Thomas Aquinas.

In other respects, however, there is no doubt that Leo XIII's pontificate was characterized by a new spirit. In his relations with civil governments, Leo XIII avoided inflamed protests and showed his preference for diplomacy. He achieved many incontestable successes through diplomacy, although his ability in this area was definitely less than is customarily asserted. The true greatness of Leo XIII was precisely that, in spite of his taste for politics, he was not exclusively a political pope. He was also an intellectual sympathetic to scientific progress and to the need for the Roman Catholic Church to demonstrate itself open to such progress, and he always remained a pastor who was concerned for the church's internal life and for the spreading of its message throughout the world.

This concern toward renewing the dialogue between the church and the world was manifested especially in his many encyclical letters giving instructions to Catholics throughout the world. In 1893 the encyclical *Providentissimus Deus* ("The Most Provident God"), now outdated but originally a pioneering work, defined in fairly broad-minded manner the principles on which Catholics should interpret the Bible. In several instructions he recommended that church and state live together in peace within the framework of modern society. The encyclical *Rerum Novarum* ("Of New Things") in 1891, though rather cautious in its approach, showed that the papacy had taken cognizance of the problems of the working class. He attempted to support the organization of the Catholic laity and was concerned about renewed dialogue with non-Catholics, as is demonstrated in the interest he showed in the attempt to create a link between the Anglican Church and Rome and in his respect for the traditions of the Eastern churches.

During the last years of Leo XIII's pontificate, until his death on July 20, 1903, there was a hardening of church policy and a more reserved attitude toward Christian democracy. Nevertheless, Leo XIII succeeded in gaining great prestige for the papacy, as was shown by the increase in countries having diplomatic relations with the Vatican, even non-Christian countries. He was a man gifted with a superior intelligence, an energetic temperament, a keen awareness of his personal worth, and a discriminating sense for public relations. Although his pontificate did not bring about many immediate changes in the relationship of the Roman Catholic Church to society, it did initiate many new attitudes that began to mature in succeeding decades.

BIBLIOGRAPHY. There is no biography that is completely scientific and objective. C. DE T'SERCLAES, *Le pape Léon XIII*, 3 vol. (1894–1906), gives the official Vatican viewpoint (the first two volumes were reviewed by the Pope himself); M. SPAHN, *Leo XIII* (1905), offers critical evaluations (in German) from the point of view of German "Reformkatholizismus" ("Critical Catholicism"); a French point of view is found in F. HAYWARD, *Leon XIII* (1937). The essay by R. AUBERT, "Léon XIII," in *I cattolici italiani dall'800 ad oggi*, pp. 189–220 (1964), utilizes the most recent works. On the various aspects of the work of Leo XIII as pope, see E.T. GARGAN (ed.), *Leo XIII and the Modern World* (1961); and E. SODERINI, *Il pontificato di Leone XIII*, 3 vol. (1932–33).

(R.F.M.A.)

Leo III the Isaurian

Founder of the so-called Isaurian, or, more properly, Syrian, dynasty, Leo III was an important Byzantine emperor because of his success in halting Arab expansionism and because of his inauguration of Iconoclasm (destruction of religious images) as the state religious policy.

Leo III the Isaurian, gold solidus, 8th century. In the British Museum.
Peter Clayton

Military accomplishments. Born about 675–680 at Germanicea (Mar'ash) in northern Syria (modern Maraş, Turkey), as a youth he was taken by his parents (who apparently were prosperous) to Mesembria, in Thrace. Emperor Justinian II appointed the young man to the prestigious rank of *spatharius* (attendant) as a reward for assisting him in the recovery of his throne in 705. But the Emperor soon developed a distrust of him and therefore sent him to perform a perilous mission among the Alani on the remote eastern frontier, anticipating that he would never return. Despite the danger, Leo accomplished his assignment, managed to preserve his life, and ultimately, at the hands of a subsequent emperor, Anastasius II (713–715), received appointment as commander of the Anatolikon, the largest theme, or military-district army, in Asia Minor. As the result of a military revolt in 715, Anastasius was deposed, exiled to a monastery, and replaced by Theodosius III. Leo, in alliance with Artavasdos, the commander of the Armeniakon theme (the second largest in Asia Minor), refused to recognize the new emperor and continued to champion the cause of Anastasius. Meanwhile, Arab armies had invaded Asia Minor. Leo deceived them into believing that he would subjugate the empire for them, and thus he won their goodwill and support for his own attempt to seize the throne. After persuading them to spare Byzantine territory, he marched on Constantinople. The feeble Theodosius III abdicated, realizing that opposition was futile. Leo became emperor on March 25,717.

Defense against the Arabs

Leo's first task as emperor was the organization of the defense of Constantinople against the Arab troops under Maslamah ibn 'Abd al-Malik, who angrily perceived the deception. They besieged the city by land and sea from August 15, 717, to August 15, 718. Leo's skillful defense, which was aided by Greek fire (an igneous petroleum mix), a severe winter, desertions from the Arab fleet, and a Bulgarian assault upon those Arabs who had encamped in Thrace, compelled Maslamah to abandon the siege, which was the second and supreme Arab effort to capture Constantinople. Leo's victory marked an important check to Arab expansionism, preventing their establishment of a bridgehead in southeastern Europe. His complicated negotiations with the Arabs, as well as those earlier with the Alani, brought him a contemporary reputation for cleverness.

Leo consolidated his authority by crushing a rebellion in Sicily and a plot of army officers and officials to restore former emperor Anastasius II to the throne. Leo then sealed an alliance with his associate Artavasdos by marrying his daughter Anna to him. Throughout the reign, Artavasdos remained the second most powerful man in the empire by virtue of his control of several important military posts. Leo's wife, Maria, bore him a son, Constantine, whom he crowned in 720. An able diplomat, Leo married Constantine in 733 to a daughter of the Khagan of the Khazars; the marriage brought Leo a valuable military alliance with the Khazars in the trans-Caucasus against the Arabs.

Leo maintained peaceful relations with the Bulgarians to the north, enabling him to concentrate his military abilities against the Arab menace to Asia Minor. In 740 he won a major victory over the Arabs at Akroïnon (Afyonkarahisar). This victory freed Asia Minor from any immediate serious threat of Arab conquest, and it made possible the forceful counteroffensive and reconquest of some lost territory in the subsequent reign of his son Constantine V (741–775). He also repaired the extensive walls of Constantinople.

An energetic soldier-emperor, who personally led his troops in battle, Leo displayed great concern for the efficiency of his army. His victories improved army morale. He subdivided some of the very large military district armies, or themes, which had proved, as his own career demonstrated, so large and powerful that their commanders might attempt to seize the throne. Whether or not this administrative reorganization was itself responsible for subsequent stability, he did succeed in halting the vulnerability of the throne to military overthrow. He unquestionably left the army a more effective instrument at his death than he had found it on his accession. It no longer seems probable, however, that he planned or established a comprehensive system of social and institutional reforms.

Promotion of Iconoclasm

Religious policies. Leo's military achievements earned him great popularity with his soldiers and the people and may have given him the confidence to pursue his religious policies forcefully. He not only held firm religious opinions but he also had a profound belief in his duty as emperor to implement them as he understood them. In 722 he ordered the forcible baptism of Jews and Montanists (a Christian heretical group). He personally investigated but did not prosecute adherents of the Paulician heresy. The origins and nature of his policy of Iconoclasm, the most singular religious development in his reign, are obscure and controversial. He was deeply religious and seems to have become genuinely convinced of the sacrilegious character of religious pictures and relics as objects of veneration in worship services. It is uncertain whether any boyhood experiences in northern Syria, including contact with Muslims, influenced his Iconoclastic views, as his critics often charged. The lconoclastic opinions of certain bishops in western Asia Minor did, however, have some effect upon him. Thus, in 726 he began to speak out publicly against the use of sacred pictures. Opposition to his doctrines may have been the cause for an unsuccessful rebellion against him in the Cyclades Islands in 727. In 730 he proclaimed Iconoclasm the official policy of the empire and ordered the removal and destruction of sacred pictures in churches. When Patriarch Germanus I of Constantinople refused his demand for approval of these policies, Leo removed him and appointed a patriarch of his own choice, Anastasius. Where necessary, Leo employed harsh penalties, such as beatings and imprisonment, against recalcitrant ecclesiastics. His policies met particularly strong opposition from monastic circles. Popes Gregory II and Gregory III also strongly rejected his efforts to impose Iconoclasm upon Byzantine-controlled areas of Italy. Leo retaliated by halting financial contributions to the papacy from southern Italy, and he may also have removed the churches of Sicily, Calabria, and Illyria from papal jurisdiction and placed them under the patriarch of Constantinople. At any rate, his actions severely strained relations with the papacy, causing the popes to turn increasingly to the Frankish kings as alternative protectors of the Holy See in Rome and weakening the Byzantine position in the Italian peninsula. Other harsh taxation and administrative measures added to his unpopularity in Sicily and southern Italy. Although an able commander, Leo neglected to maintain strong naval forces in the western Mediterranean and thus further weakened Byzantine power there.

Revision of Roman law

Legal and other accomplishments. One of Leo's most important acts was the promulgation, in 726, of the Ecloga, a law code of modest length, which represented a revision of Roman legal practices as embodied in the 6th-century Corpus Juris Civilis (Body of Civil Law) of Emperor Justinian I. Consciously attempting to revise Roman law in accordance with Christian principles, Leo devoted much space in the Ecloga to the regulation of marriage and property rights. Amputation and mutilation often were substitutes in this new code for the former death penalties. Leo provided regular salaries for legal officials to discourage the corrupt custom of offering gifts or bribes to judges and bureaucrats. An important codification of military law, the so-called Soldiers' Law, is sometimes attributed to Leo, but its true ascription is uncertain.

Other than his sincere predilection for theological topics, Leo's intellectual interests are unknown. He possessed, doubtless from boyhood, a speaking knowledge of Arabic. Although there is little evidence of intellectual activity during his reign, the earlier charge that he halted higher education in Constantinople by closing an ecclesiastical academy (because of the faculty opposition to Iconoclasm) can no longer be credited with certainty. Similarly, there is little source material on economic or demographic developments during his reign, but the numerous earthquakes doubtless inflicted major damage on towns and the countryside.

Leo died on June 18,741, and was buried in the Church of the Holy Apostles at Constantinople. There is inadequate information on internal history in the last eight years of his reign, but he certainly failed to silence opposition to his Iconoclastic policies; in fact, Iconoclasm divided the empire for another century. He had instilled his Iconoclastic opinions and his grasp for military tactics in his son Constantine V, who ably followed and even intensified the policies of his father. Although Leo's memory was reviled by those later Byzantines who deplored his Iconoclasm, he was admired, especially in certain military circles, for his forceful and generally successful efforts to strengthen the state.

BIBLIOGRAPHY. KARL SCHENK, *Kaiser Leon III.* (1880), a biography of his early life; ***A Manual of Roman Law: The Ecloga, Published by the Emperors Leo III and Constantine V . . . A.D. 726,*** trans. by EDWIN HANSON FRESHFIELD (1926), contains a useful introduction; J.B. BURY, ***A History of the Later Roman Empire from Arcadius to Irene, 395 A.D. to 800 A.D.,*** 2 vol. (1889), an old but useful narrative; G. OSTROGORSKY, *Geschichte des byzantinischen Staates* (1963; Eng. trans., ***History of the Byzantine State,*** 2nd ed., 1968), including a useful bibliography; R. JENKINS, ***Byzantium: The Imperial Centuries, A.D. 610–1071,*** pp. 61–68, 74–89 (1966), a good narrative; E.J. MARTIN, ***A History of the Zconoclastic Controversy*** (1930), the only study in English.

(W.E.K.)

Leonardo da Vinci

The unique fame that the Florentine artist and scientist Leonardo da Vinci enjoyed in his lifetime and that, filtered and purified by historical criticism, has remained undimmed to the present day is based on the equally unique universality of his spirit. Leonardo's universality is more than manysidedness. True, at the time of the Renaissance and the period of Humanism, manysidedness was a highly esteemed quality; but it was by no means rare. Many other good artists possessed it. Leonardo's universality, on the other hand, was a spiritual force, peculiarly his own, that generated in him an unlimited desire for knowledge and guided his thinking and behaviour. An artist by disposition and endowment, he found that his eyes were his main avenue to knowledge; to Leonardo, sight was man's highest sense organ because sight alone conveyed the facts of experience immediately, correctly, and with certainty. Hence, every phenomenon perceived became an object of knowledge. ***Saper vedere*** ("knowing how to see") became the great theme of his studies of man's works and nature's creations. His creativity reached out into every realm in which graphic representation is used: he was painter, sculptor, architect, and engineer. But he went even beyond that. His superb intellect, his unusual powers of observation, and his mastery of the art of drawing led him to the study of nature itself, which he pursued with method and penetrating logic—and in which his art and his science were equally revealed.

Alinari

Leonardo, self-portrait, chalk drawing. In the Palazzo Reale, Turin, Italy.

LIFE AND WORKS

Early period: Florence. The illegitimate son of Ser Piero, a Florentine notary and landlord, Leonardo was born in 1452 on his father's family estate in Vinci, near Empoli. His mother, Caterina, was a young peasant woman who shortly thereafter married an artisan from that region. Not until his third and fourth marriage did Ser Piero's wives have children, the first one in 1476, when Leonardo was already an adult. Thus, Leonardo grew up in his father's house, where he was treated as a legitimate son and received the usual elementary education of that day: reading, writing, and arithmetic. As for Latin, the key language of traditional learning, Leonardo did not seriously study it until much later, when he acquired a working knowledge of it on his own. Not until he was 30 years old did he apply himself to higher mathematics—advanced geometry and arithmetic—which he studied with diligent tenacity; but here, too, he did not get much beyond the beginning stages.

Leonardo's artistic inclinations must have appeared early. When he was about 15, his father, who enjoyed a high reputation in the Florence community, apprenticed him to Andrea del Verrocchio. In Verrocchio's renowned workshop Leonardo received a many-sided training that included not only painting and sculpture but the technical-mechanical arts as well. He also worked in the next-door workshop of Antonio Pollaiuolo, where he was probably first drawn to the study of anatomy. In 1472 Leonardo was accepted in the painters' guild of Florence but remained five years more in his teacher's workshop. Then he worked independently in Florence until 1481. In the few extant works of this early period one may clearly trace the development of the artist's remarkable talent. Keenness of observation and creative imagination stand out. His early mastery is revealed in an angel and a segment of landscape executed by him in Verrocchio's painting the "Baptism of Christ" (Uffizi, Florence) and in two Annunciations (Uffizi, as well as the Louvre, Paris), both of them done in Verrocchio's workshop, as were the "Madonna and Child," the "Madonna Benois," and the portrait of Ginevra de'Benci, "Portrait of a Young Lady." This mastery reached its peak in two paintings that remained unfinished: "St. Jerome" and a large panel painting of "The Adoration of the Magi." In addition to these few paintings there are a great many superb pen and pencil drawings, in which Leonardo's mastery blazed new trails for this graphic art. Among the drawings are many technical sketches—for example, pumps, military weapons, mechanical apparatus—evidence of Leonardo's interest in and knowledge of technical matters at the outset of his career.

Evidence of early mastery

Unfolding of Leonardo's genius: first Milanese period (1482–99). In 1482 Leonardo entered the service of the Duke of Milan—a surprising step when one realizes that the 30-year-old artist had just received his first substantial commissions from his native city of Florence: the above-mentioned unfinished panel painting of "The Adoration of the Magi" for the monastery of S. Donato a Scopeto (1481) and an altar painting for the St. Bernard Chapel in the Palazzo della Signoria, which was never fulfilled. That he gave up both projects despite the commitments he had undertaken—not even starting on the second named—seems to indicate deeper reasons for his leaving Florence. It may have been that the rather sophisticated spirit of Neoplatonism prevailing in the Florence of the Medici went against the grain of his experience-oriented mind and that the more realistic academic atmosphere of Milan attracted him. Moreover, there was the fascination of Ludovico Sforza's brilliant court and the meaningful projects awaiting him there.

Leonardo spent 17 years in Milan, until Ludovico's fall from power in 1499. He was listed in the register of the royal household as ***pictor et ingeniarius ducalis*** ("painter and engineer of the duke"). Highly esteemed, Leonardo was constantly kept busy as a painter and sculptor and as a designer of court festivals. He was also frequently consulted as a technical adviser in the fields of architecture, fortifications, and military matters, and he served as a hydraulic and mechanical engineer.

In this phase of his life Leonardo's genius unfolded to the full, in all its versatility and creatively powerful artistic and scientific thought, achieving that quality of uniqueness that called forth the awe and astonished admiration of his contemporaries. At the same time, in the boundlessness of the goals he set himself, Leonardo's genius bore the mark of the unattainable so that, if one traces the outlines of his lifework as a whole, one is tempted to call it a grandiose "unfinished symphony."

The six paintings completed in the Milanese years

Painting and ***sculpture.*** As a painter Leonardo completed only six works in the 17 years in Milan: portraits of Cecilia Gallerani ("Lady with an Ermine") and a musician, an altar painting of "The Virgin of the Rocks" (two versions), a monumental wall painting of the "Last Supper" in the refectory of the monastery of Sta. Maria delle Grazie (1495–97), and the decorative ceiling painting of the Sala delle Asse in the Milan Castello Sforzesco (1498). Three other pictures that, according to old sources, Leonardo was commissioned to do have disappeared or were never done: a "Nativity" said to have belonged to Emperor Maximilian; a "Madonna" that Ludovico Sforza announced as a gift to the Hungarian king Matthias Corvinus; and the portrait of one of Ludovico's mistresses, Lucrezia Crivelli.

Also unfinished was a grandiose sculptural project that seems to have been the real reason Leonardo was invited to Milan: a monumental equestrian statue in bronze to be erected in honour of Francesco Sforza, the founder of the Sforza dynasty. Leonardo devoted 12 years—with interruptions—to this task. Many sketches of it exist, the most impressive ones discovered only a few years ago, when Leonardo's notebooks came to light again in Madrid. They reveal the sublimity but also the almost unreal boldness of his conception. In 1493 the clay model of the horse was put on open display on the occasion of the marriage of Emperor Maximilian with Bianca Maria Sforza, and preparations were made to cast the colossal figure, which was to be 16 feet (five metres) high—double the size of Verrocchio's equestrian statue of Bartolomeo Colleoni! But, because of the imminent danger of war, the metal, ready to be poured, was used for cannon instead, and so the project came to a halt. Ludovico's fall in 1499 sealed the fate of this abortive undertaking, which was perhaps the grandest concept of a monument in the 15th century. The ravages of war left the clay model a heap of ruins.

As a master artist Leonardo maintained an extensive workshop in Milan, employing apprentices and students. The role of most of these associates is unclear. Their activity involves the question of Leonardo's so-called apocryphal works, in which the master collaborated with his assistants. Scholars have been unable to agree in their attributions of these works, which include such paintings as "La Belle Ferronnière" in the Louvre, the so-called "Lucrezia Crivelli" in the Pinacoteca Ambrosiana, Milan, and the "Madonna Litta" in the Hermitage, Leningrad. Among Leonardo's pupils at this time were Giovanni Antonio Boltraffio, Ambrogio de Predis, Bernardino de' Conti, Francesco Napoletano, Andrea Solari, Marco d'Oggiono, and Salai.

Stimulus for Leonardo's scientific studies

Art and science: the notebooks. The Milan years also saw Leonardo's decided turn toward scientific studies. He began to pursue these systematically and with such intensity that they demanded more and more of his time and energy and developed into an independent realm of creative productivity. Within him there arose now a growing need to note and write down in literary form every one of his perceptions and experiences. It is a unique phenomenon in the history of art. Undoubtedly, the several treatises on art that appeared or were made available during those decades provided an external stimulus. Leon Battista Alberti's De re aedificatoria (Ten Books on Architecture) was first printed in 1485; Francesco di Giorgio's treatise on architecture was available in its first manuscript versions, and Leonardo had received a copy from the author as a gift. Moreover, Piero della Francesca in his De prospectiva pingendi ("On Perspective in Painting") had provided for his contemporaries a model text on the theory of perspective. Finally, there was the mathematician Lucas Pacioli, who had become an acquaintance of Leonardo's. In 1494 Pacioli published his Summa de arithmetica *geometria proportioni* et proportionalità, followed by his Divina proportione ("On Divine Proportion"), for which Leonardo drew figures of symmetrical bodies.

In this ambience Leonardo began to nourish the desire to write a theory of art of his own, and there arose in him the far-reaching concept of a "science of painting." Alberti and Piero della Francesca had already offered proof of the mathematical basis of painting in their analysis of the laws of perspective and proportion and thereby buttressed painting's claim to being a science. But Leonardo's claims went much further. Proceeding from the basic conviction that sight is the human being's most unerring sense organ, yielding immediate, accurate, and reliable data of experience, Leonardo—equating "seeing" with "perceiving'—arrived at a bold conclusion: the painter, doubly endowed with subtle powers of perception and the complete ability to pictorialize them, was the prime person qualified to achieve knowledge by observing and to reproduce that knowledge authentically in a pictorial manner. Hence, Leonardo conceived the staggering plan of observing all objects in the visible world, recognizing their form and structure, and pictorially describing them exactly as they are. Thus, drawing became the chief instrument of his didactic method.

In the years between 1490 and 1495 the great program of Leonardo the writer (author of treatises) began. In it, four main themes, which were to occupy him for the rest of his life, could be discerned and gradually took shape: a treatise on painting, a treatise on architecture, a book on the elements of mechanics, and a broadly outlined work on human anatomy. His geophysical, botanical, hydrological, and aerological researches also belong to this period and constitute parts of the "visible cosmology" that loomed before Leonardo as a distant goal. Over against speculative book knowledge, which he scorned, he set irrefutable facts gained from experience—from *saper* vedere ("knowing how to see").

The notebooks

All these studies and sketches were written down in Leonardo's notebooks and on individual sheets of paper. Altogether they add up to thousands of closely written pages abundantly illustrated with sketches—the most voluminous literary legacy any painter has ever left behind. Of more than 40 codices mentioned in the older sources—often, of course, rather inaccurately—21 have survived; these in turn sometimes contain notebooks originally separate and now bound together so that 31 in all have been preserved. To these should be added several large bundles of documents: an omnibus volume in the Biblioteca Ambrosiana, Milan, called Codex Atlanticus because of its size, was collected by the sculptor Pompeo Leoni at the end of the 16th century; its sister volume, after a roundabout journey, came to England and fell into the possession of the English crown. Today it constitutes the most highly prized object in the Royal Library, Windsor Castle. Finally there is the Arundel Manuscript (British Museum, MS. 263), which contains a number of Leonardo's "caricatures" on various themes.

It was during his years in Milan that Leonardo began the earliest of these notebooks. He would first make quick sketches of his observations on loose sheets or on tiny paper pads he kept in his belt; then he would arrange them according to theme and enter them in order in the notebook. Surviving are a first collection of material for the painting treatise (MSS. A and B in the Institut de France, Paris), a model book of sketches for sacred and profane architecture (MS. B, Institut de France, Paris), the treatise on elementary theory of mechanics (MS. 8937, Biblioteca Nacional, Madrid), only recently rediscovered, and the first sections of a treatise on the human body (Anatomical MS. B; Windsor Castle, Royal Library).

Two special features make Leonardo's notes and sketches unusual: his use of mirror writing and the relationship between word and picture.

Leonardo was left-handed; so mirror writing came easily and naturally to him. It should not be looked upon as a secret handwriting. Though somewhat unusual, his script can be read clearly and without difficulty with the help of a mirror—as his contemporaries testified. But the fact that Leonardo used mirror writing throughout, even in his fair copies drawn up with painstaking calligraphy, forces one to conclude that, although he constantly addressed an imaginary reader in his writings, he never felt the need to achieve easy communication by using conventional handwriting. Yet occasional examples of normal handwriting (drafts of letters, notes, and comments to be submitted to third parties) show that Leonardo was completely at home in it. In the overwhelming majority of his notes in mirror writing, therefore, one gets the strong impression of "monologues in writing." Finally, then, his writings must be interpreted as preliminary stages of works destined for eventual publication, which Leonardo never got around to completing. In a sentence in the margin of one of his late anatomy sketches, he implores his "neighbours" to see that his works are printed.

Function of Leonardo's illustration vis-à-vis text

The second unusual feature in Leonardo's writings is the new function given to illustration vis-à-vis the text. Leonardo strove passionately for a language that was clear yet expressive. The vividness and wealth of his vocabulary were the result of intense self-study and represented a significant contribution to the evolution of scientific prose in the Italian vernacular. On the other hand, in his teaching method Leonardo gave absolute precedence to the illustration over the written word; hence, the drawing does not illustrate the text; rather, the text serves to explain the picture. In formulating his own principle of graphic representation—which he himself called *dimostrazione*—Leonardo was a precursor of modern scientific illustration.

Thus, during Leonardo's years in Milan the two "action fields"—the artistic and the scientific—developed and shaped his future creativity. It was a kind of "creative dualism," with mutual encouragement but also mutual pressure from each field.

Second Florentine period **(1499/1500–06).** In December 1499 or at the latest January 1500—three months after the victorious entry of the French into Milan—Leonardo left that city in the company of Lucas Pacioli. He stopped first at Mantua, where, in February 1500, he drew a portrait of his hostess, Marchioness Isabella d'Este, and then proceeded to Venice (in March), where the Signoria (governing council) sought his advice on how to ward off a threatened Turkish incursion in Friuli. Leonardo recommended that they prepare to flood the menaced region. From Venice he returned to Florence, where, after a long absence, he was received with acclaim and honoured as a renowned native son. In that same year he was appointed an architectural expert to a committee investigating damages to the foundation and structure of the church of S. Francesco al Monte. A guest of the Servants of Mary order in the cloister of SS. Annunziata, Leonardo began there a painting of the "Virgin and Child with St. Anne," the composition of which won admiration from artists and art lovers of the city. He also painted (1501) a "Madonna with the Yarn-Winder," which has survived only in copies and which he probably never finished. Mathematical studies seem to have kept him away from his painting activity much of the time, or so the Marchioness Isabella d'Este, who sought in vain to obtain a painting done by him, was informed by her representative in Florence.

Service with Cesare Borgia

Only his omnivorous "appetite for life" can explain Leonardo's decision, in the summer of the following year (1502), to leave Florence and enter the service of Cesare Borgia as "senior military architect and general engineer." Borgia, the notorious son of Pope Alexander VI, had, as commander in chief of the papal army, sought with unexampled ruthlessness to gain control of the Papal States of Romagna and the Marches. Now he was at the peak of his power and, at 27, was undoubtedly the most compelling and at the same time most feared person of his time. Leonardo, twice his age, must have been fascinated by his personality. For ten months he travelled across the condottiere's territories and surveyed them. In the course of his activity Leonardo sketched some of the city plans and topographical maps that laid the groundwork—before the invention of printing—for modern cartography. At the court of Cesare Borgia, Leonardo also met Niccolò Machiavelli, temporarily stationed there as a political observer for the city of Florence.

In the spring of 1503 Leonardo returned to Florence to make an expert survey of a project for diverting the Arno River behind Pisa so that the city, then under siege by the Florentines, would be deprived of access to the sea. The plan proved unworkable, but Leonardo's activity led him to a much more significant theme, one that served peace rather than war; the project, first advanced in the 13th century and now again under consideration, was to build a large canal that would bypass the unnavigable stretch of the Arno and connect Florence by water with the sea. Leonardo developed his ideas in a series of studies; with panoramic views of the river bank, which are also landscape sketches of great artistic charm, and with exact measurements of the terrain, he produced a map in which the route of the canal (with its transit through the mountain pass of Serravalle) was shown. The project, considered time and again in subsequent centuries, was never carried out. Today the express highway from Florence to the sea covers the exact route Leonardo chose for his canal.

That same year (1503), however, Leonardo also received a prized commission: to paint a mural for the Hall of the Five Hundred in Florence's Palazzo Vecchio, a historical scene of monumental proportions (at 23 x 56 feet [7 x 17 metres], it would have been twice as large as the "Last Supper"). For three years he worked on this "Battle of Anghiari"; like its intended complementary painting, Michelangelo's "Battle of Cascina," it remained unfinished. But the cartoon and copies of the main scene of the battle, with standards flying, were for a long time, to quote the 16th-century art historian Giorgio Vasari, "the school of the world." These same years saw the portrait of "Mona Lisa" and a painting of a standing "Leda," which was not completed and has survived only in copies.

Scientific study during the Florentine period

The Florentine period was also, however, a time of intensive scientific study; Leonardo did dissections in the hospital of Sta. Maria Nuova and broadened his anatomical work into a comprehensive study of the structure and function of the human organism. He made systematic observations of the flight of birds, concerning which he planned a treatise. Even his hydrological studies, "on the nature and movement of water," broadened into research on the physical properties of water, especially the laws of currents, which he compared with those pertaining to air. These were also set down in his own collection of data, contained in the so-called Leicester Codex in Holkham Hall, Norfolk, England.

Second Milanese period **(1506–13).** Thus, during these years in Florence, Leonardo's productivity was also marked by his "creative dualism." Only sporadically did he work at his paintings. When, in May 1506, Charles d'Amboise, governor of the King of France in Milan, asked and was granted permission by the Signoria in Florence for Leonardo to go for a time to Milan, the artist had no hesitation about accepting the invitation. But what was originally a limited period of time became a permanent move under the stress of political circumstances. Florence let Leonardo go, and the monumental "Battle of Anghiari" remained unfinished. Unsuccessful technical experiments with paints seem to have impelled Leonardo to stop working on the mural. One cannot otherwise explain his abandonment of this great work—great both in conception and in realization.

Leonardo spent six years in Milan, interrupted only by a six months' stay in Florence in the winter of 1507–08, where he helped the sculptor Giovanni Francesco Rustici execute his bronze statues for the Florence Baptistery but did not resume work on the "Battle of Anghiari." Honoured and admired by his patrons Charles d'Amboise and King Louis XII, who gave him a yearly stipend of 400 ducats, Leonardo never found his duties onerous. They were limited to advice in architectural matters, tangible evidence of which are plans for a chateau for Charles d'Amboise and perhaps also sketches for an oratory for the church of Sta. Maria alla Fontana, which Charles funded. Leonardo also looked into an old project revived by the French governor: the Adda canal that would link Milan with Lake Como by water.

In Milan he did very little as a painter: two Madonnas, which he promised the King of France, were never painted. He continued to work on the paintings of the "Virgin and Child with St. Anne" and "Leda," which he had brought with him from Florence, as copies from the Lombard school of that period attest. Again Leonardo gathered pupils around him. With Ambrogio de Predis he completed a second version of "The Virgin of the Rocks" (1508), in the course of which protracted litigation between the purchasers and the artists had a happy ending. Of his older disciples, Bernardino de' Conti and Salai were again in his studio; new pupils came, among them Cesare da Sesto, Giampetrino, Bernardino Luini, and the young nobleman Francesco Melzi, Leonardo's most faithful friend and companion until his death.

The tomb sculpture for Trivulzio

An important commission in sculpture came his way. Gian Giacomo Trivulzio had returned victoriously to Milan as marshal of the French army and a bitter foe of Ludovico Sforza. He commissioned Leonardo to sculpt his tomb, which was to take the form of an equestrian statue and be placed in the mortuary chapel donated by Trivulzio to the church of S. Nazaro Maggiore. But after years of preparatory work on the monument, for which a number of significant sketches have survived, the Marshal himself gave up the plan in favour of a more modest one; so this undertaking, too, remained unfinished. Leonardo must have felt keenly this second disappointment in his work as a sculptor.

Compared with his almost cursory work in art, Leonardo's scientific activity flourished. His studies in anatomy achieved a new dimension in his collaboration with a famous anatomist from Pavia, Marcantonio della Torre. He outlined a plan for an overall work that would include not only exact, detailed reproductions of the human body and its organs but would also include comparative anatomy and the whole field of physiology. He even thought he would finish his anatomical manuscript in the winter of 1510–11. Beyond that, his manuscripts are replete with mathematical, optical, mechanical, geological, and botanical studies that must be understood as data for his "perceptual cosmology." This became increasingly actuated by a central idea: the conviction that force and motion as basic mechanical functions produce all outward forms in organic and inorganic nature and give them their shape and, furthermore, the recognition that these functioning forces operate in accordance with orderly, harmonious laws.

Last years (1513–19). In 1513 political events—the temporary ouster of the French from Milan—caused the now 60-year-old Leonardo to move again. At the end of the year he went to Rome, accompanied by his pupils Melzi and Salai as well as by two studio assistants, hoping to find employment there through his patron, Cardinal Giuliano de' Medici, brother of the new pope Leo X. Giuliano gave him a suite of rooms in his residence, the Belvedere, in the Vatican. He also gave him a considerable monthly stipend, but no large commissions came to him. For three years Leonardo remained in the Eternal City, off to one side, while Donato Bramante was building St. Peter's, Raphael was painting the last rooms of the Pope's new apartments, Michelangelo was struggling to complete the tomb of Pope Julius, and many younger artists such as Peruzzi, Timoteo Viti, and Sodoma were active there. Drafts of embittered letters betray the disappointment of the aging master who worked in his studio on mathematical studies and technical experiments or, strolling through the city, surveyed ancient monuments. A magnificently executed map of the Pontine Marshes (Royal Library, Windsor Castle; 12684) suggests that Leonardo was at least a consultant for a reclamation project that Giuliano de' Medici ordered in 1514. On the other hand, there were sketches for a spacious residence for the Medici in Florence, who had returned to power there in 1513. But this did not go beyond the stage of preliminary sketches and never came to pass. Leonardo seems to have resumed his friendship with Bramante, but the latter died in 1514. And there is no record of Leonardo's relations with any other artists in Rome.

Service with King Francis I of France

In a life of such loneliness, it is easy to understand why Leonardo, despite his 65 years, decided to accept the invitation of the young king Francis I to come to France. At the end of 1516 he left Italy forever, together with his most devoted pupil, Francesco Melzi. Leonardo spent the last three years of his life in the small palace of Cloux, near the King's summer residence of Amboise on the Loire. Premier peintre, architecte et *méchanicien* du Roi ("first painter, architect, and mechanic of the King") was the proud title he bore; yet the admiring King left him complete freedom of action. He did no more painting or at most completed the painting of the enigmatic, mystical "St. John the Baptist," which the Cardinal of Aragon, when he visited Amboise, saw in Leonardo's studio along with the "Mona Lisa" and the "Virgin and Child with St. Anne."

For the King he drew up plans for the palace and garden of Romorantin, destined to be the widow's residence of the Queen Mother. But the carefully worked-out project, combining the best features of Italian–French traditions in palace and landscape architecture, had to be halted because the region was threatened with malaria.

Leonardo still made sketches for court festivals, but the King treated him in every respect as an honoured guest. Decades later, Francis I talked with the sculptor Benvenuto Cellini about Leonardo in terms of the utmost admiration and esteem. Leonardo spent most of his time arranging and editing his scientific studies. The final drafts for his treatise on painting and a few pages of the anatomy appeared. Consummate drawings such as the "Floating Figure" (Royal Library, Windsor Castle; 12581) are the final testimonials to his undiminished genius. In the so-called "Visions of the End of the World," or "Deluge" (Royal Library, Windsor Castle), he depicts with overpowering pictorial imagination the primal forces that rule nature.

On May 2, 1519, Leonardo died at Cloux, near Amboise. He was laid to rest in the palace church of Saint-Florentine. But the church was devastated during the French Revolution and completely torn down at the beginning of the 19th century. Hence, his grave can no longer be located. Francesco Melzi fell heir to his artistic and scientific estate.

ANALYSIS AND EVALUATION OF LEONARDO'S ACHIEVEMENT

Painting. Leonardo's total output in painting is really not large; only 17 of the paintings that have survived can be definitely attributed to him, and several of them are unfinished. Two of his most important works—the "Battle of Anghiari" and the "Leda," neither of them completed—have only survived in copies. Yet these few creations have established the unique fame of a man whom Vasari, in his Lives, dividing art history into three ages, placed in the last "golden age of the arts." His works, unaffected by all the vicissitudes of aesthetic doctrines in subsequent centuries, have stood out in all periods and all countries as consummate masterpieces of painting.

Leonardo's art of expression

The many testimonials to Leonardo, ranging from Vasari to Peter Paul Rubens, Johann Wolfgang von Goethe, and Eugène Delacroix, make it unmistakably clear that it has been, above all, Leonardo's art of expression that has called forth the utmost admiration. It is, in fact, the core of his formation as a painter—from his earliest beginnings to his last work. This expression was nurtured by his power of invention but also by every technical means: drawing, colour, use of light and shadow. To Leonardo, expression became a key concept of art; it also included the basic demands of truth, beauty, and accuracy in everything depicted.

What Leonardo was striving for was already revealed in his angel in Verrocchio's "Baptism of Christ" (c. 1474–75): in the natural structuring of the angel's body based on movement in several directions, in the relaxation of his attitude, and in his glance, which takes in what is occurring but at the same time is directed inward. In his landscape segment in the same picture, Leonardo also found a new expression for "nature experienced," in reproducing the forms he perceived as if through a veil of mist. The landscape study dated 1473, a pen drawing, foreshadows in its treatment of transparent atmosphere by a 21-year-old his telling ability to transform verceived phenomena into convincing graphic forms.

In the "Madonna Benois" (1478) Leonardo succeeded in giving an old traditional type of picture a new, unusually charming, and expressive mood by showing the Christ child reaching for the flower in Mary's hand in a sweet and tender manner.

His "Portrait of a Young Lady" ("Ginevra de'Benci"; c. 1475–78) opened new paths for portrait painting with his singular linking of nearness and distance.

The emaciated body of his "St. Jerome" (c. 1480) is presented with realistic truth based on his sober and objective studies in anatomy; gesture and look give Jerome an unrivalled expression of transfigured sorrow.

The interplay of mimicry and gesture—"physical and spiritual motion," in Leonardo's words—is also the chief concern of his first large creation containing many figures, "The Adoration of the Magi" (1481). Never finished, the painting nevertheless affords rich insight into the master's subtle methods of work. The various aspects of the scene are "built up" from the base with very delicate, paper-thin layers of paint in chiaroscuro (the balance of light and shadow) relief. The main treatment of the Virgin and Child group and the secondary treatment of the surrounding groups are clearly set apart with a masterful sense of composition; yet thematically they are closely interconnected: the bearing and expression of the figures—most striking in the group of praying shepherds—depict all degrees and levels of profound amazement.

The two versions of "The Virgin of the Rocks"

"The Virgin of the Rocks" in its first version in the Louvre is the work that reveals Leonardo's painting art at its purest. The painting, according to Leonardo's first contract with the Confraternity of the Immaculate Conception, was to be the central panel of a large work for their chapel in the church of S. Francesco and was done in the years 1483–85. It never arrived, however, at the place it was originally destined for. It seems to have been prematurely taken from the Confraternity, perhaps by some highly placed interested party who removed it from Leonardo's workshop. Instead of this first painting, Leonardo and Ambrogio de Predis painted a second, slightly revised version, probably begun around 1494. This one gave rise to a ten-year litigation between the artist and the Confraternity regarding the price, a dispute that was not settled until 1506 in favour of Leonardo; whereupon, two years later, the painting was delivered as per contract. This second version remained in the chapel of S. Francesco until the Confraternity was dissolved (1781), and then, after changing owners frequently, it came finally in 1880 to the National Gallery in London. Art scholars are still discussing the authenticity of the two versions.

"The Virgin of the Rocks" depicts the apocryphal legend of the meeting in the wilderness between the boy John and the equally young Jesus returning home from Egypt. Leonardo's artistry makes of this theme a vision that the true believer experiences when he contemplates the devotional picture. In the visionary character of the picture lies the secret of its effect: it presents not a "reality" but a "manifestation." Leonardo uses every artistic means at his disposal to emphasize the visionary nature of the scene. The soft colour tones (his famous sfumato), the dim light of the cave from which the figures emerge bathed in light, their quiet attitude, the meaningful gesture with which the angel (the only one facing the viewer) points to John as the intercessor between the Son of God and humanity—all this combines, in a patterned and formal way, to achieve an effect of the highest expressiveness.

The "Last Supper"

Leonardo's "Last Supper" is among the most famous paintings in the world. In its monumental simplicity, the composition of the scene is masterful; the power of its effect comes from the striking contrast in the attitudes of the young disciples as counterposed to Christ. Leonardo did not choose the portrayal of the traitor Judas customary in the iconographic tradition; he portrayed, rather, that moment of highest tension as related in the New Testament, "One of you which eateth with me will betray me." All of the Apostles—as human beings who do not understand what is about to occur—are agitated, whereas Christ alone, conscious of his divine mission, sits in lonely, transfigured serenity. Only one other being shares the secret knowledge: Judas, who is both part of and yet excluded from the movement of his companions; in this isolation he becomes the second lonely figure—the guilty one—of the company.

In the profound conception of his theme, in the perfect yet seemingly simple arrangement of the individuals, in the temperaments of the Apostles highlighted by gesture and mimicry, in the drama and at the same time the sublimity of the treatment, Leonardo attained a height of expression that has remained a model of its kind. Untold painters in succeeding generations, among them great masters, such as Rubens and Rembrandt, marvelled at Leonardo's composition and were influenced by it. The painting also inspired some of Goethe's finest pages of descriptive prose. It has become widely known through countless reproductions and prints, the most important being those produced by Raffaello Morghen in 1800. Thus, the "Last Supper" has become part of humanity's common heritage and remains today one of the world's outstanding paintings.

Technical deficiencies in the execution of the work have not lessened its fame. Leonardo was uncertain about the technique he should use. He bypassed fresco painting, which, because it is executed on fresh plaster, demands quick and uninterrupted painting, in favour of another technique he had developed: tempera on a base mixed by himself on the stone wall. This procedure proved unsuccessful, inasmuch as the base soon began to be loosened from the wall. Damage appeared by the beginning of the 16th century, and deterioration soon set in. By the middle of the century the work was called a ruin. Later, inadequate attempts at restoration only aggravated the situation, and not until the most modern restoration techniques were applied after World War II was the process of decay halted.

The "Mona Lisa"

In the Florence years between 1500 and 1506, four great creations appeared that confirmed and heightened Leonardo's fame: the "Virgin and Child with St. Anne," "Mona Lisa," "Battle of Anghiari," and "Leda." Even before it was completed, the "Virgin and Child with St. Anne" won the critical acclaim of the Florentines; the monumental plasticity of the group and the calculated effects of dynamism and tension in the composition made it a model that inspired Classicists and Mannerists in equal measure. The "Mona Lisa" became the ideal type of portrait, in which the features and symbolic overtones of the person painted achieved a complete synthesis. The young Raphael sketched the work in progress, and it served as a model for his "Portrait of Madalena Doni." Similarly, the "Leda" became a model of the *figura serpentinata* ("sinuous figure")—that is, a figure built up from several intertwining views. It influenced such classical artists as Raphael, who drew it, but it had an equally strong effect on Mannerists such as Jacopo Pontormo.

In the "Battle of Anghiari" (1503–06) Leonardo's art of expression reached its high point. The preliminary drawings—many of which have been preserved—reveal Leonardo's lofty conception of the "science of painting"; the laws of equilibrium that he had probed in his studies in mechanics were put to artistic use in this painting. The "centre of gravity" lies in the group of flags fought for by all the horsemen. For a moment the intense and expanding movement of the swirl of riders seems frozen; this passing moment, the transition from one active movement to the next, is uniquely interpreted.

On the other hand, Leonardo's studies in anatomy and physiology influenced his representation of human and animal bodies, particularly when they were in a state of excitement. He studied and described extensively the baring of teeth and puffing of lips as signs of animal and human anger. On the painted canvas, rider and horse, their features distorted, are remarkably similar in expression.

The highly imaginative trappings take the event out of the sphere of the historical into a timeless realm. Thus, the "Battle of Anghiari" became the standard model for a cavalry battle. Its composition has influenced many painters: from Rubens in the 17th century, who made the most impressive copy of the scene from Leonardo's now-lost cartoon, down to Delacroix in the 19th century.

After 1507—in Milan, Rome, and France—Leonardo did very little painting. He did resume work on the Leda theme during his years in Milan and sketched a variation, the "Kneeling Leda." The drawings he prepared—revealing examples of his late style—have a curious, enigmatic sensuality. Perhaps in Rome he began the "St. John the Baptist," which he completed in France. Bursting all the boundaries of usual painting tradition, he presented Christ's forerunner as the herald of a mystic oracle; his was an "art of expression" that seemed to strive consciously to bring out the hidden ambiguity of the theme.

Last manifestation of Leonardo's art of expression

The last manifestation of Leonardo's art of expression was in his "Visions of the End of the World," a series of pictorial sketches that took the end of the world as its theme. Here Leonardo's power of imagination—born of reason and fantasy—attained its highest level. The immaterial forces in the cosmos, invisible in themselves, appear in the material things they set in motion. What Leonardo had observed in the swirling of water and eddying of air, in the shape of a mountain boulder and in the growth of plants now assumed gigantic shape in cloud formations and rainbows. The framework of the world splits asunder, but even its destruction occurs—as the monstrously "beautiful" forms of the unleashed elements show—in accordance with the self-same laws of order, harmony, and proportion that presided at its creation and that govern the life and death of every created thing in nature. Without any model, these "visions" are the last and most original expressions of Leonardo's art—an art in which his perception based on *saper vedere* seems to have come to fruition.

Sculpture. That Leonardo worked as a sculptor from his youth on is borne out by his own statements and those of other sources. In the introduction to his ***Treatise on Painting*** he gives painting precedence over sculpture in the hierarchy of the arts; yet he emphasizes that both arts exert similar effects. A small group of generals' heads in marble and plaster, works of Verrocchio's followers, are sometimes linked with Leonardo because a lovely drawing on the same theme from his hand suggests such a connection. But the inferior quality of this group rules out an attribution to the master. Not a trace has remained of the heads of women and children that, according to Vasari, Leonardo modelled in clay in his youth.

The two great sculptural projects to which Leonardo devoted himself wholeheartedly stood under an unlucky star; neither the huge, bronze equestrian statue for Francesco Sforza, on which he worked until 1494, nor the monument for Marshal Trivulzio, on which he was busy in the years 1506–11, were brought to completion. Leonardo kept a detailed diary about his work on the Sforza horse; it has recently come to light with the rediscovery of the Madrid MS. 8937. Text and drawings both show Leonardo's wide experience in the technique of bronze casting but at the same time reveal the almost utopian nature of the project. He wanted to cast the horse in a single piece, but the gigantic dimensions of the steed presented insurmountable technical problems. Indeed, Leonardo remained uncertain of the problem's solution to the very end.

The greatness of Leonardo's concept of sculpture

The drawings for these two monuments reveal the greatness of Leonardo's concept of sculpture. Exact studies of the anatomy, movement, and proportions of a live horse —Leonardo even seems to have thought of writing a treatise on the horse—preceded the sketches for the monuments. Leonardo pondered the merits of two types, the galloping or trotting horse, and in both cases decided in favour of the latter. These sketches, superior in the suppressed tension of horse and rider to the achievements of Donatello's Gattamelata and Verrocchio's Colleoni sculptures, are among the most beautiful and significant examples of Leonardo's art. Unquestionably—as ideas—they exerted a very strong influence on the development of equestrian statues in the 16th century.

A small bronze of a galloping horseman in Budapest is so close to Leonardo's style that, if not from his own hand, it must have been done under his immediate influence (perhaps by Giovanni Francesco Rustici). Rustici, according to Vasari, was Leonardo's zealous student and enjoyed his master's help in sculpting his large group in bronze of "St. John the Baptist Teaching" over the north door of the Baptistery in Florence. There are, indeed, discernible traces of Leonardo's influence in John's stance, with the unusual gesture of his upward pointing hand, and in the figure of the bald-headed Levite. Moreover, an echo of Leonardo's inspiration is unmistakable in the much-discussed and much-reviled wax bust of "Flora" in Berlin. It may have been made in France, perhaps in the circle of Rustici, who entered Francis I's service in 1528.

Architecture. Leonardo, who in a letter to Ludovico Sforza applying for service described himself as an experienced architect, military engineer, and hydraulic engineer, was concerned with architectural matters all his life. But his effectiveness was essentially limited to the role of an adviser. Only once—in the competition for the cupola of the Milan Cathedral (1487–90)—did he actually consider personal participation; but he gave this idea up when the model he had submitted was returned to him. In other instances, his claim to being a practicing architect involved sketches for representative secular buildings: for the palace of a Milanese nobleman (around 1490), then for the villa of a French governor in Milan (1507–08), and for the Medici residence in Florence (1515). Finally, there was his big project for the palace and garden of Romorantin in France (1517–19). Especially in this last named, Leonardo's pencil sketches clearly reveal his mastery of technical as well as artistic architectural problems; the view in perspective (Windsor Castle) gives an idea of the magnificence of the site.

Activity as a military engineer

Leonardo was also quite active as a military engineer, beginning with the years of his stay in Milan. But no definite examples of his work can be adduced. Not until the discovery of the Madrid notebooks was it known that in 1504, sent probably by the Florence governing council, he stood at the side of the Lord of Piombino when the city's fortifications system was repaired and that Leonardo suggested a detailed plan for overhauling it. Finally, his studies for large-scale canal projects in the Arno region and in Lombardy show that he was also an expert in hydraulic engineering.

But what really characterizes Leonardo's architectural studies and makes them stand out is their comprehensiveness; they range far afield and embrace every type of building problem of his time. Furthermore, there frequently appears evidence of Leonardo's impulse to teach: he wanted to collect his writings on this theme in a theory of architecture. This treatise on architecture—the initial lines of which are in MS. B (Institut de France, Paris), a model book of the types of sacred and profane buildings —was to deal with the entire field of architecture as well as with the theory of forms and construction and was to include such items as urbanism, sacred and profane building, and a compendium of the important individual elements (for example, domes, steps, portals, and windows).

In the fullness and richness of their ideas, Leonardo's architectural studies offer an unusually wide-ranging insight into the architectural achievements of his epoch. Like a seismograph, his observations sensitively register all themes and problems. For almost 20 years he was associated with Bramante at the court of Milan and again met him in Rome in 1513–14; he was closely associated with such other distinguished architects as Francesco di Giorgio, Giuliano da Sangallo, Giovanni Antonio Amadeo, and Luca Fancelli. Thus, he was brought in closest touch with all of the most significant building undertakings of the time. Since Leonardo's architectural drawings extend over his whole life, they span precisely that developmentally crucial period—from the 1480s to the second decade of the 16th century—in which the principles of the classical style were formulated and came to maturity. That this genetic process can be followed in the ideas of one of the greatest men of the period lends Leonardo's studies their distinctive artistic value and their outstanding historical significance.

Science. ***Science of painting.*** Notwithstanding Leonardo's abundant scientific activity, one must never lose sight of the fact that it was the intellectual output of a man who proudly and consciously felt himself an artist throughout his life. And he described himself as such. He first came in contact with science as an artist, in the task he set himself of writing a treatise on painting.

Treatise on Painting

Leonardo's famous book on painting, in the form known and read today, is not an original work by the master but a compilation of texts from various manuscripts by Leonardo, collected and arranged with loving care by his disciple and heir, Francesco Melzi. It is the Codex Urbinas Latinus 1270, now in the Vatican Library. It was prepared around 1540–50, but from its form one can see that it was still an unfinished rather than a completed manuscript. Many original texts known to exist are missing; whole sections of Leonardo's overall plan are not included.

The first printed edition of the treatise in Melzi's version, omitting the long introductory chapter concerning the "pecking order" among the arts, appeared in a luxurious binding in 1651 in Paris, published by Raffaelo du Fresne with illustrations after drawings by Nicolas Poussin. The first complete edition of Melzi's text did not appear until 1817, published by Guglielmo Manzi in Rome. The two standard modern editions are that of Emil Ludwig, three volumes, Vienna, 1882 (with German translation); and that of A. Philip McMahon, Princeton, 1956, two volumes (facsimile of the Codex Urbinas and English translation).

Leonardo's plan envisaged a much broader treatment of the theme, as his own allusions to it indicate. For, in addition to detailed practical instructions for painting and drawing, the treatise was to deal with every area involving the artist's perception and experience, which he could then convey as acquired criteria. Three main problems form the keynote of the work: the definition of painting as a science, which is briefly outlined above (see *Life and works; Art and science: the notebooks);* the theory of the mathematical basis of painting—that is, geometry, perspective, and optics—with the systematic study of light and shadow, colour, and aerial perspective; and the theory of forms and functions in organic and inorganic nature, as they are explained and made comprehensible to the painter trained in *saper vedere.* This theory of the forms and functions of the visible world sought first of all to describe the animal world, including man; next it sought to include the plant world; finally it endeavoured to explain how such phenomena of inorganic nature as water and earth, air and fire came into being.

Drawings for the *Treatise*

In the drawings for the *Treatise on Painting,* extending from the earliest Milan period to the final years of Leonardo's life in France, the progressive broadening and deepening of the theme can be followed. Many drawings were placed by the side of the text, and some of them were coloured; many studies of nature that are admired as art works, such as the famous rain landscape in Windsor Castle (12409) or the "Foliage" (Royal Library, Windsor Castle; 12431), can be identified as illustrations for the treatise. Manuscript C in the Institut de France, Paris, with its diagrams of the blending of lights and shadows, likewise represents a segment of this textbook. Leonardo's so-called grotesque heads are also closely linked with the treatise. They have often been erroneously described as caricatures; but actually, for the most part, they represent types and only occasionally individuals. They are variations of the human face in its gradations between the poles of the beautiful and ugly, the normal and abnormal, the dignified and vulgar. They are also related to anatomical-physiological studies, in which old age—with wrinkled skin and bulging tendons—is contrasted with youth. Representation of the human being was to be treated at length: his body, his proportions, his organs and their functions but also his attitudes in physical and spiritual movement. Here Leonardo's artistic and scientific aims intertwine.

Anatomical studies and drawing. Leonardo's anatomical studies are perhaps the best way of revealing the process by which, in Leonardo's mind, an increasing differentiation set in among his diverse spheres of interest; but it was a differentiation in which the seemingly divergent areas of study—likewise on a higher level—always remained interrelated. Thus, Leonardo's study of anatomy, originally pursued for his training as an artist, quickly grew into an independent area of research. As his sharp eye uncovered the structure of the human body, Leonardo became fascinated by the *figura istrumentale dell' omo* ("man's instrumental figure"), and he sought to probe it and present it as a creation of nature. The early studies dealt chiefly with the skeleton and muscles; yet even at the outset Leonardo combined anatomical with physiological researches. From observing the static structure, Leonardo proceeded to study the functions exercised by the individual parts of the body as they bring into play the organism's mechanical activity. This led him finally to the study of the internal organs; among them he probed most deeply into the brain, heart, and lungs as the "motors" of the senses and of life. He did practical work in anatomy on the dissection table in Milan, then in the hospital of Sta. Maria Nuova in Florence, and again in Milan and Pavia, where he received counsel and inspiration from the doctor-anatomist Marcantonio della Torre. By his own admission he dissected 30 corpses in his lifetime, thus acquiring an astonishing range of experience on his own. This experience was distilled in the famous anatomical drawings, which are among the most significant achievements of Renaissance science. Called *dimostrazione,* these drawings are based on a curious connection between natural and abstract representation; sections in perspective, reproduction of muscles as "strings" or the indication of hidden parts by dotted lines, and finally a specifically devised hatching system enable him to represent any part of the body in transparent layers that afford an "insight" into the organ. Here Leonardo's mastery of drawing proved most useful. The genuine value of these "demonstrations" and their superiority to descriptive words—as Leonardo proudly emphasized—lay in the fact that they were able to synthesize a multiplicity of individual experiences at the dissecting table and make the data immediately and accurately visible. The effect is unlike that of all dead anatomical preparations; in this way the "live quality" of the organism is retained.

Value of the "demonstrations"

This great picture chart of the human body was what Leonardo envisaged as a *cosmografia del minor mondo* ("cosmography of the microcosm"). From the advanced portions that have survived, it is apparent how much and how long it occupied his mind. And it provided the basic principles for modern scientific illustration. Leonardo has not sufficiently received his due in this domain. Thanks to a method of seeing that was peculiarly his own, he elevated the art of drawing into a means of scientific investigation and teaching of the highest quality.

Mechanics and cosmology. With Leonardo, mechanics also proceeds from artistic practice, with which he became quite familiar as an architect and engineer. Throughout his life Leonardo was an inventive builder; he was thoroughly at home in the principles of mechanics of his epoch and contributed in many ways to advancing them.

His model book on the elementary theory of mechanics, which appeared in Milan at the end of the 1490s, was only recently discovered in the Madrid Codex 8937. Its importance lay less in its description of specific machines or work tools than in its use of demonstration models to explain the basic mechanical principles and functions employed in building machinery. Leonardo was especially concerned with problems of friction and resistance. These elements—screw threads, gears, hydraulic jacks, swivelling devices, transmission gears, and the like—are described individually or in various combinations; and here, too, drawing takes precedence over the written word. As in his anatomical drawings, Leonardo develops definite principles of graphic representation—stylization, patterns, and diagrams—that guarantee a precise demonstration of the object in question.

Importance of primal mechanical forces to his thought

In the course of years his interest in pure mechanics merged increasingly with an interest in applied mechanics. Leonardo realized that the mechanical forces at work in the basic laws of mechanics operate everywhere in the organic and inorganic world. They determine animate and inanimate nature alike as well as man. Leonardo wrote on a page of his treatise on anatomy:

> See to it that the book of the principles of mechanics precedes the book of force and movement of man and the other living creatures, for only in that way will you be able to prove your statements.

So, finally, "force" became the key concept for Leonardo; as *virtù spirituale* ("spiritual property"), it shaped and ruled the cosmos.

Wherever Leonardo probed the phenomena of nature, he recognized the existence of primal mechanical forces that govern the shape and function of the universe: in his studies on the flight of birds, in which his youthful idea of the feasibility of a flying apparatus took shape and led to exhaustive research into the element of air; in his studies of water, the *vetturale della natura* ("conveyor of nature"), in which he was as much concerned with the physical properties of water as with its laws of motion and currents; in his researches on the laws of growth of plants and trees as well as the geological structure of earth and hill formations; and finally in his observation of air currents, which evoked the image of the flame of a candle or the picture of a wisp of cloud and smoke. In his drawings, especially in his studies of whirlpools, based on numerous experiments he undertook, Leonardo again found a stylized form of representation that was uniquely his own: breaking down a phenomenon into its component parts—the traces of water or eddies of the whirlpool—yet at the same time preserving the total picture, analytic and synthetic vision.

Thus, for all the separate individual realms of his knowledge, Leonardo's science offered a unified picture of the world: a cosmogony based on *saper vedere* ("knowing how to see"). Its final wisdom is that all the workings of nature are subject to a law of necessity and a law of order that the *Primo Motore,* the divine "Prime Mover," created. "Marvelous is Thy justice, O Prime Mover! Thou hast seen to it that no power lacks the order and value of your necessary governance."

Leonardo as artist-scientist. As the 15th century expired, Scholastic doctrines were in decline, and Humanistic scholarship was on the rise. Leonardo, however, was part of an intellectual circle that developed a third, specifically modern form of cognition. In his view the artist—as transmitter of the true and accurate data of experience acquired by visual observation—played a significant part. With this sense of the artist's high calling, Leonardo approached the vast realm of nature to probe its secrets. His utopian idea of transmitting in encyclopaedic form the knowledge thus won was still bound up with medieval Scholastic conceptions, but the results of his research were among the first great achievements of the thinking of the new age because they were based on the principle of experience in an absolutely new way and to an unprecedented degree.

Finally, Leonardo, although he made strenuous efforts to teach himself and become erudite in languages, natural science, mathematics, philosophy, and history, as a mere listing of the wide-ranging contents of his library demonstrates, remained an empiricist of visual observation. But precisely here—thanks to his genius—he developed his own "theory of knowledge," unique in its kind, in which art and science form a synthesis. In the face of the overall achievements of Leonardo's creative genius, the question of how much he finished or did not finish becomes pointless. The crux of the matter is his intellectual force—self-contained and inherent in every one of his creations. This force has remained constantly operative down to the present day.

MAJOR WORKS

PAINTINGS: "The Annunciation" (*c.* 1472–77; Uffizi, Florence); "The Annunciation" (*c.* 1472–77; Louvre, Paris); "Madonna and Child" (*c.* 1474; Alte Pinakothek, Munich); "Portrait of a Young Lady" ("Ginevra de'Benci"; *c.* 1475–78; National Gallery of Art, Washington, D.C.); "Madonna Benois" (1478–after 1500; Hermitage, Leningrad); "St. Jerome" (*c.* 1480; Vatican Museums, Rome); "The Adoration of the Magi" (1481; Uffizi); "The Virgin of the Rocks" (1483; Louvre); "The Musician" (*c.* 1490; Pinacoteca Ambrosiana, Milan); "Lady with an Ermine" ("Cecilia Gallerani"; c. 1490; Muzeum Narodowe, Kraków, Poland); "The Virgin of the Rocks" (1494–1508; National Gallery, London); "Last Supper" (1495–97; Sta. Maria delle Grazie, Milan); decoration of the Sala delle Asse (1498; Castello Sforzesco, Milan); "The Virgin and Child with St. Anne" (cartoon, *c.* 1499; National Gallery, London); "Virgin and Child with St. Anne" (1500; Louvre); "Madonna with Yarn-Winder" (1501; Duke of Buccleuch Collection, London); "Mona Lisa" ("La Gioconda"; 1503–06; Louvre); "Battle of Anghiari" (1503–06; Palazzo Vecchio, Florence); "St. John the Baptist" (before 1517; Louvre).

DRAWINGS AND NOTEBOOKS: Main collections: Institut de France, Paris; British Museum; Uffizi, Florence; Biblioteca Ambrosiana, Milan; Accademia, Venice; Royal Library, Windsor.

BIBLIOGRAPHY. MARIO POMILIO and ANGELA OTTINO (eds.), *L'opera completa di Leonardo pittore* (1967; Eng. ed., *The Complete Paintings of Leonardo da Vinci),* catalogs the paintings. The most informative account of Leonardo's workshop and pupils is W. SUIDA, *Leonardo und sein Kreis* (1929). ARTHUR E. POPHAM, *The Drawings of Leonardo da Vinci* (1945), is important for the study of Leonardo as a draftsman; the standard publication on the drawings is KENNETH CLARK and CARLO PEDRETTI, *A Catalogue of the Drawings of Leonardo da Vinci in the Collection of Her Majesty The Queen at Windsor Castle,* 2nd ed., 3 vol. (1968). A. MARIONI, "I manoscritti di Leonardo da Vinci," in *Leonardo: Saggi e richerche* (1953), is a concise summary of all manuscripts, their facsimile editions, and chronology. This collection contains other excellent essays by various authors on Leonardo as artist and scientist. LADISLAO RETI'S two articles in *Burlington Magazine,* vol. 110 (1968), discuss the two rediscovered codices in Madrid. AMOS P. MCMAHON, *Leonardo's Treatise on Painting,* is a facsimile edition of *Coder Urbinas Latinus 1270* accompanied by an English translation (2 vol., 1956). The best anthologies of Leonardo's literary heritage are JEAN P. RICHTER, *The Literary Works of Leonardo da Vinci,* 3rd ed., 2 vol. (1970); and EDWARD MCCURDY (ed.), *The Notebooks of Leonardo da Vinci,* 2 vol. (1955). For additional sources of information, see ETTORE VERGA, *Bibliografia Vinciana, 1493–1930,* 2 vol. (1931); and *Raccolta Vinciana,* fasc. 1–20 (1905–64). Two institutions devoted exclusively to the study of Leonardo are the Raccolta Vinciana, Castello Sforzesco, Milan, and the Elmer Belt Library of Vinciana at the University of California at Los Angeles. *Leonardo da Vinci,* 2 vol. folio, ed. by the ISTITUTO GEOGRAFICO DE AGOSTINI (1964), contains numerous essays and is the most richly illustrated compendium of Leonardo's artistic and scientific activity. The two standard publications on Leonardo sources are LUCA BELTRAMI, *Documenti e memorie riguardanti la vita e le opere di Leonardo da Vinci in ordine cronologico* (1919); and GOROLAMI CALVI, *I manoscritti di Leonardo da Vinci, dal punto di vista cronologico, storico e biografico* (1925). The following six monographs are among the most valuable of the countless studies: G. SEAILLES, *Léonard de Vinci: l'artiste et le savant* (1892); W. VON SEIDLITZ, *Leonardo da Vinci,* 2nd ed. (1935), accompanied by extensive documentation (in German); KENNETH CLARK, *Leonardo da Vinci: An Account of His Development As an Artist,* 2nd ed. (1952), also available in paperback; L.H. HEYDENREICH, *Leonardo da Vinci,* 2 vol. (Eng. trans. 1954); V.P. ZUBOV, *Leonardo da Vinci* (1968; orig. pub. in Russian, 1961); and PETER R. RITCHIE-CALDER, *Leonardo and the Age of the Eye* (1970). MORRIS PHILIPSON (ed.), *Leonardo da Vinci: Aspects of the Renaissance Genius* (1966), contains valuable contributions to the historical and psychological aspects of the phenomenon, Leonardo.

(L.H.H.)

Leonardo of Pisa

Leonardo of Pisa, or Leonardo Fibonacci, was the most distinguished mathematician of the Middle Ages. He helped introduce into mathematics the Hindu-Arabic numerals and the number sequence that bears his name.

Little is known about Leonardo's life beyond the few facts given in his mathematical writings. From the date of his first and best known work, the *Liber abaci* (1202; "Book of the Abacus"), it is probable that he was born in the 1170s. He was probably born in Pisa, Italy, but there is no evidence that he was. During Leonardo's boyhood, his father, Guglielmo, a Pisan merchant, was appointed consul, or chief magistrate, over the community of Pisan merchants in the North African port of Bugia (now Bejaïa, Algeria). Leonardo soon joined him. "With a view to future usefulness," the father sent his son to study calculation with an Arab master, and Leonardo later, in the *Liber abaci,* described his enjoyment in learning "the art of the nine Indian figures." Leonardo also travelled to Egypt, Syria, Greece, Sicily, and Provence, where he studied different numerical systems and methods of calculation but never found one as satisfactory as the Hindu-Arabic numerals.

The *Liberabaci*

When Leonardo's *Liber abaci* first appeared, Hindu-Arabic numerals were known to only a few European intellectuals through translations of the writings of the 9th-century Arab mathematician and astronomer al-Khwārizmī. Leonardo began his explanation of the notation by observing: "The nine Indian figures are: 9 8 7 6 5 4 3 2 1. With these nine figures and with the sign 0 . . . any number may be written, as is demonstrated below." The first seven chapters dealt with the notation, explaining the principle of place value, by which the position of a figure determines whether it is a unit, ten, 100, and so forth, and demonstrating the use of the numerals in arithmetical operations. The techniques were then applied to such practical commercial problems as profit margin, barter, money changing, conversion of weights and measures, partnerships, and interest. Most of the work was devoted to speculative mathematics—proportion (represented by such popular medieval techniques as the Rule of Three and the Rule of Five, which are rule-of-thumb methods of finding proportions), the Rule of False Position (a method by which a problem is worked out by a false assumption, then corrected by proportion), extraction of roots, and the properties of numbers, concluding with some geometry and algebra. In 1220, Leonardo produced a brief work entirely on geometry, the *Practica geometriae* ("Practice of Geometry"), which included eight chapters of theorems based on Euclid's *Elements* and *On Divisions.*

The *Liber abaci,* which was widely copied and imitated, drew the attention of the Holy Roman emperor Frederick II, whose patronage of science had earned him the epithet *Stupor mundi* (Wonder of the World). In the 1220s, Leonardo was invited to appear before the Emperor at Pisa, and there John of Palermo, a member of Frederick's scientific entourage, propounded a series of problems, three of which Leonardo presented in his books. The first two belonged to a favourite Arabic type, the indeterminate, which had been developed by a 3rd-century Greek mathematician, Diophantus. This was an equation with two or more unknowns for which the solution must be in rational numbers (whole numbers or common fractions). The third problem was a third-degree equation (*i.e.*, containing a cube), $x^3 + 2x^2 + 10x = 20$ (expressed in modern algebraic notation), which Leonardo solved by a trial-and-error method known as approximation; he arrived at the answer

$$1^{0}\,22^{\mathrm{I}}\,7^{\mathrm{II}}\,42^{\mathrm{III}}\,33^{\mathrm{IV}}\,4^{\mathrm{V}}\,40^{\mathrm{VI}}\left(1+\frac{22}{60}+\frac{7}{3{,}600}+\frac{42}{216{,}000}+\cdots\right)$$

in sexagesimal fractions (a fraction using the Babylonian number system that had a base of 60), which, when translated into modern decimals (1.3688081075), is correct to nine decimal places. Leonardo's solutions to the problems were a combination of ingenuity and accuracy.

For several years Leonardo corresponded with the Emperor and his scholars, exchanging problems with them; and he dedicated his *Liber quadratorum* (1225; "Book of Square Numbers") to the Emperor. Devoted entirely to Diophantine equations of the second degree (*i.e.*, containing squares), the *Liber quadratorum* may be considered Leonardo's masterpiece. It is a systematically arranged collection of theorems, many invented by the author, who used his own proofs to work out general solutions. Probably his most creative work was in congruent numbers—numbers that give the same remainder when divided by a given number. He worked out an original solution for finding a number that, when added to or subtracted from a square number, leaves a square number. Leonardo's statement that $x^2 + y^2$ and $x^2 - y^2$ could not both be squares was of great importance to the determination of the area of rational right triangles. Although the *Liber abaci* was more influential and broader in scope, the *Liber quadratorum* alone ranks its author as the major contributor to number theory between Diophantus and Pierre de Fermat, the 17th-century French mathematician.

The *Liber quadratorum*

In 1228, Leonardo revised the *Liber abaci,* dedicating it to Michael Scott, the Emperor's chief scholar. After that date, nothing is known of Leonardo until 1240, when Pisa granted him an annuity of 20 Pisan pounds, along with expenses, for services to the city. The date of his death is unknown.

Except for his role in spreading the use of the Hindu-Arabic numerals, Leonardo's contribution to mathematics has been largely overlooked. His name is known to modern mathematicians mainly because of the Fibonacci sequence (see below) derived from a problem in the *Liber abaci:*

> A certain man put a pair of rabbits in a place surrounded on all sides by a wall. How many pairs of rabbits can be produced from that pair in a year if it is supposed that every month each pair begets a new pair which from the second month on becomes productive?

The Fibonacci sequence

The resulting number sequence, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55 (Leonardo himself omitted the first term), in which each number is the sum of the two preceding numbers, is the first recursive number sequence (in which the relation between two or more successive terms can be expressed by a formula) known in Europe. Terms in the sequence were stated in a formula after the development of algebraic notation in the early 17th century by the French-born mathematician Albert Girard in 1634: $u_{n+2} = u_{n+1} + u_n$, in which u represents the term and the subscript its rank in the sequence. Later, the mathematician Robert Simson at the University of Glasgow in 1753 noted that as the numbers increased in magnitude, the ratio between succeeding numbers approached phi—the ancient mean and extreme ratio, or golden section—whose value is 1.6180 . . . , or $\frac{1+\sqrt{5}}{2}$. When, in the 19th century, the term Fibonacci sequence was coined by French mathematician Edouard Lucas, scientists began to discover the numbers in nature, such as in the spirals of sunflower heads, in pine cones, in the regular descent (genealogy) of the male bee, and the related logarithmic (equiangular) spiral in snail shells, in the arrangement of leaf buds on a stem, and in animal horns. A Fibonacci Association was founded in 1962 in California "to exchange ideas and stimulate research in the Fibonacci numbers and related topics."

BIBLIOGRAPHY. For general treatments, see JOSEPH and FRANCES GIES, *Leonard of Pisa and the New Mathematics of the Middle Ages* (1969), a summary of Leonardo's life and survey of his works, plus a brief overview of the history of numerical notation and of medieval Pisa; GINO LORIA, "Leonardi Fibonacci," in *Gli scienziati italiani* (1919), contains a short account of Leonardo's life and works, in Italian. For specialized treatments, see RAYMOND B. MCCLENON, "Leonardo of Pisa and His Liber Quadratorum," ***Am.** Math. Mon.*, 26:1–8 (1919); and VERNER E. HOGGATT, JR., *Fibonacci and Lucas Numbers* (1969), an introduction to the Fibonacci numbers and sequence, with propositions and exercises for mathematics students; and N.N. VOROB'EV, *The Fibonacci Numbers* (1961; orig. pub. in Russian). Leonardo's works include: *Scritti di Leonardo Pisano,* ed. by BALDASSARRE BONCOMPAGNI, 2 vol. (1857–64), in Latin; and *Le Livre des nombres carrés*, trans. and ed. by PAUL VER EECKE (1952), in French.

(F.C.Gi.)

Leopold I, Emperor

The name of the Holy Roman emperor Leopold I is linked with half a century of Austrian history. His reign (1658–1705) saw events that marked new departures in many respects and were to condition the development of the Habsburg monarchy during the next two centuries. From its struggles with the Turks and France, Austria emerged as a great power and was able to maintain this newly won position in spite of upheavals in Hungary. In internal politics, the rise of monarchical absolutism brought attempts at a more strictly centralized administration. If the Emperor, who had not been trained for the throne, rarely interfered with the course of events, he, nevertheless impressed contemporaries with an imperturbability founded in personal piety, which did not fail him even during the worst crises to his long reign. His biographer, the Jesuit Hans Jacob Wagner von Wagenfels (died 1702), quite aptly praises his magnanimity as his most conspicuous character trait. The interest Leopold took in all matters of learning, his gift for music, and his preoccupation with historiography made him a patron of renown and, notwithstanding the military conflicts of the time and his precarious finances, gave enormous impetus to learning and the arts throughout the Austrian countries and especially made Vienna a famous cultural centre. His reign saw the first flourishing of Baroque culture in Austria.

Early years

Leopold was born in Vienna on June 9, 1640. As the second son of Ferdinand III's first marriage, to his cousin Maria Anna, daughter of Philip III of Spain, he was destined for the church. He received a careful education by excellent teachers, among whom the cultured count Johann Ferdinand Portia was the leading personality. Made lord high steward by his pupil, Portia retained his influence with Leopold until his death in 1665. From an early age Leopold showed an inclination towards learning. He learned easily and became fluent in Latin, Italian,

Leopold I, portrait bust by an unknown artist, c. 1700. In the Kunsthistorisches Museum, Vienna.

BY courtesy of the Kunsthistorisches Museum, Vienna

and Spanish, but he did not like French and later would not have it spoken at court. Besides concerning himself with antiquarian studies, history, literature, natural science, and astronomy, his special interest was music, having inherited the musical talents of his father. The keynote of his personality was a deep devotion, which made him the personification of *pietas Austriaca,* the loyal Catholic attitude of his house. From his religiosity, however, also derived a fatalistic strain, which had its negative side for a ruling monarch. He rejected all political compromising on denominational questions.

Accession

When his elder brother, Ferdinand IV, died quite unexpectedly (July 9, 1654), Leopold suddenly found himself heir apparent to the Austrian Habsburg lands. In 1655 the Lower Austrian estates did homage, and he was elected and crowned king of Hungary, the Bohemian coronation following in 1656. Then, in 1657, his father died and a new imperial election was due. After long and difficult struggles against the opposition of France, Leopold was elected and crowned in the summer of 1658.

With untiring energy and a deep sense of duty, the young monarch settled down to the unwonted task of government. From the beginning he had to fight wars, first of all against the Turks. In 1683 they appeared before Vienna, and for the second time in its history the city suffered a Turkish siege. Leopold had left the capital with his court to await the outcome at Passau. An imperial army was summoned, and from the time of their repulse at Vienna the Turks were gradually forced into the defensive, especially after the military genius of Prince Eugene of Savoy appeared on the scene in 1696. In the Treaty of Carlowitz (1699), almost the whole of Hungary was freed from Turkish rule.

The Hungarian nobles, however, who were mostly Calvinists, did not want to exchange Turkish rule for a centralized Habsburg government, which threatened to introduce the Counter-Reformation. Their opposition had been a serious problem all the time, and Leopold, who usually showed clemency, took a firm stand for once, refusing to recall the cruel sentences after the so-called Nobles' Plot. Three of the most prominent Hungarian noblemen were executed, and Hungarian resistance flared up again in the fierce Kuruc risings.

The struggle with France

Though Leopold's policy toward Catholic France was undecided at first, he finally had to agree to a coalition with the Protestant naval powers: Holland and England. In the course of the long struggle with France, the empire scored several military successes; but in the end French diplomacy remained victorious, always dividing the enemy at the decisive moment. The Emperor was accused of a wavering attitude and lack of initiative, and these character traits were indeed partly responsible for the failure of his policies. The war ended in the unfavourable Treaty of Rijswijk (1697), under the terms of which Strasbourg had to be ceded to France, a great discredit to Leopold.

Apart from some contributions from the empire and subsidies from its allies, the financial burden of all these wars had to be borne by Leopold's hereditary countries, the finances of which were badly organized. During his long reign Leopold found it impossible to arrive at a sound financial basis; indeed, he was careless in these matters and for years suffered the treasury to be mismanaged by Count Sinzendorf.

Emperor Leopold was not always fortunate in the choice of his ministers. There was, for example, Count Eusebius Potting, with whom he had formed a warm friendship but who was not the right man for the post of ambassador to Madrid. On the other hand, councillors who had convinced the Emperor of their sincerity and honesty found excellent chances for a court career, even if they were middle class, like the Austrian court chancellor Johann Paul Hocher.

Leopold no longer regarded the empire as his primary responsibility; rather, in his view, concern for the power and prestige of the Habsburg dynasty and lands took the first place. By his first marriage—to his niece Margarita Teresa, daughter of Philip IV of Spain—Leopold had acquired a claim to the Spanish throne. From the outset the Spanish succession formed the central aim of his politics. What lay behind this was the idea of the unity of the House of Habsburg, the two lines being considered only as parts of the same entail. At the death of Charles II of Spain (who left no heir), Louis XIV of France claimed Spain for his grandson, Philippe, duc d'Anjou; there could be no question for Emperor Leopold that the Spanish heritage had to be defended by force of arms. On May 5, 1705, in the middle of the War of the Spanish Succession (1701–14), Leopold died. He was buried in the Habsburg mausoleum of the Capuchins at Vienna.

Personal traits

The Emperor was of medium size, rather slender in his youth but stout in later life. His face was pale, and he had dark hair and the typical Habsburg traits of a strongly developed lower lip and a protruding chin. A Turkish traveller described him as a cultivated man of extreme ugliness. His health was bad, and, when he fell dangerously ill in 1670, everybody expected the Austrian line of the Habsburgs to become extinct. He recovered, however, and in 1673 married Claudia Felicitas from the Tirolian branch of the Austrian Habsburgs. In 1676 the Emperor solemnized his third marriage, with Eleonore of Palatinate–Neuburg; this proved a happy union and produced ten children, among them the future emperors Joseph I and Charles VI.

In spite of the Emperor's great personal simplicity, the sums expended to maintain the imperial court were gigantic. At all occasions the Emperor was anxious to emphasize his imperial dignity; official journeys, such as his coronation journey to Frankfurt in 1658, as well as the numerous pilgrimages he undertook to assure divine assistance against his enemies, were used for ostentation. A special concern of the Emperor was to reshape Vienna into a worthy imperial residence. The Vienna court was famous for its costly theatricals, in which at times the emperor and empress also took part. Italian operas and ballets were lavishly staged, often with some additional music composed by Leopold himself. As the Emperor was very fond of hunting, courtly pleasures also included heron hawking and hunting wild boars and stags in the vicinity of the residence. Though Leopold undertook no more extensive journeys after 1693, he enjoyed these regular hunting expeditions until his death.

Leopold I was a devoted book collector and, in the director of the court library, Peter Lambeck, found a helper of great renown. He was known for the encouragement he extended to learning, whereby he tried to secure the services of famous scholars for his court.

BIBLIOGRAPHY. There is no modern monograph of Leopold I. The best presentation of his reign is still to be found in OSWALD REDLICH, *Weltmacht des Barock: Österreich in der Zeit Kaiser Leopolds I. (1927,* reprinted 1961) and ***Das***

Werden einer Grossmacht: bsferreich von 1700 bis 1740, 2nd ed. (1942, reprinted 1962), with references to source materials and a detailed bibliography. The most modern summary of recent research may be found in ERICH ZOLLNER, *Geschichte bsterreichs,* 4th ed. (1970), with bibliography.

(H.Di.)

Lepidoptera

The insect order Lepidoptera, with as many as 100,000 species, is second in size only to the Coleoptera, the beetles. In addition to the familiar butterflies and moths the order includes the skippers, a worldwide group intermediate between butterflies and moths. Because of their day-flying habits and bright colours, the butterflies are more familiar than the chiefly night-flying and dull coloured moths, but the latter are far more varied and abundant. With the exception of a few moths, all adult Lepidoptera have two pairs of wings. The name Lepidoptera is derived from the Greek, meaning "scaly winged," and refers to the characteristic covering of microscopic, dustlike scales on the wings.

The life cycle of Lepidoptera is that common to the higher orders of insects and consists of four stages: egg, larva (caterpillar), pupa (chrysalis), and adult (imago). The larva and adult are almost invariably plant eaters. The larvae do most of the eating, the majority of them feeding on foliage, although many species eat stems, roots, fruits, or flowers. A number of moths and a few butterflies are important pests in agriculture or forestry, invariably as larvae. The adults of many species are important pollenizers, visiting flowers for nectar. In the ecology of most land environments the Lepidoptera are important as transformers of immeasurably large amounts of plant matter into animal matter, which then serves as food for many other groups of animals.

Many members of the order, especially the butterflies, have appealed to the human imagination for thousands of years, as symbols of fragile and ephemeral beauty. References to them abound in literature; and they have been depicted in many paintings, have inspired the designs of jewelry, ornaments and textiles, and even occur in many heraldic devices and on postage stamps. They are the most popular of insects among amateur collectors and students of natural history.

GENERAL FEATURES

The members of the order show great diversity in size and development rates. The smallest moths have wing expanses as small as 4 millimetres, while the largest moths and butterflies may expand to nearly 300 millimetres (about 1 foot). Fast developing species may complete their development in as little as three weeks, while slower ones may take as much as two, or even three, years.

Both moths and butterflies occur on every continent except Antarctica. Though they are enormously more numerous and diversified in the tropics, at least some species occur at the limits of polar vegetation. There are many successful species in nearly every land environment, from arid deserts and high mountaintops to marshes and tropical rain forests. Most of the species have become adapted for life in relatively restricted ecological niches, and may be limited to only one or a small group of food plants, often to only a single part of one plant. They are therefore seldom abundant in more than one type of habitat. A few species, however, with broader food habits, may occasionally reach peaks of abnormal abundance and defoliate large areas of such environments as deciduous forest or grassland. Some of the insects most injurious to man's interests in agriculture, forestry, and households are members of the Lepidoptera.

Geographic distribution

The large family-level groups, and many smaller ones, are worldwide, with such families as Noctuidae, Geometridae, Pyralididae, and Nymphalidae being dominant elements of the insect fauna everywhere. A few families, some subfamilies and a great many genera, however, are characteristically more abundant in one faunal region than in others. Thus, the Neotropical region (mainly Central and South America) is characterized by great diversity of such moth families as the Dioptidae, Pericopidae, and Ctenuchidae and such butterfly subfamilies as the Riodininae, Heliconiinae, Ithomiinae, Morphinae, and Dismorphiinae. The North American (Nearctic) and Eurasian (Palaearctic) faunas show many evidences of close recent connections, chiefly between Asia and western North America. Each region has, however, many distinctive generic-level groups; for example, in the Palaearctic, ***Erebia*** and other satyr butterflies, and the parnassians (Papilionidae); in the Nearctic, the underwing moths *Catocala* (Noctuidae) and the silverspot fritillaries (Nymphalidae); in the Ethiopian (African) region, the butterfly genera *Charaxes* (Nymphalidae) and ***Teracolus*** (Pieridae), and many members of the papilionid subfamily Acraeinae; in the Oriental region, ***Troides*** (Papilionidae) and *Delias* (Pieridae).

Many Lepidoptera occur in isolated colonies as relict (remnant) populations, cut off from relatives elsewhere by geologic or climatic changes. Australia and New Zealand have unusually diverse relict populations of the primitive Micropterigidae and Hepialidae. In North America, Europe, and Asia many relict species have survived since the Pleistocene glaciations on isolated southern mountaintops. They include members of such genera as ***Oeneis*** and ***Erebia*** (Satyridae), *Boloria* (Nymphalidae), *Parnassius* (Papilionidae), ***Anarta*** (Noctuidae), and ***Pediasia*** (Pyralididae).

From H. Weber, *Grundriss der Insektenkunde* (1966); Gustav Fischer Verlag, Stuttgart

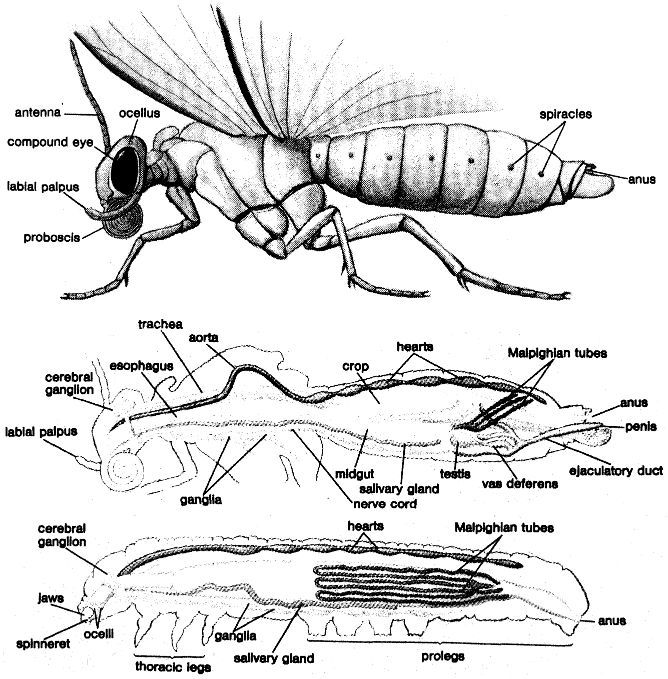


Figure 1: Body plan of generalized male adult butterfly or moth and caterpillar.

IMPORTANCE TO MAN

Many hundreds of Lepidoptera injure plants useful to man, including some of his most important sources of food, fabrics, fodder, and timber. The great majority of the injurious species are moths, and the most injurious life stage always is the larva. Unlike members of other insect orders, the Lepidoptera do not act as carriers of plant diseases, nor are any of them parasites or predators of man or domestic animals.

The list of valuable plants subject to damage by Lepidoptera is a long one, including many grains, sugar beets and cane, cotton, tobacco, some root crops and leaf crops, many fruits, and timber and shade trees. The damage may involve the leaves, stems, roots, or fruit. Woollens, furs, silk, and even feathers are eaten by tineid moths of several genera (clothes moths). The wax moth (***Galleria*** *mellonella*) causes considerable damage in beehives.

A few Lepidoptera are directly beneficial to man. Nearly all silk is obtained from the completely domesticated silk-

Lepidoptera beneficial to man

worm, Bombyx mori, originally Asiatic. Such other silks as shantung and tussah are the products of various Asiatic Saturniidae. The larvae and sometimes the adults of a few species are used for food. The larvae of a skipper, *Rhopalocampta libeon*, are collected in large quantities in the Congo, and the caterpillars of giant skippers (Megathymidae), known in Mexico as *gusanos de* maguey, are canned and exported for consumption as hors d'oeuvres.

The South American cactus moth, Cactoblastis *cactorum*, has been highly beneficial in weed control, clearing more than **60** million acres of Australia of prickly pear cactus. Doubtless man benefits also from much unrecognized weed eating by caterpillars and flower pollination by adults.

Many Lepidoptera are valuable in biological research, including work in ecology, biogeography, systematics, genetics, and physiology. Much of the present knowledge of endocrine controls of insect development has come from studies of the silkworm moth (Bombyx mori) and various saturniid moths. The study of industrial melanism (an increase in the frequency of black individuals in a population, brought on by the environmental changes that accompanied the Industrial Revolution) in the British peppered moth (*Biston* betularia) has profoundly influenced modern ideas about rates of evolutionary change (see EVOLUTION).

NATURAL HISTORY

Types of food taken

Feeding habits. The habits of Lepidoptera are extremely diverse, depending on the adaptations of the species or group to climate, environment, type of food plant, way of feeding, and many other factors. The great majority of food plants are seed plants (Spermatophyta), the dominant land plants. Primitive plants such as mosses, liverworts and ferns, and some lichens are eaten by a few groups. Nearly all parts of the plant are eaten by various specially adapted forms. Flowers are eaten by many larvae (*e.g.*, lycaenid butterflies, plume moths, some geometrid moths), and nectar by many adults. Cones and fruits and their seeds are eaten by others (*e.g.*, prodoxid and olethreutid moths, some owlet moths). Some seed eaters have become household pests, feeding on stored grains and cereals. Buds or soft, succulent stems are bored into by members of many families. Several lepidopteran groups (*e.g.*, the pine moths, Rhyacionia) specialize on the terminal shoots of conifers. Many groups feed on the turf of grasses and sedges. Three families (Cossidae, Aegeriidae, and Hepialidae) bore in woody stems and rootstocks, the Cossidae in particular tunnelling deep into hardwoods. **A** good many lepidopterans (especially members of the Tineidae, Blastobasidae, and Pyralididae) feed on dead and decaying plant matter, particularly moldy debris. In comparison with other orders of insects, relatively few Lepidoptera live in plant galls or eat animal matter. Some tineids feed on fur and feather debris, and a few exceptional species are predators or parasities of other animals.

Life cycle. Egg. The number of eggs laid varies greatly from fewer than a hundred in some species to more than a thousand in others. The eggs are almost always laid in a specific way, usually on or in a suitable foodplant. Those of many species are laid singly and widely dispersed; in others they are laid in masses that may be covered with a hardened secretion of the female's abdominal glands. In a few groups the terminal segments of the female's abdomen are greatly elongated and bladelike, and the eggs are laid in soft plant tissues or in narrow slits or crevices. In some of the primitive hepialid moths and in some butterflies the female may merely scatter the eggs in flight in the general vicinity of a suitable foodplant. Development of the embryo and emergence of the young larva is often controlled by a mechanism of physiologically enforced inactivity (diapause), which has the effect of timing the emergence of the larva to coincide with suitable conditions of weather and the growth of the foodplant. Respiration in the egg is carried on through an aeropyle, a system of air passages in the chorion (shell) that enables oxygen exchange with the environment to occur whether the egg is dry or wet. In a few species (some tineid moths and pierid butterflies) the larvae hatch in the uterus of the female.

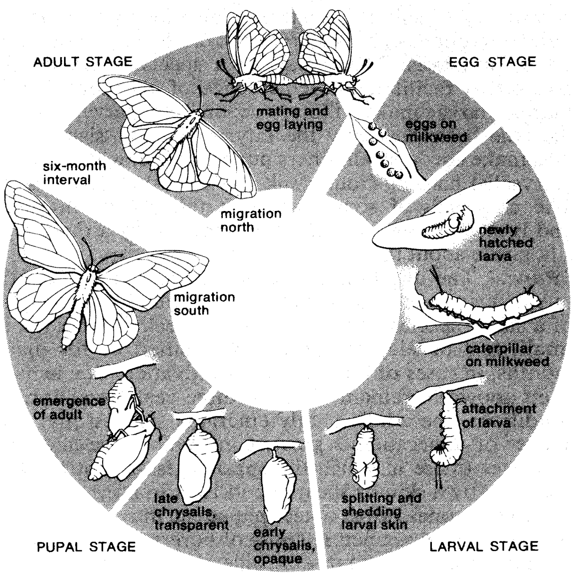


Figure 2: Life cycle of the monarch butterfly *(Danaus plexippus).*

Production of silk

Larva. The larva is the chief, often the only, nutritive stage of the life cycle, its function being simply to transform very large quantities of plant matter into animal matter, and to stay alive during the process. The larval stage is followed by the pupa, a resting stage in which the individual undergoes a major rebuilding of body tissues, to emerge as a sexually active adult. Plant foods being relatively easy to find and eat, it is not surprising that lepidopterous larvae are fundamentally quite uniform, despite their apparent diversity. Most larvae continually spin silk threads that adhere to the surfaces on which they live, thus giving them an almost unshakable grip. Silk is also used in the construction of various nests, cases, and shelters that protect the larvae from the elements and from enemies. Silk also serves in the final cocoon, in which the larva will change into the pupa. The larvae of leaf miners are mostly very small, greatly flattened, and more or less legless, well adapted for life inside a flat leaf. The larvae of borers are relatively plain and unornamented, while those that live in the open, exposed to the attacks of predators, have evolved a multiplicity of protective adaptations of form, colour, texture, and habit. Some possess chemical repellants. The proper timing of the larval development and activities is important in species in which hibernation, or estivation, is necessary. In some species (*e.g.*, nymphalid butterflies of the genus *Limenitis*) the larvae of the summer generation grow right on through without pause, but those of the next year's generation hibernate while still very young. In certain groups larval development is very slow, for reasons not understood, so that there may be only one generation a year. The larvae of the carpenter moths (Cossidae), feeding on poorly nutritious wood, may take two years to develop. Larvae of many Arctic and northern species, regularly have two-year cycles. The usual number of molts (ecdyses) is four or five, but some of the small leaf miners molt only twice. When starved, clothes moth larvae (Tineola) have been known to have a dozen molts, sometimes accompanied by a decrease in size.

Construction of the cocoon

Larvae seldom travel far from the plants where they began life. In some species there is dispersal of very young larvae, which, hanging from threads, are blown by the wind. Swarms of army worms (Pseudaletia, Noctuidae) may travel large distances, driven by crowding and lack of food. Just before pupation many larvae stop eating and crawl some distance before settling down to pupate. Many species pupate in the soil, with little or no cocoon; many others form cocoons in the soil, in trash, or under loose bark. Some cocoons are fastened to twigs or branches or rolled in leaves. The cocoons of leaf miners are usually formed in the "mine" or near it. Cocoons are commonly of silk alone but may also incorporate larval hairs, secretions or waste products, or chewed wood pulp, trash, or leaves. There may be a special seam on the cocoon to facilitate the emergence of the adult moth. Skipper larvae make a flimsy cocoon in the larval nest. Of the true butterflies only the satyr and parnassian butterflies make cocoons; all others pupate naked, with the pupa (chrysalis) hanging from a silk pad by a stalk (cremaster). The chrysalids of some butterflies (pierids, swallowtails, and lycaenids) are additionally supported by a threadlike silk girdle about the body.

Pupa. The duration of the pupal stage differs greatly in various groups and sometimes in different generations of a single species. It may last only a week or ten days, in small, fast developing species in summer, or for many months, in cases of hibernation or estivation. Pupae have been known to remain alive for three years in abnormal conditions, the adults finally emerging successfully. The danger of desiccation is greatest in small pupae and exarate ones (those in which the appendages are not fixed to the body by a skin or sheath) and least in large or obtect (compact) ones. Most obtect pupae can move only one or two abdominal segments. Those of borers can often wriggle along the tunnel to its exit, aided by rows of backward pointing abdominal spines that give purchase on the walls of the cocoon.

The adult. Although fully formed as an adult, the individual may remain quiescent within the pupal case for a long time, until conditions are right for its emergence. In some groups, the pupa has cocoon cutters, such as the movable mandibles of certain primitive moths, or saw-toothed structures on the head (leaf blotch miner moths). In saturniid moths the adult has bladelike structures on the wing bases. In puss moths (Cerura) and others the cocoon is partially dissolved by alkaline secretions.

Once out of the cocoon, the adult crawls upward to where it can hang head up or back down. Small pads on the thorax contain the incompletely developed wings in a compressed state. The insect then forces its body fluids into the thorax by contracting its abdomen and pumps blood out into the small wing pads, expanding them to form the full wings. The adult may be able to fly in a few minutes or may have to hang for several hours until its wings have stiffened sufficiently for flight.

As the larvae is the nutritive stage of the life cycle, the adult is the reproductive stage. Its mobility is necessary to bring the males and females together for copulation and to disperse the species into new areas. Nutrition is also an essential adult function in many primitive groups, such as the Micropterigidae. In the majority of lepidopterans it is important, but not essential, although in most of the highly mobile species much of the energy necessary for flight is obtained from nectar or other liquid foods taken in by the adult. The most advanced condition in this respect is found in the groups in which the adult mouthparts have become so reduced that they cannot function; here all the nutrition is unquestionably concentrated in the larval stage.

Behaviour. Food selection by the *adult*. Adults locate their food sources by both sight and scent, the former being especially important in the diurnal species, the latter in many diurnal species and in the nocturnal ones. The chief source is floral nectar, but sap (especially if fermenting), overripe fruits, homopteran honeydew, fecal matter, and carrion are sometimes used, usually by special groups. There are mutualistic relationships of a broad sort between the species of Lepidoptera with flower-visiting adults and the plants whose flowers they visit and pollenate, but these are seldom specific or obligate, since only rarely are the plant and the lepidopteran mutually dependent. Exceptions are some members of the orchid and morning glory families with very deep, tubular flowers. These appear to be pollinated only by certain hawkmoths with very long proboscises ("tongues"). The mutualism of the yucca moths (Prodoxidae) and yucca plants is obligate; the moth larvae feed only in yucca fruits and the latter can develop only from moth-pollinated flowers. The female yucca moth has special tentacles on the first maxillae, with which it gathers and carries balls of pollen.

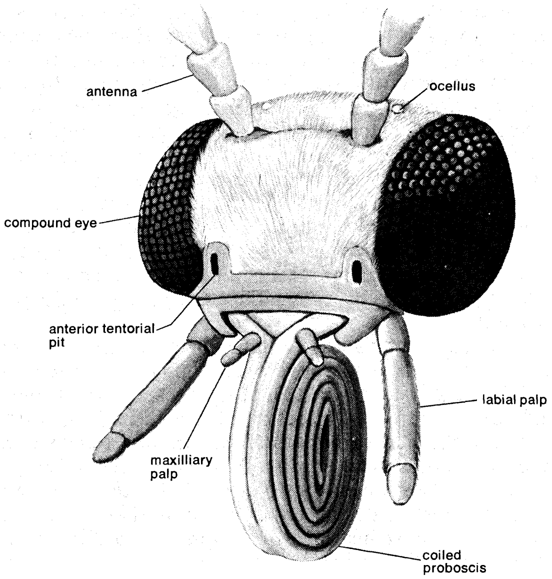


Figure 3: **Head and** mouthparts of **Lepidoptera.**
From A. Imms, *Textbook of General Entomology,* Methuen & *Co.* Ltd.

Use of odours in courtship

Courtship and mating. The sequence of mating activities is usually initiated by the female, who gives off specific odorous substances (pheromones), sometimes even before she emerges from the pupa. These are detected by scent sensilla on the male's antennae. Males with very large, feathery antennae, such as saturniid moths, can locate females from some miles away and may form courting swarms about them. A species may have a particular time of day or night for this release of the female pheromones. Males of some species, such as the European ghost moth, Hepialus humuli, may at dusk form dancing swarms, into which the females fly.

Once the male has located the female, a more or less elaborate courtship may ensue. The sight of characteristic colours and patterns may be a requisite for this, as in the bright butterflies. Male pheromones also play an essential part. Distributed from special scent scales (androconia) on the wings, body or legs, the pheromones ensure the receptivity of the female. Finally the accessory genitalic structures must fit together, not only mechanically, but also so as to stimulate sensory nerve organs of the female. Some butterfly courtships are very active (*e.g.*, in sulphurs) the male and female flying high in the air, with elaborate behavioral sequences in which each act serves as a releaser, a stimulus for the next act. Such highly specific sequences are important barriers to hybridization between otherwise similar species.

The males of many species show definite territorial behaviour, defending a particular perch and area against the intrusion of other males, and darting out at passing females. Among butterflies, certain admirals (Limenitis), coppers (Lycaena), and hackberry butterflies (*Asterocampa*) are noted for this. Congregations of males of many butterflies about hilltops may have a certain territorial and premating function. So may the aggregations, sometimes very large, of many butterflies in wet places. In many instances these consist of young, unmated males that come together because of a mutual attraction, not merely to sip water.

Migration. Many Lepidoptera are famous migrants. The American monarch butterfly (Danaus plexippus) is the only species known to perform an annual two-way migration, the same individuals flying southward in the autumn and northward in the spring. The species has also crossed the Pacific Ocean, colonizing Hawaii and Australia, and occasionally reaches Africa and Europe. The cosmopolitan painted lady (Vanessa cardui) stages mass flights nearly everywhere it occurs; these are one-way flights with no returns and must be classed as emigrations, rather than true migrations. Many other species in Europe and North America fly northward in large numbers, often reaching regions in which they cannot survive the winters.

In Europe the painted lady and many moths reach Britain and Scandinavia from central Europe. In North America the painted lady, the cloudless sulphur butterfly (Phoebis eubule), and many noctuid moths often reach Canada. Many spectacular migrations occur in the tropics, in which swarms numbering many millions may fly out to sea and be lost. The best known species are pierid butterflies, but large migrations of the large diurnal moths *Urania leila* and *U. fulgens* have also been recorded. The usual explanation of such mass population movements—*i.e.*, the advantage of possibly extending the range of the species—is far from adequate in many such instances.

Ecology. Environmental hazards. As primary consumers of green plant material, the Lepidoptera are enormously important in food chains, not only because of the very large number of species in the order and the diversity of their food habits, but also because of the abundance of individuals. In most land environments they, in turn, are eaten by a host of other animals: predators, parasites, and scavengers. All stages in their life cycles are thus under continual attack.

The major invertebrate predators on Lepidoptera include centipedes, spiders, mantids, bugs, ground beetles, ants, and both social and solitary wasps. Important predators among vertebrates include toads and tree frogs, lizards, birds, rodents, bats, and monkeys. Some of the invertebrates locate their prey by scent, others by sight; most of the vertebrates hunt by sight, except for the bats, which hunt by acoustic echolocation (the so-called bat "sonar").

Predators and parasites

The chief groups of parasites that attack Lepidoptera are tachinid flies (Tachinidae) and many families of wasps, chiefly of the Ichneumonoidea, Chalcidoidea, and Cynipoidea. More precisely called parasitoid predators, they are probably more important than the direct predators. The female parasitoids locate suitable hosts, chiefly by scent, and lay their eggs in, on, or near them. The parasitoid larvae live inside their hosts, gradually feeding on the host tissues and almost invariably consuming them almost completely. Lepidoptera seem to have evolved few defenses against parasitoids, unless some of their toxic or repellent secretions serve to discourage them. The high reproductive rate is important as a counter against loss to parasitoids, as well as against other adverse factors.

Small, red trombiculid mites often ride about on adult Lepidoptera, probably doing them no harm. A few other mites, however—*e.g., Myrmonyssus*—live and breed in the tympanic cavities of noctuid moths, destroying their auditory structures. Curiously, they regularly settle in only one of a moth's two tympanic cavities and thus only half deafen it. It is believed that, by leaving the moth with one good ear, the mite reduces the likelihood of the moth (and hence the mite itself) being captured and eaten by a bat. Lepidoptera are subject to attack by a considerable number of protozoa, roundworms, bacteria, viruses, and fungi that affect the larvae chiefly during peaks of abundance and crowding. Some of these organisms have been used by man as a means of controlling injurious species.

Protection against danger. It is chiefly against the sight-hunting predators that the Lepidoptera have evolved a multiplicity of defense mechanisms that are unequalled by those of any other group of animals.

The adults of many groups, such as skippers, many butterflies, hawk moths and many underwing moths, have fast erratic flight. When escaping they dart or fall to cover and often remain immobile for some time. They have a good chance of escaping, especially if they are coloured like their surroundings. Larvae, especially when small, drop suddenly when disturbed, either dangling from a silk thread or falling to concealment on the ground. The larvae of some noctuids can jump several inches. Dense, loose hairs and scales make many moths slippery and may facilitate escape from sticky spider webs.

Certain noctuid moths and possibly some geometrids and pyralids, subject to predation by bats, are able to receive and identify the navigational sound pulses of the bats. Upon hearing bat pulses, these moths perform violent evasive flight movements (when the bat sound is loud, hence close) or dive to the ground (when the bat pulse is weaker, indicating that the danger is further away).

Targets, such as prominent colour spots or tails on the hindwings, attract the attention and focus the attacks of predators onto parts of the prey less vulnerable to injury. Such spots are likely to be seized and torn off, but this does the moth or butterfly no real harm and gives it time to escape without vital injury.

Concealment and camouflage

Many species manage to hide very effectively from predators. Many cutworms (Noctuidae) and other larvae hide in litter by day, feeding only at night. Many moths hide in crevices, often under loose bark, and some seem to have especially flattened bodies for this purpose. Hibernating butterflies spend the winter in hollow trees or hanging immobile among dead leaves. The larvae of a great many moths, of most skippers and of many butterflies, live in individual nests of rolled, folded, or webbed leaves or grass. Sod webworms live in silk-lined tunnels in turf. Wood borers, especially those in rootstocks and deep tunnels, are relatively secure. Many larvae aggregate in communal nests such as those of the tent caterpillars (Malacosoma), the larvae of the ermine moth (*Yponomeuta*), and those of the Mexican social white butterfly (*Eucheira* socialis). The larvae of the bagworms (Psychidae) and casebearers (Coleophoridae) live and pupate in individual, portable cases that are often masked with bits of leaf or twig. Some larvae, such as those of the green geometrid moths (*Synchlora*), attach bits of leaves or flower petals to themselves. Cocoons are frequently masked with leaves or debris; those of the puss moths and some dagger moths (Apatela) are hard, woody, and inconspicuous.

A cryptic (camouflaged) appearance conceals something only in appropriate surroundings, and only when accompanied by proper behaviour, usually immobility. Great numbers of larval, pupal, and adult Lepidoptera are thus protected in their natural environments, chiefly on or about plants. Leaf-eating larvae usually blend into leafy environments. Many caterpillars (*e.g.*, Sphingidae, Geometridae) have stripes that simulate leaf veins. The saw-toothed elm caterpillar (Nerice) has a jagged outline resembling the edge of an elm leaf. A great many measuring worms (Geometridae) are notably twiglike, with long, slender, stiffly held bodies. Many other caterpillars, especially Notodontidae, have irregular shapes that resemble twisted dead leaves. Likewise, many adult moths that rest during the day among leaves or on bark are cryptically coloured and patterned; moreover, their behavioral mechanisms lead them to select matching backgrounds on which to rest.

Many larvae and adults are disruptively marked; *i.e.*, have bold, contrasting patches or bands of colour that "break up" their outlines into two or more seemingly unrelated masses. Many adult moths and butterflies have "flash coloration," which serves to startle, and thus to delay momentarily, an attacker. Moths with cryptic forewings, and butterflies with cryptic undersides, show only these surfaces when they are at rest. If, when disturbed, they suddenly expose brilliantly coloured and marked hindwings, such as in undenving moths (*Catocala*), or upper surfaces (for example, morpho, hairstreak, and anglewing butterflies), the effect is startling. When the animal lands, the bright surfaces are suddenly hidden, causing it to "disappear." A similar "startle" effect protects larvae that have prominent spots simulating large eyes.

Startling sounds are produced by members of many groups. Larvae of hawkmoths (Sphingidae) and the pupae of many lycaenid butterflies make squeaking or grating sounds when disturbed. The adult death's head moth (Acherontia atropos) makes a loud chirping sound. *Ageronia* butterflies (Nymphalidae), when startled into flight, make a loud, clicking sound by means of a structure on the wings. This may have a startling, and therefore a delaying, effect on a predator.

Chemical defenses

Certain butterflies and moths possess repellent or toxic substances that provide protection against predators. Sometimes these are secured directly from the plant on which the larva feeds, such as the toxins (glycosides) that occur in high concentration in many milkweeds eaten by danaid butterflies. More often, the toxin is secreted by the insect itself. The toxin often occurs generally in the blood (*e.g.*, hydrogen cyanide in zygaenid moths) or in the gut, or may be the product of special glands, which release it at the time of an attack. Tiger moths (Arctiidae) give off bubbling drops of repellent from glands on the prothorax. Many groups show autohemorrhization, or reflex bleeding, from leg and body joints when disturbed. The larvae of swallowtail butterflies (Papilio) and tussock moths (Liparidae) give off strong smelling, volatile substances from extrusible scent organs (osmeteria). Many notodontid moth larvae spray formic acid from ventral prothoracic glands. Many larvae and some adults introduce toxins that cause severely painful nettling effects by means of hollow, barbed hairs; the slug caterpillar moths (Eucleidae), flannel moths (Megalopygidae), io moths (Saturniidae), and some liparid moths are noted for this. A few adult moths (*e.g.*, Arctia *caja*, some Sphingidae) inject toxins through sharp spines on their hind legs. Besides those already mentioned, the chief groups that are chemically protected are the moth families Ctenuchidae, Pericopidae, and Dioptidae and the butterfly groups Heliconiinae, Ithomiinae, Acraeinae, and the swallowtails (*Battus*) that feed on birthwort. The great majority of these and other protected forms are aposematic; *i.e.*, have markings, shapes, and behaviour that draw attention. They are thus easily recognized and remembered by predators which, after seizing only one or two individuals, will thereafter leave other similar ones alone.

The protective advantage gained by a distasteful or dangerous insect and accompanied by aposematic coloration or acoustic warning may be utilized by harmless and edible insects, through the evolution of close resemblances, known as mimicry. (For background information of this phenomenon, see MIMICRY.) The distasteful insect, known as the model, may even be a member of an insect order different from that of the mimic. Members of various lepidopteran families mimic wasps, bees, and beetles. The clearwing moths (Aegeriidae) are particularly effective mimics of certain stinging wasps, the resemblance being carried to details of the shape and coloration of the wings, abdomen, and legs. Mimicry occurs widely in the moth families Dioptidae, Chalcosiidae, Callidulidae, and Zygaenidae and in the butterfly families Nymphalidae, Papilionidae, Pieridae, and Riodinidae.

Not all warning mechanisms are visual. Inedible arctiid and pericopid moths make high-pitched grating sounds by means of timbal (drumlike) organs. These sounds, above the frequency range audible to man, but within that audible to bats, may function to warn the bat of the moth's inedibility, avoiding capture and tasting that would be injurious to the moth. Some authorities believe that these sounds may interfere with the bats' acoustic orientation system, preventing the detection of the moth.

Polymorphism

The occurrence in a population of two or more distinct hereditary variants, or morphs, is known in many Lepidoptera. Each morph may have a different adaptive value, linked with such physiological features as resistance to cold, or to toxins in the environment. Striking variation in appearance may have great adaptive value by confusing predators, making it more difficult to learn the appearance of the prey. Mixed populations of both light and dark (melanic) individuals may survive better in mixed light and dark environments, and actually be prepared for a shift of the environment to a predominance of dark, through the normal succession of forest growth or through man-made industrial pollution. In England investigations of the peppered moth Biston betularia have abundantly documented the evolution of "industrial" and "natural" melanism and shown major genetic population changes taking place very rapidly.

Striking polymorphisms occur in some mimetic species, notably the African swallowtail, Papilio dardanus. The occurrence of different species of inedible butterflies ("models") in various regions has been accompanied by the evolution of correspondingly different mimetic females of the single species of swallowtail. The North American tiger swallowtail, P. glaucus, has mostly black females where it coexists with the black distasteful pipevine swallowtail, *Battus* philenor. Where B. philenor does not occur, however, the P. glaucus females tend to be all nonmimetic yellow like the males, black having no protective significance. Some very striking mimetic polymorphism occur among Neotropical Heliconius butterflies and their various models and mimics.

FORM AND FUNCTION

Within the limitations imposed by their almost entirely planteating habits, the Lepidoptera show a great diversity of size, structure, and many other characteristics. Furthermore, some members of the order have retained many primitive features, while others have evolved very advanced and specialized ones. This broad range, from primitive to advanced and from generalized to specialized, is evident not only in the adults but also in the larvae and pupae, and to some degree in the eggs.

The adult. Nearly all external surfaces of the adult animal are covered with scales, which may be broad and flat or very long and hairlike. Each of these is the outgrowth of a single epidermal cell. Similar scales occur in a few groups of other insects, but never to the same extent.

Like most insects, adult lepidopterans have three distinct body segments, or tagmata, the head, the thorax, and the abdomen, each with special functions (Figure 1). The head bears the main sensory organs and those of food-getting and ingestion. The thorax is chiefly concerned with locomotion. The abdomen contains the chief organs of digestion, excretion, and reproduction and bears the external accessory reproductive structures.

Head. The head is relatively small and round or elliptical. Evolutionarily, it is derived from the six anterior somites or metameres (the primitive body segments), but these have become so coalesced that none of the primitive segmentation is evident. The antennae are prominent and many segmented, with many microscopic receptors for detecting odours. In most moths the antennae are slender and tapering; in a few moths the surface area of the antennae is greatly enlarged by many side branches that make them comblike or featherlike. In skippers and butterflies the terminal part is broadened, forming a "club," on which most of the sensilla are concentrated. In many families there is also a cluster of sensory bristles, the chaetosema, on each side of the head, near the eye. On either side of the head is a large compound eye, sometimes consisting of thousands of units (ommatidia). Most moths have, in addition to the compound eyes, a pair of very small, simple eyes (ocelli), which have a very limited function.

Sense organs

The compound eyes are very efficient for distinguishing motion, but not for resolving clear images of distant objects. Diurnal species may be able to distinguish flower shapes, and courting individuals respond to specific wing patterns. Colour perception, especially in the blue-violet end of the spectrum, is acute in some species. Moths are generally able to see by ultraviolet, to which they are often strongly attracted at night.

The sense of smell is acute in some groups, especially in males with large antennal surfaces. The sense of taste is also acute, especially for sugars, through receptors located chiefly on the palpi and on the soles of the tarsi (feet).

Hearing is acute in the moths, with special membranes (tympana) most of which are located in the thorax, some of which react to sound waves above the range of human hearing. Mechanical information, including sound, is also received by chordotonal (stretch) receptors in a number of locations, especially in the antennae.

A relatively primitive set of chewing mouthparts occurs in the adults of the most primitive moths (Micropterigidae) as well as with larvae of nearly all Lepidoptera. Basically these consist of an anterior flap, the labmm; a pair of chewing jaws, the mandibles; a pair of complex first maxillae; and a pair of similar second maxillae joined together behind the mouth to form a structure called the labium. Each of the first and second maxillae bears a jointed, sensory appendage, or palpus, the maxillary and labial palpi respectively. Such mouthparts function for chewing and manipulating solid foods. In the vast majority of adults the mandibles are either vestigial and nonfunctional or absent. Parts of the first maxillae (the galeae) are elongated to form the two halves of a tubular proboscis, or haustellum, through which liquids may be sucked. The segmented palpi of the first and second maxillae are present and function as sensory organs. Not all adults, however, have all of these parts fully formed and functional. In numerous families the proboscis has become considerably reduced and even vestigial, so that the adult cannot feed. In advanced moths and in skippers and butterflies, the maxillary palpi are vestigial or lost, so that only the labial palpi remain functional.

Mouthpart reductions appear to have occurred independently in a number of evolutionary lines, the result being to limit feeding to the larval stage and to enable the adult to concentrate on reproduction and dispersal.

Thorax. The thorax consists of three segments, the prothorax, mesothorax, and metathorax, each derived from a primitive segment (metamere). The prothorax bears the first pair of legs and a pair of respiratory apertures (spiracles). The much larger mesothorax bears the second pair of legs and the pair of forewings, The metathorax bears the third pair of legs and the pair of hindwings. In many moths the metathorax bears a pair of complex auditory organs (tympana), and in many of these it also bears a pair of sound-producing timbal organs.

The wings begin development in the maturing larva as sacs infolding (invaginating) from the epidermis. As the pupa is formed these outfold (evaginate) to lie externally and become large and flat. Within them, branching tubes (tracheae) carry an air supply and also mark the pathways along which will develop the tubular "veins" that support the fully formed wings. The wings are expanded to full size when the adult emerges from the pupa. On the wing the scales lie in rows, overlapping like shingles. The usual scale is a flattened, rigid, air-filled sac (Figure 4) attached by a peglike base. It is usually ribbed longitudimally and toothed terminally. Many males have special, glandular scent scales (androconia) scattered or concentrated in patches (brands or stigmata) on the wings, sometimes forming expansible, hairlike tufts. These have a scent-distributing function essential in specialized courtship. Very primitive moths have the wing membrane aculeate; *i.e.*, more or less covered with microscopic bristles (setae or aculeae).

H.E. Hinton

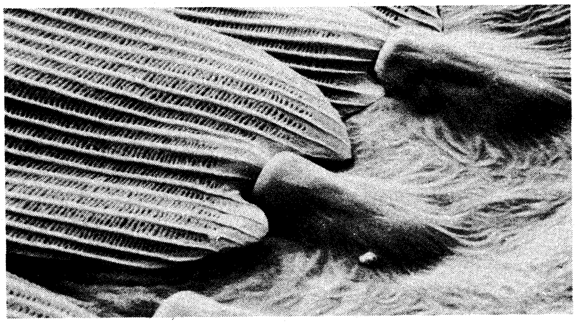


Figure 4: Detail of scales from the wing of butterfly, *Danaus plexippus* (magnified about 2,600 X).

The scales and hairs of the adults, especially on the wings, are responsible for colours and patterns. There are many different pigments of several chemically diverse types such as melanins, uric acid derivatives, and flavones. In addition, the microstructure of hairs and scales refract light in various ways so as to produce colours as well as metallic, pearly, iridescent, and white effects. Very fine parallel ridges in scales may produce an iridescence by acting as diffraction gratings. More common are the effects of very thin, superimposed layers (laminae) in the walls of the scales, a colour phenomenon comparable to that produced by a thin film of oil on water. Many of the most striking effects are caused by combinations of pigmental and structural colours.

Coupling of the fore and hind wings

The fore and hind wings on each side are coupled together in various ways. In primitive moths a fingerlike lobe (jugum, or fibula) on the forewing overlaps the base of the hindwing. In most moths a strong bristle or cluster of bristles (frenulum) near the base of the hindwing engages a catch (retinaculum) on the forewing. In some moths and in the skippers and butterflies the frenulum mechanism has been lost, and the wings are coordinated by the friction of the overlapping areas. In the most primitive moths the forewings and hindwings are similar in size, shape, and veins. In most moths and in the skippers and butterflies reduction of the hindwings has taken place; they have become shorter and more rounded, with reduced veining except in the anal cell area (the posterior wing section). The costal (anterior) edge of the forewing is thickened, with stronger veins, while the outer and anal (posterior) margins are thinner and weaker. This accords with the function of the wings as airfoils, the costal margin being the leading edge of the wing.

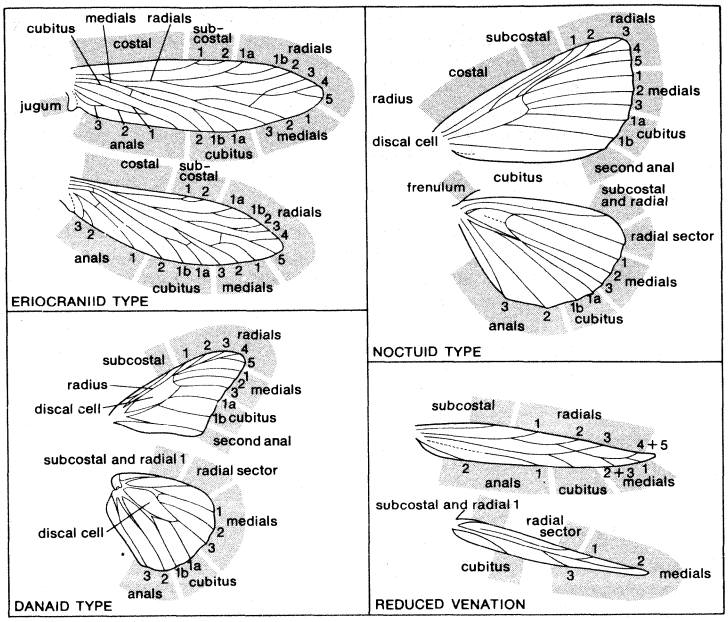


Figure 5: Types of wing venation.

From (top left) J. Comstock, *The Wings of Insects*, Comstock Publishing Co.; (top right) *An Introduction to the Study of Insects*, 3rd ed. by Donald J. Borror and Dwight M. DeLong, copyright © 1964, 1971 by Holt, Rinehart and Winston, Inc., copyright 1954 by Donald J. Borror and Dwight M. DeLong, reproduced by permission of Holt, Rinehart and Winston, Inc.; (bottom right) J. Comstock, *An Introduction to Entomology* (1956), Comstock Publishing Co.

The wings are characteristically considerably modified in various families. In many moths of the superfamily Tineoidea both wings have become extremely narrow, with reduced, degenerate venation but with long, dense fringes along the margins, maintaining the functional wing area. In one family (Pterophoridae) the wings are deeply cleft into two or three narrow plumes; and in the Orneodidae each wing is cleft into six plumes. In a number of moth families the females are wingless, although the males are fully winged (Psychidae, some Liparidae). In the aquatic pyralid genus ***Acentropus*** some females are wingless, while the others and the males are winged.

There are eight main wing veins, each with a characteristic pattern. These are usually designated according to the modified Comstock-Needham system. The names of the veins (with their symbols in parentheses) and the usual number of branches of each (subscript designations) are as follows, in progression from the costal margin of the wing to the anal, or inner, margin:

Costa (C), along the anterior margin, no branches;

Subcosta (Sc), usually with no branches, rarely with two (Sc_1, Sc_2);

Radius (R), typically with five branches (R_1 - R_5);

Media (M), with three branches (M_1 - M_3; four in some other insects);

Cubitus (Cu), primitively with three branches (Cu_{1a}, Cu_{1b}, Cu_2);

Anal Veins, First (1A), Second (2A), and Third (3A), all unbranched.

Abdomen and genitalia. The abdomen has ten segments, although the posterior ones are indistinct. Each of the first eight segments bears a pair of spiracles. The first or second segments bear paired auditory organs in the pyralid and geometrid moths. Segmental appendages are absent except for vestiges that may form parts of the genitalia. Various segments may bear special structures that produce and disperse pheromones. The genitalia of both sexes are often complex and bear characteristic spines, teeth, setae, and scale tufts. These structures are important in complex courtships and matings, preventing hybridization between unsuitable males and females.

In males (Figure 6) a ringlike structure (vinculum) is the base of attachment for a number of dorsal structures (tegumen, uncus, and gnathos) and a pair of lateral clasping organs (valvae). In copulation a median, tubular organ (the aedeagus) is extended through an eversible sheath (vesica) to inseminate the female. These structures evolutionarily derive from the body-wall plates, or sclerites (lateral tergites and ventral sternites), of segments eight and 10 and from vestiges of abdominal appendages.

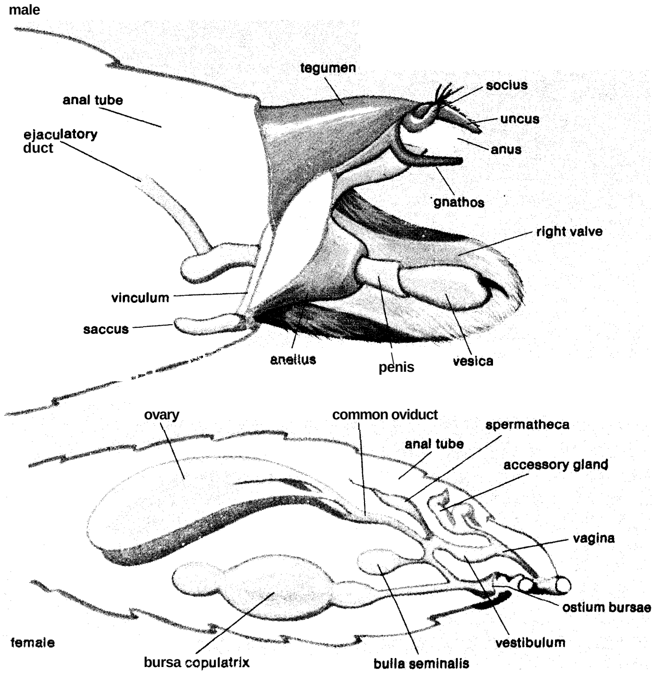


Figure 6: Genitalia and associated structures of the male and female Lepidoptera.

The female genitalia exhibit a number of different patterns of the internal ducts and the openings, varying from a condition in which there are no special genital openings, insemination and egglaying (oviposition) taking place through a single aperture, shared with the excretory system, to one in which there are two specialized openings, one for insemination and one for oviposition, both distinct from the anus.

Internal reproductive system

The internal reproductive systems of both sexes contain the organs typical of most insects. The testes of the male are paired in primitive lepidopterans but fused into a single organ in advanced forms. In both cases, the sperm ducts are paired. As in other insects, the sperm pass from the testes down paired ducts (vasa deferentia) for storage in sacs called seminal vesicles. Accessory glands, providing fluids that lengthen the life of the sperm, open into the vasa deferentia.

The female reproductive system consists of paired ovaries, paired accessory glands that provide the yolks and shells of the eggs, and a system of receptacles and ducts for receiving, conducting, and storing sperm. The individual oviducts join to form a common oviduct that leads to the vagina. In copulation, the male deposits a sperm capsule (spermatophore) in a receptacle (bursa copulatrix) of the female. The spermatophore releases the sperm, which swim into the oviduct and thence to the seminal receptacle (bulla seminalis) where they are stored until egg laying, which may be hours, days, or months after mating.

The egg. The egg is enclosed in a protective shell, the chorion, through which a system of tiny canals, the micropyle, permits the entrance of sperm. In some groups the micropyle is at the side, in others on the surface away from the substrate. Formed in one of the paired ovaries, the egg passes along the individual oviduct and through the common oviduct to the vagina. Here, just before it is laid, it comes into contact with a droplet of seminal fluid that has been stored in the female. Fertilization now takes place. Most eggs are more or less spherical, but some (Eucleidae, Lycaenidae) are flat, while others (Pieridae) are long and tapered at the ends. The surface may be strongly sculptured with pits, sharp projections, or raised ridges.

The larva. Compared with the highly specialized adult, the larva is simple and primitive. Many of the primitive characters retained in the larva are important in the classification of the suborders, superfamilies, and families, and in phyletic speculations. The head bears a pair of very short antennae and on each side a cluster of minute, simple eyes (ocelli or stemmata). A short liplike labrum lies anterior to the mouth. Behind the labrum are paired jaws (mandibles), short, broad and powerful. Next is a pair of small first maxillae, each with a segmented palp; and then, more or less connected with the maxillae, the labium-hypopharynx, a complex structure with a pair of labial palps and a median, tubular spinneret (a protrusion through which silk is extruded). Each of the three thoracic segments bears a pair of short, segmented legs. The remaining ten segments constitute the abdomen. Each of segments three through six and ten bears a pair of fleshy appendages called prolegs, which may be homologous with the primitive segmental appendages. Each proleg has one or two curved rows of minute hooklets (crochets) and an eversible, soft end, the planta. The prothorax and abdominal segments one to eight have each a pair of spiracles.

External ornamentation of the larva

The larval epidermis bears a number of bristles, known as primary setae, on the head and each body segment. These setae are constant in position and number in each species and important in classification. There are often many secondary setae, sometimes forming a dense, hairy vestiture. Larvae that live and feed as borers, burrowers, or miners are mostly plain. Those that live and feed in the open often show a great diversity of shape, colour, pattern, and ornamentation. Many have wartlike projections (verrucae) that may bear tufts of setae or spiny projections; or there may be prominent hornlike or spiny processes (*e.g.*, Saturniidae, Sphingidae) or fleshy filaments (Danainae).

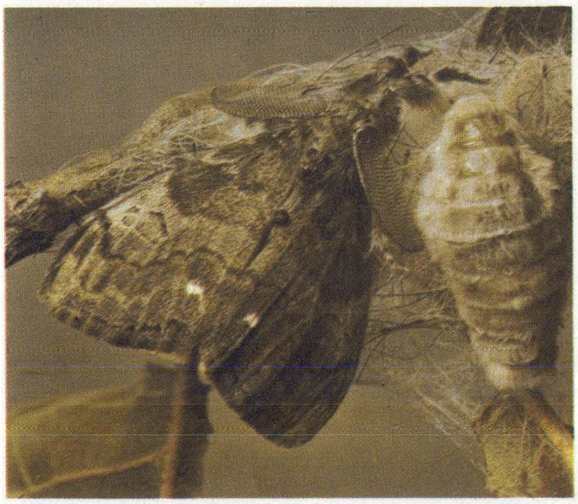
In some families the number of the prolegs has been modified. In the measuring worms (larvae of the Geometridae) the prolegs of segments three, four, and five are missing, and in some noctuid moths one or two pairs have been lost. In Drepanidae and some Noctuidae the last pair has been lost; and in the puss moth caterpillars this pair has evolved into a pair of long, eversible whiplashes (stemmatopoda). The larvae of some leaf miners (Gracillariidae, Lyonetiidae) have lost some or all of the prolegs. Micropterigid larvae have eight pairs of abdominal legs, which are more like true legs.

Internally, the larva is relatively simple, the very large digestive tract being the most prominent organ. The paired silk glands are often very large, extending far back into the abdomen. The ovaries and testes, which begin to develop during embryonic life, continue to develop in the larva, as do the wings. Many special glands secrete repellant or toxic substances, which may circulate in the blood or be extruded from special openings as a means of defense.

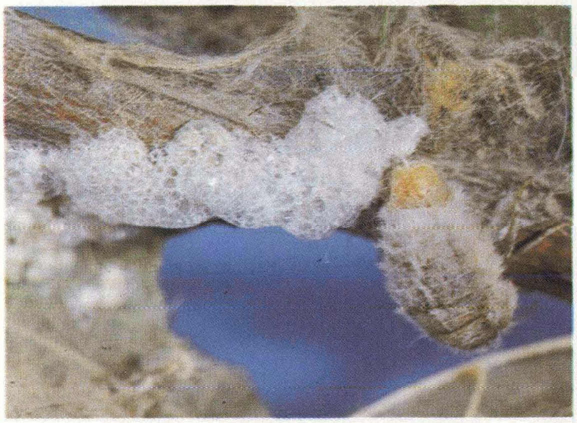
Larval vision can detect little more than differences between light and darkness. Taste is acute, depending on delicate sensory receptors in the antennae and palpi. Food discrimination is keen and many larvae will starve rather than eat abnormal food plants. The sense of touch depends on setae well distributed over the outer surface. Some of these appear to react to sound waves of low pitch, well within the limits of human hearing.

The pupa. Lepidopterous pupae show the same sort of evolutionary gradation from primitive to advanced as do larvae and adults. In the primitive Micropterigidae and

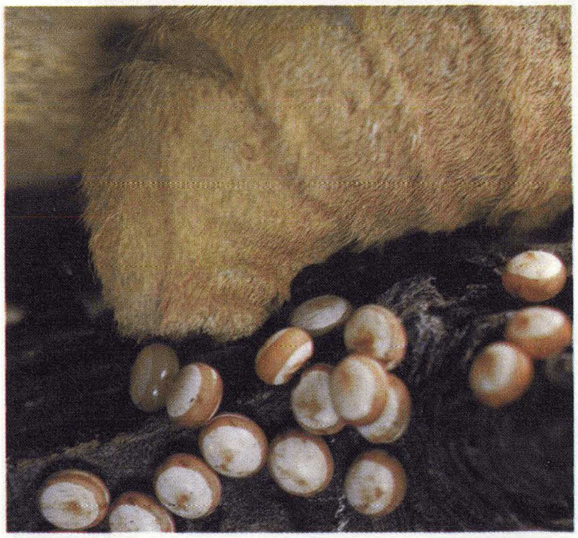
Mating and egg laying



Tussock moths (Notolophus *leucostigma)* on female cocoon. The wingless adult female lives solely for the purpose of oviposition.

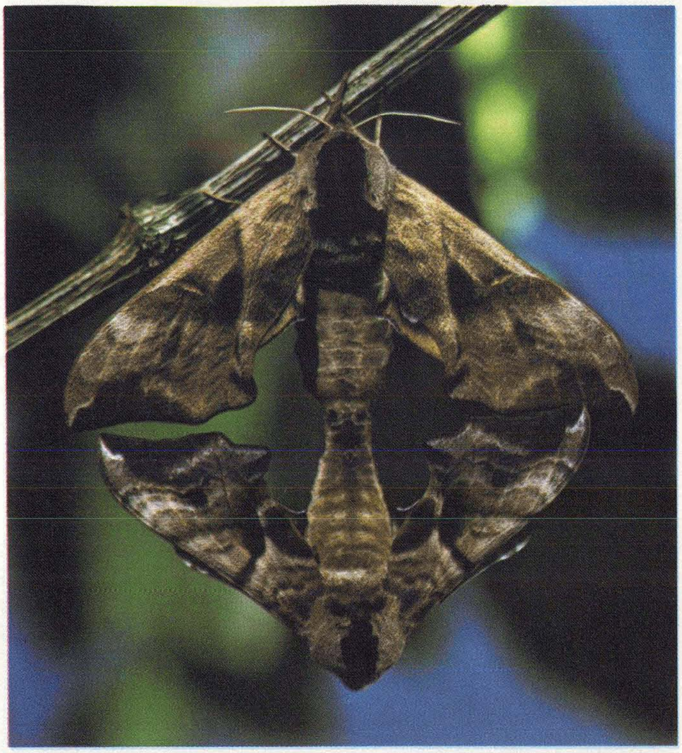


Female tussock moth laying eggs in masses on the surface of her old cocoon. She covers the eggs with a sticky secretion that becomes dry and brittle.

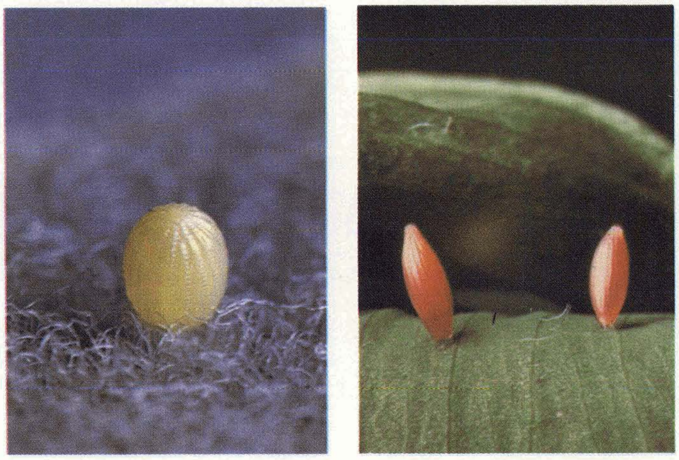


Polyphemus moth depositing eggs. Because the larvae feed on a variety of trees and shrubs, site selection for egg deposition is haphazard.

Plate 1: (Top left, centre, centre right, bottom right) Alexander B. Klots, (top right) Toni Angermayer—Photo Researchers, (bottom left) Louis Quitt—Photo Researchers. (centre left) Josephine Cikel—National Audubon Society



Eyed hawk moths *(Smerinthus ocellatus)* copulating. The female is at the top.



Lepidoptera eggs: (above left) monarch butterfly (Danaus *plexippus);* (above right) clouded sulphur butterfly *(Colias philodice);* (below) handmaden moth *(Datana ministra).*

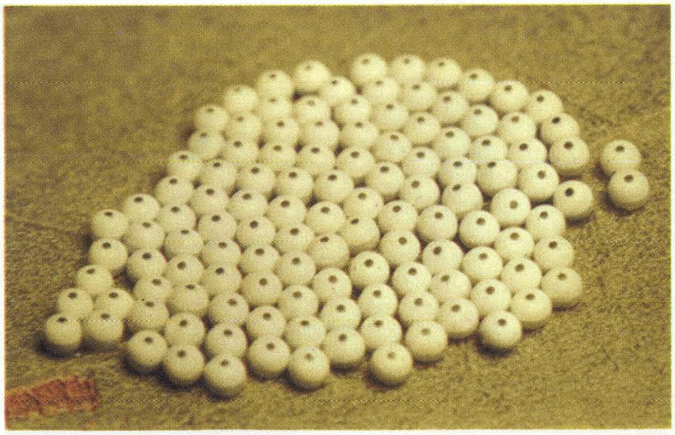
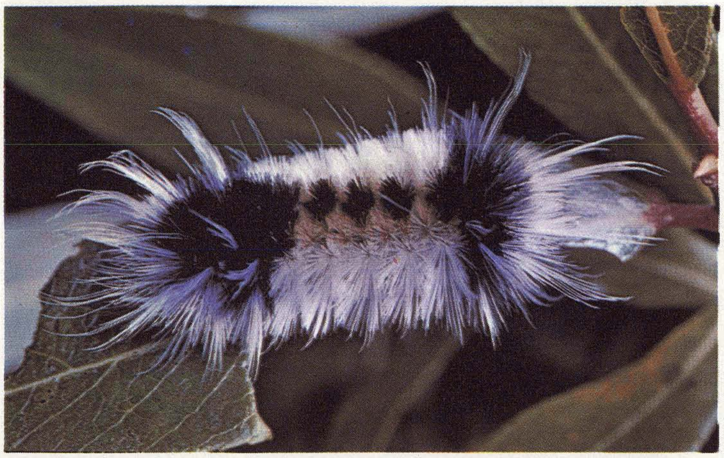
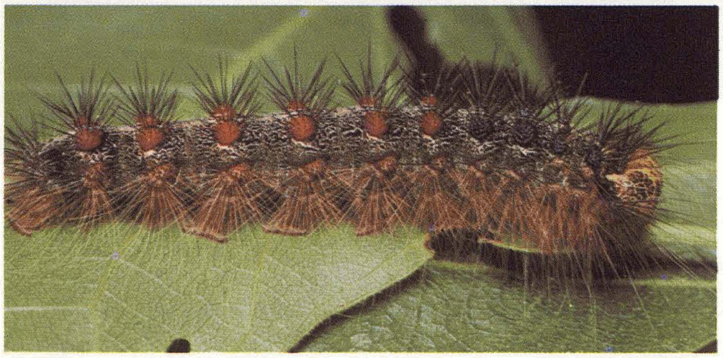


Plate 2 Lepidoptera



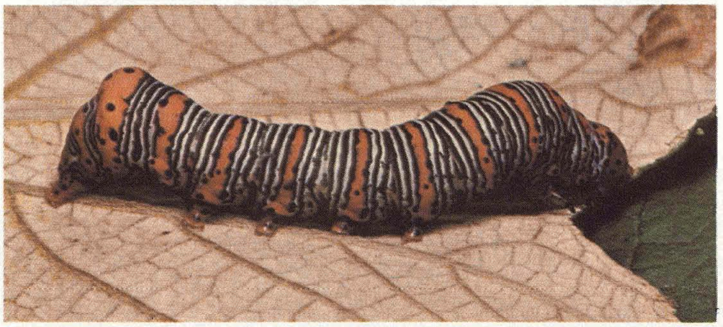
Woolly bear caterpillar (*Halisidota* argentata).



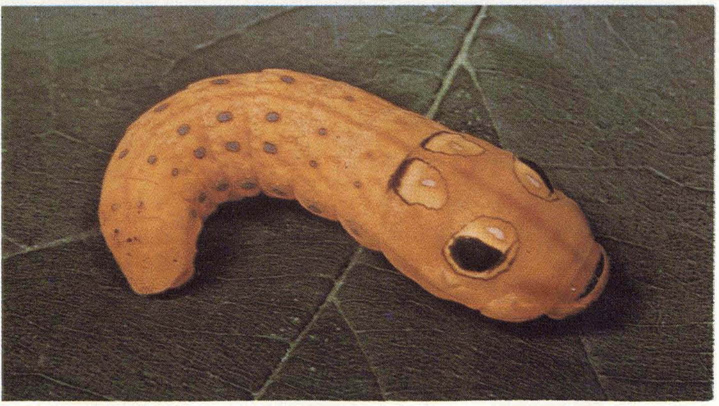
Gypsy moth (Porthetria, or *Lymantria, dispar)* larva, a species injurious to orchards and forests in the United States.

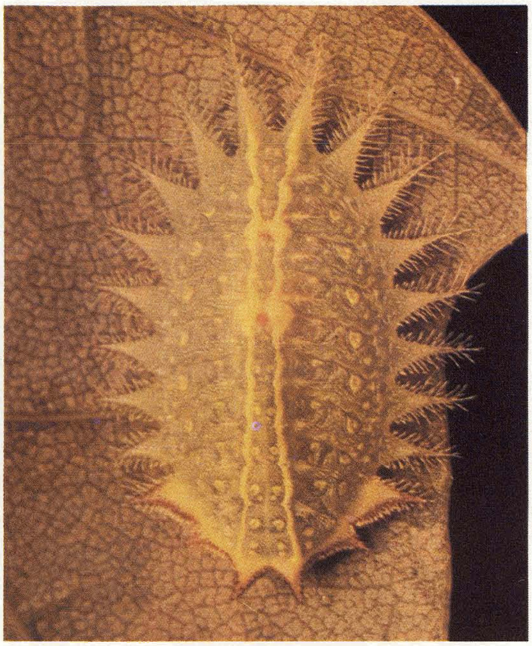
Larvae

Wood nymph (*Eudryas* grata) caterpillar.

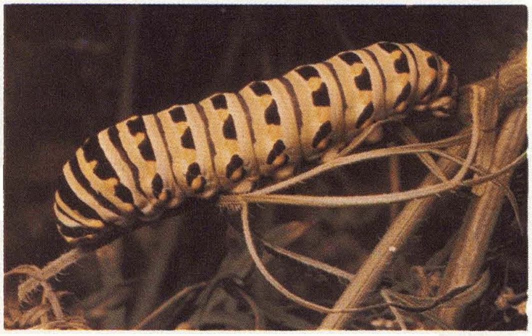


Caterpillar of the spicebush swallowtail (*Papilio troilus)*.



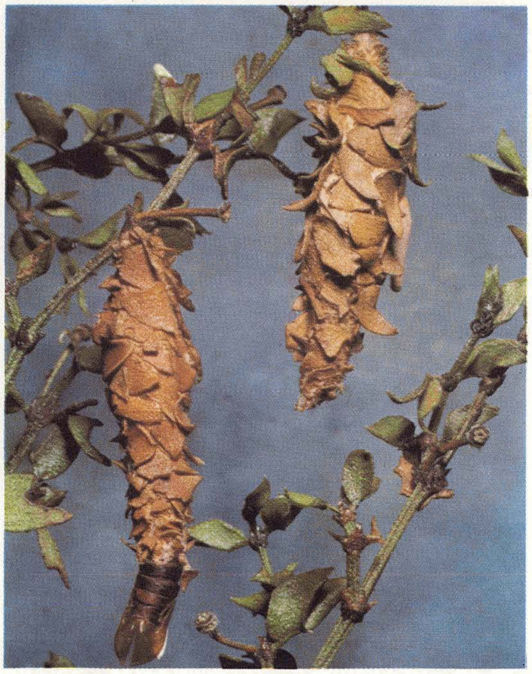


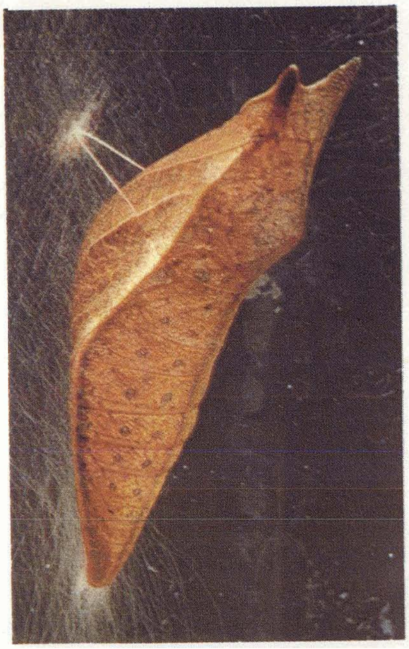
Slug caterpillar moth (*Sisyrosea textula)*.



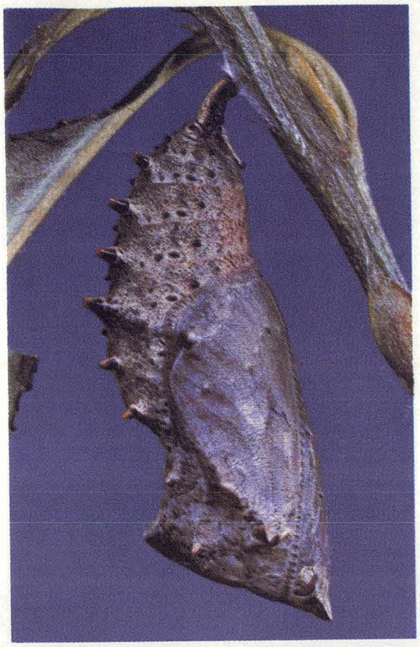
Black swallowtail (*Papilio* polyxenes *asterius)* larva.

Bags of the creosote bush bagworm moth (Thyridopteryx *meadi)*. Larvae feed inside cases attached to their backs with silk and covered with bits of leaves, grass, and twigs.

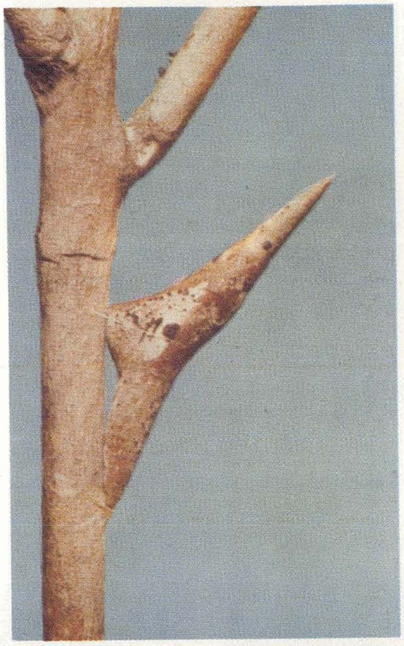




Spicebush swallowtail (*Papilio troilus)* pupa supported head upward by a girdle and cremaster (terminal abdominal spine).

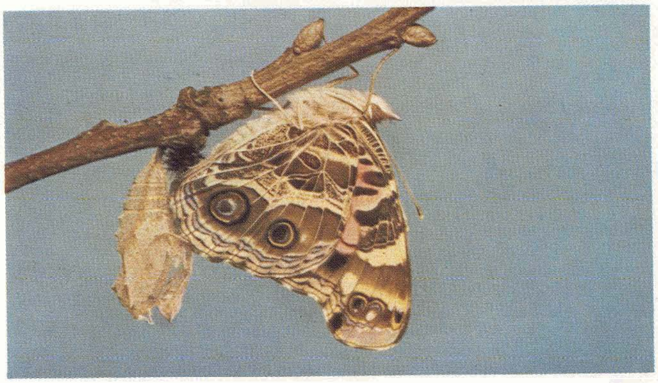


Pupa of the mourning cloak (*Nymphalis antiopa)* suspended head downward by only the tail hooks.

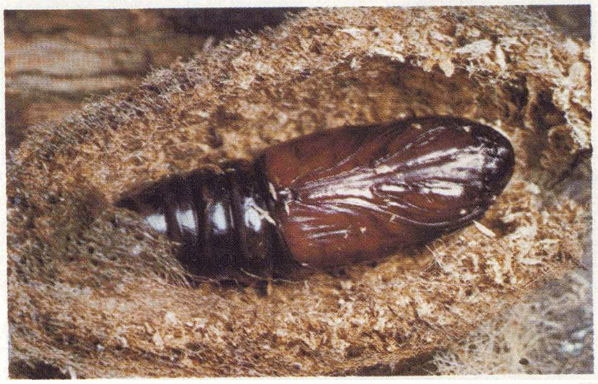


Pupa of an orange tip butterfly (*Anthocaris* midea) suspended by both the cremaster and a girdle.

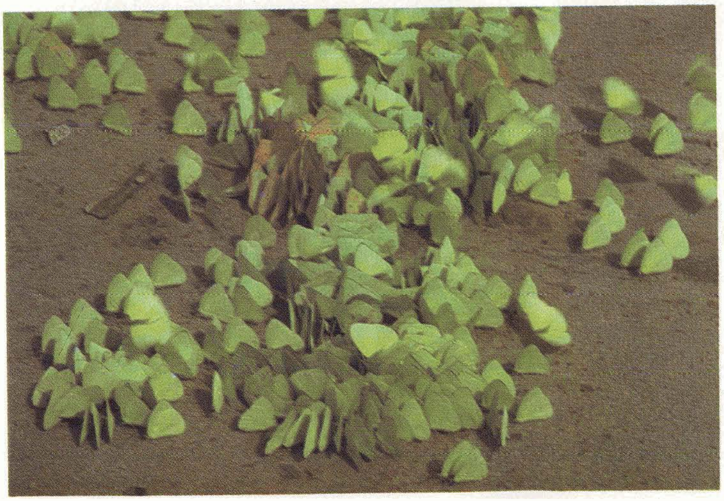
Pupae and aggregations

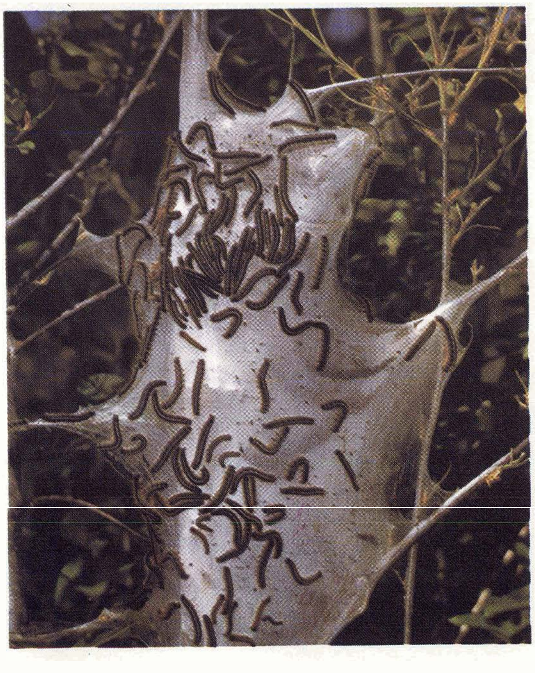


American painted lady (Vanessa *virginiensis)* emerging from its chrysalis.



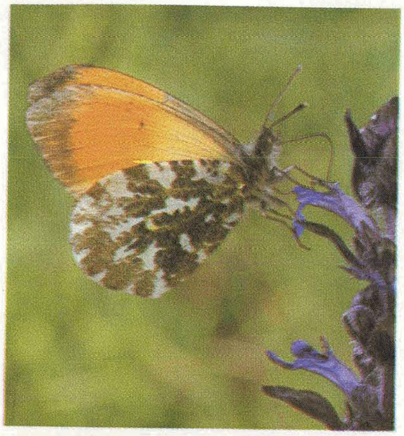
Cocoon of the Isabella tiger moth (*Isia isabella)* open to show pupa.



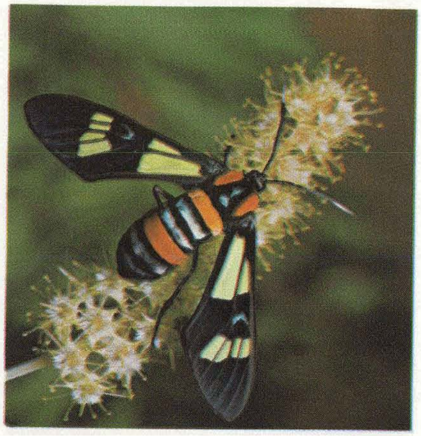


(Above) Aggregation of tropical American sulphur butterflies on a sandbar; (right) American eastern tent caterpillars (Malacosoma *americanum)* constructing a silken nest for shelter.

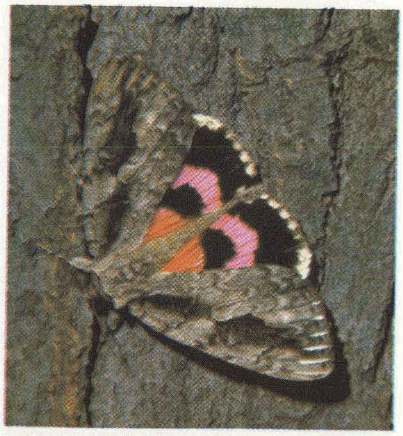
Plate 2 (Top left) E S. Ross, (top right, above centre left below centre left, centre right) Alexander B. Klots, (bottom left) John Gerard—National Audubon society. (bottom right) William E. Ferguson
Plate 3: (Top left) John Gerard, (top centre, centre right, bottom right) Louis Quitt—Photo Researchers, (bottom left) John Verde—Photo Researchers, (top right, centre left) Alexander B. Klots



An orange tip butterfly (Anthocaris cardamines).

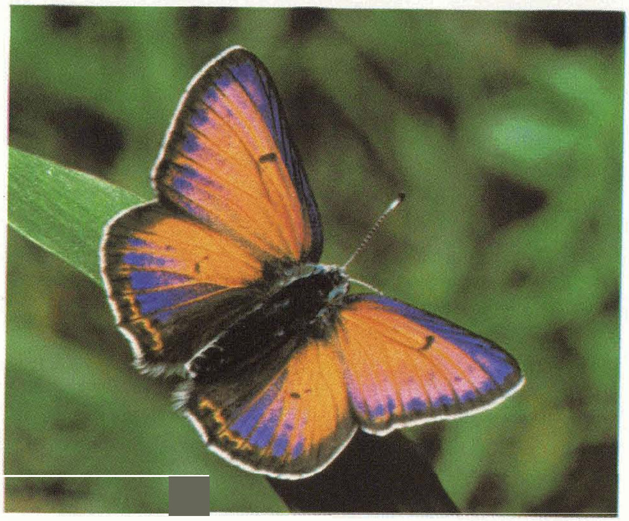


South African day-flying moth (*Euchromia formosa*).

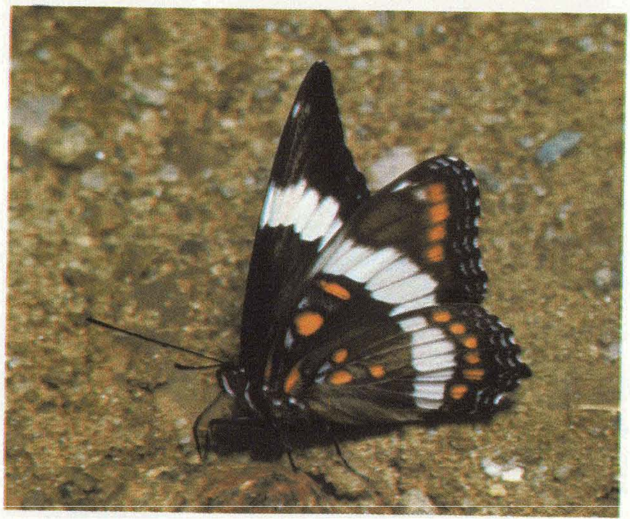


Underwing moth (Catocala).

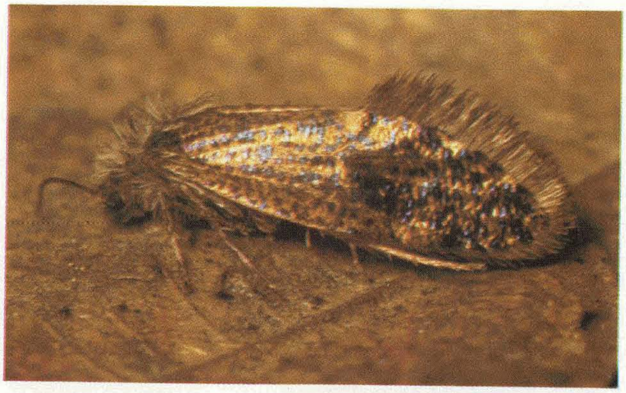
Colourful and unusual adults

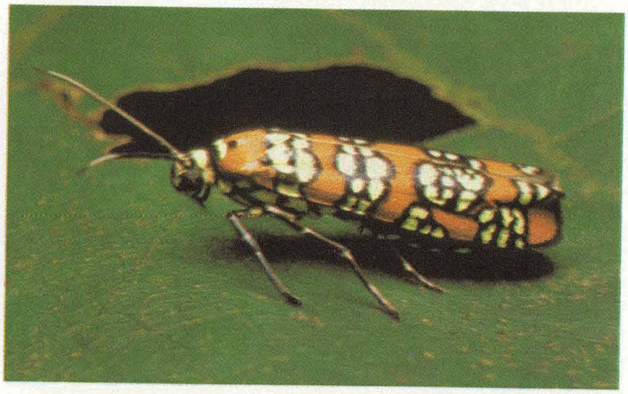


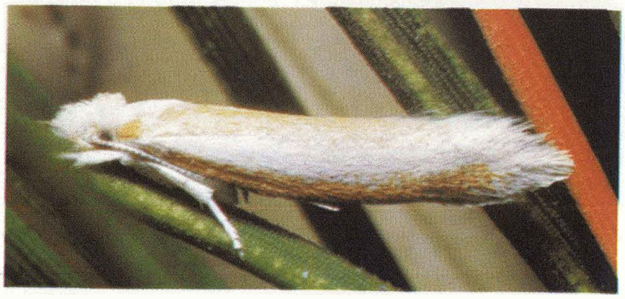
Palaeochrysophanus hippothoe, an iridescent butterfly.



Admiral butterfly, a common North American species.







(Above left) *Eriocrania* auricyanea, a primitive moth; (above) calico moth (Attava *punctella*); (left) ermine moth (*Zelleria* heimbachi).

Plate 4: (Top left, above centre left) Hermann Eisenbeiss—Photo Researchers, (above centre right) Ken Brate—Photo Researchers, (top centre) Anthony Bannister from The Natural History Photographic Agency—EB Inc., (top right) E.R. Degginger, (below centre left, below centre right) Alexander B. Klots, (bottom left) William E. Ferguson

Eriocraniidae the pupa is exarate, having the appendages free and movable, and decticous, having functional mandibles. In some less primitive groups the pupa, called incomplete, has retained considerable powers of motion of some appendages. In the higher moths and the butterflies the pupa, called obtect, has all appendages tightly fastened and immobile and is able to wriggle only one or two abdominal segments. In a few groups (Sphingidae, Lycaenidae) the pupa has special stridulating rasps for sound production. Nearly all of the external structures of the adult can be seen on the pupa. The wings are prominent, folded down flat along the ventral surface, with the proboscis halves, the legs, and the antennae between them. At the posterior end is a spiny pad or spike, the cremaster, which in many groups attaches the pupa to silk fibres spun by the larva.

Except for a very slight respiratory exchange and a little water loss, the pupa is physiologically almost self-sufficient. Within it most of the cells and tissues of the larva undergo considerable histolysis (breaking down), as the adult structures are built up from the existing rudiments. Some structures begin developing as far back as the first larval stages.

Growth, molting, and metamorphosis. As in other insects, growth and its structural changes are controlled by an interacting set of hormones. These are chiefly secreted by the corpora allata and other parts of the brain, and by paired prothoracic glands. The prothoracic gland hormone is necessary for larval molting, metamorphosis to the pupa, and formation of adult characteristics. On the other hand, a hormone secreted by the corpora allata inhibits metamorphosis until late larval development. A hormone secreted by neurosecretory cells in the pupal brain stimulates the prothoracic glands and thereby brings about differentiation of the adult and the end of the obligatory resting stage (diapause) of the pupa.

EVOLUTION AND PALEONTOLOGY

The Lepidoptera belong to an important group of insect orders, the panorpoid complex, the ancestral stem of which began in the Permian. This stem split into a number of branches, from which evolved the modem orders Mecoptera (scorpion flies), Megaloptera (dobsonflies), Neuroptera (lacewings), Diptera (true flies), Trichoptera (caddisflies), and the Lepidoptera. The nearest living relatives of the Lepidoptera are the caddisflies and, in fact, the very primitive moths, the Micropterigidae, have been grouped with the caddisflies by some students. As the Lepidoptera developed into a distinctive group certain major trends began to be emphasized. The adult mandibles disappeared and were replaced by the proboscis formed from the galeae of the maxillae, broadening the food base exploitable by the species to include not only sap and other plant juices but nectar and fruit juices as the flowering plants evolved. Various groups, however, independently lost the proboscis and concentrated on the larval stage as the sole source of nutrition. The characteristic vein pattern evolved, with the long, veinless discal cell in each wing, and an emphasis on the strength of the forewing and a de-emphasis of the hindwing, forming a particular pattern of aerodynamic efficiency. Related, also, was the replacement of the primitive jugum, for wing coupling, with the more efficient frenulum mechanism. However, some efficient flyers, such as the bombycoid moths, the skippers, and the butterflies, subsequently lost the frenulum; and quite a few groups of small moths, settling into restricted ecological niches where strong flight is not a necessity, evolved a great deal of wing reduction.

Evolution of the proboscis

Concurrent with changes in the adults, the larvae were evolving a multiplicity of different ways of feeding on the evolving seed plants. As both larval and adult differentiations accumulated, it became more and more necessary that the pupal stage be a passive, resting phase, to allow time for the metamorphosis of larva into adult.

Very few fossil Lepidoptera are known. The earliest undoubted ones, small moths related to the Eriocranioidea and Tineoidea, are from the Early Tertiary Period. The order must have existed long before then, however. The closely related caddisflies are known from the Jurassic Period of the Mesozoic Era. The best known fossil butterflies are relatively recent, being from the shales of Florissant, Colorado, generally considered to be Middle Tertiary. Some of these are not only undoubted Nymphalidae but are very close to modern genera.

CLASSIFICATION

Distinguishing Taxonomic Features. The chief characters used in the ordinal, subordinal, and family-level classification of the Lepidoptera are: the mouthparts, ranging from mandibulate to haustellate (with a proboscis) or degenerate; the venation and shape of the wings, homoneurous (the venation of the forewings and hindwings alike) or heteroneurous (forewings and hindwings different), aculeate (more or less covered with specialized bristles called microsetae) or nonaculeate, and type of coupling (jugum or frenulum); the anatomy of the female reproductive system; the exarate or obtect condition of the pupa; and the larval structure and pattern of the primary setae.

Wing venation

The venation of the wings is perhaps the most important single criterion for establishing both differences and relationships in the classification. It must be considered in terms of the evolution of venation changes from primitive to advanced conditions on a number of different phyletic lines. The most primitive groups (Figure 5, top left) tend to have the maximum number of veins and branches in each wing. More advanced groups (Figure 5, top right and bottom left) lack some veins or branches in the forewing and have the hindwing venation considerably reduced, especially by the reduction of free radial branches, the unbranched stem of R_2 - R_5, being called the radial sector (R_s).

Genitalia

The structures of the genitalia are extremely important in classification, often serving better than any other characteristics to distinguish the species. The various parts have

From A. Imms, Textbook of General *Entomology;* Methuen & Co. Ltd.

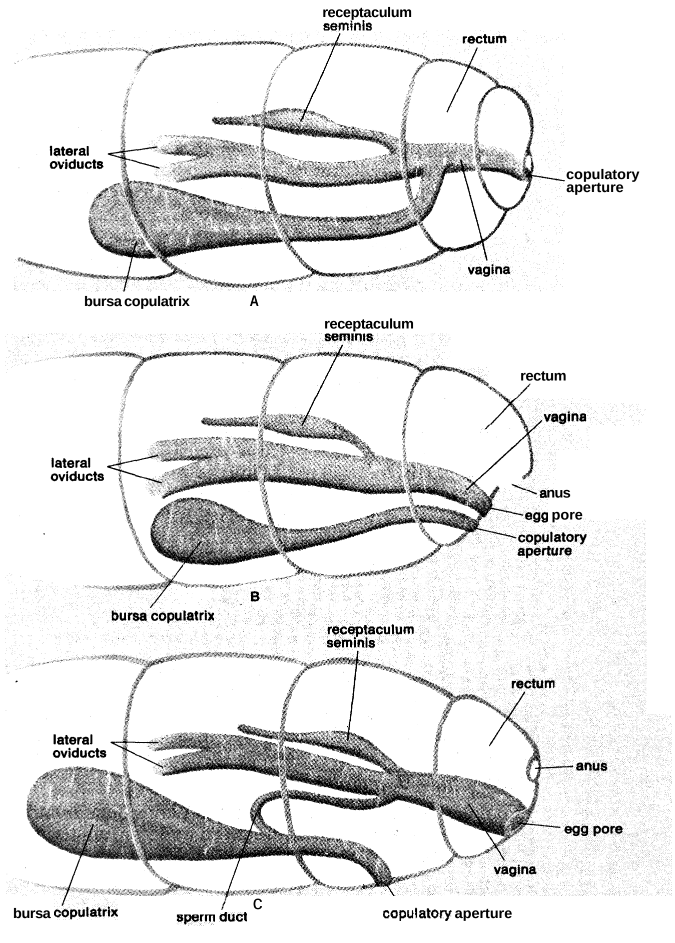


Figure 7: ***Relationships of female reproductive ducts and rectum in Lepidoptera.***
(A) Micropterigid type. (B) ***Hepialid type.*** (C) Ditrysian ***type.***

been homologized with each other, thus enabling comparative studies of families and genera as well as studies of species.

Types of genitalia

The female genitalia (Figures 6 and 7) show a number of different types of organization of the internal genital ducts and the openings. These are considered so fundamental that the subordinal classification is largely based on them, in correlation with characteristics of the mouthparts, wings, and early developmental stages. These genitalic types are as follows:

1. No special genital openings. Both insemination and oviposition take place through the cloaca, a chamber common to both the reproductive and excretory systems, the aperture being the anus.
2. A separate genital opening for insemination (reception of seminal fluid) but oviposition through the cloaca and anus.
3. Separate genital openings for insemination and oviposition, both close together and both close to the anus.
4. Separate openings for insemination and oviposition, the latter still close to the anus but the former removed some distance ventrally, in the eighth segment.

Type 1 characterizes the extremely primitive suborder Zeugloptera, the mandibulate Micropterigidae. Type 2 occurs in the superfamilies Eriocranioidea, Nepticuloidea and Incurvarioidea of the generally primitive suborder Monotrysia.

Type 3, sometimes called "exoporian," characterizes the monotrysian superfamily Hepialoidea, also quite primitive in other ways. Type 4 characterizes the most advanced suborder Ditrysia, containing the great majority of Lepidoptera.

Annotated Classification. The classification given here largely follows that of the British entomologist A.D. Imms, itself a synthesis of the suggestions of several earlier authors. This classification incorporates changes on the family level that have been suggested recently. There is no clear evidence for the phyletic relationships of numerous families, especially in the greatly reduced tineoid moths.

ORDER LEPIDOPTERA

Insects with complete metamorphosis; the wings covered with flat scales; larval mouthparts mandibulate; adult mouthparts mandibulate to haustellate, sometimes vestigial; wings with venation ranging from primitive (complex) to considerably reduced, especially in the hindwings; pupae exarate to obtect; almost all phytophagous (herbivorous).

Suborder Zeugloptera

Female with no special genital openings. Larvae, pupae, and adults functionally mandibulate; wings aculeate, homoneurous, jugate.

Family Micropterigidae (mandibulate moths). Small, a few species in the Palaearctic and Nearctic, more in Australia and New Zealand. The adults eat pollen; the larvae eat mosses and liverworts.

Suborder Monotrysia

Females with 1 or 2 genital openings near anus. Adults at least partly haustellate. Wings homoneurous to heteroneurous, at least partly aculeate. Jugum present in some, frenulum in others. As a group intermediate between Zeugloptera and Ditrysia.

Superfamily Eriocranioidea

Females with 1 genital opening, a long cloaca, and a piercing ovipositor. Wings homoneurous and aculeate. Adults with a short proboscis. Pupae with functional mandibles.

Family Eriocraniidae (eriocraniid moths). A small, widely distributed family, chiefly Nearctic and Palaearctic. Adults nectar feeding, often brilliantly coloured. Related families: Mnesarchiidae (New Zealand), Mesopseustidae (India and Taiwan).

Superfamily Hepialoidea

Females with 2 genital openings. Wings homoneurous and aculeate. Adult mouthparts reduced and nonfunctional, antennae very short.

Family Hepialidae (swifts and ghost moths). Medium-sized to very large moths, some brilliantly coloured. Larvae mostly borers in turf or wood. About 300 species; worldwide, chiefly in Australia and New Zealand. Related families, less known: Prototheoridae (Africa and Australia), Palaeosetidae (Australia), Anomosetidae (Australia).

Superfamily Nepticuloidea

Females with 1 genital opening and a soft ovipositor. Wings somewhat aculeate, heteroneurous, and somewhat narrowed and with reduced venation.

Family Nepticulidae (midget moths). Very small to minute moths. Antennae with broad "eyecaps" basally. Larvae mostly leaf and bark miners, a few gall makers. Worldwide. Also called Stigmellidae.

Families Opostegidae and Tischeriidae. Worldwide families of small moths with narrow, long-fringed wings. Larvae leaf, stem, or bark miners.

Superfamily Incurvarioidea

Family Incurvariidae (fairy moths). A small but worldwide family of small moths, many being brilliantly coloured diurnal flower visitors. Male antennae often several times as long as forewings. The mutualistic relationships of the yucca moths (Prodoxinae) with their food plants are biologically famous. Family sometimes split into families Incurvariidae, Adelidae and Prodoxidae. The related Heliozelidae (shield bearers) are worldwide.

Suborder Ditrysia

Females with 2 genital openings, the receptive one on the 8th abdominal segment. Wings strongly heteroneurous, nonaculeate, mostly with frenulum but this is degenerative or lost in some. Contains the great majority of Lepidoptera: higher moths, skippers, and butterflies.

Superfamily Zygaenoidea

Wings broad to very broad with short fringes and some primitive venational and larval features. Pupae quite primitively mobile.

Family Eucleidae (slug caterpillar moths). Larvae broad and flat ventrally, with reduced prolegs; move glidingly with head hidden beneath prothorax; many with toxic, irritant setae. Adults with heavy hairy bodies and vestigial proboscises. Also called Limacodidae.

Family Megalopygidae (flannel moths). Larvae like those of Eucleidae, but with prolegs normal, plus traces of some additional ones; setae very toxic and nettling. Almost limited to New World.

Family Zygaenidae (burnet and forester moths). Adults usually diurnal flower visitors, with bright colours and strong proboscises. Strongly protected by high concentrations of hydrogen cyanide in blood. Larvae, leaf skeletonizers. Related families: Aididae and Chalcosiidae (Old World tropics); Pyromorphidae and Dalceridae (New World).

Superfamily Tineoidea

A large group of families of mostly small moths of diverse habits. All have some primitive venation features and life cycles. Wings narrow to very narrow.

Family Tineidae (clothes moths and others). Small, narrow-winged moths with rough, hairy heads and often long, folded maxillary palpi. Larvae often casemakers, feeding on debris and fungi. Worldwide. Clothes moths (*Tineola, Tinea, Trichophaga)* often serious household pests. Related family: Acrolophidae, burrowing sod webworms.

Family Psychidae (bagworms). Larvae live and pupate in often elaborate cases. Adult males with broad, thin scaled wings; females wingless, often greatly degenerate and never leaving larval cases. Worldwide.

Family Epipyropidae (parasitic moths). Small family, chiefly Oriental. Larvae live as external parasites on plant hoppers (Homoptera). Related family: Cyclotornidae, Australian, larvae living similarly when young, then moving to ants' nests.

Family Coleophoridae (casebearer moths). Small, very narrow-winged moths. Larvae mostly mine leaves when young, then live in portable cases with distinctive shapes. Worldwide. Some are pests of fruit trees.

Families Gracillariidae, Lyonetiidae, Elachistidae and Douglasiidae. A group of worldwide families, totalling about 2000 species. Adults with narrow, long-fringed wings often with metallic markings. Larvae mostly leaf miners, sometimes greatly flattened, with degenerative legs and mandibles.

Superfamily Gelechioidea

A group of worldwide families totalling more than 12,000 species. Adults mostly larger and broader winged than Tineoidea. Larvae seldom leaf miners. Pupae obtect, relatively immobile.

Family Gelechiidae (gelechiid moths). Small to minute moths. Larvae diverse, mostly leaf tiers, rollers, and crumplers; stem, fruit, and tuber borers; a few gall makers. Hindwing often concave below apex. Worldwide. Some are serious economic pests; *i.e.*, the pink bollworm of cotton (Pectinophora *gossypiella*) and the Angoumois grain moth (Sitotroga *cerealella*).

Family Cosmopterygidae (cosmopterygid moths). Many adults are very narrow winged with bright, often metallic markings. Small moths. Larvae leaf miners; borers in stems, fruits, or seeds; leaf tiers; or scavengers. Related families; Lavernidae, Momphidae, Walshiidae.

Family Oecophoridae (oecophorid moths). Adnlts tend to be somewhat broader winged than related groups, and flat bodied; all small. Related families: Ethmiidae, Stenomatidae, Xylorictidae.

Family *Blastobasidae* (blastobasid moths). The adults small, often drab; reduced mouthparts. Larvae mostly scavengers, a few predators on scale insects.

Superfamily Yponomeutoidea

A limited, not very distinctive superfamily, the adults with close scaling and venation peculiarities. the larvae with distinctive primary setation.

Family Aegeriidae (wasp or clearwing moths). Almost worldwide: adults diurnal, flower visitors, often brightly coloured with yellow, orange, or scarlet, the wings usually mostly transparent, often very striking mimics of wasps. Larvae borers in stems, twigs, and rootstocks, often injurious; *e.g.*, the peach tree borer (*Sanninoidea exitiosa*). Also called Sesiidae.

Family Yponomeutidae (ermine moths). Adults brightly coloured, especially in tropics. Larvae often webbing leaves or living in shoots, buds, or fruits. Pupae of some make lacework cocoons. Widespread and not uncommon. Related families: Plutellidae, Glyphipterygidae, Heliodinidae, Scythrididae.

Superfamily Cossoidea

Family Cossidae (carpenterworm or goat moths). Medium to large moths. Adults with some primitive venation features, close-scaled wings, very heavy bodies, ranging up to 250 mm wing expanse, the heaviest Lepidoptera. Larvae wood borers, penetrating even solid heartwoods; may require at least 2 years to mature. Some species very destructive; *e.g.*, the coffee borer (Zeuzera *coffeae*) and leopard moth (Z. pyrina).

Superfamily Casfnioidea

Family Castniidae (castniid moths). A small family of medium sized to large diurnal moths of the New World and Indo-Australian tropics. Adults powerful, heavy bodied and broad winged; with clubbed antennae, bright colours, and often a strong mimetic resemblance to protected butterflies and diurnal moths. The larvae are often stem borers.

Superfamily Tortricoidea

The families included here comprise one of the major, worldwide groups. They have fairly broad, short-fringed wings that seldom span more than *25* mm. Maxillary palpi missing; labial palpi short. Most have cryptic colours and patterns. Larvae mostly leaf folders and rollers, but many bore in fruits and seeds and soft stems.

Family Tortricidae (leaf roller moths). Characteristics of the superfamily. Separated from Olethreutidae (below) by wing venation. Larvae of the green Tortrix viridana of Europe defoliate oak forests: those of the spruce budworm (*Choristoneura fumiferana*) are the worst forest pests of North America. Some species of Peronea and *Acleris* show most extraordinary polymorphisms of colour and pattern.

Family Olethrerrtidae (olethreutid moths). Separated from other tortricoid families by wing venation. Many are serious pests; *e.g.*, the pine shoot moths (Rhyacionia), the apple codling moth (Laspeyresia pomonella), and the oriental fruit moth (L. *molesta*). Chief subfamilies, the Olethreutinae, Eucosminae, and Laspeyresiinae, are often ranked as families. Related smaller families: Phaloniidae (chiefly Northern Hemisphere), Chlidanotidae (Indo-Australian and Neotropical) and Carposinidae (chiefly Indo-Australian, sometimes considered a separate superfamily).

Superfamily Pyralididoidea

An enormous group of about 20,000 species. Worldwide. Veins Sc and R_1 of the hindwings are typically close together or fused to beyond the cell. Most members have a pair of tympanal organs on the 1st abdominal segment. Adults usually slender bodied with long legs; many with narrow forewings and broad, often folded, hindwings.

Family Pyralididae (pyralid moths). Small moths, mostly plain, often abundant. Separated from other families by wing venation. Palps often large and held forward. Family contains many important pest species. Third largest family of the Lepidoptera. The subfamily Pyraustinae is the largest; most of the larvae live in rolled or tied leaves, or bore in soft stems or roots. The small subfamily Nymphulinae has aquatic larvae with tracheal gills, living in still or running freshwaters. The larvae of the Pyralidinae are mostly scavengers, as are those of the Galleriinae, many of which live in bee or wasp nests. Larvae of the Crambinae are either sod webworms or bore in soft stems. The larvae of the large subfamily Phycitinae have very diverse habits: most are leaf tiers or wehbers, but many are borers, others are scavengers, others pests in stored foods, and some are predators on scale insects. Other important subfamilies are: Chrysauginae, Schoenobiinae, Endotrichinae, and Ancylolominae. Many subfamilies have been separated as families. Probably related is the small, chiefly tropical family Thyrididae.

Family Pterophoridae (plume moths). Adults with very long, slender legs and bodies, the wings usually deeply cleft into plumes. Larvae spin webs on and eat the leaves of various plants.

Family Orneodidae (feather-winged moths). Small, but worldwide, family of uncertain relationships. Each wing is very deeply cleft into *6* or more narrow plumelike divisions.

Superfamily Bombycoidea

Adults large to very large. Frenulum usually lacking; proboscis reduced and nonfunctional; tympanic organs and maxillary palpi lacking; male antennae pectinate.

Family Bombycidae (silkworm moths). A very small family consisting of the domesticated silkworm (*Bombyx* mori) and a few relatives, all Asian. Related family: Eupterotidae.

Family Lasiocampidae (tent caterpillar and lappet moths). Larvae usually hairy and brightly coloured, some living gregariously in silk nests. Adults medium sized, stout bodied, short winged, exceptionally hairy. Widespread.

Family Saturniidae (giant silkworm moths). Every continent has some well known species—*e.g.*, the North American cecropia (Hyalophora cecropia), io (Automeris io), polyphemus (Antheraea polyphemus), and luna (Actias *luna*); European giant peacock or emperor (Saturnia pavonia); and Indo-Australian Attacus atlas and Coscinoscera hercules, the last with wing expanses up to *275* mm (11 in.). Males with broadly branched antennae; often with bright colours and striking wing shapes and patterns. Larvae very large and fleshy, often with brightly coloured knobs and spines. Most species spin firm cocoons of brown, green, or silvery silk. Wingspread up to 275 mm (in hercules moth, Coscinoscera hercules). Worldwide.

Family Citheroniidae (regal moths). Medium to very large, heavy bodied moths, often brightly coloured. Larvae often with long spines; *e.g.*, the North American Citheronia (hickory and pine horned devils). Pupation in the ground, with no cocoon. New World only. Related families: Brahmaeidae (African and Indo-Australian) and Lacosomidae (New World).

Superfamily Geometroidea

Adults with frenulum and abdominal tympana. Some authorities place each of the 3 major families as a separate superfamily.

Family *Geometridae* (measuring worm moths). Larvae ("measuring worms" or "inch worms") long, slender, lacking most of posterior abdominal prolegs, crawling with characteristic looping gait. Pupation usually in the soil, with no cocoon. Adults and larvae commonly very cryptic, resembling bark, dead leaves, and twigs. Many species destructive to foliage; *e.g.*, cankerworms (Alsophila and Palaeacrita) and the winter moth (Operophtera brumata). The second largest moth family, worldwide and abundantly represented nearly everywhere.

Family Uraniidae (uraniid moths). Chiefly tropical. Some adults—*e.g.*, Chrysiridia (Madagascar) and Urania (Neotropits)—are large, brilliantly iridescent diurnal moths. The Oriental Epicopeia (sometimes separated as the family Epicopeiidae) mimic protected swallowtail butterflies.

Family *Drepanidae* (hooktip moths). A small, worldwide family, chiefly Indo-Australian. Many of the adults have the forewing apexes strongly hooked. Larvae usually with last pair of prolegs lacking. Related families: Thyatiridae and Epiplemidae, both worldwide.

Superfamily Sphingoidea

Family Sphingidae (hawk or sphinx moths). Adults medium sized to large; powerful, with long, narrow forewings, thick antennae and usually strong proboscises; feed by hovering before flowers and drawing nectar through the extended proboscis. Some are diurnal, most are active at twilight. Larvae large, fleshy, mostly smooth; most with a single, long, posterior, dorsal horn; typically rest with head and thorax reared up, fancifully like the Sphinx in profile. Worldwide.

Superfamily Noctuoidea

Adults with frenulum and a pair of complex tympanic organs on metathorax.

Family Noctuidae (owlet moths). The largest family of Lepidoptera, with abundant members everywhere in a great diversity of size, colour, and habit. Sizes range from a wing expanse of 10 mm *(Hypenodes)* to 275 mm *(Thysania).* Many species brightly coloured, but the majority plain and cryptic. Nearly every type of plant food is exploited: foliage, flowers, fruits, stems, rootstocks, and woodland litter. A few genera (*e.g., Cosmia)* cannibalistic. Family includes the army worms and cutworms, among the most injurious of all moth larvae to man's interests.

Family Agaristidae (forester moths). A small, worldwide group; many adults brightly coloured, diurnal, and aposematic or mimetic.

Family Notodontidae (prominents). Adults mostly dull, cryptic. Larvae very diverse: cryptic, disruptive, or aposematic; some highly protected by toxic secretions; posterior prolegs often modified or aborted; chiefly foliage eaters.

Family Arctiidae (tiger moths). Adults, especially in tropics, mostly brightly coloured, aposematic, with protective and toxic secretions. Timbal organ for making very high pitched sounds on each side of metathorax. Larvae often with much secondary hair. Worldwide. Several other families, chiefly tropical, are closely related, also have timbal organs.

Families Dioptidae and Pericopidae. Similar to Arctiidae, separable by wing venation. Many are boldly coloured and marked, diurnal, members of mimicry complexes, often mimicking wasps. New World only.

Family Ctenuchidae (ctenuchid moths). Similar to Arctiidae; separable by wing venation. Includes many diurnal mimics of protected wasps and beetles. Worldwide. Related Lithosiidae also contain many aposematic species; their larvae feed on lichens.

Family Liparidae (tussock moths). Many larvae have prominent tussocks and pencils of hair, poison spines, and osmeteria (protrusions that emit repellant odours). Adults are broad winged, the females heavy bodied, sometimes wingless. Some species—*e.g.*, the gypsy moth *(Lymantria dispar)* and the nun moth *(L. monacha)*—are very destructive to forests. Closely related Thaumetopoeidae of the Old World are best known for the habits of the gregarious larvae (processionary caterpillars), which move in orderly columns.

Superfamily Hesperioidea (skippers)

The skippers (like the true butterflies) are distinguished from moths by diurnal habits, clubbed antennae, a functional proboscis, lack of ocelli, and lack of a frenulum. Adult skippers are fast flying, with short, usually pointed forewings, broad heads, antennae usually hooked beyond the club, and all veins arising from the forewing cell simple. Larvae usually have a pronounced, necklike collar, often with a dorsal sclerotized area.

Family Euschernonidae (regent skipper). A single anomalous Australian species; the male has a frenulum, like a moth; the female lacks it.

Family Megathymidae (giant skippers). A small, New World family of relatively large skippers, whose larvae bore in stems of agaves and yuccas, constructing silk-lined tubes. Adult wingspan to about 90 mm.

Family Hesperiidae (typical skippers). With more than 3,000 species, the adults range from small, drab species to large, often iridescent, tropical ones. Larvae mostly living concealed in individual leaf nests or in webs among grasses, forming flimsy cocoons. Worldwide.

Superfamily Papilionoidea (true butterflies)

Adults with antennae clubbed, but not hooked, flight slower and not darting; larvae lacking "neck". General evolutionary trend from a primitive condition with the forelegs fully developed in both sexes (Pieridae, Papilionidae) to one in which they are greatly reduced and useless for walking (Nymphalidae). Intermediate conditions occur in Lycaenidae. Pupae ("chrysalids") often brightly coloured and irregularly shaped, not enclosed in cocoons; either hanging head down (Nymphalidae) supported from cremaster, or additionally supported by a silk girdle (Pieridae, Papilionidae, Lycaenidae). All families worldwide.

Family Pieridae (whites and sulphurs). Small to medium sized; mostly white, yellow, or orange, often with dark tips on wings. Pupae usually with a frontal horn on head. Many species noted for mass migrations.

Family Papilionidae (swallowtails and parnassians). Medium to large; often brightly coloured. Many have the "swallowtails" on the hindwings. Some females of the Indo-Australian genus *Troides* are the largest butterflies. Larvae often curiously patterned, with eversible, protective scent organs (osmeteria). Adults of many are highly distasteful and much mimicked. Worldwide. Parnassians sometimes placed in a separate family, Parnassiidae.

Family Lycaenidae (blues, coppers and hairstreaks). A large diverse family of small butterflies. Many species are iridescent blue, green, or, in the coppers (Lycaeninae), metallic coppery. Larvae somewhat sluglike, hairy, many secreting honeydew and having mutualistic relationships with ants; some prey on the ant brood or on other insects. Many pupae have stridulatory structures on the abdomen.

Family Riodinidae (metalmarks). Small to medium sized. Closely allied to the Lycaenidae. Family has an unequalled diversity of wing shapes and colors. Many South American species have complex mimicry relationships. Chiefly tropical American.

Family Libytheidae (snout butterflies). Long, protruding palpi. Some are strongly, but unpredictably, migratory. Worldwide, but with only about a dozen species.

Family Nymphalidae (nymphalid butterflies). The largest family of butterflies, so diverse that it is often split into several families, here considered subfamilies. The Nymphalinae is the main group, with many familiar species such as the fritillaries, admirals, checkerspots and anglewings. Many tropical species brilliantly iridescent. Satyrinae contains the familiar wood nymphs, meadow browns, and heaths, usually with eyespots on the wings; the larvae distinctively pointed posteriorly and spinning crude cocoons. The very large Brassolinae and iridescent Morphinae are Neotropical, as are the highly distasteful, aposematic Heliconiinae and Ithomiinae that, with the worldwide Danainae, are models in many mimicry complexes. Most of the pantropical Acraeinae are also highly protected and aposematic models.

Critical Appraisal. The terms microlepidoptera and macrolepidoptera are applied, for convenience, to two nontaxonomic divisions of the order Lepidoptera. The microlepidoptera include the suborders Zeugloptera and Monotrysia, and the superfamilies Zygaenoidea through Pyralididoidea of the Ditrysia. The remaining superfamilies of the Ditrysia, consisting of the bombycid, geometrid, sphingid, and noctuid moths, the skippers, and the true butterflies, comprise the macrolepidoptera. Although, in general, the microlepidoptera are smaller than the macrolepidoptera, some of the former, such as certain hepalids and cossids, are among the largest moths, with adult wingspans up to *250* mm, while some macrolepidoptera have adult wingspans of only 10 mm.

BIBLIOGRAPHY. The literature on Lepidoptera is enormous and includes books and articles published in nearly all countries. The following gives a representative sample, including major works with extensive bibliographies.

General works: "Lepidoptera," in A.D. IMMS, *A General Textbook of Entomology*, 9th ed. rev. by O.W. RICHARDS and R.G. DAVIES (1957); and "Lepidoptera," in D.J. BORROR and D.M. DE LONG, *An Introduction to the Study of Insects*, rev. ed. (1964), give a key to families of Lepidoptera in North America; E.B. FORD, *Butterflies* (1945) and *Moths* (1955), are replete with details of natural history, ecology, physiology, and evolution, chiefly of British species; M. HERING, *Biologie der Schmetterlinge* (1926), covers life histories and ecology, particularly of European groups; A.B. KLOTS, *The World of Butterflies and Moths* (1958), covers natural history and ecology, with striking illustrations. See also P.J. PORTIER, *La biologie des Lépidoptères* (1949), for many details of the biology; C.B. WILLIAMS, *The Migration of Butferflies* (1930); and F.A. URQUHART, *The Monarch Butterfly* (1960). *Lepidopterorum Catalogus*, published in many parts by various authors, is the definitive scientific listing of all genera and species, with bibliographies.

Regional Works: (North America:) P.R. EHRLICH *et al.*, *How to Know the Butterflies* (1961), is an excellent handbook of the North American species; W.T.M. FORBES, *Lepidoptera of New York and Neighboring States,* 4 pt. (1923–1960), is the definitive work for eastern North America for technical classification, especially larvae, life histories, and foodplants; W.J. HOLLAND, *The Moth Book* (1914, reprinted 1968), and *The Butterfly Book*, rev. ed. (1931), are the long-

time standards, with many colour illustrations; A.B. KLOTS, *Field Guide to the Butterflies of North America, East of the Great Plains* (1951), covers all the species of butterflies and skippers of eastern North America, with much on life histories, biology, and ecology, with many illustrations. *(Europe:)* The following are more or less local in scope, covering the butterflies and moths of various countries or regions: L.T. FORD, *A Guide to the Smaller British Lepidoptera* (1949); W. FORSTER and T.A. WOHLFAHRT, *Die Schmetterlinge Mitteleuropas,* 3 vol. (1954–60); S. HOFMEYER, *Die Danske Ugler* (1962); T.W. LANGER, *Nordens dagsommerfugle* (1958); F. NORDSTROM and E. WAHLGREN, *Svenska fjarilar* (1941); R. SOUTH, *The Moths of the British Isles* rev. ed., 2 vol. (1939); A. SPULER, *Die Schmetterlinge Europas und ihre Raupen,* 4 vol. (1908–10); and M. STURANI, *Vita delle Farfalle* (1947). *(other regions:)* A. SEITZ *et al.* (eds.), *Macrolepidopfera of the World* (1906–), English, French, and German editions, 16 vol. (1–4 Palaearctic, 5–8 American, 9–12 Indo-Australian, 13–16 Ethiopian), was issued in parts, by many authors, with supplements; although out of date in much nomenclature and taxonomy, it is still the standard for most of the world for identification. The following are books covering specific regions giving many details about life histories and biologies and mostly illustrated very fully: D.A. SWANEPOEL, *The Butterflies of South Africa: Where, When, and How They Fly (1953);* G. VAN SON, *The Butterflies of Southern Africa* (1949–*64);* G.A. WATERHOUSE and G. LYELL, *The Butterflies of Australia* (1914); and M.A. WYNTER-BLYTH, *Butterflies of the Indian Region* (1957).

(A.B.K.)

Lermontov, Mikhail

Mikhail Yuryevich Lermontov, the foremost Russian Romantic poet and novelist, is recognized for the technical perfection and unusual emotional depth of his poetry and for his novel *A Hero of Our Time,* a work basic to the development of Russian psychological realism.

BY courtesy of the State Tretyakov Gallery, Moscow

Lermontov, oil painting by Pyotr Yefimovich Zabolotsky, 1837. In the State Tretyakov Gallery, Moscow.

Life and works. He was born in Moscow on October 15 (3, old style), 1814, the son of Yury Petrovich Lermontov, a retired army captain, and Mariya Mikhaylovna, *née* Arsenyeva. At the age of three he lost his mother and was brought up by his grandmother, Yelizaveta Alekseyevna Arsenyeva, on her estate in Penzenskaya province. Russia's abundant natural beauty, its folksongs and tales, its customs and ceremonies, the hard forced labour of the serfs, and stories and legends of peasant mutinies all had a great influence in developing the future poet's character. As the child was often ill, he was taken to spas in the Caucasus on three occasions, where the exotic landscapes created lasting impressions on him.

In 1827 he moved with his grandmother to Moscow, and, while attending a boarding school for children of the nobility (at Moscow University), he began to write poetry and also studied painting. In 1828 he wrote the poems *Cherkesy* and *Kavkazsky plennik* in the vein of the English Romantic poet Lord Byron, whose influence then predominated over young Russian writers. Two years later his first verse, *Vesna,* was published. The same year he entered Moscow University, then one of the liveliest centres of culture and ideology, where such democratically minded representatives of nobility as Aleksandr Herzen, Nikolay Platonovich Ogarov, and others studied. Students ardently discussed political and philosophical problems, the hard fate of serf peasantry, and the recent Decembrist uprising. In this atmosphere he wrote many lyrical verses, longer, narrative poems, and dramas. His drama *Stranny chelovek* (1831) reflected the attitudes current among members of student societies: hatred of the despotic tsarist regime and of serfdom. In 1832, after clashing with a reactionary professor, Lermontov left the University and went to St. Petersburg (Leningrad), where he entered the cadet school. Upon his graduation in 1834 with the rank of subensign (or cornet), Lermontov was appointed to the Life-Guard Hussar Regiment stationed at Tsarskoye Selo, close to St. Petersburg. As a young officer, he spent a considerable portion of his time in the capital, and his critical observations of aristocratic life there formed the basis of his play *Maskarad* (1835; publ. 1842). During this period his deep—but unreciprocated—attachment to Varvara Lopukhina, a sentiment that never left him, was reflected in *Knyaginya Ligovskaya* and other works.

Exile to the Caucasus

Lermontov was greatly shaken in January 1837 by the death of the great poet Aleksandr Pushkin in a duel. He wrote an elegy that expressed the nation's love for the dead poet, denouncing not only his killer but also the court aristocracy, whom he saw as executioners of freedom and the true culprits of the tragedy. As soon as the verses became known to the court of Nicholas I, Lermontov was arrested and exiled to a regiment stationed in the Caucasus. Travel to new places, meetings with Decembrists (in exile in the Caucasus), and introduction to the Georgian intelligentsia—to the outstanding poet Ilya Grigoryevich Chavchavadze, whose daughter had married a well-known Russian dramatist, poet, and diplomatist, Aleksandr Sergeyevich Griboyedov—as well as to other prominent Georgian poets in Tiflis (now Tbilisi) broadened his horizon. Attracted to the nature and poetry of the Caucasus and excited by its folklore, he studied the local languages and translated and polished the Azerbaijanian story "Ashik Kerib." Caucasian themes and images occupy a strong place in his poetry and in the novel *Geroy nashego vremeni* (1840, *A Hero of Our Time),* as well as in his sketches and paintings.

Return to St. Petersburg

As a result of zealous intercession by his grandmother and by the influential poet V.A. Zhukovsky, Lermontov was allowed to return to the capital in 1838. His verses began to appear in the press: the romantic poem *Pesnya pro tsarya Ivana Vasilyevicha, molodogo oprichnika i udalogo kuptsa Kalashnikova* (1837), the realistic satirical poems *Tambovskaya kaznacheysha* (1838) and *Sashka* (written 1839, published 1862), the drama *Maskarad* (1835–36, published 1842)—a remarkable work of Russian dramatic composition that exposes the evils and vices of fashionable society—and the romantic poem *Demon.* Soon Lermontov became popular; he was called Pushkin's successor and was lauded for having suffered and been exiled because of his libertarian verses. Writers and journalists took an interest in him and fashionable ladies were attracted to him. He made friends among the editorial staff of *Otechestvennye zapiski,* the leading magazine of the Western-oriented intellectuals, and in 1840 he met the prominent progressive critic V.G. Belinsky, who envisioned him as the great hope of Russian literature. Lermontov had arrived among the circle of St. Petersburg writers.

At the end of the 1830s, the principal directions of his creative work had been established. His freedom-loving sentiments and his bitterly skeptical evaluation of the times in which he lived are embodied in his philosophical lyric poetry ("Duma," "Poet," "Ne ver sebye . . .") and are interpreted in an original fashion in the romantic and fantastic images of his Caucasian poems, *Mtsyri* (1840) and *Demon,* on which the poet worked for the remainder of his life. Finally, Lermontov's mature prose showed a critical picture of contemporary life in his novel ***A Hero of Our Time,*** containing the sum total of his reflections on contemporary society and the fortunes of his generation. The hero, Pechorin, is a cynical person of superior accomplishments who, having experienced everything else, devotes himself to experimenting with human situations. This realistic novel, full of social and psychological content and written in prose of superb quality, played an important role in the development of Russian prose.

Second exile

In February 1840 Lermontov was brought to trial before a military tribunal for his duel with the son of the French ambassador at St. Petersburg—a duel used as a pretext for punishing the recalcitrant poet. On the instructions of Nicholas I, Lermontov was sentenced to a new exile in the Caucasus, this time to an infantry regiment that was preparing for dangerous military operations. Soon compelled to take part in cavalry sorties and hand-to-hand battles, he distinguished himself in the heavy fighting at Valerik River, which he describes in "Valerik" and in the verse "Ya k vam pishu. . . ." The military command made due note of the great courage and presence of mind displayed by the officer-poet.

As a result of persistent requests by his grandmother, Lermontov was given a short leave in February 1841. He spent several weeks in the capital, continuing work on compositions he had already begun and writing several poems noted for their maturity of thought and talent ("Rodina," "Lyubil i ya v bylye gody"). Lermontov devised a plan for publishing his own magazine, planned new novels, and sought Belinsky's criticism. But he soon received an order to return to his regiment and left, full of gloomy forebodings. During this long journey he experienced a flood of creative energy: his last notebook contains such masterpieces of Russian lyric poetry as "Utes," "Spor," "Svidanye," "Listok," "Net, ne tebya tak pylko ya lyublyu," and "Prorok," his last work.

On the way to his regiment, Lermontov lingered on in the health resort city of Pyatigorsk for treatment. There he met many fashionable young people from St. Petersburg, among whom were secret ill-wishers who knew his reputation in court circles. Some of the young people feared his tongue while others envied his fame. An atmosphere of intrigue, scandal, and hatred grew up around him. Finally, a quarrel was provoked between Lermontov and another officer, N.S. Martynov; the two fought a duel on July 27 (15, O.S.), 1841, that ended in the poet's death. He was buried two days later in the municipal cemetery, and the entire population of the city gathered at his funeral. Later, Lermontov's coffin was moved to the Tarkhana estate, and on April 23, 1842, he was buried in the Arsenyev family vault.

Assessment. Only 27 years old when he died, Lermontov had proved his worth as a brilliant and gifted poet-thinker, prose writer, and playwright, the successor of Pushkin, and an exponent of the best traditions of Russian literature. His youthful lyric poetry is filled with a passionate craving for freedom and contains calls to battle, agonizing reflections on how to apply his strengths to his life's work, and dreams of heroic deeds. He was deeply troubled by political events, and the peasant mutinies of 1830 had suggested to him a time "when the crown of the tsars will fall." Revolutionary ferment in western Europe met with an enthusiastic response from him (verses on the July 1830 revolution in France, on the fall of Charles X), and the theme of the French Revolution is found in his later works (the poem ***Sashka).***

Central themes of his work

Lermontov's creative work grew out of the historical reality of the 1830s, though the heroic traditions of Russian history, the Romantic poetry of Pushkin's era, and the influence of Byron (which Pushkin had also felt) all nourished it. Byronism satisfied a particular quality in Lermontov, for it helped him to express his passionate ideals and to assert the idea of personal freedom, the central theme of his work. But Lermontov's Romanticism, also includes realistic concepts of the world; these are particularly characteristic of his poems based on historical events and real subjects. Lermontov gradually discarded the romantic stock phrases and images that had filled his early works, overcoming, as he stated, their "false tinsel" and "deafening language." In the late works, noted for their richness of genre and style, the Romantic approach to reality gave way to objective depiction, to the reproduction of a specific social environment, daily life, and characters from the people.

Civic and philosophical themes as well as subjective, deeply personal motifs were closely interwoven in Lermontov's poetry. He introduced into Russian poetry the intonations of "iron verse," noted for its heroic sound and its unprecedented energy of intellectual expression. His enthusiasm for the future responded to the urgent spiritual needs of Russian society and the emancipation movement. His creativity promoted a new golden age of Russian literature, and his influence can be seen in the work of Ivan Turgenev, Leo Tolstoy, Fyodor Dostoyevsky, and N.A. Nekrasov, as well as in the poetry of A.A. Blok and V.V. Mayakovsky. Lermontov's legacy has found varied interpretations in the works of Russian artists, composers, and theatrical and cinematic figures. His dramatic compositions have played a considerable role in the development of theatrical art, and his life has served as material for many novels, poems, plays, and films.

MAJOR WORKS

NOVEL: ***Geroy nashego vremeni*** (1840; ***A Hero of Our Time,*** trans. by P. Longworth, 1964).

VERSE: "Parus" (1832), lyric; ***Pesnya pro tsarya Ivana Vasilyevicha, molodogo oprichnzka i udalogo kuptsa Kalashrzikova*** (1837; ***A Song About Tsar Ivan Vasilyevitch, His Young Body-Guard, and the Valiant Merchant Kalashnikov,*** trans. by J. Cournos, 1929); ***Mtsyri*** (1840); ***Demon*** (1841; Eng. trans. by G. Shelley, 1930). Lermontov's poems are collected in English in ***Selected Poetry,*** trans. by C.E. L'Ami and Alexander Welikotny (1965).

BIBLIOGRAPHY

Collections of works: *Сочинения* ***М.Ю. Лермонтова*** 6 vol. (1889–91), the first complete edition; ***Полное собрание сочинений*** 5 vol. (1910–13), the complete works; ***Полное собрание сочинений*** 5 vol. (1935–37), illustrated complete works; ***Стихотворения*** 3 vol. (1940–50), verse; ***Избранные произведения в двух томах*** (1949), selections; ***Three Russian Poets: Selections from Pushkin, Lermorztov and Tyutchev in*** new translations by VLADIMIR NABOKOV (1944); ***A Lermontov Reader,*** ed. and trans. by GUY DANIELS (1965).

Biographical and critical studies: Most of the works on Lermontov have been published in Russian—V.G. BELINSKIY (1940), articles and reviews; N.G. CHERNYSHEVSKIY (1947), an outline of the Gogol period of Russian literature; A.I. GERTSEN (1937), on the development of revolutionary ideas in Russia; N.L. BRODSKIY (1945), a biography; and V.A. MANUYLOV (1950), on Russian dramatic compositions. English-language studies include JANKO LAVRIN, ***Lermontov*** (1959); JOHN MERSEREAU, ***Mikhail Lermontov*** (1962); and C.E. L'AMI and ALEXANDER WELIKOTNY, ***Michael Lermontov: Biography and Translation*** (1967). See also A. HEIFETZ (comp.), ***Lermontov in English: A List of Works by and About the Poet*** (1942).

(V.V.Z.)

LeSniewski, Stanislaw

Stanisław Leśniewski, a Polish logician and mathematician, was a cofounder of the Warsaw school of logic, its leading representative, and one of the greatest logicians of all time. Working along lines laid down late in the 19th century by Gottlob Frege in Germany and early in the 20th century by Bertrand Russell in Great Britain, Leśniewski developed one of the most complete, precise, and rigorous systems of logic and of the foundations of mathematics that have so far been devised. In collaboration with the logician Jan Lukasiewicz and the mathematicians Wactaw Sierpiński and Stefan Mazurkiewicz, he made the University of Warsaw, during the period between World Wars I and II, a renowned centre for research in logic, mathematics, and metalogic (on characteristics of formal systems and logical calculi). Despite the disruption caused by World War II, which killed many members of the school and destroyed their manuscripts, leading representatives were able to reach the United States and Great Britain, where they have continued the work begun by Leśniewski, as has a new generation of students in the revived universities of Poland.

His life. Leśniewski was born March 30, 1886, in Serpukhov, Russia, where his father was one of the civil engineers chiefly responsible for the construction and supervision of the trans-Siberian railroad.

Influence of Twardowski and Łukasiewicz

Education. After preliminary schooling in Russia and the gymnasium in Siberia, he attended—as was the custom of the time—several universities in continental Europe, finally taking his doctoral degree in 1912 at the Polish University of Lwów (now Lvov, U.S.S.R.), then a part of Austria. His dissertation was approved by Kazimierz Twardowski, who, for his wide-ranging influence on Polish intellectual life, is known as the father of contemporary Polish philosophy. Twardowski, like Edmund Husserl (the founder of Phenomenology), was a student in Vienna of Franz Brentano, an Aristotelian and Scholastic philosopher who, although not himself interested in formal logic, was noted for his precise and thorough analysis of philosophical problems. Under his influence, Leśniewski's first scholarly interests focussed upon problems of philosophical logic, such as those that had concerned John Stuart Mill in 19th-century England and Husserl and others of the Austrian school. Thus, his doctoral dissertation of 1911 dealt with the analysis of existential propositions.

Period of his early writings. The intellectual activity of Leśniewski divides into three distinct periods. The first extends from his dissertation to the appearance in 1916 of his first work on the theory of collective sets. Leiniewski attributed the discovery of his true intellectual vocation to the influence of Jan Łukasiewicz, also a pupil of Twardowski and then a *privat dozent* at the University of Lwów. Already learned in the history of logic, to which he was to make outstanding contributions, Lukasiewicz was at the time studying the work of the German logicians Gottlob Frege and Ernst Schroder, the importance of which he was mainly responsible for making known in Poland, and teaching his first course in mathematical logic. It was Łukasiewicz' book *O Zasadzie Sprzeczności u Arystotelesa* (1910; "On the Principle of Contradiction in Aristotle") that awakened Leśniewski from his dogmatic slumber. From it he became interested in the problem posed by the discovery of the antinomies, or paradoxes, in logic and mathematics that threatened to undermine the foundations of all deductive science. His efforts to overcome and solve these antinomies, with which Frege and Russell were also wrestling, eventually led to the great discoveries for which he is known.

Turn from distrust to use of symbolic logic

Although Leiniewski then definitely turned his back on philosophy in favour of logic—he later spoke of himself as a renegade from philosophy—his initial impression of mathematical logic was not at all favourable. He distrusted its technical formal notation, the scant attention given to its relation to ordinary language, and the resulting equivocations in the use of such terms as class, implies, true. In attempting to clear away the equivocations in the work of Russell, however, he soon became convinced that the formal and artificial language of mathematical logic was essential for his work, that ordinary language was too clumsy and imprecise. The writings of this period were completed, however, before he had adopted the rigorous methods of mathematical logic and were all later repudiated by him.

Period of intensive research. The period from 1916 to 1927 was one of intensive and creative research in which he accumulated a mass of results but refrained from publishing them.

In 1915, upon the reopening of the University of Warsaw, Lukasiewicz had been called from Lwów to become professor of philosophy. Leśniewski, after teaching for two years at a Warsaw gymnasium and spending the war years in Moscow, followed his friend and colleague to Warsaw in 1919 as professor of the philosophy of mathematics. They soon established a thriving centre of research that attracted talented students from all sides.

Period of maturity and publication. LeSniewski finally felt constrained to start publishing some account of his findings, even though they were not yet in as perfect a form as he would have desired. Beginning with the publication, in 1927, of his first mature work on the foundations of mathematics and extending until his death, in 1939, he published a series of papers expounding the main lines of his theories of logic and mathematics. These publications gained a worldwide reputation for the Warsaw school. Yet, just as it was reaching its height, Leśniewski died suddenly and unexpectedly on May 13, 1939, on the eve of the war that shortly engulfed the school in the common fate of Poland.

Many of Leiniewski's findings remained unpublished at his death. Although all of his manuscripts were destroyed by the war, many of the unpublished results of his researches have since been made known through the work of his students, particularly that of Bolesław Sobociński and Alfred Tarski.

The notion of distributive and collective classes

His three logical systems. The distinctive and original contribution of Leśniewski consists in the construction of three interrelated logical systems, to which he gave the names, derived from the Greek, of protothetic, ontology, and mereology. The logical basis of the whole theory, and hence its name (*prōtos*, "first"), is provided by protothetic, which is the most comprehensive theory yet developed of the relations between propositions. The other two systems are based on a distinction the lack of which, Leiniewski claimed, was the source of Russell's difficulties with the antinomies: that between a distributive and a collective class. In its distributive use, a class expression is identical with a general name; thus, to say that a person belongs to the class of Poles is to say that that person is a Pole. Hence, ontology (*on* "being") is the logic of names; and, combined with protothetic. it yields all of the theorems of syllogistic (traditional Aristotelian logic) and of logical algebra, as well as of the logic of sets and relations. Mereology (*meros*, "part") is the logic of a whole conceived as though physically constituted by its parts; *i.e.*, of the collective class, as the class of all automobiles in Chicago consists of the entire collection of them. Hence, mereology is a general theory of the relation between part and whole.

In developing these theories, LeSniewski gave great care to the statement of their metalogic and, for this purpose, elaborated a general theory of semantical categories, which is analogous, on the one hand, to the traditional doctrine of the parts of speech and, on the other, to Husserl's "meaning categories."

His intuitive formalism

Leśniewski developed the systems with a clarity and precision that established a new standard for mathematical rigour. In their powers of implication, they are strong enough to provide a logical foundation for all of classical mathematics. They also overcome the antinomies in a way that Leiniewski claimed is better and truer than any other solution. In his opinion, modern mathematicians and logicians are often too neglectful, if not contemptuous, of man's naïve and basic intuitions of the way things are. For this very reason, Alfred Tarski, one of his students who later went to the United States, described his position as "an intuitive formalism." Leśniewski was openly critical of a pure formalism that would consider logic and mathematics as nothing more than a game of symbols. It is true that he advocated and employed formalist methods for their rigour and precision, but he maintained that a theory ultimately must be judged for its accord with reality. Nevertheless, he maintained that his logical systems are neutral in that they make no metaphysical assumptions and are equally well adapted to diverse and even conflicting philosophical interpretations.

BIBLIOGRAPHY. Z.A. JORDAN, *Philosophy and Ideology: The Development of Philosophy and Marxism-Leninism in Poland Since the Second World War* (1963), a comprehensive background history of the Warsaw school, with much information about Leśniewski and his associates and a full bibliography; TADEUSZ KOTARBINSKI, *Elementy Teorii Poznania, Logiki Formalnej i Metodologii Nauk*, 2nd ed. (1961; Eng. trans., *Gnosiology: The Scientific Approach to the Theory of Knowledge,* 1966), a Polish classic, which for more than two decades provided the only elementary account of Leśniewski's work and influence, and is still of value, having been written by a lifelong friend and colleague of Leśniewski, in a position to supply information otherwise unpublished; CZESLAW LEJEWSKI, "Leśniewski" entry in the *Encyclopedia of Philosophy*, vol. 4, pp. 441–443 (1967), with good bibliog-

raphy; EUGENE C. LUSCHEI, *The Logical Systems of Leśniewski* (1962), the only book-length study of LeSniewski and his work, contains a comprehensive analytical bibliography of sources to 1962; STORRS MCCALL (ed.), *Polish Logic, 1920–1939* (1967), original papers in English translation with an introduction by TADEUSZ KOTARBINSKI, "Notes on the Development of Formal Logic in Poland in the Years 1900–39," including papers by Leśniewski and his associates, together with a slightly revised version of Z.A. JORDAN, "The Development of Mathematical Logic in Poland Between the Two Wars"; BOLESLAW SOBOCINSKI, "L'Analyse de l'antinomie Russellienne par LeSniewski," in *Methodos*, 1:94–107, 220–228, 308–316 (1949) and 2:237–252 (1950), containing, besides the analysis, backgound information about Leiniewski's theories, especially his ontology and mereology; and "In Memoriam Jan Lukasiewicz," in *Philosophical Studies* (Ireland), 6:3–49 (1956), containing background information about Leiniewski's outlook and research; ALFRED TARSKI, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938* (1956), containing Leśniewski's ideas on semantics and related topics.

(O.A.B.)

Lesotho

The Kingdom of Lesotho in southern Africa has an area of 11,720 square miles (30,355 square kilometres) and a population of about 1,000,000. It is one of only two independent states in the world (the other is the Republic of San Marino in western Europe) completely encircled by a single country on whom it must depend exclusively for access to the outside world. It forms an enclave within the Republic of South Africa, bordering on three of the latter's provinces—Natal, the Orange Free State, and Cape Province. This physical dependence on its neighbour is further accentuated by Lesotho's own grave lack of resources which makes it necessary for about 20 percent of its able-bodied male population to live and work, mainly as migrants, in South Africa. For this reason, it is sometimes described as a "hostage state." It is a member of the Commonwealth of Nations and of the Organization of African Unity. Before its independence on October 4, 1966, it was one of the three British High Commission Territories—the other two being Bechuanaland (now Botswana) and Swaziland. Its constitution was suspended after its first postindependence crisis in January 1970.

"Lesotho" is the name of the country and "Basotho" the name of the people; a single individual is referred to as a "Mosotho"; Sesotho is the language. The capital is Maseru. For an associated physical feature see ORANGE RIVER; for historical aspects see SOUTHERN AFRICA, HISTORY OF.

The plateau

The landscape. *Relief.* Two-thirds of the country consists of mountains, which rise to its highest peak, the Mont Aux Sources, which is 11,425 feet above sea level. The Drakensberg Range forms the eastern boundary with Natal. The Maloti spurs, running north and south, join the main range in the north, where they form a plateau between 9,000 and 10,500 feet in altitude. This plateau is the source of South Africa's two largest rivers—the eastward-flowing Tugela and the westward-flowing Orange River, as well as tributaries of the Caledon—hence, Lesotho's claim to be the "sponge" of South Africa. The mountain area divides into the highlands and the foothills. The foothills, which have average altitudes of between 6,000 and 7,000 feet, descend in undulating slopes to the west, where the lowlands bordering on the Orange Free State average 5,000 to 6,000 feet in altitude.

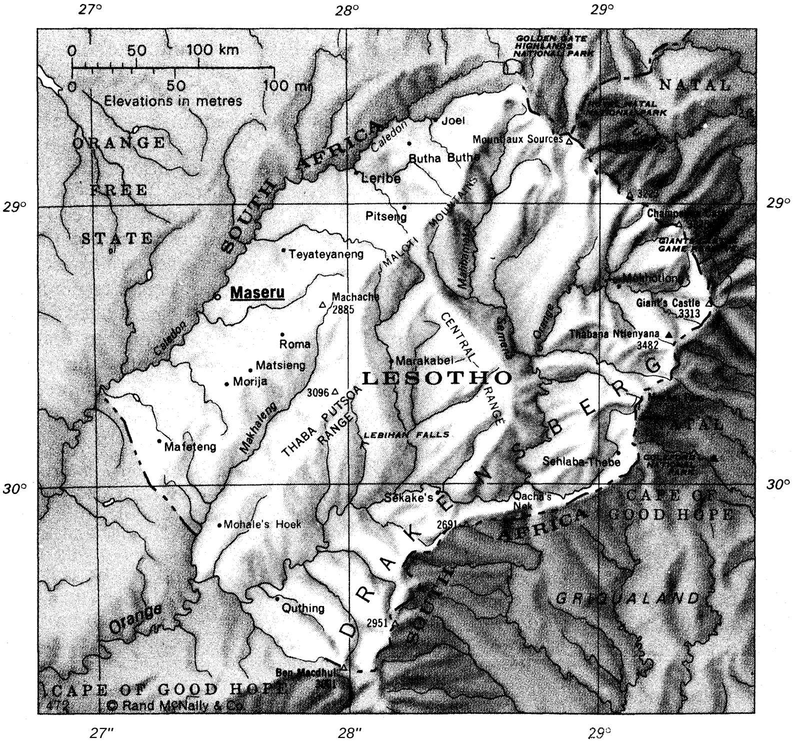
The mountain soils are of basaltic origin and are shallow but rich. The soils of the lowlands derive mainly from the underlying sandstone. Extensive erosion has severely damaged soils all over the country.

Climate. The rainfall, brought by the prevailing winds, occurs mostly between October and April; it is variable, averaging about 28 inches a year over most of the country. Although droughts are rare, their periodic occurrence is devastating. Temperatures in the lowlands vary from 90° F (32" C) in the summer to 20° F (−7° C) in the winter. In the highlands the temperature range is much wider, and below-zero readings are not unusual. Frost occurs widely in the winter, when the Maloti are usually snow-capped. Hail is a frequent summer hazard.

Vegetation. The grasslands of the highlands are mainly covered by *Themeda triandra.* Indigenous trees and shrubs have survived only in the more remote and sheltered valleys. The most common types are wild olives, wild garlic, wild willow, brushwood, thatching grass (*Hyparrhenia hirta*), hypoxis, and aloes.

Animal life. Wildlife is now almost extinct, except for occasional small antelopes, hares, and reptiles.

The landscape under human settlement. Although not permanently inhabited, the mountain grasslands found on the slopes of the high plateau as well as in the valleys provide excellent grazing for sheep and cattle, which are tended by herdsmen in isolated cattle posts. Some of the deep valleys, like the Senqunyane, produce good crops of



LESOTHO

wheat, peas, and beans. About 30 percent of the population live in the highlands; 20 percent live in the foothills, which good grazing and cultivation have made the most productive legion. Half the population live in the lowlands, which form a narrow corridor along the Caledon River, which averages only 25 miles in width.

The system of migratory labour

The Sotho (Basotho, Basuto) combine modernism and traditionalism to an unusual degree; this provides stability in a society that, possibly more than any other in existence, is disrupted by a permanent system of migratory labour. Traditional authority is still firmly exercised through a system of chieftaincy extending from the paramount chief (the king) and his court, down through senior chiefs and subchiefs, to headmen and subheadmen at the village level. Their authority now rests largely on the fact that they are responsible for the working and distribution of all land that belongs to the nation, with the king acting as trustee.

Social cohesion is strengthened by the persistence of clan and family loyalties. Families and clans still cluster together as units in the numerous small rural villages. There are no large towns in the country. The villages range in size from one large to four or five extended families, with an average of from 30 to 50 immediate families. The villages are often picturesque, offering fine views of the rocky highlands; on the plains they are often surrounded by aloes and trees, with the walls and doors of the houses frequently being covered with patterned designs. The villages themselves consist of clusters of circular or rectangular huts solidly built of turf, Kimberley brick (unburned clay), or dressed stone, mostly with thatched roofs; latterly, however, corrugated iron has come into vogue as a roofing material.

Initiation ceremonies

The average household usually has two or three huts, the larger one being used as a living and dining room and as the parents' bedroom; the smaller ones are used for kitchen and storage purposes and as sleeping quarters for the children. The hut of the chief, or headman, is usually in the centre of the village, flanked by that of the principal wife and ringed by those of the junior wives. The open court (the *lekhotla)* is in front of the chief's hut; besides it are the enclosures (kraals) for the cattle and stables for horses. Village life centres largely on the fields, the chief's court, the kraals, the school, the church, and the initiation lodge. Circumcision forms an integral part of the ritualized initiation ceremonies that train boys to take their place as full members of the family, clan, and nation—the three centres of social cohesion. Young boys still spend a large part of their lives as herdsmen, while women and young girls do much of the hard work in the fields, a practical necessity because the able men are usually absent, although they contrive to return home briefly for the plowing or the harvest.

Two distinctive features of social life are the people's love for horses and blankets. The small, sturdy Sotho pony is renowned for its sure-footedness in rough country; it is often the only means of rural transport. Because of the sharp variations in climate, both men and women invariably wear blankets, which they use as cloaks; a great deal of care is taken in choosing a blanket, which, especially for the men, is usually multicoloured. Men and women also wear the typical Sotho hat, which is woven from reed into conical shapes, with an unusual topknot.

Maseru, the capital, in the northwest on the banks of the Caledon River, is itself a small rural town with about 14,000 inhabitants, mostly civil servants, professional people, and traders. On most days of the week, however, it is a bustling little town crowded with people who are either from the surrounding lowlands, down from the mountains, or migratory workers passing to and from South Africa.

People and population. *Ethnic groups.* The Sotho comprise a cluster of tribes of the southern Sotho linguistic stock, who are united by a common loyalty to the royal house of Moshoeshoe (Moshesh). The group that formed the nucleus of the nation is that of the Kwena, who are made up of the Molibeli, Monaheng, Hlakwana, Kxwakxwa, and Fokeng tribes. They still regard themselves as the true nation and tend to look down on the other two main groups that were politically absorbed into their society but were not always culturally assimilated, although they speak the same language. The first of these groups is formed by the Natal (North) Nguni, who include the Phetla, Polane, and Phuti; they are largely indistinguishable from the Kwena. The second group comprises the Mahlape tribe of the Natal Nguni and the Cape (South) Nguni (Tembu).

There are about 1,000,000 Sotho (1969 census), 2,000 Europeans, 300 Asians, and 1,000 Coloureds (of mixed racial origin). The Europeans are mainly traders, businessmen, technicians, government officials, missionaries, and teachers, while the Asians are mainly traders.

Emigration

As the country is severely overpopulated, both temporary and permanent emigration has taken place. No exact figures exist for the number of Sotho working or living in South Africa. The 1966 Lesotho census showed that 97,000 men and 20,000 women were absent, while a South African census at the same time recorded about 130,000 males and 70,000 females who claimed they were born in Lesotho.

Religious groups. Christianity and traditional religions (which are based on a belief in ancestral spirits) co-exist uneasily in a country where the great majority of people are active members of Christian churches. Over 30 percent are Roman Catholics, and just under 40 percent are Protestants (mainly French Protestants, Anglicans, Baptists, and adherents of the Dutch Reformed Church).

Demography. About half the people are concentrated in just over one-quarter of the land. Although the average density of the country is only 87 per square mile, there are from 100 to 303 persons per square mile in the lowlands, as against an average of 34 per square mile in the highlands. The annual population growth is 3 percent.

Lesotho, Area and Population

	area		population	
	sq mi	sq km	1966 census	1972 estimate
Districts				
Berea	830	2,150	100,000	132,000
Butha Buthe	700	1,815	55,000	70,000
Leribe	1,280	3,315	139,000	180,000
Mafeteng	800	2,070	103,000	133,000
Maseru	2.270	5,880	182.000	224,000
Mohale's Hoek	1,435	3,715	97,000	123,000
Mokhotlong	1,760	4,560	55,000	67,000
Qacha's Nek	1.520	3,935	57,000	70,000
Quthing	1,125	2,915	65,000	81,000
Total Lesotho	11,720	30,355	852,000*†	1,081,000*

'Figures do not add to total given because of rounding.
†Excludes absentee workers amounting to 12% of the population at the 1966 census.
Source: Official government figures.

Economic links to South Africa

The national economy. Lesotho is basically a poor country with few natural resources; these are certainly insufficient to sustain even the present population. Lesotho's gross national product is only about $80,000,000. (The monetary unit is the South African rand; R 0.71 = $1 U.S.; R 1.71 = £1 sterling on December 1, 1970.) The condition of the country was well summed up in 1963 by the *Financial Mail* of Johannesburg, South Africa, as that of "a waif, sustained, not developed, by the South African economy." But its economy could not be sustained at all without such benefits as it derives from its neighbour, with which it forms part of a customs union, shares a single currency and an integrated communications system, besides depending wholly on it for the export of its surplus working population.

Agriculture. Only one-tenth of the land—about 750,000 acres—is suitable for arable agriculture; this amounts to about one acre per head of the resident population. Agriculture provides about two-thirds of the gross domestic product (GDP); it is devoted largely to subsistence crops. The main crops are maize (corn), sorghum, wheat, peas, beans, and barley. Crop yields are low. The large cattle herds have no significant commercial value; about 70 percent die annually from disease, exposure, malnutrition, or old age. The principal cash crops are wool and mohair.

Mining and industry. Geological surveys have so far shown little promise of mineral wealth, although small quantities of diamonds have been found; production is valued at about $1,600,000 a year. Other minerals discovered, such as iron ore, gypsum, quartz, and calcite (limestone chalk), are commercially uneconomic.

Secondary industry is of recent growth and is confined to such small-scale manufacturing as milling, brickmaking, and furniture making.

Tourism. Tourism is promising for a country that has justifiably been called the South African Switzerland. Modern hotels have been built, mountain roads and pony trails developed, and trout streams stocked with fish. There is a casino in Maseru.

Irrigation and lzydroelectricity. The country's most promising prospects for development lie in the harnessing of the rivers for irrigation and hydroelectric power. The Oxbow project for a dam to generate power in the Maloti mountains would produce over 350,000,000 kilowatt-hours of energy a year. The scheme, estimated to cost about $35,000,000, would benefit both Lesotho and South Africa, which has indicated interest in the project.

The Oxbow project

Trade. The economy runs on a heavy adverse balance of trade, importing in any normal year roughly five times more than it exports. Its budget and capital-aid deficits have, in the past, been made up by grants and loans from the United Kingdom (about $5,000,000 in 1970 to 1971 alone) and by remittances from migrants abroad. These amount to about $6,000,000 a year—an amount larger than the total exports, which total about $5,600,000.

The public sector. The government is the largest employer of labour in the country, over half its annual budget being made up of payments to its public employees. Its budget of about $8,000,000 has a deficit of almost 50 percent. Direct taxation produces almost $3,000,000, and $1,500,000 is provided by revenues from the customs agreement with South Africa.

The private sector. About 90 percent of all trade is controlled by the 72 members of the Chamber of Commerce, only two of whom are Sotho. Asians are not members. Two European firms control almost half of the wholesale trading outlets, which number about 140.

Taxation. Three forms of direct taxation are levied: a basic tax on all adult males; a graded tax on all income earners; and an income tax.

Transport. A single main road runs for 228 miles along the western and southern boundary. A mountain road from Maseru reaches into the interior. These two main arteries are served by short-distance feeder roads. Villages in the mountains are served by 1,600 miles of bridle paths.

A one-mile railway line links the capital to the South African transport network.

Although considerable use is made of light aircraft for passengers and for transporting mail and freight to the interior, only Maseru has a regular airport suitable for small commercial planes. There are 28 rough airstrips.

As mentioned earlier, ponies are much used in rural areas, and in mountainous areas they are often the only means of transport.

Administration and social conditions. *The governmental structure.* The country's constitution (suspended under the state of emergency in 1970) provides for a bicameral parliamentary system.

The National Assembly has 60 members, each elected to represent a single constituency; the elections are based on universal adult suffrage. The Senate's 33 members are nominated partly by the government and partly by the king. Bills must pass through both houses to become law and must be assented to by the king. There is also a College of Chiefs which deals with matters relating to the appointment and control of chiefs.

The constitution was suspended following the contested results of a general election in which the opposition claimed to have won a majority of the seats. The country has since been ruled by a cabinet which operates under a system of emergency regulations.

Defense and justice. There are no armed forces, but there are about 1,500 police and palamilitary personnel.

The legal system is based on Roman and Dutch law. There are magistrates' courts and a High Court. Appeals of certain cases of a constitutional nature may be appealed to the Privy Council in the United Kingdom.

Literacy

Education. The country has the highest literacy rate in Africa of about 80 percent. It has more girls than boys in school, due to the latter's employment as herdsboys. Although about 90 percent of all school-age children are in primary schools, less than 9 percent advance beyond the lower primary educational level, and less than 3 percent advance to secondary education. Most of the schools are run by the Christian churches, which receive over 80 percent of the cost of teachers' salaries from government subsidies. There are about 1,300 primary schools and 29 secondary schools; one agricultural school; and seven teacher-training colleges. The University of Botswana, Lesotho, and Swaziland is based at Roma, 22 miles from Maseru. Fewer than 30 percent of its students are Sotho. The university has an extension department.

Health. The country's healthy climate contributes largely to the comparatively low rate of sickness. The main incidence of ill health is the result of food-deficiency diseases, venereal disease, chronic rheumatism, infections of the respiratory tract, and dyspepsia. There are fewer than 20 hospitals, about half of which are operated by the government, and a number of clinics for maternity, child health, and venereal diseases, as well as a few health centres and mountain dispensaries. There are almost forty doctors in the country—about one doctor for every 22,820 people. Some teachers and corporate empolyees are paid by the state.

Employment. The country offers few opportunities for earning a salary. Out of approximately 20,000 people in employment outside the rural sector, 40 percent are government employees, 35 percent are employed by schools or missions, and 30 percent are employed in commerce and by cooperative societies.

Cultural life and institutions. Traditional institutions remain strong despite the modernization and politicizing of Sotho society. The paramount chief and the system of chieftaincy remain a strong focus of loyalty, especially since some of its more venal aspects were reformed in the 1960s; there still is, however, considerable criticism of the power of chiefs and headmen, especially of their manner of administering the distribution of land and of the proliferation of officeholders within the chieftaincy system. Historical traditions and the philosophy of the founder of the nation, Moshoeshoe I (who was born at the end of the 18th century and died in 1870), exercise a profound influence in sustaining the Sotho nation's sense of independence and national pride. There is a strong literary tradition, which has been stimulated by modem education. Folklore flourishes, both in oral and written forms.

Importance of the chieftains

The initiation schools continue to play an important role in the villages and are a source of keeping alive a sense of the traditional values of the clans and the nation.

Simple village crafts, such as the decoration of walls, woodwork, clay making, weaving, and basketwork, flourish. Choir singing is popular.

The press. There are no daily papers in the country. In normal times there are ten regular weekly or monthly publications. Under the emergency government, only one governmental bulletin and a few religious publications are available.

Broadcasting. A national broadcasting service is provided by a 10-kilowatt medium-wave transmitter and a 5-kilowatt shortwave transmitter, both located at Maseru. There is no television.

Prospects for the future. Losotho has poor prospects of ever being able to achieve true economic viability, since it has inadequate resources of its own, and its economy and communications are inextricably linked to South Africa. Its future depends upon the answer to the question of how, under these circumstances, it can best preserve the political independence of statehood, to which the Sotho are resolutely committed.

BIBLIOGRAPHY

Physical geography: HARM J. DE BLIJ, *A Geography of Sub-saharan Africa (1964);* ANTHONY SILLERY, *Africa: A Social*

Geography (1961); ALAN BERTRAM MOUNTJOY and CLIFFORD EMBLETON, *Africa: A New Geographical Survey* (1967); WILLIAM A. HANCE, *The Geography of Modern Africa* (1964).

Land and water resources: JEFFARES and GREEN, *A Water Resources Survey of Basutoland* (1951); VERNON G SHEDDICK, *Land Tenure in Basutoland* (HMSO 1954); G.M. STOCKLEY, *Report on the Geology of Basutoland* (1947).

Demography: R.R. KUCZYNSKI, *Colonial Population* (1937); a primary source of population information about Africa.

People, institutions, culture, and history: HUGH ASHTON, *The Basuto* (1952), the accepted standard work on the Basotho people; D.F. ELLENBERGER and J.C. MacGREGOR, *History of the Basuto* (1912); H.V. MEYEROWITZ, *A Report on the Possibilities of the Development of Village Crafts in Basutoland* (1936); JACK HALPERN, *South Africa's Hostages: Basutoland, Bechuanaland and Swaziland* (1965); VERNON G. SHEDDICK, *The Southern Sotho* (1953); C.W. DE KIEWIET, *A History of South Africa: Social and Economic* (1941); ERIC WALKER, *A History of Southern Africa*, 3rd ed. (1957); RICHARD STEVENS, *Lesotho, Botswana and Swaziland* (1967); G.P. MURDOCK, *Africa: Its Peoples and Their Cultural History* (1959); C.G. SELIGMAN, *Races of Africa*, 4th ed. (1966).

Administration, and social, political, and constitutional affairs: LORD W.M. HAILEY, *The Republic of South Africa and the High Commission Territories* (1963), *An African Survey* (1938); COLIN LEGUM and JOHN DRYSDALE (eds.), *AfricaContemporary Record* (annual); COLONIAL OFFICE, LESOTHO, *Annual Reports* (HMSO, to 1966), useful official reports reviewing all aspects of policy and development; *Basutoland, the Bechuanaland Protectorate and Swaziland: History of Discussions with the Union of South Africa, 1909–39*, Cmd. 8707 (HMSO 1952); BASUTOLAND CONSTITUTIONAL COMMISSION, *Report* (1963); J.E. SPENCE, *Lesotho: The Politics of Dependence* (1967).

Description: AUSTIN COATES, *Basutoland* (1966), a well-written useful introduction to the country.

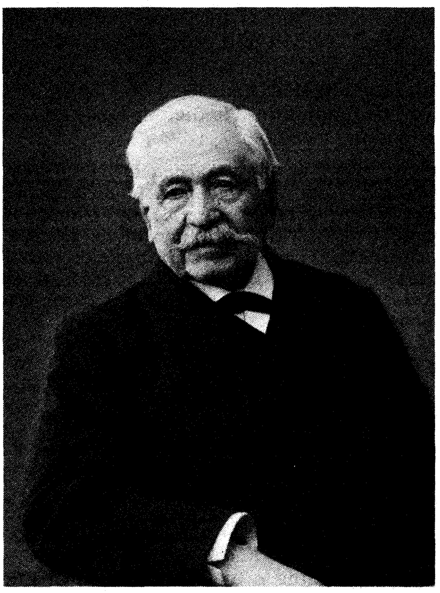
Economy: G.M.E. LEISTNER, *Lesotho: Economic Structure and Growth* (1966); *Basutoland Development Plan, 1963/66* (1963); *Basutoland, Bechuanaland and Protectorate and Swaziland: Report of an Economic Survey Mission* [the Morse report] (HMSO 1960).

(Co.L.)

Lesseps, Ferdinand de

Diplomatic career

Ferdinand-Marie, vicomte de Lesseps, a French diplomat for 24 years, achieved lasting fame by realizing his dream of building a canal across the Isthmus of Suez, thus creating the shortest maritime route between Europe and the countries bordering the Indian and Pacific oceans. He was born at Versailles on November 19, 1805, of a family long distinguished in government service. Appointed assistant vice consul at Lisbon in 1825, he was sent to Tunis in 1828 and to Alexandria in 1832. At Alexandria the survey report of J.-M. Le Pkre, one of Napoleon's chief engineers, on the Isthmus of Suez, and his friendship with Muhammad ʿAlī, the Turkish viceroy of Egypt, and his son, Sa'id Pasha, led Lesseps to hope that he might one day finish the canal that Le Pkre had begun. For the time, however, he could not pursue his plans. From 1833 to 1837, Lesseps was consul at Cairo, where he gained distinction in combatting an outbreak of plague. Two years later he was transferred to Rotterdam. Subsequently he served at Málaga and at Barcelona, where he was promoted to consul general. From 1848 to 1849, after the proclamation of the Second Republic, he was minister of France at Madrid. In May 1849 he sent a mission to Rome from where Pope Pius IX had fled and where Mazzini had proclaimed the republic. This mission was ambiguous: it was a question of "placing a limit on the pretensions of Austria . . . of ending by arbitration . . . the differences which divided . . . the peninsula . . . ." Lesseps tried to reconcile the irreconcilables: the papacy and the republic. But at the end of May, when the French Legislative Assembly, conservative by nature, followed the Constituent Assembly, which held republican views, he was recalled, handed over to the Council of State, and censured. French troops re-established pontifical power in Rome. The diplomatic career of Lesseps was shattered. But in 1854, an invitation from Sa'id Pasha, newly appointed viceroy, or khedive, of Egypt, revived his ambitions. On November 30, 1854, Saʿīd Pasha signed the first act of concession authorizing Lesseps to pierce the isthmus of the Suez.



Lesseps.
Culver Pictures

Scheme for the canal

A first scheme, directed by Lesseps, was immediately drawn up by the surveyors Linant Bey and Mougel Bey (L.-M. Linant de Bellefonds and E. Mougel) providing for direct communication between the Mediterranean and Red sea; and, after being slightly modified, it was adopted by an international commission of engineers in 1856. Encouraged by this approval, Lesseps allowed no obstacles to retard the work, and he succeeded in rousing the French people to subscribe more than half the capital needed to form the company, which was organized in 1858. The first blow of the pickax was given by Lesseps at Port Said on April 25, 1859; and ten years later, on November 17, 1869, the Suez Canal was officially inaugurated by the Empress Eugénie who had been invited by the host of the celebrations, the khedive (viceroy), Ismāʿīl Pasha. In 1875 the British government, on the initiative of the prime minister, Benjamin Disraeli, purchased the khedive Ismāʿīl's Suez Canal shares and became the largest shareholder. Lesseps cooperated loyally with the British (in spite of the fact that they had earlier tried to block the building of the canal because of their suspicions of the French) and facilitated the transfer of ownership. Though he usually tried to keep out of politics, Lesseps stood as a Bonapartist candidate for a seat in the Chamber of Deputies at Marseille in 1869 but was defeated by Léon Gambetta, later one of the founders of the Third Republic.

In 1879, when the International Congress of Geographical Sciences met in Paris and voted in favour of the construction of a Panama canal, the 74-year-old Lesseps undertook to carry out the project. His despotic temper and stubbornness, however, made him fail to appreciate the difficulties of the task: at first he thought that it would be possible to pierce a canal without locks, even though the route was barred by the Culebra cut and by the torrential Chagres River. The task proved to be beyond the capacities of a private company, so that eventually, in 1889, the company that Lesseps had formed had to liquidate. After an official inquiry in 1892, the French government instituted the prosecution of the company's administrators; and in February 1893 Lesseps and his son Charles (1849–1923) were sentenced to five years' imprisonment. Only Charles, however, was imprisoned, and in June an appeals court reversed the decision. On the other hand, the fact that members of the government and parliamentarians were accused of having accepted bribes from the company made the Panama scandal a political affair as well as a financial one, with important repercussions in the history of the Third French Republic.

Lesseps was a member of the French Academy, of the Academy of Sciences, and of numerous scientific societies. He was also decorated with the grand cross of the

Legion of Honour and the Star of India and received the freedom of the City of London. His great gifts, unselfishness, and social charm made him everywhere respected, and the scandal that clouded his last years has done nothing to tarnish his reputation. He died at La Chenaie, near Guilly (Indre), on December 7, 1894.

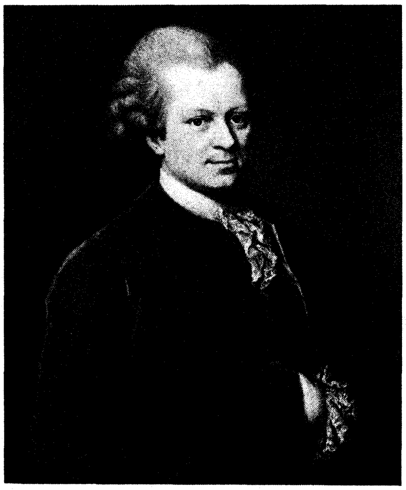
BIBLIOGRAPHY. G. EDGAR BONNET, *Ferdinand de Lesseps*, vol. 1, *Ferdinand de Lesseps, le diplomate, le créateur de Suez* (1951), vol. 2, *Ferdinand de Lesseps, après Suez, le pionnier de Panama* (1959), written by one of de Lesseps' successors, a director of the Compagnie Universelle du Canal Maritime du Suez; P. BUNAU-VARILLA, *Panama: la création, la destruction, la Resurrection*, 3rd ed. (1913), 200 pages are devoted to the undertaking of de Lesseps; R. COURAU, *Ferdinand de Lesseps: de l'apothéose du Suez au scandale de Panama* (1939), a biography with reprints of relevant economic and financial documents.

(A.Da.)

Lessing, Gotthold Ephraim

The first German dramatist of lasting importance, Gotthold Ephraim Lessing was also a social, dramatic, and religious critic and a writer on philosophy and aesthetics. His work is a plea for tolerance and humanity and everywhere reveals the two guiding principles of his own life — truth and reason.

By courtesy of the Gleimhaus, Halberstadt, East Germany



Lessing, oil painting by Georg May, 1768.
In the Gleirnhaus, Halberstadt, East Germany.

Lessing was born on January 22, 1729, at Kamenz, in Upper Lusatia, Saxony. His father, a highly respected theologian, was hard put to support his large family even though he occupied the position of *pastor primarius* (chief pastor). At the age of 12, Lessing, even then an avid reader, entered the famous *Fürstenschule* (Elector's school) of St. Afra, in Meissen. He was a gifted and eager student and acquired a good knowledge of Greek, Hebrew, and Latin. These were to stand him in good stead for his later work, while his admiration for the plays of the Latin dramatists Plautus and Terence fired him with the ambition to write comedies himself.

Early fascination with the theatre

Education and first dramatic works. In the autumn of 1746, Lessing entered the University of Leipzig as a student of theology. His real interests, however, lay toward literature, philosophy, and art. He also attended lectures in mathematics by A.G. Kastner, who was an epigrammatic poet of some renown and who may be considered as Lessing's first literary mentor. Lessing's cousin, the writer Christlob Mylius, and the dramatist Christian Felix Weisse introduced him to the lively social life of Leipzig, Germany's busy publishing capital and the site of a famous annual fair. Above all, he was fascinated by the theatre there, which had recently been revitalized by the work of a talented and energetic actress, Caroline Neuber. She had collaborated for many years with Johann Christoph Gottsched, an influential dramatic critic, moral philosopher, and educationalist (whom Lessing was later to attack for his narrow brand of rationalism). The "Neuberin" took an interest in the young poet and in 1748 successfully produced his comedy *Der junge Gelehrte* ("The Young Scholar"). In this, Damis, a merchant's son, a bigoted and self-centred young man, fancies that he deserves an academy prize for his verbose literary "masterpieces" and is so absorbed in them that he is quite unaware of the intrigues of marriage and inheritance taking place about him. The play is a delightful satire on a shallow know-it-all and ends with the audience laughing at the incurable arrogance of the vain and easily offended scholar. Lessing mocks his own bookishness in the play but is mainly concerned to criticize the middle class preoccupations of narrow-minded burghers and merchants, contrasting these characters with their much cleverer and warmer hearted servants. The other comedies belonging to this Leipzig period of 1747–49 (*Damon*, *Die alte Jungfer* ["The Old Maid"], *Der Misogyn* ["The Misogynist"], *Die Juden* ["The Jews"], *Der Freigeist* ["The Free Thinker"]) are witty commentaries on human weaknesses — bigotry, prejudice, nagging, fortune hunting, matchmaking, intrigue, hypocrisy, corruption, and frivolity. Set against this background are the virtuous men and women who are considerate and selfless, sensitive and helpful, forthright, and faithful in love. In *Die Judert* Lessing praised unappreciated nobility of mind and thus struck a blow for tolerance toward the Jews at a time when they were still confined to a ghetto life. Lessing had set himself the goal of becoming the German Molière: in these comedies he most interestingly begins to draw his characters as recognizable individuals, breaking away from the traditional dramatic "types" (although he was to create his finest comic individual only in his last comic masterpiece, *Minna von Barnhelm*).

Early work in Berlin

Early in 1748 Lessing's parents, who disapproved of his life in Leipzig, summoned him home. But he managed to win their consent to begin studying medicine and was soon allowed to return to Leipzig. He quickly found himself in difficulties because he had generously stood surety for some members of the Neuber company — although himself heavily in debt. When the company folded, he had no choice but to flee from Leipzig in order to avoid being arrested for debt. He eventually reached Berlin in 1748, where he hoped to find work as a journalist through his cousin Mylius, who was by this time an established editor. In the next four years he undertook a variety of jobs, mainly translating French and English historical and philosophical works into German. But he also began to make a name for himself through his witty criticism for the *Berlinische privilegirte* (later *Vossische*) *Zeitung*, on which he was book review editor, especially for his contributions to a monthly supplement, *Das Neueste aus dem Reicke des Witzes* ("The Latest from the World of Wit"). He also launched a periodical of his own, *Beiträge zur Historie und Aufnahme des Theaters* ("Contributions to the History and Improvement of the Theatre"). Despite its topical content, it did not long survive and had to be discontinued in 1750.

Rising reputation as dramatist and critic. After his flight from Leipzig, Lessing had first gone to Wittenberg, where he enrolled as a student of medicine. Now, from 1751 to 1752, Lessing went back there and took his degree. He then returned to Berlin, where he started another periodical, *Theatralische Bibliothek* ("Theatrical Library"), but this too had to be closed down after only four volumes were published (1754–58). The most significant event about this time, however, was the publication, in 1753–55, of a six-volume edition of his works. Apart from some witty epigrams, which Lessing called *Sinngedichte* ("mind poems") after those of the 17th-century poet Friedrich von Logau (whose epigrams he had earlier edited), the edition contained the most important of his Leipzig comedies. It also contained *Miss Sara Sampson*, which is the first major *bürgerliches Trauerspiel*, or domestic tragedy, in German literature. Middle class writers had long wanted to do away with the traditional class distinctions in literature, whereby heroic and tragic themes were played out by aristocratic figures, while middle class characters appeared only in comedy. Lessing was, in fact, not the first German writer to chal-

Attempt to create German middle class tragedy

lenge this tradition, but it is fair to say that his play marks the decisive break with the classical French drama that still dominated the German stage. *Miss Sara Sampson* was inspired by George Lillo's *London Merchant* (1731) and by the novels of Samuel Richardson—with their praise of middle class feminine virtue—and to a lesser degree, by the sentimental *comédie larmoyante* ("tearful comedy"), originated in France by the early 18th-century dramatist Pierre-Claude de La Chaussée. It is the first German play in which *bürgerlich* characters bear the full burden of a tragic fate, and it received its successful premiere at Frankfurt an der Oder in 1755. Its reflective prose skillfully lays bare the psychology of the situation —a conflict between the demands of virtue and the heart, between conscience and passion—and its characters are finely drawn. The plot centres on an innocent, sensitive heroine of a bourgeois family; she becomes the victim of Lady Marwood, her vampire-like rival in love, who disregards all restraints and inhibitions, and of Mellefont, a weak man who vacillates between the two women but finally atones for his guilt by his death.

Characteristic of Lessing's writings at this period are his *Rettungen* ("Vindications"), outstanding for their incisive style and clarity of argument. Here, in the spirit of the Age of Enlightenment—that was ruled by the conviction that right reasoning, rather than faith or revelation, would lead mankind to truth and happiness—he aimed to defend independent thinkers such as the Reformation-period writers Johannes Cochlaeus and Gerolamo Cardano, who had been unjustly slandered and persecuted. His scintillating and biting polemic *Vade Mecum für den Herrn Samuel Gotthold Lange* (1754) is a "vindication" of the poet Horace, directed against the carelessly corrupt translations of the arrogant scholar Lange, whose literary reputation was demolished by Lessing's attack. From this time, Lessing was justly feared as a literary adversary. His style increasingly became a finely honed weapon. But although he hit hard at those he considered incorrigible, he always fought for a cause and fought for truth, not victory.

The philosopher Moses Mendelssohn and the writer and publisher C.F. Nicolai stand out among his Berlin friends. With these men, Lessing conducted a truly epoch-making correspondence *(Briefwechsel über das Trauerspiel,* 1756–57) on the aesthetic of tragic drama. Tragedy, Lessing maintained, should not preach morality; it should arouse admiration and pity in the audience as evidence of their emotional involvement. Somewhat earlier, in the essay *Pope ein Metaphysiker!* (1755, "Pope a Metaphysician!"), he had joined forces with Mendelssohn to define the borders between philosophy and the theatre. At this time, Lessing also had many contacts with the poets Johann Wilhelm Ludwig Gleim and Karl Wilhelm Ramler, whose work he tended to overrate because of personal regard and friendship.

Between November 1755 and April 1758 Lessing was once again living at Leipzig, making brief excursions to Dresden, Halberstadt, Wolfenbüttel (where he was to live during the last years of his life), and Hamburg, there making the acquaintance of the poet Friedrich Gottlieb Klopstock. Lessing also planned to visit England with a young merchant, Gottfried Winkler, but they only got as far as Amsterdam when the outbreak of the Seven Years' War forced them to abandon the journey. In Leipzig Lessing became friendly with Ewald von Kleist, author of the poem *Der Frühling* (1749, "Springtime"). Kleist, a Prussian officer who fell in the Battle of Kunersdorf, was immortalized by Lessing as the character Major Tellheim in his comic masterpiece *Minna von Barnhelm.*

Friendship with the poet Ewald von Kleist

Lessing moved back to Berlin in May 1758 and remained until November 1760. There, besides writing occasionally for Nicolai's *Bibliothek der schönen Wissenschaften* ("Library of the Fine Arts"), he also contributed regularly, in 1759–60, to Nicolai's weekly, *Briefe, die neueste Literatur betreffend* ("Letters Concerning the Latest Literature"), writing a number of essays on contemporary literature. Here he praised the poems of Kleist as well as the *Kriegslieder von einem preussischen Grenadier* ("Battle Songs of a Prussian Grenadier") of his friend Gleim (whose unreflected patriotism, however, strikes most readers today as somewhat naïve). Lessing also stressed the importance of the poetry of Klopstock and C.M. Wieland, even while analyzing and criticizing their weaknesses. The central point of his essays, however, was a vigorous attack on the influential theatre critic J.C. Gottsched for his advocacy of a theatre modelled on French drama, especially that of the 17th-century tragedian Pierre Corneille. Lessing maintained that the courtly, mannered drama of France was alien to the German mentality. Instead, he demanded a truly national drama, belonging to the people, based on faithfulness to nature and reality. He urged German playwrights to take Shakespeare as their model. In the 17th *Literaturbrief* he published a stirring scene from his own fragmentary Faust drama. In this scene, Lessing sketches out a "Faust without evil" whose relentless spirit of inquiry is justified before God, notwithstanding his pact with the devil. He thus paved the way for his young contemporary Johann Wolfgang von Goethe and his great dramatic version of the Faust story.

In 1759 Lessing published some masterly prose fables, largely social criticism, and with them an essay on the fable form itself, in which he formulated the particular laws of the genre by analyzing its didactic and allegorical structure. At this time he was made an associate member of the Berlin Academy of Sciences; this was an unusual distinction to be bestowed on a free-lance writer. It did not, however, provide him with any economic security. In 1760 Lessing went to Breslau as secretary to General Tauentzien, a position in which he became familiar with the garrison life of Prussian officers. He later made good use of this knowledge in his play *Minna von Barnhelm.* He also studied philosophy and aesthetics in the Breslau libraries, the result being the great treatise *Laokoon: oder über die Grenzen der Malerei und Poesie* (1766, "Laocoon; or, On the Limits of Painting and Poetry"). Here he took issue with the contemporary art historian Johann Winckelmann, specifically over his interpretation of the "Laocoon," a famous *(c.* 1st-century-BC) sculpture of Hellenistic times, which shows the priest Laocoon and his sons as they are about to be killed by the serpents that hold them entwined. Winckelmann had admired above all the artistic control of suffering in the sculpture, and this provided the starting point for a discussion in which Lessing aimed to define the separate functions of painting and of poetry. He thought that whereas painting is bound to observe spatial proximity—and must, therefore, select and render the seminal and most expressive moment in a chain of events—poetry has the task of depicting an event organically and in its temporal sequence. The essence of poetry thus lies not in description but in the representation of the transitory, of movement. In his exposition Lessing displays a great understanding of Greek poetry without, however, recommending its imitation.

Lessing's definition of the separate functions of painting and poetry

The poetic fruit of his stay in Breslau was the play *Minna von Barnhelm* (1767), which marks the birth of classical German comedy. Goethe was to praise it for its contemporary relevance and for its central theme (the struggle between Prussia and Saxony in the Seven Years' War), which was an event of national significance. The central characters are a Prussian officer, Major Tellheim, and a young gentlewoman from Thuringia, Minna. The upright officer's conscientiousness and rigid interpretation of the code of honour has endangered his relationship with Minna; she, charming and spirited, takes matters into her own hands and, prompted by her heart's perceptions, resolutely overcomes the obstacles that war and occupation have placed in the way of their union. Clearheaded, and yet warmhearted, she resolves the conflict between the claims of conscience and those of happiness. Thus, in thinking and acting like true representatives of the Enlightenment, the two eventually behave like ordinary people and so bear witness to Lessing's concept of humanity. The two protagonists are supported by forcefully drawn secondary characters: the rough, adventure-seeking sergeant, the faithful manservant, the nosy innkeeper, the quick-witted lady's maid, the French gambler. Lessing's dialogue in this play enhances a lively dramatic action that can still hold an audience's attention.

Attempt to start a German national theatre

On returning to Berlin in 1765 he tried for the post of director of the royal library; but since he had quarrelled with Voltaire, who lived as a favourite at Frederick the Great's court, the King (who in any case thought little of German authors) rejected his application. Lessing then accepted the offer of some Hamburg merchants to act as adviser and critic in their privately funded venture of a national theatre. The project collapsed within a year, however, and Lessing recognized with some bitterness that the time for a German national theatre was not yet ripe, for the Germans did not then form a nation, either in the cultural or political meaning of the term. Even so, his reviews of over 50 performances were published, in the form of 104 brief essays on basic principles of the drama, under the title of *Hamburgisclze Dramaturgie* (1767–69). Here, too, Lessing argued against tragedy modelled on that of Corneille and Voltaire, although he praised the realism of the contemporary French writer Denis Diderot's descriptions of middle class life. Lessing interpreted Aristotle's famous and much-discussed concept of tragic catharsis (purging) as meaning the emotional release that follows tension generated in spectators who witness tragic events; he concludes that the sensations evoked by pity and fear should afterward exert a moral influence on the audience by being transformed into virtuous action. He also pleaded for acceptance of the principle of artistic unity in the drama. In 1768–69 he published *Briefe antiquarischen Inhalts* ("Letters of Antiquarian Content"), an attack on the pretentious learning and elitist attitudes of the Halle professor C.A. Klotz. Another result of this dispute was the perceptive essay *Wie die Alten den Tod gebildet,* in which Lessing contrasts the ugly medieval skeleton figure of death with the youthful spirit carrying a lowered torch, whom the Greeks thought of as the brother of sleep.

Final years at Wolfenbiittel. Being extremely poor, in 1770 Lessing had no choice but to accept the badly paid post of librarian at Wolfenbiittel, which he had earlier visited in 1766. This post was offered him by a German princeling, the dullish Duke of Brunswick. His years there were unhappy but rich in achievement. In 1771 he published "Zerstreute Anmerkungen uber das Epigramm" ("Scattered Observations on the Epigram"), in which he took up the cudgels on behalf of the Roman poet and epigrammatist Martial and also defended the artistic integrity of the epigram as a literary genre. In *Wolfenbiitteler Beiträge* ("Wolfenbiittel Contributions"), which appeared from 1773, he reported on discoveries he had made in the Wolfenbiittel library, bringing to light the late medieval fable poet Ulrich Boner. He also published extracts from the papers of the Hamburg biblical critic and scholar H.S. Reimarus, under the title of *Fragmente eines Ungenannten* (1774–77, "Fragments of an Unknown").

Controversy with the orthodox church

Theologians viewed this as a serious challenge to orthodoxy, even though Lessing himself had taken up a mediating position toward the radical theses of Reimarus, who had rejected the basic tenets of the Christian faith. Lessing went into battle against the orthodox clergy, involving himself in violent controversies with their leader, the chief pastor of Hamburg, J.M. Goeze. Against this rigid dogmatist, who was a man of almost pharisaical narrow-mindedness, Lessing launched some of his most cutting polemics, notably "Eine Parabel," "Axiomata," and "Anti-Goeze" (1778). In the equally militant pamphlet "Eine Duplik" ("A Rejoinder"), the subject of which is the Resurrection, he expounded his belief that the never-ending search for truth is more valuable than the certainty of having found it. *Ernst und Falk* (1778–SO), profound dialogues that plead for men to prove themselves humane in their everyday activities, had to be published anonymously, since the Duke had by this time forbidden Lessing to engage in any further polemic.

The tragedy *Emilia Galotti* had meanwhile been performed in 1772. Written in intense and incisive prose, this brilliantly constructed play deals with a conflict of conscience at the court of an Italian prince. Under the influence of the scheming courtier Marinelli, the prince, a shallow man of the world, abducts Emilia from her fiancé and her patrician parents. The girl, confused by the tumult of her own awakening sensuality, prefers death to dishonour; unable to do away with herself, she begs her father, Odoardo, to kill her. He is determined to save her from shame and stabs her with a dagger that the prince's own discarded mistress, in a fit of jealousy and enraged by the corruption of the court, had handed to him—intending that he should kill her faithless lover, the prince, for lusting after Emilia. But despite the fiery speeches that Lessing gives to Odoardo, the patrician cannot bring himself to raise the dagger against his own sovereign; instead of fighting against a social order that countenances such wantonness, he would rather kill his own innocent daughter. The timelessness of this play was demonstrated in Fritz Kortner's 1970 production in Vienna, where the play was brilliantly revealed as charged with contemporary force and significance.

Publication of *Nathan der Weise*

In 1779 appeared the "dramatic poem" in iambic verse *Nathan der Weise,* in which Lessing, according to his own statement, preaches once more from his old pulpit, the stage. It is a didactic play of a theological and philosophical nature, combining ethical profundity with many comic touches, and is a work of high poetic quality and dramatic tension. *Nathan der Weise* symbolizes the equality of three great religions in regard to their ethical basis, for the play celebrates man's true religion—love, acting without prejudice and devoted to the service of mankind. Among the representatives of the three religions—Islām (Saladin), Christian (the Templar), and Jewish (Nathan)—only the Jew, in whose character Lessing paid tribute to his old friend Moses Mendelssohn, lives up to the ideal of full humanity; he alone is capable of complete self-abnegation and has the courage to speak the truth even to the mighty. The fact that the main characters discover that they are related to one another serves to underscore their common membership in the larger family of mankind. Men whose religion guides them to a life of practical humanity are thus bound to each other by brotherly solidarity and mutual tolerance. Lessing maintains that God truly manifests himself in those deeds of charity, motivated by a sense of human responsibility, which man performs for his fellowman.

This conviction is also found in Lessing's last work, *Die Erziehung des Menschengeschlechts* (1780, *The Education of the Human Race),* a treatise that closely reflects the working of his mind and expresses his belief in the perfectibility of the human race. In the history of the world's various religions, Lessing saw a developing moral awareness that would, he believed, eventually attain the peak of moral freedom. After Lessing's death it became known that, in conversations with the Christian philosopher F.H. Jacobi, Lessing had expressed his belief in the pantheistic philosophy of Spinoza, after reading Goethe's poem "Prometheus." Thus he evidently made a decisive break even with the Deism of the Enlightenment, which did acknowledge a Creator. In *Die Erziehung des Menschengeschlechts,* mankind's salvation and the individual soul's eternal life do still beckon, but from an infinite distance, and the treatise closes with the question: "And what, then, am I in danger of missing? Is not all eternity mine?"

Thus the last decade of his life spent at Wolfenbiittel produced a rich harvest of philosophical and literary works. But his life there was otherwise full of tribulations. His health had begun to give way, and it was a lonely existence, with only a few trips to brighten the monotony. In 1775 he went to Vienna, where he was kindly received by the emperor Joseph II, and thence to Italy as companion to one of the Brunswick princes. Although he was there for nine months, the visit made little impression on him. Later, when visiting Mannheim in 1777, he tried to obtain a position at that city's theatre, but in vain. Also during this time (in 1776) he had finally been able to marry Eva Konig, the widow of a Hamburg merchant and a friend of long standing. But his happiness did not last; she died in 1778, giving birth to their only child. In a touching letter to a friend, Lessing wrote that "for once, I wanted as good a life as that which is granted to others; but I had to pay dearly for it." Eva's two children from her first marriage were a comfort to him, and he overcame his misery once again in order to defend his convictions in many critical and creative writings. Yet his last years were lonely and poor, and he died of a stroke on February 15, 1781, at the age of 52, in Braunschweig, and was there buried in a pauper's grave at public expense.

MAJOR WORKS

DRAMATIC WORKS: *Der junge Gelehrte* (produced 1748, printed 1754); *Damon* (written 1741–46); *Die alte Jungfer* (written 1746–48); *Der Misogyn* (written 1746–48); *Die Juden* (1749); *Der Freigeist* (1749); *Der Schatz* (1750); *Miss Sara Sampson* (1755); *Philotas* (1759); *Minna von Barnhelm oder das Soldatenglück* (1767); *Emilia Galotti* (1772); *Nathan der Weise* (1779).

OTHER WORKS: *Kleinigkeiten* (1751), poems and epigrams; *Ein Vade Mecum für den Herrn Samuel Gotthold Lange* (1754), criticism of Lange's translation of Horace; *Rettungen* (1754); *Fabeln* (1759); *Laokoon: oder über die Grenzen der Mahlerey und Poesie* (1766); *Hamburgische Dramaturgie* (1767–69), articles on the theory of drama; *Briefe antiquarischen Inhalts* (1768–69); *Wie die Alten den Tod gebildet* (1769); *Betengarius Turonensis* (1770); "Zerstreute Anmerkungen über das Epigramm" (1771); "Eine Paralbe," "Axiomata," and "Anti-Goeze" (1778), polemical writings directed against J.M. Goeze; *Ernst und Falk* (1778–80; *Lessing's Masonic Dialogues*, 1927); *Die Erziehung des Menschengeschlechts* (1780; *The Education of the Human Race*, 1858).

Many of Lessing's works have been translated, most of them during the 19th century. The most notable translations are *Selected Prose Works* and *The Dramatic Works*, ed. by E. Bell in "Bohn's Standard Library" (1819ff.); *Laocoon, Nathan the Wise*, and *Minna von Barnhelm* in "Everyman's Library" (1930 and 1959); *Theological Writings* (1956).

BIBLIOGRAPHY

Editions: (Miscellaneous publications): Sämmtliche Schriften, the first volumes that were published by Lessing during his lifetime until his death, ed. by his brother C.G. LESSING, J.J. ESCHENBURG, and C.F. NICOLAI, 31 vol. (1771–1825). *(Complete works): Lessings samtliche Schriften*, ed. by K. LACHMANN and F. MUNCKER, 23 vol. (1886–1924), remains the standard and most complete historical–critical edition; *Lessings Werke*, likewise a complete edition, ed. by J. PETERSEN and W. VON OLSHAUSEN, 25 vol. (1925–35). Selected works that are editorially sound throughout and that contain commentaries: *Gesammelte Werke*, ed. by P. RILLA, 10 vol. (1954–58), includes all the letters; *Lessings Werke*, ed. by G. WITKOWSKI, 7 vol. (1911); *Lessings Werke*, ed. by T. HOHLE, 5 vol. (1959 and 1964); *Lessings Briefe in einem Band*, ed. by H. GREINER-MAY (1967).

Biography: E. SCHMIDT, *Lessing: Geschichte seines Lebens und seiner Schriften*, 4th ed. rev. by F. SCHULTZ, 2 vol. (1923), the most complete biography written in the positivistic method; T.W. DANZEL and G.E. GUHRAUER, *Gotthold Ephrairn Lessing: Sein Leben und Seine Werke*, 2nd ed. rev. by W.V. MALTZAHN and R. BOXBERGER, 2 vol. (1880–81), a basic work, complementary to Schmidt-Schultz; H.B. GARLAND, *Lessing: The Founder of Modern German Literature*, 2nd ed. (1962), compressed and to the point; P. RILLA, *Lessing und sein Zeitalter* (1960), the first total presentation from the Marxist viewpoint; K.S. GUTHKE and H. SCHNEIDER, *Gotthold Ephraiin Lessing* (1967), a most important study.

Critical studies: F. MEHRING, *Die Lessing-legende*, 9th ed. (1926; abridged Eng. trans., *The Lessing Legend*, 1938), the first Marxist polemic appraisal of Lessing's alleged Prussianism; W. DILTHEY, "G.E. Lessing," in *Das Erlebnis und die Dichtung, Lessing, Goethe, Novallis, Hölderlin* (1906 and later editions), on the intellectual and philosophical–historical foundations; B.V. WIESE, *Lessing: Dichtung, Asthetik, Philosophie* (1931); A.M. WAGNER, *Lessing: Das Erwachen des deutschen Geistes* (1931); H. LEISEGANG, *Lessings Weltanschauung* (1931); O. MANN, *Lessing: Sein und Leistung*, 2nd ed. (1961); *Lessing Yearbook*, issued annually by the American Lessing Society since 1969.

(J.Mu.)

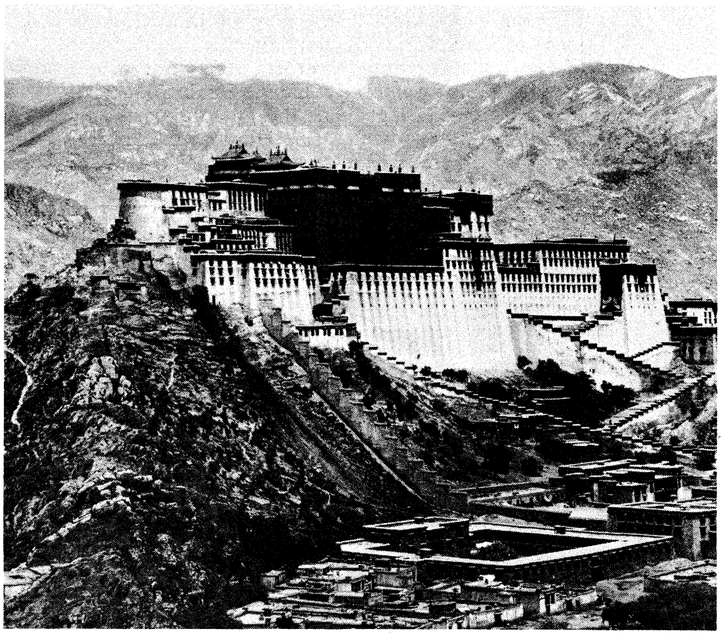
Lhasa

The former national capital of Tibet, Lhasa (La-sa in Pin-yin romanization) is the capital of the Tibetan Autonomous Region of the People's Republic of China. Perched high in the Tibetan Himalayas near the Skyid Chu (Skyid, or Happy, River; La-sa Ho in Chinese), a tributary of the Upper Brahmaputra, the city was long isolated from the outside world. It served as the national religious centre of Tibet, and much of its population was composed of Buddhist monks and laymen. The city has grown tremendously since the 1960s, and most of its estimated 400,000 inhabitants are now Chinese. Though its physical facilities have been improved under Chinese administration, most of Lhasa's Tibetan population has reportedly fled in the face of cultural and political persecution (for related articles see TIBET; BUDDHIST MYSTICISM).

History. Lhasa had been designated as the capital of Tibet by the 9th century AD. National power was decentralized following the assassination of the Tibetan king in 842, and Lhasa lost its position as the country's capital, though it gained in religious importance in succeeding centuries. In 1642 Lhasa was again the seat of the central government, a position it held into the 20th century. Although Lhasa and Tibet came under Chinese occupation in 1951, the city and the country remained under the Tibetan government until 1959, when direct Chinese administration was imposed. (For a more detailed discussion of history, see TIBET, HISTORY OF.)

The city site. Lhasa stands on a level valley plain at an altitude of 11,830 feet (3,606 metres) and is surrounded by dark and barren hills. The Skyid Chu flows through the valley on the south about a half mile from the city. The plain is fertile, though marshy in parts, and gardens scattered around the city are planted with fine trees. Lhasa is screened from view from the west by a lofty and narrow ridge with two high summits, between which is the main entrance to the city. The ridge to the north, called Potala Hill or Dmar-po Ri (Red Mountain), is crowned by the majestic buildings of the Potala, the winter palace of the Dalai Lamas. On the southern peak, called Lcagspo Ri (Iron Mountain), stands a monastic school devoted to the teaching of medicine. Gardens, groves, and open ground intervene between the ridge and the city for about one mile.

Ewing Galloway



The Potala Palace, former fortress of the Dalai Lamas, Lhasa.

The environment. The climate of Lhasa is generally dry, the city receiving only slightly more than 18 inches of precipitation annually. Humidity is low, and fog is practically nonexistent. The mean yearly temperature is approximately 49° F (9° C), with a January mean of 18° F and a July mean of 63" F.

Vegetation in the city includes weeping and Alpine willows, white and black poplars, junipers, elms, and thorn trees. There are fruit-bearing trees such as the peach, walnut, apple, pear, and apricot. Various vegetables are grown, and flowers include the rose, hollyhock, marigold, petunia, sweet pea, hydrangea, pansy, nasturtium, tiger lily, poppy, sunflower, carnation, zinnia, and gladiolus.

Domestic animals, such as the horse, mule, donkey, sheep, goat, yak, and cattle, are used for transportation, food, and clothing. Among the birds are the vulture, eagle, cuckoo, dove, owl, raven, crow, sparrow, lark, and hawk.

The city plan. The centre of the city is occupied by a four-story temple of Gtsug-lag-khang, considered to be the holiest in Tibet. Now converted by the Chinese into a guest house, it is ringed by an extensive mall, which held the main shopping area. A road leading west from the temple to the People's City Hall is lined with shops and is known as the New Market. In the western section of the city stands the former summer palace of the Dalai Lama—the Nor-bu-gling-ka (in Chinese, Na-pu-lin-k'a [Jewel Palace])—which is now the People's Pleasure Park.

The northwestern portion of the city is dominated by the imposing Potala Palace. At the foot of the palace stands a large exhibition hall built by the Chinese, military installations, and a prison. To the north of the palace is a lake, the temple of Klu-khang, and a large sports field, while between the palace and the city are located the People's Hospital, the Commerce Department, and an observatory. About four miles west of the city lies the monastery of 'Bras-spungs (Drepung; in Chinese, Che-pang), the largest in Tibet, which is now used as a military installation. The monastery of Se-ra (in Chinese, Se-la), the second largest in the country, lies about two miles north of the city and is now also used for military purposes.

Transportation. Before the Chinese occupation, the city was composed of narrow, unplanned streets lined with houses and shops. There was one wide road from the Nor-bu-gling-ka Palace to the Gtsug-lag-khang, by way of the Potala Palace. Since the 1960s new roads have been built, and an airport has been constructed near the city. A highway system now connects Lhasa with the major cities in the neighbouring Chinese provinces of Szechwan, Tsinghai, and the autonomous region of Sinkiang. The Skyid Chu has been bridged to accommodate motor vehicles, which were introduced by the Chinese. During the 1960s the mountain passes leading to India and the neighbouring Himalayan countries were sealed, and traffic was restricted within Tibet and between Tibet and China.

The population. Before the Chinese occupation, Lhasa had a permanent population of between 60,000 and 70,000 people from all areas of Tibet. The population was of homogeneous ethnic origin, and the Lhasa dialect of the Tibetan language was the national lingua franca. During the annual Smon-lam (Monlam) festival, the city population would swell to approximately 100,000 with Tibetan monks and pilgrims. The small groups of resident Chinese, Nepalese, and Indian and Chinese Musilms were concentrated in the mall area around the central temple.

Demographic changes

Since the 1960s the city has grown to more than 400,000 inhabitants, most of whom are Chinese civil servants and military personnel. Many of the city's inhabitants have gone to neighbouring countries, and the Chinese government has allegedly drafted most of the able-bodied men for forced labour on various construction projects. Lhasa is divided into eight administrative zones, and the small Tibetan population is restricted in movement from one zone to another.

Housing. Most houses in Lhasa were built of stone or brick; they were rectangular and topped with flat roofs. Buildings were whitewashed, and the doors and windows were outlined in black and often shaded with white canopies. Private buildings did not exceed three storys; the first story often served as a shop or stable. Every house had a Buddhist chapel or shrine, and prayer flags flew from every roof.

Economy. Before the Chinese occupation, the city's economy was based on the historic trade routes that converged on Lhasa from China, India, Nepal, and Bhutan. Except for handicrafts, the only industries were those of the ammunitions factory and mint, which were located between the city and the monastery of Se-ra. The Chinese administration has restricted trade to that with China and has established experimental farms outside the city. Small-scale industries include tanning and wool processing.

Administration and social conditions. Before the 1960s Lhasa served as the national capital of Tibet and the seat of its theocratic government. The city government was composed of both religious and secular civil servants.

Under Chinese administration Lhasa is the capital of the Tibetan Autonomous Region (*tsu-chih-ch'ü*) and also of the Lhasa Municipality (*shih*), which extends from beyond Na-mu Hu (Na-mu Lake), in the west, to Mo-t'o, in the east. The city is the seat of the autonomous region's revolutionary committee, which is headed jointly by the local Communist Party chief and Tibet's military commander.

Improved public utilities

Before the 1960s there was no water supply system in Lhasa, despite the proximity of the Skyid Chu. The only pipeline led to the central temple, and water for daily use was drawn from wells scattered around the city. Refuse was collected for use as fertilizer, and electricity, which was generated by hydroelectric power at Dokde to the north, was restricted to official use. The wealthy used kerosene and gas lamps and the poor, candles and oil lamps. The Chinese administration has apparently improved public utilities, and a large hydroelectric plant has been constructed on the Skyid Chu at Nachin, about 10 miles east of Lhasa.

Health facilities were updated with the operation of Lhasa's first hospital, which largely serves the Chinese community. Medical services were previously provided by the medical college and monasteries, and the government frequently ordered religious rites for the prevention of disease. The homeless, handicapped, and aged were housed and fed by the government. There were no fire-prevention services because Lhasa's houses were built of nonflammable materials. Police services were provided by the government for most of the year and by the religious community during the celebration of the Smon-lam festival.

The educational system, which was run largely by religious institutions, has been taken over by the Chinese government. There are now almost 40 primary and middle schools, and institutes of scientific research in agronomy and animal husbandry. Both the Tibetan and Chinese languages are taught.

Cultural differences with China

Cultural life. Cultural differences have perhaps been the greatest cause of strife between Lhasa's Tibetan population and its Chinese rulers. Buddhism had formerly formed the central core around which all of city life revolved, but the new government is attempting to eradicate it.

The weekly newspaper, the *Tibet Mirror*, which was published in India, and other Indian publications have all been banned, and all publications now arrive from China. Radio broadcasts in Tibetan and Chinese likewise are beamed only from China.

The city's various parks and open spaces are used for recreational purposes. Tibetans enjoy sports, and kite flying, horse racing, soccer, and wrestling were popular. Betting on sporting events was common, as was gambling with cards, dice, and Mah-Jongg (a game of Chinese origin, which is played with tiles); gambling is no longer allowed.

BIBLIOGRAPHY. The present situation and the conditions of the Tibetan people under the Chinese Communists are clearly given in KUNSANG PALJOR, *Tibet: The Undying Flame* (1970). Earlier descriptions of Lhasa are provided by S.C. DAS in *Journey to Lhasa and Central Tibet* (1902, reprinted 1970); LAURENCE A. WADDELL, *Lhasa and Its Mysteries* (1905); F. SPENCER CHAPMAN, *Lhasa: The Holy City* (1938); and GIUSEPPE TUCCI, *A Lhasa e oltre* (1950; Eng. trans., *To Lhasa and Beyond*, 1956).

(T.W.D.S.)

Liaoning

Liaoning (Liao-ning in Pin-yin romanization) is the southernmost of the three Chinese provinces, called by the Chinese the Northeast, that form the region of Manchuria. With an area of 58,300 square miles (151,000 square kilometres), Liaoning is bounded on the northeast by the province of Kirin, on the east by North Korea, on the south by the Yellow Sea, on the southwest by the province of Hopeh, and on the northwest by Inner Mongolian Autonomous Region. Its population was estimated at about 28,000,000 in 1968. The provincial capital is Mukden (Shen-yang).

The area, a region of early Chinese settlement in the Northeast, was known as Sheng-ching in Manchu times (1644 to 1911). The area was redefined in 1903 and named Feng-t'ien; in 1928 the boundaries were altered and it was renamed Liaoning. From 1947 to 1954 the territory was divided into a western province, Liao-hsi, and an eastern province, Liao-tung. In 1954, however, a northern zone was detached and it was re-established as a single province. It achieved its present form in 1956, when the former province of Jehol was partitioned and a portion added to Liaoning. Liaoning. Liao-hsi, and Liao-tung all take their names from the Liao Ho (Liao River). Precedents for the names go back to Han times.

Industrial importance

The economy of Liaoning is by far the strongest in the Northeast and is one of the strongest provincial economies in China. Liaoning is one of the country's principal industrial provinces, producing (with less than 4 percent of China's population and less than 2 percent of its area) about one-eighth of its total industrial production and 40 percent of its steel. One reason for the high level of development in Liaoning is the level of capitalization—very high by Chinese standards—which is based both on investments made under the People's government since 1949 and on important foreign investments made between 1896 and 1945, mainly by the Japanese. Equally, the high level of development in Liaoning cannot be separated from the province's wealth of readily accessible and varied resources, most of them still relatively unexploited at the beginning of the 20th century. Nor can either be separated from the growth of the Chinese population from a total (for the whole Northeast) of about 10,000,000 people in 1900 to more than 66,000,000 in 1968.

No part of contemporary Liaoning is densely populated by comparison with other parts of China where resources are strongly concentrated, and there is no reason to think that economic growth has yet approached its limits. (See MUKDEN; for an associated physical feature, see YALU RIVER; for historical aspects, see CHINA, HISTORY OF; and MANCHURIA, HISTORY OF.)

History. Most of the present Liaoning fell within the confines of the earliest Great Wall of China, built during the reign of the first emperor (221–209 BC), and hence formed part of China from early times. The environment and traditional Chinese civilization of the central plain in Liaoning continue into the North China Plain and into Shantung to the south. Chinese immigration from the south also has a long history, but, until the last years of the 19th century, it was on a modest scale. The traditional economy of Liaoning was one of Chinese peasant farm production in the plains; on the peripheries the economy was based in various places on herding, forestry, fishing, mining, and estate farming. Neither the aristocratic holders of the landed estates nor the people who depended on animals or on the products of the forests or rivers were generally Chinese; as elsewhere in the Northeast, they were of Mongol or Manchu stock.

Under the Ch'ing dynasty (1644–1911), whose own origins lay in the Manchu frontier aristocracy, official efforts were made to protect the Northeast from Chinese encroachment, except for the old Chinese-settled area in the Liao Valley. This policy was gradually abandoned, partly because of the pressure of Russian influence in the north.

Toward the end of the 19th century and on into the 20th, two intersecting forces together created a radically novel situation in Liaoning. One was foreign interference—Russian and, later, Japanese. The other was rapidly increasing Chinese immigration. These two forces resulted in the expansion of the economy of Liaoning at a wholly unprecedented rate.

Patterns of Chinese immigration

Immigrants travelled to Liaoning both by the Shan-hai-kuan land corridor in the southwest and from Shantung by sea to the Liaotung Peninsula in the east. The second group were the more numerous and have been the more successful, both because of the relative ease of travel by sea and because of the better opportunities for both work and settlement in Liao-tung. During the first half of the 20th century, during which time the population of the province rose to a size proportionate to its present figure, about two-thirds of the migrants travelled by sea. The population movement was based on seasonal migration for farm work, with about half of the migrants remaining in the Northeast after the harvests each year. Many strong Shantung communities grew up in Liao-tung, with traditions of mutual dependence and help.

The all-important South Manchurian Railway was constructed by the Russians between 1896 and 1903. This railway linked the new Liao-tung port of Dairen (now Ta-lien) with Ch'ang-ch'un, in Kirin Province, as well as with Harbin, in Heilungkiang Province, and with the then new Chinese Eastern Railway branch of the Trans-Siberian Railway. The South Manchurian Railway passed close to Mukden, replaced navigation on the Liao and much of the old cart transport, and bypassed the old port of Ying-k'ou. The foundations of the modern geography of Liaoning were laid by this railway. In 1907 the Russian railway, port, and territorial privileges were transferred to Japan at the conclusion of the Russo-Japanese War of 1904–05. From that time, Japan continually strengthened its hold on the economic life of Liaoning and the whole Northeast, partly through physical control but also through an active and successful policy of investment and economic expansion. From 1932 to 1945, Liaoning was part of the Japanese-dominated, "independent" state of Manchoukuo. Throughout this period, which included the Sino-Japanese War and World War II (1937 to 1945), Japanese policy aimed to develop the resources of Liaoning in a manner that would complement the economic strength of Japan. Heavy industry was particularly developed. Agriculture was in some degree diverted to soybeans and industrial crops, at the expense of grains. Engineering and light industries also developed but relatively weakly. The concentration on heavy industries represented a degree of imbalance that has been criticized; but, in the face of China's general lack of heavy industrial capacity, these installations have stood her in good stead since 1949.

The period of Japanese domination

Mukden fell to the Chinese Communists in 1948. The industrial installations of Liaoning had suffered heavily from war damage and from Soviet seizures of stockpiles and machinery. Industrial production in 1949 was only 28 percent of that of 1943. The People's government made the restoration of the Northeast one of its first priorities, and, by 1952, production had risen 28 percent above the previous peak of 1943.

After the defeat of the Japanese in 1945, the Soviet Union took over from Japan, and retained, residual rights to share with China the use of the naval base facilities at Lii-shun (Port Arthur) and some rights on the former South Manchurian Railway. These rights were all finally given up in 1955.

The landscape. Liaoning consists essentially of a central lowland, with Mukden at its centre, flanked by mountain masses to east and west. A southward extension of the eastern highlands forms the Liaotung Peninsula. There are four main topographical regions: the central plains, the Liaotung (Kwantung) Peninsula, the western highlands, and the eastern mountain zone.

The four main regions

Climate. The latitude of Mukden in central Liaoning is 42° N; mean temperatures there are January, 9° F (—13" C); April, 46° F (8" C); July, 76° F (24" C); October, 48° F (9° C).

Average annual precipitation is about 26 inches, three-quarters of it falling in the months of June, July, August, and September, and almost none during December, January, and February. At Mukden, there are about 160 to

180 frost-free days a year. The summer rainfall is often torrential; rainfall reliability is about 80 percent. Strong winds occur in spring.

Rainfall in Liaoning as a whole diminishes consistently from southeast to northwest. Everywhere the scarcity of spring precipitation tends to leave growing crops short of water.

Soils and animal life

The central plains. The central plains are the most important area in the province. Structurally, the depression that it occupies is continuous with that of the North China Plain, but, topographically, the Liaoning plains are erosional, rather than depositional, in character. The relief of the plain is undulating but low, and natural drainage is inadequate in many places, creating swamps, some of which have been redrained. Most of the landscape of the central plains is man-made, consisting of cultivated fields. Undeveloped areas include swamps and sand formations. The soils of the middle of the Liao lowland are of the calcareous alluvial type; those of the peripheries to east and west, of brown-forest types; those of the northern peripheries, red earths. The swamps have gley soils (having a sticky layer of clay under the waterlogged surface). Wild animals are scarce, apart from rodents. Locusts are the most important pest. The sea freezes at the coast in winter.

The Liaotung Peninsula. The Liaotung Peninsula is a rugged, mountainous area with a rocky coast. The usual height of the country is 1,000 to 1,500 feet above sea level. The rock types are very mixed, a fact that tends to create a complex and varied topography. Structurally, Liaotung represents a part of the same fold system as Shantung. The coastline is now experiencing submergence. Winters are milder than at Mukden. The January mean temperature at Dairen is 23° F (−5° C). There are about 200 frost-free days. The sea does not freeze at Dairen, though there is floating ice.

The soils of the peninsula, like the rock types and the topography, are very mixed and varied. Most of the best soils are of brown-forest type or of red or yellow loess (an unstratified wind-borne loamy deposit). There has been serious soil erosion, and skeletal soils occur on the steeper slopes. Generally speaking, soils are chemically neutral. The natural vegetation is not well preserved, because of the extent of cultivation and settlement. The forests that remain, mostly on the eastern sides of the hills, contain birch, lime, elm, and pine, together with typical Manchurian trees—oak, apple, and ash. On the western sides, trees are scarce. Wild animal life is now meagre, being almost limited to rodents.

The western highlands. Western Liaoning, fringing the northern shore of the Gulf of Liaotung (Po Hai) between Shan-hai-kuan and Chin-chou, is predominantly a highland area. These highlands comprise the broken and eroded fringe of the Mongolian plateau. They rise in Liaoning to general heights of about 1,500 feet. Toward the sea, the mountains have been intensely eroded by fast-flowing rivers, so that a complex mass of valleys and ridges has been formed. The section of these highlands that lies within Liaoning faces toward the southeast and usually receives adequate rainfall. Winter cold is moderated by the proximity of the sea. The frost-free period is about 200 days. The soils of the western highlands are brown earths in general, but there is great diversity, including thick loess deposits in some places. Vegetation is also very mixed and includes oaks, birches, pines, limes, and spruces. In former times, especially between 1911 and 1948, there was much indiscriminate cutting and thinning of the forests, not necessarily for permanent cultivation. Many areas now have scattered woodland where formerly thick forest stood. The animal life of the western highlands is impoverished by the extent of both forest clearance and human settlement; but it includes the wolf, fox, and marmot and some kinds of deer.

The eastern highlands. The eastern mountain zone lies to the east of Mukden. The least developed part of the province, it consists of a complex mountain mass, extending northward into Kirin Province, with elevations averaging about 1,500 feet. Average precipitation is plentiful, reaching 40 inches per annum in much of the area. There are only 120 to 140 frost-free days. Natural vegetation is predominantly a mixed coniferous and broad-leaf forest. Coniferous trees are mainly cedar, fir, and spruce species; deciduous trees include birch, oak, elm, ash, and walnut species. Soils are predominantly brown-forest types, slightly podzolized (with leached upper leaves), with podzolic–gley and other similar soils in ill-drained areas, often associated with peat.

Demography

Population. The population of Liaoning was officially recorded at 24,095,000 in 1957, the last year for which an official figure has been published. Because of changes in the area of the province since the previous census was taken in 1953, a direct projection of the present population from these figures is not possible. The natural increase of the population in Liaoning reached very high rates in the 1950s, probably reaching about 35 per thousand. Immigration contributed about 20 percent of population increase in the generation before 1957. Either or both of these sources of increase may have slackened since 1957. These figures, projected roughly to 1972, justify an estimate for population in Liaoning in that year of between 33,000,000 and 40,000,000 people, and probably about 35,000,000.

The urban population of the province was 42 percent of the total in 1958, the highest provincial figure of this kind in China; the figure was due, no doubt, to the high degree of industrialization in Liaoning and also to the relatively high capitalization of agriculture. In 1940, the corresponding figure was only about one-tenth. All the big towns are industrial. They include the following (populations for 1960 are given where available): Mukden or Shen-yang (earlier called Feng-t'ien or Sheng-ching), 2,400,000; (Lü-ta; comprises Lü-shun, formerly called Port Arthur, and Dairen, Ta-lien or Dalny), 1,500,000; Fu-shun, 1,000,000; An-shan 800,000. Other major towns are: Liao-yang, Pen-ch'i (Pen-hsi or Penki), Ch'ao-yang, Chin-chou (Chinchow), Ying-k'ou, and Tan-tung (Antung), all with populations below 500,000 in 1957. Some of the larger towns experienced spectacular growth in the earlier 1950s. Between 1948 and 1958, for example, the population of Mukden increased by 54 percent, while that of Fu-shun increased by 50 percent.

Population densities in Liaoning vary considerably. The highest densities (1957 figures), which occur between Mukden and the sea at Ying-k'ou, were upward of 780 per square mile. The usual figure in the developed areas with rail transport was about 520 per square mile. Thinly populated parts, such as the eastern mountains, had about 130 persons per square mile or less. The population-density figure for the whole province was 415 persons to the square mile—about the same as Hupeh to the south.

Ethnic composition and distribution

In Liaoning, 93 percent of the total population is recorded as Han Chinese. The bulk of the national minority population is Manchu, 70 percent, with 1,121,000 Manchu out of the total minority population of 1,600,000. The Manchus are located mainly in the Liao Valley and around Mukden, in the southeast around Tan-tung, and in the southwest around Chin-chou. The second significant minority is that of the Mongols, who are located toward the frontier of the Inner Mongolian Autonomous Region to the west. They number about 235,000. Broadly speaking, the Hui, or Chinese Muslim minority, follow the Manchus in their distribution. There are two autonomous *hsien* (administrative subdivisions representing minority nationalities). They are Fou-hsin (Fu-hsin) Mongol Autonomous Hsien, centred on the coal town of Fou-hsin; and Kharchin East Wing Mongol Autonomous Hsien, in the southwest. A small Korean minority is located near the Korean frontier.

Apart from the registered minority populations, many of the Chinese people of modern Liaoning have origins that are wholly or partly non-Chinese, usually Mongol or Manchu. Many of them are now totally assimilated into the Chinese sector of the population, both in terms of language and custom, as well as in terms of the adoption of contemporary Chinese life-styles.

Resources and economy. *Mining and industry.* Liaoning Province is rich in mineral resources, especially iron ore and coal. Most of the iron ores of Liaoning are concentrated in the triangle formed by the towns of Pen-ch'i, Hai-ch'eng, and Feng-ch'eng, all of which lie to the south of Mukden. These ores are generally easy of access, but their iron content is low—only about 20 to 35 percent; and their quartz content is high. Ores of better quality occur to the north of T'ieh-ling.

Mineral resources and mining

Coal is more widespread; its distribution partly overlaps that of iron. Coal is exploited in three main areas that may be said to radiate from Mukden: one to the north, including Fa-k'u and K'ang-p'ing; one to the east and southeast, including Fu-shun, Pen-ch'i, An-shan, and Tan-tung; and one to the west, extending as far as Ch'ao-yang and including Chang-wu, Fou-hsin, and Pei-p'iao. In 1957 Liaoning produced about 20 percent of China's coal output, although because of growth elsewhere, this proportion is likely to have fallen rather than risen, since that time. The coal reserves of Liaoning form less than 4 percent of national reserves. More than 80 percent of the coal reserve is of bituminous quality, with some anthracite in the southeastern section.

Fu-shun and Fou-hsin have two of the most important collieries in China. Both were exploited under the Japanese but have been expanded since the Communists came to power. Both use opencast (strip-mining) methods. Apart from its use as fuel and in smelting, coal is used in Liaoning to produce synthetic petroleum. Petroleum is also produced from oil shale, which occurs in the Fu-shun area and at Ch'ao-yang, generally overlying coal seams.

Rich reserves of manganese ore occur in the Ch'ao-yang region of western Liaoning and in the southeast. In the eastern mountain area, there are substantial deposits of copper, lead, and zinc; smaller similar deposits also occur in the Chin-chou area in the west, together with an important deposit of molybdenum. Important concentrations of magnesium ore are found at the southern corner of the central plain, at Ying-k'ou and Ying-k'ou-hsien. There are also reserves of other minerals, including bauxite and gold. Sea salt is produced for use in food and in the chemical industry.

Heavy industry

During the first five-year plan of the People's government (1953 to 1957), capital investment in Liaoning was about one-seventh of that in the whole of China. The bulk of this and subsequent investment appears to have taken place in heavy industry, and especially in the industrial network centred on the An-shan iron and steel complex, which forms the industrial heart of Liaoning. In heavy industrial production, Liaoning ranks first among the provinces of China, producing, in addition to steel, over one-third of China's cement and crude oil and about one-fifth of its electrical power.

Apart from iron and steel, the industries of Liaoning include nonferrous-metal processing; machinery manufacture, including electrical, agricultural, mining, and transport machinery, and machine tools; textiles, including both cotton and silk; foodstuffs; paper; and cement. Some hydroelectricity is generated at the Shui-feng (Sup'ung) plant, on the Yalu River, and at the Ta-huo-fang plant, to the east of Fu-shun.

About 75 percent of Liaoning's industrial production by gross value in 1957 came from five cities: Mukden, An-shan, Dairen, Fu-shun, and Pen-ch'i.

An-shan

An-shan is China's principal steel centre. There are abundant local iron-ore resources and supplies of manganese and other metals. Coal is brought there from Fu-shun and Fu-hsin. The first blast furnace (1919) was Japanese, but An-shan has had a long history of smelting from Han times onward. After 1949, An-shan was restored and reconstructed, partly with Soviet aid, under the first five-year plan. It became a principal industrial centre for China, as well as one of the principal suppliers of producer goods within a wide technological spectrum: 5,000 steel products are now manufactured at An-shan. Equipment and trained men have also been supplied by An-shan to other steel centres in China, such as Wu-han. Taken as a whole, An-shan is the biggest single enterprise in industrial investment ever established in China. In 1958 it produced 40 percent of the Chinese iron and steel output. A number of other iron and steel works in Liaoning have been drawn into complementary relationship with An-shan; these are located at Pen-ch'i, Fu-shun, Dairen, Chilin, and T'ung-hua.

Mukden has a wide and varied range of industries, including electricity, metallurgy, machinery and machine tools, chemicals, and building materials; light industries manufacture textiles and paper among other things. Fu-shun is part of the Mukden complex; but its industries. which include artificial petroleum, are all based upon coal. Pen-ch'i, lying to the southeast of Mukden, is another important iron and steel centre, and also manufactures cement.

Dairen lies near the tip of the Liaotung Peninsula and, in the Yellow Sea, is of obvious strategic importance. It has the best harbour in the Northeast. In the Japanese period, Dairen was developed as a port for trans-shipment, servicing, and repairs, and the foundations were laid for the development of its modern engineering industries, including shipbuilding and locomotive production. Chemical industries form the second big industrial grouping in Dairen. There are also a steel mill, an oil refinery and textile mills, together with a food-processing industry and a fishery industry.

Western Liaoning, centred on Chin-chou, is less advanced than the Mukden–An-shan area, or Dairen; but it has valuable mineral resources and some industries. The minerals include coal, oil shale, manganese, and rare metals, for which Chin-chou forms a centre. Coal is mined, particularly at Fou-hsin and Pei-p'iao. Petroleum and other synthetic products are made at Chin-chou and Chin-hsi. There are also cotton and paper mills. Since 1949, industry in Chin-chou has expanded, with emphasis on heavier types of manufacturing.

Eastern Liaoning is also less advanced. In this area, there are important mineral and timber resources. The chief commercial products, apart from farming products, are nonferrous metals, lumber, paper, and wild (tussah) silk. Tan-tung is the chief port of the area, as well as the chief centre for silk reeling, paper, and other light industries such as chemicals.

Agriculture. Agricultural advance in Liaoning has been less spectacular than industrial development. There are several reasons for this. Investment has always been much heavier in industry than in farming. The province's inheritance from the Japanese phase was much less valuable in agriculture than in industry. Liaoning also suffers from both natural calamities such as spring droughts, and, at least until 1958, from backward cultivation methods in many places that resulted in low yields. Exceptional opportunities for employment in industry also tend to deprive agriculture of much of the best labour, in spite of policies designed to prevent this. Yet, in much of Liaoning, topography and soils, and even climate, are quite favourable to agriculture, and the degree of farm mechanization is very high by Chinese standards; Liaoning has more tractors than any other province. One-third of the area of the province is under cultivation, and, of this amount, 60 percent lies on the central plains. The average area of cultivated land per farming family in 1957 was almost four acres. About 15 percent of the cultivated area is used for industrial crops or for export crops like apples; the rest is used for grain crops, vegetables, and soybeans. Liaoning is, nevertheless, a food-deficit province. It must depend partly on food from Kirin and Heilungkiang to the north and partly on imported grain.

Agricultural problems

The summer in Liaoning is not long. In few places, consequently, are two crops per year produced. Corn (maize) and kaoliang (grain sorghum) are the principal crops, followed by soybeans—which are used both as a food and as an industrial crop—and by millet and rice, cotton, and vegetables. Higher yielding corn and rice, formerly grown mainly in the east and the southeast, have tended to supplant millet and kaoliang in the plains. Industrial demand for cotton is very high and competes with the demand for surplus-grain production to feed the industrial workers. The central plain is the best farming area, growing kaoliang, corn, and wheat, with soybeans and vegetables. In the Liaotung Peninsula, with its shorter winter, there is a diversified agriculture, with less grain growing. Three-quarters of the total Chinese output of apples comes from this area; much of the apple crop is exported. Vegetable farming is important, with some double-cropping.

Western Liaoning grows millet and kaoliang as staple crops. Apple growing has been successfully practiced since 1949. In the eastern region, less than 20 percent of the land is cultivated. Rice, sweet potatoes, and corn are characteristic of this area, together with tobacco and the forest products of the mountains. The chestnut-leaved oak feeds the tussah silkworm. The forests support commercial lumber exploitation, but the supply of mature trees is limited, due to previous overexploitation, and the output of lumber is low.

Livestock raising is of minor importance. There are some 8,000,000 head of domestic animals in Liaoning, half of which are pigs. Pigs are bred mainly in the south and central parts, and in the west other animals are raised.

Transportation. The rail transportation facilities of Liaoning are the best in China, and the tonnage transported is also the highest for any province (about one-fifth of the total). The backbone of the transportation system is the Ch'ang-ch'un-Dairen railway (formerly called the South Manchurian Railway), which passes through Mukden and which was double-tracked in 1954. Railroads in Liaoning total (1963) 1,920 miles. Rail traffic primarily comprises either industrial freight or food products in bulk.

Highways in the province are extensive, amounting to about 15,000 miles, but many are of poor quality; the total length of good roads is probably about the same as that of railroads. About 70 percent of the goods transported by road are carried in carts by animals in the traditional style.

Shipping

Navigation, by sea or river, carries less than 1 percent of the internal traffic of Liaoning, but sea navigation is of great importance for transport to other parts of China. Dairen is the greatest port, handling 70 percent of traffic, followed by Ying-k'ou.

Administration and political development. At least until 1958, the Northeast in general and Liaoning in particular stood in the forefront of development in mainland China. The Communists came to power in the Northeast in 1948, and land reform took place here earlier than elsewhere. Communist experiments in economic planning and the accounting systems were also first in the Northeast. For a few administrative purposes, Liaoning has been granted exceptional treatment; for instance, in 1958 the province was grouped with Peking, Shanghai, and Tientsin in respect to revenue provisions. Industrial investment in Liaoning was particularly heavy.

BIBLIOGRAPHY. Material on Liaoning appears in most indexes and bibliographies under "Manchuria." There is very little literature specifically on Liaoning.

SUN CHING-CHIH (ed.), *Economic Geography of Northeast China* (1962; orig. pub. in Chinese, 1959), contains a long chapter on Liaoning. Because of its relatively recent date, and its origin in mainland China as one of a series of standard regional texts, this work carries special authority from the standpoint of information. E.M. MURZAYEV, USSR *Survey of Northeastern China* (1965; orig. pub. in Russian, 1955), is a formal physical geography; O. LATTIMORE, *Manchuria: Cradle of Conflict*, rev. ed. (1935), a wide-ranging and perceptive study of the northeast in historic and modern times. K.C. SUN and R.W. HUENEMENN, *The Economic Development of Manchuria in the First Half of the Twentieth Century* (1969), is a narrower and more technical historical study that utilizes Japanese materials.

(F.A.L.)

Liberalism

Liberalism does not lend itself to easy definition. A major difficulty is that, with some exceptions, liberals have shunned dogma, preferring generally a pragmatic to a doctrinaire approach to social problems. Another, which has been a prolific source of misunderstanding, has been liberals' own frequently opposing views concerning the scope of government. The confusion thus engendered is sometimes compounded by a tendency to identify liberalism exclusively with its 18th- and 19th-century variant, or with the program of this or that liberal party, in a formulation that has on occasion led many to announce the "decline" or "end" of liberalism and to compose obituaries that have been quite misleading. Through the centuries liberalism has changed drastically in content, but it has maintained a constant form. Those who note the first and neglect the second understandably find the term confusing and its application inconsistent.

HISTORICAL BACKGROUND

Liberalism is the culmination of a development that goes back to the Hebrew prophets, the teachings of the pre-Socratic philosophers, and the Sermon on the Mount, from all of which there emerged a sense of the importance of human individuality, a liberation of the individual from complete subservience to the group, and a relaxation of the tight hold of custom, law, and authority.

Throughout much of his history, man as an individual has been submerged in his group. His emancipation as an individual can be understood as a unique achievement of Western culture, perhaps its very hallmark. If this be so, then the emergence of liberalism was, in an important sense, inseparable from Western man's quest for freedom; for liberalism, in the broadest sense, seeks to protect the individual from arbitrary external restraints that prevent the full realization of his potentialities.

Mercantilism and the rise of the middle class

Medieval society did not provide a soil in which the first seeds of liberalism might easily germinate. The Middle Ages produced a society of status in which the rights and responsibilities of the individual were determined by his place in a stratified, hierarchically ordered system. Such a closed, authoritarian order, however grandiose in outline and noble in aspiration, was bound to place great stress upon acquiescence and conformity. As new needs and interests, generated by the slow commercialization and urbanization of Europe, gained strength, the medieval system was modified to accommodate the ambitions of national rulers and the requirements of an expanding industry and commerce. The ensuing policies and arrangements came to be known as mercantilism, a policy of state intervention that, in theory at least, might be extended to regulation of the most minute details of economic life (*cf.* Eli Heckscher, ***Mercantilism***, 1935). However, as such intervention came more and more to serve established interests and to inhibit enterprise, it was challenged by the members of the newly emerging middle class. The challenge took the form of revolt, first against the Universal Church, and later against mercantilist states, presided over by absolute monarchs. The former manifested itself in the Protestant Reformation and the quest of Calvinists and Calvinist sects for freedom of conscience; the latter in the great revolutions that rocked England and France in the 17th and 18th centuries, notably the Glorious Revolution of 1688, the French Revolution a century later, and the successful revolt of England's American colonies. Classical liberalism as an articulated creed is a product of those great collisions.

The fortunes of liberalism differ with the historical conditions in each country—with the strength of the crown, the élan of the aristocracy, the pace of industrialization, and the circumstances of national unification. Thus, by contrast with England, the character of liberalism in France reflected the decadence of its nobility and the absolutism of the Bourbons. The failure of liberalism in Germany in the 19th century was attributable in great part to the dominant role of a militarized and Lutheran Prussia and the reactionary influence of Austria. The advent of liberalism in Italy was delayed by the armies of Austria and of Louis Napoleon and the opposition of the Vatican. Whatever the variations, the liberal impact on authoritarianism reverberated throughout Europe and its dominions, voiced by a Kossuth in Hungary, a Mazzini in Italy, a Thorbecke in the Netherlands, a Bolivar in South America. For a moment even Russia, in 1905 and again between March and November in 1917, heard the echo and, had there been a sufficiently numerous middle class to listen, the course of modern history would no doubt have been changed.

ECONOMIC AND POLITICAL FOUNDATIONS

The authors of the liberal creed vary widely, even in the countries in which liberalism was cradled. But their agreements sufficiently exceed their differences to permit their being included in the same tradition—a tradition whose main manifestations are economic and political. That the classical liberals, perhaps more perceptively than their successors, regarded these as only abstractly separable is indicated in the very title of their science—political economy.

"Private vices, public benefits"

The economics of the "free" market. On the economic side 18th- and 19th-century liberalism based itself on the sovereignty of the market and the "natural harmony of interests." On this view, if individuals are left free to pursue their self-interest in an exchange economy based upon a division of labour, the welfare of the group as a whole will necessarily be enhanced. Classical liberal economists describe a self-adjusting market mechanism free from all teleological influences. While moral goals are invoked and ethical criteria presupposed in passing ultimate judgment on the system, they play no part in determining the sequence of events within it. The one propelling force is the selfishness of the individual, which becomes harnessed to the public good because in an exchange economy he must serve others in order to serve himself. It is only in a *free* market, however, that this happy consequence can ensue; any other arrangement must lead to regimentation, exploitation, and economic stagnation. The most celebrated formulation of this doctrine is to be found in Adam Smith's *Wealth of Nations:*

> He generally, indeed, neither intends to promote the public interest, nor knows how much he is promoting it. . . . by directing. . . industry in such a manner as its produce may be of the greatest value, he intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention. Nor is it always the worse for the society that it was no part of it. By pursuing his own interest he frequently promotes that of the society more effectually than when he really intends to promote it. I have never known much good done by those who affected to trade for the public good. It is an affectation, indeed, not very common among merchants, and very few words need be employed in dissuading them from it.

The same thought had been expressed more tersely by Bernard de Mandeville (*c.* 1670–1733) in the phrase, "private vices publick benefits." A poet (Pope) said it best:

> Thus God and Nature link'd the gen'ral frame,
> And bade Self-love and Social be the same.

Every economic system must be concerned with at least two basic problems: first, arrangement must somehow be made for determining *what* shall be produced—*i.e.*, how "scarce means" shall be allocated—and, second, there must be some way of apportioning what is produced. In a controlled economy this is accomplished by a planning agency acting at the behest of the government. In the economy envisaged by the classical economists of the 18th and 19th centuries, and those conservative economists who today may be called neoclassicists, this is accomplished in the free market through the price mechanism. In such a market the theoretically free choices of individual buyers and sellers determine how the resources of society (labour, goods, capital) shall be employed. These choices manifest themselves in bids and offers that in their totality determine the price at which a commodity will sell. Theoretically, when the demand for a commodity is great, prices will rise, making it profitable for producers to increase the supply; as supply approximates demand, prices will tend to fall until producers divert productive resources to other uses. In this way the closest possible coincidence is said to be achieved between what is wanted and what is produced. Moreover, in the distribution of the wealth thereby produced, the system is asserted to assure a reward in proportion to merit. The assumption is that in a freely competitive economy in which no one is barred by status from engaging in economic activity, the income received from such activity is a fair measure of its value to society.

Man—the rational economic animal

Presupposed in the foregoing is a conception of man as an economic animal rationally engaged in minimizing costs and maximizing gains (wages, profit, interest). Egoistic and hedonistic assumptions about human nature, which were taken for granted, led easily to an emphasis on man as a forward-looking and end-seeking creature. "When matters of such importance as pain and pleasure are at stake . . . who is there that does not calculate," Bentham asked. In Pope's descriptive couplet, although

> Self-love, the spring of motion, acts the soul;
> Reason's comparing balance rules the whole.

If, as the Enlightenment liberals assumed, "reason's comparing balance" rarely fails man in any of his activities (Bentham went so far as to construct an entire theory of penology based on the assumption that even would-be lawbreakers carefully balance the pleasure to be derived from their contemplated crime against the pain of punishment), it could be affirmed *a fortiori* that they would also meticulously balance utilities against costs at the marketplace. Since rational men best know their own interest, it must follow that interference by agencies of government could only diminish that "greatest happiness of the greatest number" that the followers of Bentham claimed to desire.

Assumed also by the classical creed is a conception of the consumer as sovereign, decreeing by his purchases how the resources of society shall be allocated. No one has celebrated this coronation of the consumer more eloquently than Ludwig von Mises in his *Omnipotent Government:*

> Within the market society the working of the price mechanism makes the consumers supreme. They determine through the prices they pay and through the amount of their purchases both the quantity and quality of production. They determine directly the prices of consumers' goods, and thereby indirectly the price of all material factors of production and the wages of all hands employed.
>
> . . . In that endless rotating mechanism [i.e., a market society] the entrepreneurs and capitalists are the servants of the consumers. The consumers are the masters, to whose whims the entrepreneurs and capitalists must adjust their investments and methods of production. The market chooses the entrepreneurs and the capitalists and removes them as soon as they prove failures. The market is a democracy in which every penny gives a right to vote and where voting is repeated every day.

The justification of private property

Implied by the logic of this economic creed is a functional justification of private property, often buttressed, to be sure, by a doctrine of natural right to shield manifestly functionless claims to property. John Locke's *Two Treatises of Government* (1690) pointed the way by defining property as "whatsoever . . . [man] hath mixed his labour with. . . . " Adam Smith found that "The property which every man has in his own labour. . . is the original foundation of all property. . . ." And Bentham noted that "It is this right that has overcome the natural aversion to labour. . . ." Since acquisitiveness and indolence were regarded as inborn and ineradicable human traits, security of property had to be preserved if incentive was not to be destroyed and the producton of goods discouraged. Both French revolutionaries and English gentry could be rallied to such a defense of property; and the American Constitution as well as the French Declaration of the Rights of Man and of the Citizen, both of them liberal testaments, charge government with basic responsibility for its protection.

The role of government. On the political side the guiding principle of historical liberalism has been an undeviating insistence on limiting the power of government. Inspired by the need to remove the state from destructive interference with the economic life of the community, the principle degenerated, under the influence of men like Herbert Spencer, into a doctrinaire form of antistatism. Spencer would even forbid government, either local or national, to assume responsibility for the paving, lighting, and sanitation of cities. Even the less doctrinaire Bentham's sole advice to the state was to "Be quiet," and Edmund Burke—who in this context must be reckoned a liberal—declared that "It is in the power of the state to prevent much evil; it can do very little positive good. . . ." Thomas Paine, eloquent pamphleteer for the American Revolution (of which Burke approved), wrote in his *Common Sense* that "Government, even in its best state, is but a necessary evil . . . ," and generations of Americans almost ritualistically repeated Thomas Jefferson's advice that that government is best that governs least.

The prevailing view was perhaps best spelled out in Adam Smith's enumeration of government functions. Smith assigned the "sovereign" three tasks: to protect the group from outside violence; to protect individual members of society from the injustices or oppressions of their fellow citizens; and, finally, to erect and maintain "those public institutions and those public works, which, though they may be in the highest degree advantageous to a great society, are however of such a nature, that the profit could never repay the expense to any individual, or small number of individuals. . . ." An agenda more faithful to Adam Smith's intent would provide a central place for the enforcement of contracts, without which voluntary individual arrangements could hardly replace government fiat in the conduct of economic affairs.

In general, liberals believed that government must not do for the individual what he is able to do for himself. Legislation like Britain's Ten Hours Act (1847), which limited the labour of women and children to 58 hours a week, was denounced by the English jurist A.V. Dicey as late as 1905 as socialistic. Criticizing the Adulteration of Food or Drink Act and the Sale of Food and Drugs Act of 1899, he contended that such laws safeguard individuals from mistakes "which often may be avoided by a man's own care and sagacity" and therefore "rest upon the idea . . . that the State is a better judge than a man himself of his own interest . . ." (*Law and Public Opinion in England,* 2nd ed.; 1914; pp. 237–238; 263–264). Such views were even more prevalent in the United States.

Separation of powers. The institutional devices by means of which liberals sought to limit government to the exercise of its proper functions were numerous: federalism (though not in France and Britain), bicameralism, the separation of powers. The last of these, that is, the distribution of power among such functionally differentiated agencies of government as the legislature, the executive, and the judiciary, and the system of checks and balances by which it was accomplished, was given its classic embodiment in the U.S. Constitution and its political justification in that brilliant apology for the Constitution *The Federalist* (1787–88), by Hamilton, Madison, and Jay. Montesauieu had already formulated such a doctrine in his famous De *l'esprit des loix* (1748; Eng. trans., *The Spirit of Laws,* 1750), receiving the idea from Locke, who had not developed it fully.

Such a separation of powers could have been achieved, of course, through a "mixed constitution," that is, by having a monarch, a hereditary chamber, and an elected assembly share in power with some appropriate differentiation of function. The Greek historian Polybius hailed such a mixed government as the glory of the Roman constitution, and Blackstone (misled perhaps by Montesquieu) in his famous *Commentaries on the Laws of England* (1765–69) extolled its embodiment in the constitution of England, as Burke was to do later with greater eloquence. But it was precisely despotic kings and functionless aristocrats (more functionless in France than in England) who thwarted the interests and ambitions of the middle class, which turned, therefore, to the principle of majoritarianism.

Liberal conflicts of interest. The greatest check on government is, of course, the threat of dismissal by its constituents. But in determining the crucial question of who the constituents are to be, classical liberalism fell victim to an ambivalence, torn between the great emancipating tendencies generated by the revolutions with which it was associated, and middle class fears that democracy would undermine private property. Most 18th- and 19th-century liberal spokesmen feared popular sovereignty and for a long time suffrage was limited to property owners. In Britain even the important Reform Act of 1867 did not completely abolish property qualifications. France, for her part, bore revolutionaries but suckled reactionaries: although the Revolution of 1789 proclaimed the ideal of universal manhood suffrage and the Revolution of 1830 reaffirmed it, there were no more than some 200,000 qualified voters in a population of about 30,000,000 during the reign of Louis-Philippe, "citizen king" installed by the ascendant bourgeoisie in 1830. And, in the United States, Jefferson's brave language in the Declaration of Independence notwithstanding, it was not until 1860 that universal white male suffrage prevailed.

Although by the time of the Revolution he had changed his opinion, Benjamin Franklin spoke for the Whig liberalism of the Founding Fathers when he observed that "as to those who have not landed property the allowing them to vote is an impropriety." John Adams in his famous *Defense of the Constitutions of Government of the United States of America* (1787) was more explicit, finding that, if the majority were to control all branches of government, "Debts would be abolished first; taxes laid heavy on the rich, and not at all on others; and at last a downright equal division of everything be demanded and voted." Thomas Babington Macaulay spoke for English Whigs when he found universal suffrage "incompatible, not with this or that form of government, but with all forms of government." French statesmen like François Guizot (1787–1874) and Louis Adolphe Thiers (1797–1877) expressed similar sentiments and on the rest of the continent universal suffrage was for the most part a remote ideal until the turn of this century.

The fear
of majority
tyranny

Objections to democracy were not limited to misgivings about the fate of private property. Many liberals genuinely feared the potentialities for tyranny latent in democracy. If the will of the majority is to be supreme, everyone will be at its mercy. Benjamin Constant (1767–1830), noted French diplomat, expressed a general concern when he observed that from the point of view of the individual it makes no difference whether he is tyrannized by a single despot or by the totality of individuals composing his society; he is oppressed just the same. Indeed the latter could well be worse: tyrannicide might rescue him from an individual oppressor and, in any event, others would commiserate with him; whereas if oppressed by a large majority, he would have no comparable recourse or comfort.

Despite such misgivings, the voices of Thomas Paine and Jefferson, of Rousseau and de Tocqueville, and of the English philosophical radicals led by Bentham and James and John Stuart Mill finally prevailed among the liberal philosophers. But the democratic principle that they espoused had first to be reconciled with the liberal requirement that power be limited. The problem was to accomplish this in a manner consistent with the democratic ideal, that is, without doing violence to the majoritarian principle and in conformity with Bentham's dictum that each person is to count as one and no one as more than one. How, since hereditary elites were discredited, could the power of a majority be checked without giving greater voice to property owners or some other kind of "natural" elite?

The liberal formula. The essence of the liberal solution is twofold. It lies partly in making the decision of any given majority subject to the concurrence of other majorities distributed over a period of time. The majority that elects a president of the United States, for example, is different from the majority that two years before elected one-third of the Senate and two years later elects another third. Likewise two-thirds of the senators are elected by a majority other than the majority by which members of the House of Representatives are elected. These groups, in turn, are checked by the Constitution, which was approved and later amended by majorities no longer alive. While an act of Britain's Parliament is forthwith a part of the British constitution, Parliament before acting on a highly controversial issue of great concern to the country will seek a mandate from the people, that is to say, a majority other than the majority that elected it. Thus, the "people" to be consulted in a liberal democracy, if they are not quite that partnership envisaged by Burke ". . . between those who are living, those who are dead, and those who are to be born," are not either, except in a revolutionary crisis, a "momentary aggregation" whose sovereign will may be ascertained by a plebiscite. Burke's characterization of democracy as "the vulgar practice of the hour" may well apply to those spontaneous assemblies of the people to which Rousseau would assign plenary power; it can be denied that it applies to the mainstream version of liberal democracy that is essentially *constitutional* democracy, that is to say, democracy in which the power of a current majority is checked by the verdicts of majorities that preceded it.

The second part of the solution is related more directly to the initial inspiration of liberalism, its basic commitment to the autonomy and integrity of the individual, which the limitation of power is, after all, intended to preserve. In the liberal understanding, the individual is not only a citizen who shares a social compact with his fellows; he is a person and, as such, he possesses rights that the state may not invade. Even as a citizen, he must, if majoritarianism is to be meaningful, possess such rights. Majorities do not form in a vacuum. Unless a majority verdict is some miraculous coincidence or spontaneous merger of individual judgments—which it is not—it can come about only if individuals are free to some extent to formulate and express and exchange their views. This involves, beyond the right to speak and write freely, the freedom to associate and organize and, above all, freedom from fear of reprisal. But the individual has rights apart from his role as citizen. These are rights that secure his personal safety and hence his protection from arbitrary arrest and punishment. And beyond these are those rights that preserve large areas of *privacy.* In a liberal democracy there are affairs that do not concern the state, even if the state's interest in them reflects the will of an overwhelming majority. Such affairs may range from the worship of God, to works of art, to how parents raise their children. And for the liberals of the 18th and 19th centuries they included, above all, most of those activities in which individuals engage in production and trade.

Classical liberals differed in their interpretation of individual rights—whether to understand them as "natural" (Locke) and therefore "inalienable" and "indefeasible," to justify them as functional (Bentham) or traditional (Burke). But eloquent and persuasive declarations affirming such rights were embodied in the English Bill of Rights of 1689, the American Declaration of Independence and Constitution, the French Declaration of the Rights of Man and of the Citizen of 1789, and the basic documents of nations throughout the world that later used these as their models. Freedom thereby became more than the right to make a fractional contribution in an intermittent mandate to government; it designated the right of men to live their own lives.

The problems of the dispossessed

Rhetoric and reality rarely coincide. Liberals were not usually troubled over the small numbers by whom, in fact, such freedom was enjoyed, and the masses were still too inarticulate to remind them. In the dawn of the Industrial Revolution, the blessings of liberty must have seemed remote to millions of brutally exploited factory workers huddled together in the slums of their dreary cities. Liberals have been accused of concentrating exclusively on the property right to the neglect of such abuses, and in general, few liberals paused to ponder how the freedom they cherished would be used, and it did not occur to them, as it has to a later generation of libertarians, that such freedom brings with it a heavy burden of choice and decision. Finally, in their concern for the individual, liberals depicted him too simply as an isolated monad. and society as a collection of such monads, failing to reckon with the myriad social relationships through which individuals achieve identity.

ACHIEVEMENT AND FAILURE

In historical perspective it can be seen that the complex of forces, of which classical liberalism was the rationalization, wrought great changes. The feudal system was destroyed. Capitalism replaced the static society of the Middle Ages. A functionless aristocracy was deprived of its privileges. Tyrants were challenged and curbed. The middle class was left free to employ its energies in expanding the means of production and vastly increasing the wealth of society. Representative government came into its own. As they set about limiting the sovereign power, liberals converted the ideal of constitutional government into a reality, and they developed a doctrine of individual rights, including the right to worship freely, the right of a free press, of free speech, and of free assembly, which lies at the heart of modern democracy. J.S. Mill's essay *On Liberty* (1859) is justly celebrated as one of the great testimonials to civil liberties.

However, vast economic changes, first in Great Britain and later in the United States, led increasingly at the turn of this century to disenchantment with the principal economic basis for liberalism—the ideal of a market economy. (This ideal had never evoked comparable loyalties on the continent of Europe.) The advent on a large scale of absentee ownership made it increasingly difficult to invoke a functional justification for many forms of private property. Multitudes whose real bargaining power fell far short of what it was in theory—especially those seeking work—did not actually enjoy the "free choices" postulated by the economists. Economic man's "pecuniary sagacity" (Veblen) fell far short of the idealizations of textbook writers; and modern psychology suggests that he is at least as much a creature of impulse, habit, and custom as he is a rational calculator with his eye on the main chance. The often impenetrable complexity of goods offered on the market in an economy where transactions are no longer simple affairs of horse-trading, not to mention the mendacity of much advertising, seemed to make the consumer more a subject than a sovereign. Reality failed to approximate the state of free competition envisaged by the classical economists. Orthodox liberal economists have accordingly referred with increasing frequency to "frictions" and "exceptions" and have employed qualifications ("in the long run"; "other things being equal") to save their generalizations. The result has been an abstract science that a new generation of liberals and multitudes of socialists came to regard as having only limited relevance to the real world.

Concentration of wealth and irrational economies

Worst of all, according to its critics, the profit system concentrated vast wealth in relatively few hands, with several decisively adverse consequences. First, great masses of people failed to benefit from the wealth flowing from mines and mills and lived under impoverished conditions that became increasingly anomalous in an affluent society. Second, since those who alone could consume the output of a vastly expanded productive system lacked the purchasing power, the system, after other markets were glutted, endemically failed to realize its productive potentialities and recurrently came to a near halt in periods of stagnation that have come to be called "depressions." Finally, those who owned the means of production, or their managers, were endowed with vast power that could be used to overwhelm the individual as surely as the power of a 17th-century despot, thanks to the device of incorporation on a scale far beyond the simple requirements of efficiency and economy in production. In short, the near atrophy of government achieved in the 19th century had left a vacuum that private interests readily filled. Businesses were effectively organized and some of them used their power to influence and control government, to manipulate an inchoate electorate, to limit competition, and to obstruct substantive social reform. Some of the same forces that had once released the productive energies of Western society now restrained them; some of the very energies that had demolished the power of despots now nourished a new despotism. Such, at any rate, was the verdict of 20th-century liberals and such were the conditions that led them to oppose private collectivism by supporting a positive role for government and encouraging the formation of power centres outside business and government.

CONTEMPORARY LIBERALISM

The contemporary variant of liberalism is even more amorphous than the classical. There are no "fathers" such as John Locke or Adam Smith. In Germany and the Scandinavian countries where socialists, even though they trace their ancestry to Marx, are overwhelmingly revisionist, it is difficult to distinguish between their program and programs elsewhere designated as liberal. British socialism never embraced Marx and, as the Labour Party grows less doctrinaire, it is increasingly difficult to distinguish it from the Liberal Party, which largely accounts for the decline of the latter. U.S. liberal social legislation since the crash of 1929 has been notably formless. But even so, the outlines of contemporary liberalism are fairly discernible.

Critique of the market. Cognizant of the real achievements of the profit system, liberals do not seek its abolition, only its modification and control. They find no fixed line laid up in heaven eternally dividing the private and the public sectors of the economy; the determination, they contend, must be by reference to what works. The spectre of regimentation in completely planned economies and the dangers of bureaucracy even in mixed economies deter them from jettisoning the market and substituting an omnicompetent state. On the other hand—and this is a basic difference between classical (or neoclassical) and contemporary liberalism—most liberals now believe that the dispensations of the market, as it has in fact operated, must be supplemented and corrected in substantive ways. They hold that the rewards dispensed by the market are too crude a measure of the contribution many or most people make to society, and that the needs of those who lack opportunity or are physically handicapped are ignored. They contend that enormous social costs incurred in production are not reflected in market prices, and that resources are used wastefully. Not least, liberals charge that the market biasses the allocation of human and physical resources in the direction of satisfying superficial wants (for oversized motor cars in annual models, changing fashions in attire, and unnecessary gadgets), while basic needs (for schools, housing, rapid public transit, sewage treatment plants) go unmet. Finally, although liberals believe that prices, wages, and profits should continue to be subject to negotiation among the interested parties and responsive to conventional market pressures, they insist that price–wage–profit decisions affecting *the economy* as a *whole* must be reconciled with public policy.

The liberal program. To achieve a juster distribution of rewards, liberals have relied on two major strategies. First, they have promoted the organization of workers and consumers to improve their power to bargain with employer-producers. Such a redistribution of power has had political as well as economic consequences, making possible a party system in which at least one party is responsive to the interests of wage earners and consumers. Second, enlisting the political support of the economically deprived, liberals have evolved the so-called welfare state, with its panoply of social services "from the cradle to the grave."

Foundations of the welfare state

Social legislation, beginning with free public education and workmen's accident insurance, now includes support for all who are physically and mentally handicapped, programs of old-age, unemployment, and health insurance, minimum-wage laws, and—for the most part still in the blueprint stage—guaranteed annual incomes. Such legislation is most comprehensive in the Scandinavian countries and, among countries with mature economies, possibly least comprehensive in the United States where social legislation at the federal level was virtually ignored until passage of the Social Security Act of 1935.

Liberals have been less successful in correcting what they regard as an irrational schedule of priorities by using the taxing powers of government to obtain greater allocations for the public sector of the economy. They have been least successful, at any rate in the United States, in persuading the business community and labour unions to accept government participation in pricing and wage decisions.

Initially, the quest for these objectives was viewed as requiring a redistribution of wealth to be achieved by steeply graduated income and death (inheritance) taxes—a program likely, if it were to become confiscatory, to evoke the unlimited resistance of high income groups. Increasingly, as modern technology promised mature economies miracles of abundance, attention shifted to the institutional failures that kept such economies from fully realizing their productive potentialities, especially during periods of mass unemployment and depression. The culmination of this effort was the White Paper on Employment Policy produced by Britain's wartime coalition government and, in the United States, the Employment Act of 1946, which goes well beyond the British, declaring it to be "the continuing policy and responsibility of the Federal Government to use all practicable means . . . to promote maximum employment, production, and purchasing power." Thereafter the old rhetoric about "sharing the wealth" gave way to concentration on growth rates as liberals, inspired by J.M. Keynes' General Theory of Employment, Interest and Money, urged the use of fiscal policy—*i.e.*, use of the government's power to borrow, tax, and spend—not merely to counter contractions of the business cycle but to encourage expansion of the economy. Here, clearly, was a program less disruptive of class harmony and the basic consensus essential to a democracy than the old Robin Hood method of "taking from the rich and giving to the poor."

Industrialism reconsidered

In the 1960s the emphasis of liberals again shifted as it became more and more evident that expanding production is not an unmixed blessing. The litany is now a familiar one. The same industry that produces our wealth pollutes our rivers, lakes, and atmosphere. Its gleaming artifacts become mountains of junk and it menaces the ecological balance of nature. It concentrates millions in drab cities that, in the United States at least, are being evacuated by those who can afford to escape to the suburbs. The result is a daily movement of traffic, choking the highways, poisoning the air, jangling the nerves. The inventory could be extended indefinitely. Much of the "wealth" produced by industry hardly justifies such consequences.

Some criticism goes deeper. It is urged that, quite apart from its disastrous impact on the environment, modem technology deracinates man, depriving him of his sense of identity, chaining him to a treadmill, trapping him in a depersonalized, regimented, over-organized society. The voice of the romanticist, denouncing preoccupation with production as a bourgeois aberration or an archaic Protestant ethic and urging a return to the relaxed, simple life, is once again heard. That voice is often found persuasive, it so happens, by a new generation of youthful intellectuals and vagabonds.

Most liberals contend that environmental disruption is not a necessary consequence of expanding industry but of a failure to subordinate the quest for quick profits to the requirements of intelligent planning. The "fallacy of romantic regression," as the American psychologist Kenneth Keniston has called it, is also rejected. Nevertheless, there is increased respect for Freud's reminder in *Das Unbehagen* in der *Kultur* (Eng. trans. Civilization and Its Discontents, 1930) that "power over nature is not the only precondition of human happiness," and more and more concern is expressed for the "quality" of life. However, Freud went on to observe that, if happiness requires more than power over nature, this "is no ground for inferring that technical progress is worthless from the standpoint of happiness." Mainstream liberals have agreed.

THE FUTURE OF LIBERALISM

At present the traditional coalition on which liberals have relied is in disrepair in many countries. Ethnic minorities have become confused and divided by the separatist creed of their more militant leaders; the academic community is traumatized and in a state of turmoil; and labour unions lack a good deal of their old dynamism; even the once engagé artist has been lost to aestheticism and esotericism. The increasing polarization between right and left finds the liberals attacked on both flanks, but the appearance of beleaguerment may well be deceptive. Much of the liberal program has in fact become public policy. When, in the 1930s, Pres. Franklin D. Roosevelt established the National Resources Planning Board, it was called un-American by conservatives, its budget was slashed, and it was finally abolished. Today, although most conservatives might resist liberal proposals to establish a council on national goals, they would no doubt support a national agency to plan the use of natural resources. Despite appearances, Western society could be on the eve of less, rather than more, polarization.

All of this suggests a new source of strength for liberalism epitomized in what has been called the "end of ideology." Whether the movement of Western capitalist nations toward mixed economies and the Soviet Union's flirtation with the profit motive is indicative of such a trend need not be explored here. But, clearly, both socialists and conservatives in western Europe (and the British Commonwealth) are much less prone to apply pre-established dogma to new problems. This is equally true of Conservatives and Labourites in Britain. And, except for the far right and left, there is a growing tendency in the United States to reject doctrinaire approaches to social problems. Ideology may not be dead, but it appears to be in decline. If the result is a pragmatic, experimental temper in which old principles are adjusted to new needs and dissent is not identified with disloyalty, this is very much in the liberal mood.

Changing policies, enduring values

The expansion of government power and responsibility sought by liberals today is clearly opposed to the contraction of government power and responsibility sought by liberals yesterday. The content of liberalism varies with varying conditions: liberals may one day challenge and another day cherish the church; in one age they may seek less government intervention in economic affairs, in another age more; they have been hospitable to the interests and ambitions of the business community, under changed circumstances they may be hostile; for decades they have preached the virtues of labour unions, they may one day consider their vices. But in every case the inspiration is the same: a hostility to concentrations of power that threaten the freedom of the individual and prevent him from realizing his potentialities; a willingness to re-examine and reconstruct social institutions in the light of new needs. This willingness, tempered by aversion to sudden, cataclysmic change, is what sets the liberal off from the radical, who often ignores its hazards. Also the very eagerness constantly to entertain and encourage useful change distinguishes the liberal from the conservative. If the content of liberalism varies, the above listed characteristics constitute its distinctive and enduring form.

The 18th-century French philosopher Condorcet could write that "human perfectibility is in reality indefinite" and that "the progress of this perfectibility, henceforth independent of any power that might wish to stop it, has no other limit than the duration of the globe upon which nature has placed us." Another French philosopher of the Enlightenment, Helvétius, wrote that "to be happy and to be powerful is only a matter of perfecting the science of education." In his ***Autobiography***, John Stuart Mill, reminiscing about the elder Mill, writes:

> So complete was my father's reliance on the influence of reason over the minds of mankind, whenever it is allowed to reach them, that he felt as if all would be gained if the whole population were taught to read, if all sorts of opinions were allowed to be addressed to them by word and in writing, and if by means of the suffrage they could nominate a legislature to give effect to the opinions they adopted. He thought that when the legislature no longer represented a class interest, it would aim at the general interest, honestly and with adequate wisdom . . .

Commenting on the faith of the mid-19th century, Henry Adams noted: "Education was divine, and man needed only a correct knowledge of facts to reach perfection . . ."

Sobered by the tragic events of this century and informed by new psychological insights into man's biasses and perversities, a chastened and far more sophisticated liberalism no longer shares the naive confidence of most classical liberals in man's rationality and perfectibility and in the inevitability of progress. Most of today's liberals are more likely to heed those who warn that human nature is ineradicably flawed than heed those who hope that man may be persuaded to apply the scientific method to the solution of social problems and thereby find contentment. There is a strong suspicion that if man had no serious social problems he would invent them. Nevertheless, the strong commitment of liberals to social reform suggests a persistent optimism and a belief that man can control his fate and build a better world. To this extent the Heavenly City of the Philosophers still beckons them.

BIBLIOGRAPHY

General works: R.D. CUMMING, ***Human Nature and History: A Study of the Development of Liberal Political Thought*** (1969); H.K. GIRVETZ, ***Evolution of Liberalism***, rev. ed. (1963), a comparison of classical and contemporary liberalism that examines the psychological assumptions underlying the political and economic views of the classical liberals; L. HARTZ, ***The Liberal Tradition in America: An Interpretation of American Political Thought Since the Revolution*** (1955); K. MARTIN, ***The Rise of French Political Thought***, 2nd ed. K.R. MINOGUE, ***The Liberal Mind*** (1963), on the contention that liberalism now reflects a moral and political consensus; W.A. ORTON, ***The Liberal Tradition: A Study of the Social and Spiritual Conditions of Freedom*** (1945); F.M. WATKINS, ***The Political Tradition of the West: A Study in the Development of Modern Liberalism*** (1948).

Classical liberalism: A.V. DICEY, ***Lectures on the Relation Between Law and Public Opinion in England During the Nineteenth Century*** (1905); E. HALEVY, ***La formation du radicalisme philosophique***, 3 vol. (1901–04: Eng. trans., ***The Growth of Philosophical Radicalism***, 1928); L.T. HOBHOUSE, ***Liberalism*** (1911); H.J. LASKI, ***The Rise of Liberalism: The Philosophy of a Business Civilization*** (1936); GUIDO DE RUGGIERO, ***Storia del liberalismo europeo*** (1925; Eng. trans., ***The History of European Liberalism***, 1927); LESLIE STEPHEN, ***The English Utilitarians*** (1900).

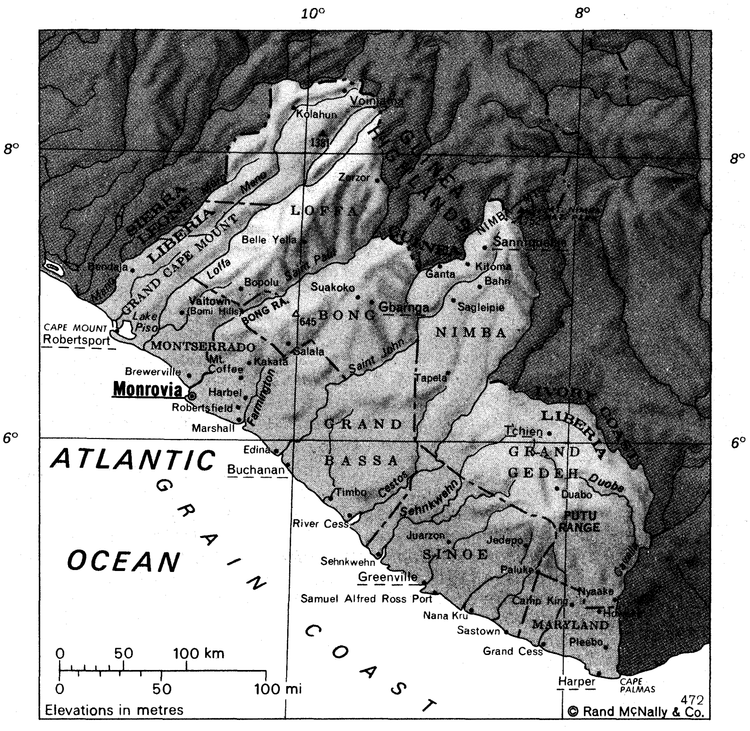
Contemporary liberalism: A.A. BERLE and G.C. MEANS, ***The Modern Corporation and Private Property*** (1933), the classic account of how the private corporation has generated "managerial" power; W.H. BEVERIDGE, ***Full Employment in a Free Society*** (1945), by the chief English architect of the "welfare" state; G.D.H. COLE, ***Economic Planning*** (1935); C.A.R. CROSLAND, ***The Future of Socialism*** (1957), by a leading theoretician of the British Labour Party who believes that the Keynesian revolution has made many traditional socialist attitudes obsolete; JOHN DEWEY, ***Human Nature and Conduct: An Introduction to Social Psychology*** (1922), the best criticism of the psychological preconceptions of the classical liberals, ***Liberalism and Social Action*** (1935); JOHN KENNETH GALBRAITH, ***American Capitalism: The Concept of Countervailing Power***, 2nd ed. rev. (1956), ***The Affluent Society*** (1958), one of the first works to contend that liberal emphasis should shift from greater productivity to the question of priorities, ***The New Industrial State*** (1967); A.H. HANSEN, ***Economic Policy and Full Employment*** (1947), by a leading American exponent of Keynesianism, ***Fiscal Policy and Business Cycles*** (1941); S.E. HARRIS, ***Economic Planning*** (1949), an examination of economic planning in ten European countries as well as in the U.S., Argentina, Japan, and India; JOHN KEYNES, ***The End of Laissez-faire*** (1926), ***General Theory of Employment, Interest and Money*** (1936), the most influential economic tract of the first half of this century; W.A. LEWIS, ***The Principles of Economic Planning*** (1951), on the difference between direct and indirect planning, and its important implications; JAMES E. MEADE, ***Planning and the Price Mechanism*** (1949), a liberal socialist approach to the problems of planning and use of the price mechanism in postwar Britain; GUNNAR MYRDAL, ***Beyond the Welfare State*** (1960), on the social impact of economic planning in highly industrialized states and the economic relations between them and less developed economies, ***Challenge to Affluence*** (1963); E.V. ROSTOW, ***Planning for Freedom*** (1959), an analysis of the impact of the Keynesian "revolution" on government intervention in the economic sphere.

(H.K.G.)

Liberia

The oldest republic on the African continent, and the only black African state never to have been subjected to colonial rule, Liberia is situated on the West African coast in the tropical rain forest region. It has an area of about 43,000 square miles (111,400 square kilometres) and a population of about 1,500,000. It is bounded by Guinea to the north, Sierra Leone to the northwest, the Ivory Coast to the east, and the Atlantic Ocean to the south and west. The capital is the port city of Monrovia (population about 100,000).

Formerly relying heavily upon its rubber production, Liberia, since World War II, has also become a major producer of iron ore. Liberia has preserved a traditional friendship with the United States, which, during the 19th



LIBERIA

century, played a role in the country's foundation; Liberia's currency is the United States dollar, with Liberian coins being used in smaller denominations. Under the presidency of William V.S. Tubman (1944–71), an Open Door economic policy encouraged Western investment. Liberia is a founding member of the United Nations as well as of the Organization of African Unity (OAU).

THE LANDSCAPE

Physical features

Physiography. *Relief.* Liberia has a somewhat complex relief, and its physical features are not uniform. The coastal terrain, which is 350 miles long and which varies from 15 to 25 miles in width, is sandy, low, and interspersed with lagoons and swamps, as well as with occasional rocky promontories. Behind this region is rolling hill country. Farther inland there occurs a thin strip of steep scarps and hills, about 20 miles wide, and trending northeast to southwest. This is followed by a relatively flat highland about 1,500 to 2,000 feet in height, cut into hills and valleys by erosion. Finally, farthest from the coast, in the northwest and centre, is a mountainous region of which one of the most striking features is Mt. Nimba, at the Guinea–Liberian border, which rises to a height of 4,500 feet (1,372 metres) on the Liberian side and which contains rich deposits of iron ore. The Mano and Morro rivers form the northwestern boundary of the country, and the Cavalla River forms the eastern boundary. Apart from these, there are five major rivers, all of which drain into the Atlantic Ocean—the Lofa, St. Paul, Farmington, St. John, and Cestos rivers. The rapids on the St. Paul River, 17 miles northeast of the capital, have been harnessed as a source of hydroelectric power.

Rocks and soils. Liberia forms part of the Precambrian Shield, an ancient crystalline rock formation 2,700,000,000 to 3,400,000,000 years old, composed of granite, schist (a coarse-grained rock containing mica), and gneiss (another coarse-grained rock, in which bands of granular minerals alternate with bands of schistose minerals). The shield underlies the entire African continent and lies exposed over about one-third of its surface. Specimen granitic rocks from western Liberia have been found to be between 2,500,000,000 and 3,600,000,000 years old. In Liberia, the shield has been subjected to intense folding and faulting; it is interspersed with iron-bearing formations known as itabirites. Along the coast lie beds of sandstone, while intrusive rocks crop out, forming such promontories as Cape Mount, which is 1.050 feet (320 metres) high, Cape Mesurado, and Cape Palmas. Monrovia, the capital city, stands on a ridge of diabase, a basaltic (dark-coloured, fine-grained) rock, about four to six miles (six to nine kilometres) long and one to two miles wide. Diabase and gabbro (a dark-coloured, coarse-grained igneous rock) occur in the interior of the country where, as a result of geological eruptions, they have intruded into fissures in the older rock in long and narrow masses.

The soil is an evenly disposed bed of laterites (rust-red aluminum and iron oxides), which have been eroded and repeatedly leached as a result of heavy rainfall. Because of its thin top soil it is more suited to the growth of tree crops than of food crops.

Temperature and rainfall

Climate. The climate, especially on the coast, is warm and humid all year round. It is dominated by a dry season from November to April, and by a wet season from May to October. A minor "middle dry" season occurs from mid-July to August. The dusty and dry desert winds, known as the harmattan, which blow from the Sahara to the coast in December, also influence the climate. Temperatures range between 65° and 85° F (18" to 29° C). The greatest amount of rainfall, 205 inches (5,200 millimetres) occurs on the Cape Mount Promontory; rainfall diminishes inland to about 70 inches (1,775 millimetres) on the central plateau. The interior has hot but pleasant days and cool nights during the dry season.

Vegetation. Liberia has year-round evergreen vegetation. Its tropical rain forest contains some trees that attain heights of 200 feet (61 metres). Many trees are of commercial value, but natural stands of a single species are rare, thus preventing easy exploitation by the lumber industry. Forestry operations are therefore concentrated on scattered species. Lumber is put to a variety of uses. Termite-resistant woods, such as red ironwood, camwood, and whismore (a large tree resembling mahogany, but tougher), are used for construction, while several African mahoganies are used for cabinet and veneer woods. Farmers, in order to practice crop rotation, cut or burn 50,000 acres (20,000 hectares) of often valuable forest trees each year. As a countermeasure, a Forest Conservation Act was passed in 1953. Other trees of value are rubber, cocoa, coffee, and the raffia palm.

Animal life. Liberia's tropical rain forest is not conducive to an abundant animal life. The monkey, chimpanzee, some species of small antelope, and various rodents are to be found. There is also a pygmy hippopotamus and several species of anteater. Elephants, bush cows (short horned buffaloes), and leopards are present, but gradually disappearing. There are many reptiles and at least eight poisonous snakes, among them the black cobra and the causus viper, or night adder. Three types of crocodiles have been identified. Scorpions and lizards are numerous, and there are several unique species of bats and birds. Fish are also numerous. Pond-fish culture is being encouraged.

The landscape under settlement. There are more than 2,000 villages, the majority of which are concentrated in Bong County in central Liberia, Lofa County in the northwest, and in the coastal county of Montserrado. Different architectural styles are used by different tribes. The huts in the villages of the Vai people, for example, are both round and rectangular and are handsomely decorated with clay and geometric designs. Traditional activities in the villages include farming, hunting, and fishing, although weaving and carving are also practiced.

Monrovia

The capital of Liberia, Monrovia, founded in 1822, is situated on the left bank of the St. Paul River. Standing on the ridge formed by Cape Mesurado, it commands an imposing view of the Atlantic Ocean. The city proper, although not precisely defined, occupies an area of about one square mile, while its outlying districts and suburbs occupy another four square miles. The population of Monrovia is about 100,000. The old style of architecture that once characterized it, reminiscent of that of the southern United States before 1860, is now fast disappearing, giving way to contemporary styles of housing. Monrovia was founded by Liberian settlers from the United States after an earlier attempt to settle near Sierra Leone had failed. It is the political, administrative, and industrial centre of Liberia. All of the tribes of Liberia are represented in its population, as also are other groups, including refugees, African nationals from other countries, and Europeans. Elsewhere, some mining towns have been built in rural areas since World War II.

People and population. The group most closely associated with the founding of the Liberian state were Negroes from the New World, known historically as Americo-Liberians. Most of them migrated to Liberia between 1820 and 1865, during and immediately after the suppression of the transatlantic slave trade. Other Negroes from the Americas have since then continued to migrate intermittently. Out of Liberia's population of about 1,500,000, however, they number fewer than 50,000. The government is still controlled by this group, although a unification and integration policy that encourages equal opportunities for all citizens, tribal and nontribal, is slowly effacing divisions.

The African tribes may be classified into three linguistic groups, the Mande, the Kru, and the Atlantic (West Atlantic) groups. The Mande are themselves divided into two groups, the Mande-tan (so named because the number ten is *tan* in their language), called the Nuclear Mande by anthropologists, who are located not only in northwest Liberia but also in Senegal, Mali, Guinea, and Sierra Leone; and the Mande-fu (so named because the number ten is *fu* in their language), known to anthropology as the Peripheral Mande. Prominent among the Mande-tan are the Malinke (Mandingo). The Mande-fu group, who inhabit the northern and central region of Liberia, include the Loko, Gbande, Gio, Kpelle, Toma (also known as Buzi or Loma) Mano, Mende, and Vai tribes. The Kru include the Bassa, De, Grebo, and Kru tribes of the coast, and the Kran Padebu, Sapo, Sikon, and other tribes occupying the interior and southern half of the country. The Atlantic group includes the Gola, Kissi, and other tribes in the north. Traditionally, different tribes have shown a preference for different shapes of huts, the Mande, for example, preferring round dwellings with conical thatched roofs, while the Kru prefer rectangular houses. Architectural borrowings have, however, occurred in both directions. The Vai are renowned for having invented an alphabet in the 19th century; during World War II this alphabet was borrowed and improved on by the Germans, who, having acquired knowledge of it from German traders in the Cape Mount area, used it for a code. The De, Kpelle, Mano, Mende, and Vai, among others, have secret societies known as the Poro school, for men, and the Sande school, for women. These institutions exercise educational, political, legal, and religious functions. In effect, only the Bassa, Kru, and Malinke are reputed not to have had Poro in the past. The Kru tribes are organized by age sets, or generations; long distinguished by militarism, they opposed the Liberian government in Monrovia until the 1930s.

Tribal dwellings

Liberia, Area and Population

	area*		population	
	sq mi	sq km	1962 census	1970 estimate
Counties				
Bong	3,650	9,454	132,000	
Grand Bassa	5,075	13,144	132,000	
Grand Cape Mount	2,250	5,828	32,000	
Grand Gedeh	6,575	17,029	59,000	
Lofa	7,475	19,360	123,000	..
Maryland	1,675	4,338	63,000	...
Montserrado†	2,550	6,605	259,000	...
Nimba	4,650	12,044	161,000	...
Sinoe	4,350	11,267	56,000	...
Total Liberia	38,250 43,000	99,068‡ 111,400	1,016,000‡	1,523,000‡

*Area figures given purportedly represent land area, although some inland water appears to be included. Second total is generally accepted total area of Liberia. †All area and population figures for the commonwealth district of Monrovia are included in Montserrado County. ‡Figures do not add to total given because of rounding.
Source: Official government figures.

There is no adequate reporting of births and mortality rates. In 1969 only 12,000 births and fewer than 2,000 deaths were reported. Immigration from other West African countries and elsewhere has occurred in recent years. No immigrants were admitted in 1969. Emigration, especially among the population under 45, appears to be increasing.

Population distribution is markedly uneven, amounting to less than the national average of 40 persons per square mile in many areas. There is a concentration of about 100,000 per square mile in the capital and about 6,000 per square mile in mining centres. This uneven distribution is expected to continue as settlement follows the siting of industries, resulting in either permanent or seasonal migration from rural areas.

The national economy. Liberia's buoyant economy has been attributed to the Open Door Policy, which encourages foreign investment, promulgated by the Tubman administration. Government expenditure increased from barely $1,000,000 on the eve of World War II to $60,000,000 in 1970. Liberia's rapid rate of growth has been second only to that of Japan in recent years. The fact that Liberia's currency is based on the dollar has, however, placed it in an isolated economic zone. Liberia is not a member of any major African economic or free-trade bloc, although it is a member of a not-fully-operative free-trade grouping to which Sierra Leone, Guinea, and the Ivory Coast also belong. It is also not a member of any international trade grouping.

Iron ore production

Liberia is the leading producer of iron ore on the African continent and is one of the principal exporters of iron ore in the world. The registration of foreign shipping under a Liberian "flag of convenience" has established Liberia as the foremost nation in terms of registered shipping tonnage. Liberia nevertheless remains a primarily agricultural and underdeveloped country.

The country is, however, rich in natural resources, although the distribution of wealth after production is uneven. The coastal districts receive a far greater share of economic benefits than the hinterland, after which the county capitals are the next beneficiaries.

Mineral resources. Since 1951 Liberia has been a producer of rich iron ore. Reserves amounted to 20,000,000 tons of magnetite (magnetic iron ore) and hematite (the principal iron ore), containing 66 percent of pure iron, in the Bomi Hills; 100,000,000 tons of red and black itabirites (iron-bearing formations), with an iron content of 55 percent, 50 miles away from Monrovia at Mano River; between 200,000,000 and 250,000,000 tons of poorer ores, with an average iron content of 38 percent, in the Bong Range; and 300,000,000 tons of ore, with a content of from 60 to 70 percent of iron, in the Nimba Range in northeastern Liberia. Other reserves exist at Kitoma Mountain, south of the Nimba range; in the Wologizi range in northwest Liberia, and in the Putu range in eastern Liberia. Other minerals including diamonds, gold, cyanite (a silicate of aluminum, with thin bladelike crystals), and barite, while there are possible oil reserves off the coast.

Biological resources. About half of the land area is suitable for cultivation. It is estimated that 2,000,000 acres (800,000 hectares) at a time can be devoted to growing food crops, and a further 5,000,000 acres (2,000,000 hectares) of low land can produce either tree crops or food crops, while yet another 5,000,000 acres on rolling or hilly terrain is suited to growing tree crops. The climate favours rubber production. The rain forest type of vegetation is unsuitable for the raising of livestock but produces fine hardwood timber, especially in the east of the country, but also in the centre and in the west.

Agriculture, forestry, and fishing. Both rice and cassava, the principal food crops, are cultivated. More than 500,000 acres (200,000 hectares) are planted with these crops each year; after use the land is left to lie fallow for a decade or more, so that altogether more than 10,000,000 acres (4,000,000 hectares) are in periodic use. About 200,000 tons of rice are produced annually, but this is insufficient to meet the growing demand. The deficit is sometimes met by imports of rice from the United States and Egypt. Taiwanese rice experts are practicing experimental farming in order to increase production, and rice strains from the Philippines and Sri Lanka (Ceylon) have been introduced.

Rubber production

Rubber is the main cash crop. In 1926 the Firestone Tire and Rubber Company of the United States obtained the concession of up to 1,000,000 acres (400,000 hectares) for rubber cultivation. The concession consists of two estates—a large estate at Harbel in Montserrado County, 15 miles (24 kilometres) inland, and a small estate near the Ivory Coast frontier. More than 80,000 acres (30,000 hectares) are in production on these sites alone, apart from additional acreage elsewhere, which is either in use by other U.S. concerns, notably the B.F. Goodrich Company, or by the 4,000 Liberian rubber farmers, who occupy about 130,000 acres (50,000 hectares). Altogether over 250,000 acres (100,000 hectares) of rubber trees are under cultivation. Coffee, both of the robusta variety and of the local variety known as liberica, is also grown, as are cocoa, kola nuts, palm oil, palm kernels, sugar cane, peanuts (groundnuts), and cotton. Cattle, goats, pigs, and poultry are also raised. Although between 1,000,000 and 1,500,000 cubic feet (28,000 and 42,000 cubic metres) of timber are produced a year, exploitation of the forest resources is difficult, chiefly because of the scarcity of good roads and shortage of labour.

Kru and Fanti fishermen, the latter from Ghana, have traditionally been the suppliers of fish to coastal areas; two Liberian fishing companies now supplement these tribal fishing activities. Inland fish-breeding ponds have been introduced in order to provide a source of protein.

The Nimba Range deposits

Mining. To export the iron ores, iron interests have built special railroads. The Bomi Hills mine, which is American-operated, is linked to the port of Monrovia by a narrow-gauge railroad 43 miles (69 kilometres) long. A further 49-mile (79-kilometre) length of line connects the Mano River workings to the Bomi Hills. The Bong Range mine, which is operated by a German company, is linked to Monrovia by a 50-mile (80-kilometre) standard-gauge line. The Nimba Range reserves, which constitute one of the largest and richest iron ore deposits in the world, are operated by the Liberian American-Swedish Minerals Company (LAMCO) in conjunction with the Bethlehem Steel Corporation of the United States. The reserves have almost no overburden (other material covering the deposits), impurities are insignificant, and the ore can be used untreated. It is exported from the port of Buchanan, to which the ores are sent via a 168-mile (270-kilometre) standard-gauge railroad.

Manufacturing. The number of manufacturing enterprises has increased, particularly since 1960. Most of them serve the local market. Near the port of Monrovia there is a petroleum refinery as well as a cement plant. Other industries include an explosives factory, to supply the iron mines, and plants engaged in tire retreading, paint manufacturing, tuna processing, milk reconstituting, and manufacturing pharmaceuticals. Bricks, tiles, cement blocks, lumber and furniture, soap, and footwear are also manufactured, and there are several distilleries.

Energy. Power is provided by public and private installations, which are either hydroelectric or employ petroleum. Altogether they have a total capacity of over 250,000,000 kilowatts. The rubber-growing and iron-mining enterprises generate their own power. A hydroelectric plant at Mt. Coffee on the St. Paul River, which supplies power to Monrovia and to adjacent districts, has a 102,000-kilowatt capacity.

Financial services. Financial services are provided by eight banks, of which six are foreign. These are the Bank of Monrovia (an affiliate of the First National City Bank of New York), the International Trust Company of Liberia, the Chase Manhattan Bank, the Union National Bank (Liberia), Inc., the Bank of Liberia, Inc. (under Liberian management), the Commercial Bank of Liberia, the Liberian Trading and Development Bank Ltd. (Tradevco), and the government-owned Liberian Bank for Industrial Development and Investment, which provides medium- and long-term loans to private firms. There are also two credit corporations—the Liberian Agricultural Credit Corporation and the Liberian Development Corporation. Several Lebanese merchants provide short-term loans.

Foreign trade. Liberia enjoys a favourable balance of trade; its trade surplus reached a record amount of over $80,000,000 in 1969. Rubber and iron ore account for almost 90 percent of the value of all exports. Imports include machinery and transport equipment, food, beverages and tobacco, miscellaneous manufactured goods, mineral fuels, lubricants, and chemicals. The United States is the principal trading partner, followed by West Germany.

The Open Door Policy

Taxation. The tax structure is designed to attract foreign companies and capital to the country. Under the terms of the Open Door Policy, as operated since World War II, 80-year leases for tracts of undeveloped land may be obtained, and the flow of capital and dividends from the country is not restricted. Imports of machinery for industries are exempt from customs duty. A graduated income tax is applicable both to individuals and to corporations. Corporate incomes of more than $50,000 are taxed at a rate of 25 percent.

Trade unions. The two principal industrial unions are the Labor Congress of Liberia and the Congress of Industrial Organizations (CIO). Under the Labor Code, enacted in 1963, labour disputes are to be submitted to the government-established Labor Practices Review Board before a strike can be called.

Economic policies. The Open Door economic policy, which encourages foreign investment, has already been referred to. In addition, an Operation Production Policy aims at making Liberia self-sufficient in the production of staple food crops and at raising living standards in rural communities.

Problems and prospects. While Liberia has experienced little difficulty in obtaining international loans because of its abundant mineral resources, a substantial percentage of government revenue is allocated to reducing outstanding debt. An austerity program was proclaimed in the early 1960s in order to hasten the rate at which the national debt is liquidated. The reported location of further minerals resources, including the possibility of oil in the shallow waters off Monrovia, plus the existence of the further iron ore deposits at Mt. Kitoma, Wologizi Range, and Putu Range have led to the expectation that the economy will continue to expand.

Transportation. There is no national transportation network. Nearly all transport facilities are controlled by private enterprise; Liberian National Airways is an exception. There are about 4,200 miles of roads, of which about 325 miles are paved.

Four ports—Monrovia, Greenville, Harper, and Buchanan—serve Liberia's 350-mile coastline. The free port of Monrovia, the principal commercial port, with a draft of 35 feet (11 metres), can accommodate eight ships at a time. It also has facilities for transshipping iron ore and liquid latex. Nimba Range iron ore is shipped from Buchanan, while the port at Greenville and also a partially sheltered pier at Harper used by coastal craft are used primarily for the shipment of rubber and forest products. A fifth port, Robertsport, is planned for the export of iron ore.

Roberts International Airport is 27 miles from Monrovia, with which it is connected by road. It is used by 12 airlines and has a runway 9,000 feet (2,745 metresj long. In 1968 Pan American World Airways, in agreement with the Liberian government, took over the management of all airfields and is engaged in improving James Spriggs Payne Airfield at Monrovia to meet international standards.

A telecommunications network links Monrovia and 21 other towns by microwave relays for telephone and telegraph. There are automatic telephones in Monrovia, and several larger towns have manual exchanges.

Tourism is promoted by the government. There are several hotels and motels in Monrovia and its vicinity. Beaches provide popular recreation.

Administration and social conditions. Liberia has a republican form of government patterned after that of the United States and divided into legislative, executive, and judicial branches. The legislature is bicameral. The Senate has 18 members representing the nine counties, and the House of Representatives has 52 members representing the four territories and nine counties. Senators are erected for a six-year period and members of the House for four years. The president and vice president are elected for eight-year terms and may be re-elected for additional four-year terms. County superintendents, appointed by the president, manage county affairs, while district commissioners and tribal chiefs administer local affairs in the hinterland.

Political, cultural, and economic factors long separated the descendants of the original Liberian immigrants, also known as Arnerico-Liberians, or settlers, from the African tribal populations. In 1944 a national unification program was launched to speed the integration of the two groups, and in 1964 the old division between coastal and hinterland rule was abolished. In addition to the original five coastal counties of Grand Cape Mount, Montserrado, Grand Bassa, Sinoe, and Maryland, and the four territories (tribal enclaves) of Marshall, River Cess, Sastown, and Kru Coast, four new hinterland counties—Grand Gedeh, Nimba, Bong, and Lofa—were created. Tribal peoples are allowed, as far as possible, to govern themselves according to customary law.

National unification program

The constitution provides for a Supreme Court and for one chief justice assisted by four associate justices, all of whom are appointed by the president and who serve during good behaviour (*i.e.*, so long as they uphold the constitution to the satisfaction of the president). Subordinate courts established by legislative enactment serve the nine counties.

All citizens are entitled to vote. Since the creation of the four inland counties, tribal members have begun to participate in politics. Liberian politics, which, during the early years, took the form of a two-party system representing a conflict of interest between mulatto and black, has since 1878 been controlled by the True Whig Party, which is black-dominated and characterized by strong executive control. The Reformation Party (formed in 1951) and the Independent True Whig Party (formed in 1954) were both banned by the National Legislature in 1955.

Defense. The Liberian National Guard, numbering 3,200, is maintained at various strategic points. The militia, a voluntary body of officers and men numbering about 10,000, constitutes the reserve force. There is a small coast guard fleet of patrol cutters. A constabulary force of about 1,000 men and women comprises the national police force. These services are administered and controlled by the departments of justice and national defense.

Education. Formerly supplied mostly by Christian missions, educational services were greatly expanded by the government after 1945. The budget for national education was about $6,000,000 in 1970. Total enrollment at all levels was 150,000 by 1970. Since 1939 education has been legally compulsory for children between the ages of 6 and 16. By 1970, however, only about three out of ten children of school age were receiving education. Education is usually free at the primary level, although in the interior, but not in Monrovia, books must be bought. There are three vocational schools—the Booker Washington Agricultural and Industrial Institute at Kakata, which is a government school, and two more that are concession-operated. Advanced training is provided at the University of Liberia at Monrovia, at Cuttington College and Divinity School at Suakoko (Episcopalian), and at Maryland College of Our Lady of Fatima at Harper (Roman Catholic). Tubman University, was preparing to open in Harper. A medical school, the Monrovia Torrino Medical College, trains paramedical students. Liberian students abroad, amounting to nearly 500 in any given year, receive advanced training under a government foreign scholarship program. Grants from the United Nations enable Liberia to attract teachers from Europe and the United States to assist in the teaching of economics, science, and mathematics, in addition to which the government also sponsors the employment of foreign teachers. The United Nations supported the government's national literacy program, and ten of the major tribal languages have now been written down. Middle-level education has been adapted to the needs of the economy. Six years of primary school education are followed by three of middle school and three of high school education. In 1968 Liberia became an associate member of the West African Examinations Council so as to provide an international yardstick for measuring the quality of its education.

Vocational schools, colleges, and universities

Health and welfare. Conditions remain poor, although much progress has been made in providing better health facilities. The incidence of malaria is close to 100 percent, and that of leprosy 2 percent. Yaws is no longer a serious problem, following a control program carried out by the World Health Organization, but tuberculosis has been difficult to eradicate. Dysentery, malaria, and diarrhea continue to be the most important causes of infant mortality. Influenza, hernias, and intestinal worms, as well as trypanosomiasis (sleeping sickness), schistosomiasis (a parasitic disease), and elephantiasis (a skin disease in which the skin resembles an elephant's hide) are also prevalent. Most modem homes are screened against malaria mosquitoes, tsetse flies, and other dangerous insects. The government conducts inoculation campaigns to combat smallpox and yellow fever, and mobile units employ insecticides, usually DDT (dichloro-diphenyltrichloroethane), against malaria-carrying mosquitoes and other insects.

Major diseases

The government, foreign Christian missions, and concessions together sponsor 31 hospitals, of which the most modem are the 250-bed John F. Kennedy Memorial Hospital in Monrovia, sponsored by the U.S. Agency for International Development, the Roman Catholic Hospital at Monrovia, and Phebe Hospital in Bong County, all of which have nursing and medical schools attached to them. There are 143 clinics, some specializing in the problems of the blind and the aged, and others specializing in tuberculosis, sleeping sickness, and leprosy.

The Department of Public Works supervises low-cost housing projects and is also responsible for city planning and zoning. Housing is expensive, and most building equipment is imported. On the outskirts of Monrovia, tribal communities exist side by side with modern communities. Most of the people in these districts are squatters who are liable to be vacated from empty town lots by the authorities.

Living conditions. Living conditions throughout Liberia, while poor, are improving. Increasingly the hinterland farmers are growing such cash crops as rubber, coffee, and palm produce, as well as subsistence crops. The money economy is consequently gradually expanding throughout the interior. Some 80,000 men and women are employed on the various plantations and in iron-mining enterprises. There is a standard 48-hour working week in industry and a 40-hour week in most government departments. A wage law establishes the minimum unskilled industrial wage at 15 cents an hour and the minimum unskilled agricultural wage at 8 cents an hour. Many of the concessions provide family housing, as well as free medical and educational facilities, and provide partially subsidized rice, fish, and palm oil to their employees.

Social conditions. Intermarriage and economic progress have both been important factors in breaking down social divisions in Liberia. Mines and plantations are distributed throughout the country and have contributed to the slow trend toward a more equitable distribution of income. Particularly in the coastal districts, government jobs, foreign businesses, and local markets also provide greater opportunities for economic and social advancement. Economic and social divisions are more keenly felt betweeen the coastal and rural areas because the cash economy has only recently spread inland. Transportation, communications, and commerce, however, all serve slowly to lessen differences.

Cultural life and institutions. In the dance halls of Monrovia and in towns elsewhere, both Western and African music and dancing styles are in vogue. In rural areas, African musicians and drummers use traditional rhythms at weddings, burials, and feasts, as well as at the graduation ceremonies of boys and girls from the Poro and Sande schools.

Musical festivals and dramas are staged several times a year at Cuttington College and the University of Liberia, which are also visited by foreign drama and ballet troupes.

Several institutions instruct young men and women in the legends, traditions, songs, arts, and crafts of African culture. The government encourages the preservation and promotion of African culture through such agencies as the National Cultural Center, which exhibits house styles representing Liberia's 16 different tribal cultures. The sculpture of masks, representing both humans and animals, is an artistic pursuit that is also related to the social structure of some tribes.

There are two daily newspapers and one biweekly paper, ***The Liberian Age,*** which is government-owned. There are several radio stations, including the government-owned ELBC station, ELWA (a religious broadcasting station for Africa), and a Voice of America station. There are also some concession-operated newspapers and radio stations. There is also a television station (ELTV), although relatively few receiving sets are as yet available. In addition to western television programs, locally produced programs are shown for several hours each week.

Prospects. The Open Door Policy has been the capstone of the nation's economic growth since the 1940s. No other national policy has, since the foundation of the Liberian state, exerted such an influence upon the economy as well as upon the course of political, cultural, and social development. As a result, life in Liberia has undergone a transformation in less than three decades. Unless the unpredictable should occur, the prospect—particularly since further mineral discoveries have been made—is for continued economic growth. With the passage of time and the increased pace of modernization, the old division between the coastal regions and the interior may be expected slowly to be eroded.

BIBLIOGRAPHY. R.E. ANDERSON, ***Liberia: America's African Friend*** (1952); G.W. BROWN, ***The Economic History of Liberia*** (1941); R.W. CLOWER ***et al., Growth Without Development: An Economic Survey of Liberia*** (1966); R.J.H. CHURCH, ***West Africa,*** 6th ed. (1968); M. FRANKEL, ***Tribe and Class in Monrovia*** (1964); C.H. HUBERICH, ***The Political and Legislative History of Liberia,*** 2 vol. (1947); SIR HARRY JOHNSTON, ***Liberia,*** 2 vol. (1906); J.G. LIEBENOW, ***Liberia: The Evolution of Privilege*** (1969); H.R. LYNCH, ***Edward W. Blyden: Pan Negro Patriot, 1832–1912*** (1967); I. MARINELLI, ***The New Liberia*** (1964); P.J. STAUDENRAUS, ***The African Colonization Movement, 1816–1865*** (1961); UNITED STATES DEPARTMENT OF STATE, ***Handbook on Liberia*** (1968). See also the annual reports of the various departments of the Liberian Government.

(A.B.J.)

Library

The word library, which in English refers to a collection of books gathered for purposes of reading, study, or reference (and which derives from the Latin *liber*, "a book"), has a long history, occurring in a prose translation of the Roman philosopher Boethius' ***Consolation of Philosophy*** that Geoffrey Chaucer made in about 1374. The word *librairie* in French (and its counterpart in other Romance languages) does not have the same meaning, being used to denote a bookshop or, by extension, a publisher; the word used in many other countries to signify a collection of books, public or private, is derived from a Latinized Greek word, ***bibliotheca:*** hence *bibliothèque* in French, ***biblioteca*** in Italian and Spanish, *Bibliothek* in German, ***biblioteka*** in Russian. In Japanese the word is *tosho-shitsu.* The use of the word library to denote a building, room, or set of rooms in which a collection of books is contained also has a lengthy history, which goes back to the 15th century.

The basic function of all libraries is to preserve material by collection, but they have come to serve many other purposes as well. The material they stock now often includes films, slides, phonograph records, and tapes, as well as books and manuscripts. Thus, by housing readily available information and the means of its convenient transmission to students and others and by providing a source of instructive and creative reading, viewing, and listening for the general public of all ages, libraries play a major role in modern social organization. The following article deals with the development of libraries from earliest times and describes the functions of different kinds of libraries; it outlines some of the principles of planning involved in library architecture and concludes with an explanation of some problems faced by libraries in a technological age. The detailed operations of libraries and library procedures are dealt with in the article LIBRARY SCIENCE.

HISTORICAL PERSPECTIVES

The ancient world. In earliest times there was no distinction between a record room (or archive) and a library, and in this sense libraries can be said to have existed for almost as long as records have been kept. A temple in the Babylonian town of Nippur, dating from the first half of the 3rd millennium BC, was found to have a number of rooms filled with clay tablets, suggesting a well-stocked archive or library. Similar collections of Assyrian clay tablets of the 2nd millennium BC were found at Tell el-Amarna in Egypt. Ashurbanipal (668–627 BC), the last of the great kings of Assyria, maintained an archive of some 25,000 tablets, comprising transcripts and texts systematically collected from temples throughout his kingdom. The majority carry his name, and some state that the king had "edited" them and caused them to be collected in his palace at Nineveh.

Ashurbanipal's library

Greece and Alexandria. The idea of book collecting, however, and hence of libraries as they are now understood had its origin in the classical world. Most of the larger Greek temples seem to have possessed libraries, even in quite early times; many certainly had archive repositories. The tragic playwright Euripides was known as a private collector of books, but the first important institutional libraries came in Athens during the 4th century BC with the great schools of philosophy. The Stoics, having no property, had no library; the schools of Plato and of the Epicureans did possess libraries, whose influence lasted for many centuries. But the most famous collection was that of the Peripatetic school, founded by Aristotle and systematically organized by him with the intention of facilitating scientific research. Aristotle's library went to his successor and heir, Theophrastus. After that it was owned by Apellicon of Teos, from whom it was claimed as war booty by Sulla, when he conquered Athens in 84 BC, and taken back by him to Rome (where the orator Cicero, one of Rome's great book lovers, enjoyed its use).

Aristotle's library is said by Strabo, a traveller and historian of the late 1st century BC, to have formed the model for that established at Alexandria, which became the greatest in antiquity. It was planned by Ptolemy I Soter (died 283 BC) and brought into being by his son Ptolemy II Philadelphus (308–246 BC) with the collaboration of Demetrius of Phalerum, their advisor. The founders of this library apparently aimed to collect the whole body of Greek literature in the best available copies, arranged in systematic order so as to form the basis of published commentaries. Its collections of papyrus and vellum scrolls are said to have numbered hundreds of thousands. Situated in a temple of the Muses called the Mouseion, it was staffed by many famous Greek writers and scholars, who included the grammarian and poet Callimachus (died 240 BC), Eratosthenes (died c. 194 BC), the philosopher Aristophanes of Byzantium (died c. 180 BC), and Aristarchus of Samothrace (died 145 BC), the foremost critical scholar of antiquity.

Pergamum. In Asia Minor, a library rivalling that of Alexandria was set up at Pergamum during the reigns of Attalus I (died 197 BC) and Eumenes II (died 159 ac). Parchment (charta pergamena) was said to have been developed there in order to counteract the decision of Ptolemy Philadelphus to forbid the export of papyrus from Egypt, thereby impeding the copying of books for its library. The library was bequeathed with the whole of the kingdom of Pergamum to the Roman people in 133 BC, and Plutarch records a rumour that Antony gave its 200,000 volumes to Cleopatra, to become part of the Alexandrian library.

Rome. There were many private libraries in classical Rome, including that of Cicero. Indeed, it became highly fashionable to own a library, judging from the strictures of the moralizing statesman Seneca and the spiteful jibes by the poet Lucian on the uncultured "book clown." Excavations at both Rome and Herculaneum have revealed what were undoubtedly library rooms in private houses, one at Herculaneum being fitted with bookcases around the walls. A Roman statesman and general, Lucullus, who became known as the richest man in the ancient world, famous for his luxurious way of life, acquired as part of his war booty an enormous library, which he generously put at the disposal of those who were interested. His biographer, Plutarch, speaks appreciatively of the quality of his book collection, and Cicero tells of an episode when he visited the library to borrow a book to find his friend, Cato, already ensconced there surrounded by books of the Stoic philosophy.

The library planned by Julius Caesar

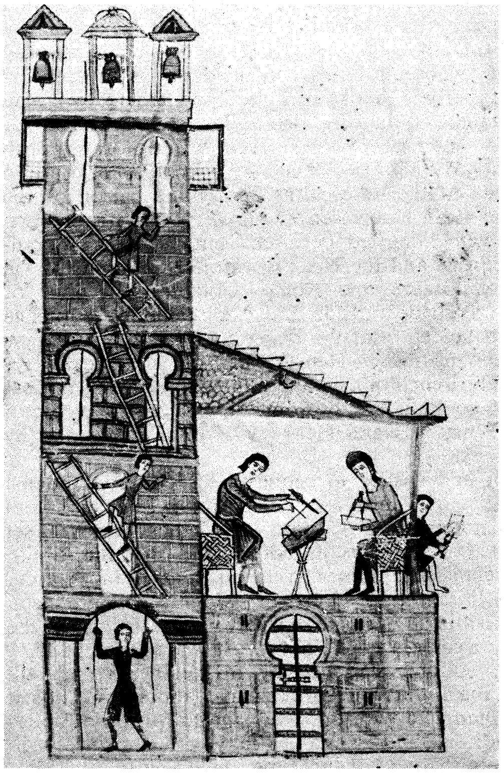
Julius Caesar planned the creation of a public library and entrusted the implementation of his plans to an outstanding scholar and writer, Marcus Terentius Varro (died 27 BC), also the author of a treatise on libraries, De bibliothecis (which has not survived). Caesar died before his plans were carried out, but a public library was built within five years by the literary patron Asinius Pollio (died AD 4). Describing its foundation in his Natural History, Pliny coins a striking phrase which has application to libraries generally: *ingenia* hominum *rem publicam* fecit ("He made men's talents a public possession"). Libraries were also set up by Tiberius, Vespasian, Trajan, and many of the later emperors; the Bibliotheca Ulpia, which was established by Trajan in about AD 100 and continued until the 5th century, was also the Public Record Office of Rome.

The Imperial library at Byzantium, set up by Constantine the Great in the 4th century, was an important centre for Christian literature; but it also played a part in the preservation of the literature of Greece and Rome. In the 5th century, when the Roman Empire was being destroyed by the onslaught of Germanic barbarian incursions, libraries both public and private went up in flames or were left to decay. Fragments of the cultural heritage of Greece and Rome were, however, preserved in churches and monasteries, from which later blossomed the culture of the Renaissance.

Destruction of ancient libraries

Middle Ages and Renaissance. Role of the monasteries. As monastic communities were set up (from as early as the 2nd century AD) books were found to be essential to the spiritual life. The rule laid down for observance by several monastic orders enjoined the use of books: that of the Benedictine Order, especially, recognized the importance of reading and study, making mention of a "library" and its use under the supervision of a precentor, one of whose duties was to issue the books and make an annual check of them. Scriptoria, the places where manuscripts were copied out, were a common feature of the monasteries—again, especially in those of the Benedictine Order, where there was a strict obligation to preserve manuscripts by copying them. Many—Monte Cassino (529) and Bobbio (614) in Italy; Luxeuil (c. 550) in France; Reichenau (724), Fulda (744), and Corvey (822) in Germany; Canterbury (597), Wearmouth (674), and Jarrow (681) in England—became famous for the production of copies. Rules were laid down for the use of books, and curses invoked against any person who made off with them. Books were, however, lent to other monasteries and even to the secular public against security. In this sense, the monasteries to some extent performed the function of public libraries.

BY courtesy of the Pierpont Morgan Library, New York



"The Scriptorium in the Bell Tower of San Salvador at Tavara," from the Beatus of Liebana's Commentary on the Apocalypse (M. 429 folio 183), 1220. In the Pierpont Morgan Library, New York.

Contents of monastic libraries

The contents of these monastic libraries consisted chiefly of the Scriptures, the writings of the early Church Fathers and commentaries on them, chronicles, histories such as Bede's Ecclesiastical History of the English Peoples, the philosophical writings of St. Thomas Aquinas and Roger Bacon, and possibly some secular literature represented by the Roman poets Virgil and Horace and the orator Cicero. After the universities were founded in the 13th century, monkish students, on returning to their monasteries, deposited in the libraries there the lecture notes they had made on Aristotle and Plato, on law and medicine, and so forth, and in this way expanded their contents.

The new learning. The libraries of the newly founded universities—along with those of the monasteries—were the main centres for the study of books until the late Middle Ages; books were expensive and beyond the means of all but a few wealthy people. The 13th, 14th, and 15th centuries, however, saw the development of private book collections. Philip the Good, duke of Burgundy, Louis IX, and Charles V (who may be looked upon as the founder of what is now called the Bibliothèque Nationale in Paris) were great collectors, as were also such princes of the church as Richard de Bury, bishop of Durham (died 1345), who wrote a famous book in praise of books, *Philobiblon* (first printed in Cologne, 1473). But new cultural factors—including the growth of commerce, the new learning of the Renaissance (that was based on newly discovered classical texts), the invention of printing, and a substantial expansion of lay literacy—widened the circle of book collectors to include wealthy merchants whose libraries contained books of law and medicine, herbals, books of hours, and other devotional works. Italian humanists, such as Petrarch, searched for and copied manuscripts of classical writings to establish their scholarly libraries. Niccolo Niccoli (librarian to Cosimo de Medici, a wealthy Florentine patron of the arts), and Gian Francesco Poggio Bracciolini were two scholars who inherited Petrarch's enthusiasms and ransacked Europe and the Near East for manuscripts of the writers of Greece and Rome. Notable collections of books were made outside Italy, too (though Florence remained the centre of the rising book trade): by Diane de Poitiers, mistress of Henry II of France; by Jean Grolier, a high French official and diplomat, who was a great patron of bookbinders; by John Tiptoft, earl of Worcester; by Henry VII and Henry VIII of England and by many others.

Renaissance public libraries

Petrarch had wished to bequeath his collection to the municipality of Venice as a public library, but his intention was not realized. Cosimo de' Medici, however, set up, on the basis of Niccolo Niccoli's library, the Biblioteca Marciana in Florence, in the convent of San Marco. The rich library of Lorenzo the Magnificent, grandson of Cosimo and an even greater patron of learning and the arts, also became a public library. It was opened in 1571 in a fine building designed by Michelangelo and still exists as the Biblioteca Laurenziana (though in 1808 it was amalgamated with the Marciana to form the Biblioteca Medicea-Laurenziana). Many other princely libraries were formed at this time, including that of Matthias I Corvinus of Hungary. Like the Medicis, Corvinus employed agents to purchase manuscripts in the Levant, and he was also a customer of a well-known Florentine dealer in manuscripts, Vespasiano da Bisticci. Corvinus' library was destroyed when the Turks overran his capital of Buda in 1526 (a few books that survived are in the Austrian National Library at Vienna). The library of the Escorial in Madrid, founded in 1557, was based on the collections of Philip II. The Vatican library also dates its foundation from this time, the real founder of the collections as they are today having been Pope Nicholas V (reigned 1447–55), who dispatched agents to Germany, England, and Greece in search of manuscripts; Sixtus IV (reigned 1471–84) reassembled the Vatican collection in its present location.

Effects of the Reformation and religious wars. In England the end of the monastic libraries came in 1536–40 when the religious houses were suppressed by Henry VIII and their treasures dispersed. No organized steps were taken to preserve their libraries: it seems likely that a few—though only a very small proportion of the whole—were taken for the Royal Library. Even more wholesale destruction came in 1550: Henry VIII and Edward VI aligned with the "new learning" of the humanists; and university, church, and school libraries were purged of books embodying the "old learning" of the Middle Ages. The losses were incalculable. A change of attitude came about with Elizabeth's reign when those in authority realized that the kind of books that had been dispersed would form useful propaganda for the government's policy; the archbishop of Canterbury, Matthew Parker, and the Queen's principal secretary, Robert Cecil, took the lead in seeking out and acquiring the scattered manuscripts. Many other collectors were also active, including Sir Robert Bruce Cotton and Sir Thomas Bodley. As a result, some considerable portion of the libraries that had been scattered at the suppression was, by 1660, reassembled in collections—Parker's eventually went to Corpus Christi College at Cambridge; Cotton's to the British Museum; and the Bodley's to form the Bodleian Library at Oxford—where they remain to this day.

Sistine Hall of the Vatican Library, built by Domenico Fontana. 1588.
Ewing Galloway

Elsewhere in Europe, the period of the Reformation also saw many of the contents of monastic libraries destroyed, especially in Germany and the northern countries. The Reformation leader Martin Luther, however, did himself passionately believe in the value of libraries, and in a letter of 1524 to all German towns he insisted that neither pains nor money should be spared in setting up libraries, or book houses, particularly in the larger towns. As a consequence, many town libraries in Germany, including those at Hamburg (1529) and Augsburg (1537), date from this time. These, and the libraries of the newly created universities (such as Konigsberg, Jena, and Marburg), were partly, at any rate, built up on the basis of the old monastic collections. In Denmark, similarly, some books from the churches and monasteries were incorporated with the new university library, though many were destroyed.

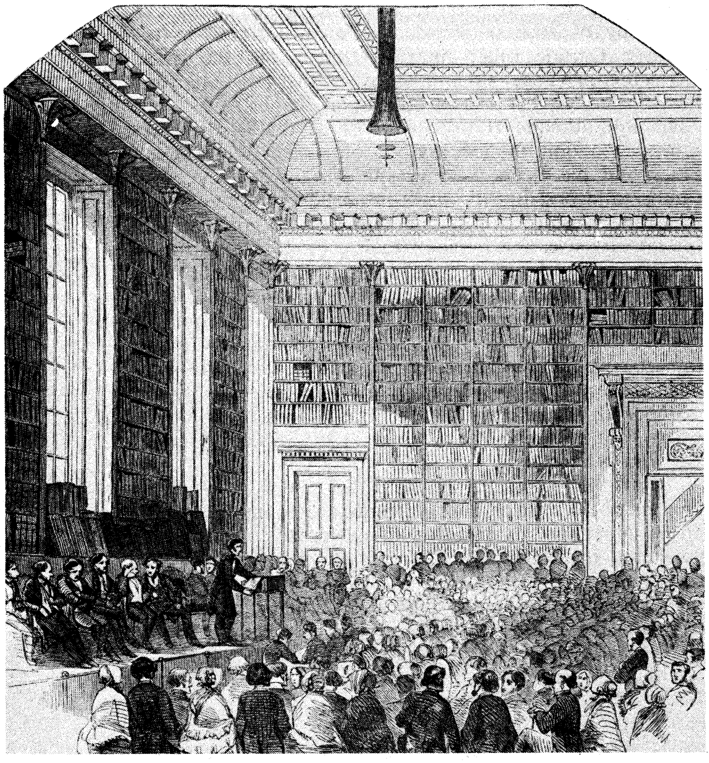
Libraries in Germany suffered severely in the Thirty Years' War. The Bibliotheca Palatina at the University of Heidelberg, for example, which had been founded in 1386, was taken as the spoil of war by Maximilian I of Bavaria, who offered it to Pope Gregory XV in 1623; and Gustavus Adolphus sent whole libraries to Sweden, most of them to swell the library of the University of Uppsala, which he had founded in 1620. The collections of the Royal Library in Stockholm were similarly enriched by the war booty that fell to Sweden during the reigns of Queen Christina and Charles X.

In France, Italy, Southern Germany, and Austria, where the Catholic faith remained unshaken, the old libraries remained and were supplemented by new ones set up, for educational purposes, by the Society of Jesus (the Jesuits).

17th and 18th centuries: the great national libraries. A notable feature of the 17th and 18th centuries was a widespread enthusiasm for book collecting, sometimes out of ostentation, but often the result of genuine scholarship. Several fine private collections were assembled, many of which were eventually to become the core of today's great national and state libraries—for this was also the period that saw new national and university collections springing up all over Europe. There were, of course, other developments. In England there was established a number of parish libraries, attached to churches and chiefly intended for the use of the clergy (one of the earliest, at Grantham in Lincolnshire, was set up as early as 1598, and some of its original chained books are still to be seen there). They were sometimes the result of lay donation: a Manchester merchant, Humphrey Chetham, left money in 1653 for the foundation of parish libraries in Bolton and Manchester and also for the establishment of a town library in Manchester (which still exists, housed in its original bookcases, in its original building). Later, in the 18th century, especially in England and America (though also elsewhere in Europe), there was a great vogue for the circulating and "subscription" libraries—book clubs that provided an agreeable reference service and lending collections for their members and had much influence on the formation of popular literary taste, especially in fiction.

Parish, town, and subscription libraries

@ The Illustrated London News



"Opening of the Manchester Free Library," engraving by an unknown artist, from The *Illustrated* London News, 1852.

Library planning. 'The private libraries of powerful and influential collectors such as Cardinal Mazarin in France were so large that a new approach to library organization was needed. The Escorial library in Madrid, erected in 1584, had been the first to do away with book bays on the medieval pattern and to arrange its collection in cases lining the walls. The old practice of chaining books to their cases was gradually abandoned; and the change to the present arrangement of standing them, spines facing outward, began in France, probably with the library of the historian Jacques-Auguste de Thou (died 1617). Mazarin's library was in the charge of Gabriel Naudé, who produced the first modern treatise on library economy, ***Advis pour dresser une*** *Bibliothèque* (1627). This work had great influence on the collectors and scholars of the Restoration period: widely translated, it marked the transition to the age of modern library practice. One of its first fruits, the library of the diarist Samuel Pepys, organized with meticulous care and beautifully housed, is still at Magdalene College, Cambridge. Naudé's concept of a scholarly library, systematically arranged, displaying the whole-of recorded knowledge and open to all scholars without fear or favour, took root. It was above all absorbed by the philosopher Gottfried Wilhelm Leibniz (died 1716), the greatest librarian of his age, who conceived the idea of a national bibliographical organization that would provide the scholar with easy access to all that had been written on his subject. He realized that scientific progress above all depends on ease of communication between scholars and that the purpose of great libraries and museums is to provide and keep open the channels of communication.

Emergence of national collections. The boundaries of knowledge were expanding rapidly during the 17th and 18th centuries, especially in the field of historical studies and philosophy, resulting in a vast outpouring of books that went to swell the libraries of private collectors. The lawyer, historian, and councillor of state Jacques-Auguste de Thou was in charge of the Bibliothkque du Roi (founded by Charles V and largely reorganized during the 15th century by Louis XII). Mazarin's library was scattered when he was compelled to leave France during the movement known as the Fronde, but it was reassembled when he returned to power in 1661. Rehoused in a new building, it was opened to the public in 1691 and remained one of France's great libraries until after the French Revolution when it was incorporated with other collections (including the Bibliothkque du Roi) to form the Bibliothkque Nationale, today one of the world's great libraries. August, duke of Brunswick, established a library in 1604 that later became the Herzog August Bibliothek at Wolfenbiittel, one of the finest libraries in Europe (Leibniz was its librarian from 1690–1716). A library assembled by the elector Friedrich Wilhelm of Brandenburg was founded in 1659 and later became the Prussian State Library. The collections of the English book collectors, Sir Hans Sloane, Sir Robert Cotton, and Edward and Robert Harley, earls of Oxford, formed the basis of the British Museum collection, founded in 1753, which from the start was regarded as the national library; it was further enlarged by the addition in 1757 of the Royal Library, containing books collected by the kings of England from Edward IV to George II.

The great national collections

The effects of the French Revolution. The anti-clerical movement on the continent of Europe that found expression in revolution sealed the fate of many monastic and church libraries: those in France, for example, were expropriated in 1789; in Germany in 1803; in Spain in 1835. The dispersal of books was, however, better organized than that of the English monasteries under Henry VIII. In France they were collected in the main towns of the *départements* in what were called *dépots littéraires.* In 1792 the same fate befell the collections of aristocratic families, and these, too, were added to the *dépots.* The enormous accumulations caused problems, and many books were lost, but the plan of coordinating library resources throughout the country was carried out. The Bibliothèque Nationale received some 300,000 volumes; and new libraries were set up in many important provincial cities. Napoleon's successful campaigns in Europe also brought acquisitions, in the form of war booty, to the Bibliothkque Nationale, but these turned out to be only temporary; most of the books were returned after the treaty of Vienna in 1815. In Germany the Bavarian State Library was greatly enriched by the contents of over 150 confiscated libraries, and many of the provincial libraries were similarly enlarged. In Austria, as a result of confiscations, Studienbibliotheken (study libraries) were set up at Linz, Klagenfurt, and Salzburg, the university libraries at Graz and Innsbruck were substantially enlarged, and many valuable acquisitions accrued to the Hofbibliothek in Vienna.

19th- and 20th-century developments. The 19th century saw a significant awakening. Libraries had grown in size, but their growth had been haphazard; administration had become weak, standards of service almost nonexistent; funds for acquisition were inadequate, the post of librarian was looked on as a part-time one, cataloging was in arrears and lacked proper method.

The university library at Gottingen was a notable exception. Johann Gesner, the first librarian, working in close association with the curator of the University, G.A. von Miinchhausen, and proceeding on the principles laid down by Leibniz, made strenuous efforts to cover all departments of learning; the library provided good catalogs of carefully selected literature and was available to all as liberally as possible. The next director, C.G. Heyne, enthusiastically followed the same principles, with the result that Gottingen became the best organized library in the world.

The revolution in library administration

By the middle of the century new conceptions of the purpose and scope of learned libraries were everywhere taking shape. Within a few years the picture of library service had been transformed, and a leading figure in this development was Antonio (afterward Sir Anthony) Panizzi, a political refugee from Italy who began working for the British Museum in 1831 and was its principal librarian in 1856–66. From the start he revolutionized library administration, demonstrating that the books in a library should match its declared objectives and showing what these objectives should be in the case of a great national library. He perceived the importance of a good catalog and to this end elaborated a complete code of rules for catalogers. He also saw the potentiality of libraries in a modern community as instruments of study and research, available to all, and showed how—by his planning of the British Museum reading room and its accompanying bookstacks—this potential might be realized. His ideas have dominated library thought in the field of scholarly—or, as they are now called, research—libraries up to the present day, and may have achieved their major expression so far in the Library of Congress in Washington, D.C.

By the middle of the 19th century the idea had been accepted that libraries might be provided by local authorities at public expense. This proved a significant stage in the development of library provision. Panizzi had stated that he wanted the facilities of a great library to be available to poor students so that they could indulge their "learned curiosity"; now, in 1840, an Act of Parliament was passed in England enabling local councils to levy a rate for the provision of free library facilities. From the first tentative beginnings there has been continuous growth in the providing of reading services and in the use that is made of them by the public. Libraries now provide not only books and periodicals but also many audio-visual materials; in some cases also they provide facilities for lectures and dramatic performances and art exhibitions and are able even to lend works of art for use at home. Such developments in the public libraries are commonest in the United Kingdom, the United States, and the Scandinavian countries. Elsewhere, the large town libraries in Europe have always been regarded as public libraries; and, since the 18th century, local libraries and reading groups have provided, to some extent, popular instructional reading. But so far there has been no comparable vitality and spirit of social purpose.

Library services in the age of specialization

The 20th century has added a new dimension to library services, as well as a new range. Developments in scientific and industrial research, with a consequent vast increase in the publication, on a worldwide basis, of specialized information—mostly in the form of periodicals—have led to a demand for rapid and easy access to wide ranges of periodical literature and an information service that could furnish references and bibliographies on specific subjects. In the applied sciences especially, books are no longer regarded as the main repositories of knowledge; existing research libraries, with their large reserves of books but comparatively deficient stock of modern scientific periodicals, with their traditional practices and administrative routines, were, generally speaking, unable to satisfy the new demands and in some quarters were looked upon as backward, out of touch with contemporary needs. An important result has been the growth of special libraries, often connected with commercial enterprises or specialized professional bodies and financed by them. These libraries have had a significant influence on the kind of service now offered in research libraries and also on the public libraries.

KINDS OF LIBRARIES

Library services available throughout the world vary so much in detail from country to country that it is impossible to present anything but the most general picture of their activities. Nevertheless, they follow a broad but discernible pattern that has evolved over the years.

Aims of a national library

National libraries. In most countries there is a national or state library or group of libraries maintained by national resources, usually bearing responsibility for publishing a national bibliography and for maintaining a national bibliographical information centre. National libraries strive principally to collect and to preserve the nation's literature, though they try to be as international in the range of their collections as possible.

Most national libraries receive, by legal right, one free copy of each book and periodical printed in the country. Certain other libraries throughout the world share this privilege, though many of them receive their legal (or copyright) deposit only upon request.

The Bibliothèque Nationale in Paris, the British Museum in London, the Library of Congress in Washington, D.C., and the Lenin Library in Moscow are the most famous and possibly the most important national libraries in the Western world. Their importance springs from the quality, size, and range of their collections, which are comprehensive in scope and attempt to maintain their comprehensiveness. This latter they achieve with diminishing success in view of the vastly increased number of publications that daily appear throughout the world and the difficulty of ensuring adequate representation of publications issued in the developing countries.

Bibliothèque Nationale. The Bibliothkque Nationale was before the Revolution known as the Bibliothkque du Roi and owes its origin (as is indicated above) to Charles V. It was the recipient during the 15th and 16th centuries of a number of important collections of manuscripts; in 1617, under the librarianship of the great collector J.A. de Thou, its right to legal deposit was reaffirmed and continued to be rigidly enforced. In the first quarter of the 18th century four of the library's departments (of Prints, Coins, Printed Books, and Manuscripts) were created; it was opened to the public in 1735. Enormous additions accrued to the library as a result of the French Revolution and the confiscation of aristocratic and church private collections. The catalog of the library on cards was completed under the librarianship (1874–1905) of Léopold Delisle, and in 1897 he made a start to the task of compiling a printed catalog in volume form. This is still incomplete (209 volumes had appeared by 1971). The library has 6,500,000 volumes of printed books; 180,000 volumes of manuscripts; and over 5,000,000 prints.

Unique character of the British Museum

The British Museum Library. The British Museum is unique in that it combines a great museum of antiquities with a great comprehensive library. It was founded in 1753 by the acceptance of the bequest of the collections of Sir Hans Sloane, physician to King George II and president of the Royal Society. The library was built up on the basis of two other important collections, that of Sir Robert Cotton and that of Edward and Robert Harley, earls of Oxford; to these were added the Royal Library, given by George II in 1757. With this collection came also the right to legal deposit of one copy of every book published in the British Isles; this right is most rigorously enforced. These four basic collections were notably enlarged during the first century of the library's history by the addition of many private collections, including the libraries of King George III (1823) and of Sir Thomas Grenville (1846). Sir Anthony Panizzi, as already indicated, reorganized the library and gave it its present modern shape; he was also responsible for its printed catalog, made between 1880 and 1905; a new edition was published between 1961 and 1966 in 263 volumes. Like the Bibliothkque Nationale, the British Museum Library is exceptionally rich in early printed books and manuscripts, of which there are a number of catalogs. It has continued to receive significant collections of valuable manuscripts and printed material; it has very fine collections of music and maps. The collections amount to 8,500,000 volumes of printed books; there are over 150,000 manuscripts. Prints and drawings, of which there are very large collections, are maintained in a separate department of the museum, as are the coins and medals.

Library of Congress. The United States' Library of Congress, located in Washington, D.C., is the largest of the national libraries and its collection of modern books is particularly extensive. It was founded in 1800 but lost many of its books by fire during a bombardment of the Capitol in 1814. These losses were to some extent made good by the purchase of Thomas Jefferson's library shortly thereafter. The library remained a strictly congressional library for many years, but as the collections were notably enlarged by purchases and by additions under the copyright acts, the library became and remained—in effect, although not in law—the national library of the United States. Library of Congress catalog cards were first printed for the use of subscribing libraries in 1902 (in the first year there were 212 subscribers; they are now used by many thousands of American libraries and institutions), a service initiated by Herbert Putnam, librarian from 1899 to 1939. In 1927 the library began printing the *National Union Catalog,* representing the stock of some 2,500 libraries, which has now reached the formidable size of 18,000,000 entries and is currently being printed in volume form. The library has collections totalling over 15,000,000 volumes and pamphlets, 29,000,000 pieces of manuscript, 176,000 prints and drawings, and 3,000,000 photographic negatives and slides. Further services to scholarship include the publication of several specialist catalogs.

National Union Catalog

Lenin Library. The State V.I. Lenin Library in Moscow, the national library of the Soviet Union, is to all intents a recent creation, founded after the Revolution of 1917 as a national library to replace the old Imperial Library in Leningrad. It is already one of the largest libraries in the world. It receives copies of all new publications issued in the Soviet Union, and its collections are given as 25,000,000 volumes of printed books and 2,500,000 manuscripts.

Other national collections. Besides these four great national libraries, there are many others with important collections and very long histories. The Bibliothkque Royale Albert I, in Brussels, founded in 1837 on the basis of the library of the dukes of Burgundy, is the national library of Belgium and the centre of the country's library network; it maintains a regular lending service with the university libraries and with the large town library of Antwerp. The Dutch Royal Library in The Hague was founded in 1798 and it, too, is the centre of a well-developed interlibrary loan system. Italy, because of its historical development as a series of city-states, has a number of national libraries, the chief being the Biblioteca Nazionale Centrale Vittorio Emanuele II in Rome, founded in 1875 and possessing more than 2,000,000 volumes, and the historically richer Biblioteca Nazionale Centrale at Florence, founded in 1747, with over 4,000,000 volumes. Other Italian national libraries are at Milan, Naples, Palermo, Turin, and Venice. Germany was equally remarkable before World War II both for the importance of its state or provincial libraries and for the lack of a recognized national library. The former Preussische Staatsbibliothek was permitted to assume this role in 1919; after World War II, under the name Deutsche Staatsbibliothek, it became the national library of East Germany with a collection numbering 3,000,000 volumes. West Germany has no national library as such, the functions of such a library being performed by the Deutsche Bibliothek in Frankfurt and the Staatsbibliothek Preussischer Kulturbesitz in Marburg and West Berlin. The National Library of Austria, founded by the Emperor Maximilian I in 1493, has rich collections—notably of manuscripts from the Austrian monasteries and from the library of Matthias Corvinus, dispersed after the capture of his capital, Buda, by the Turks in 1526. The National Library of Australia in Canberra is a comparatively new creation, having been formally created by legislation in 1960. It had, however, a long preparation for its role by its association with the Commonwealth Parliamentary Library. It is responsible for the *Australian National Bibliography* and has built up significant collections and played a large part in cooperative library developments. It has occupied since 1968 a fine new building on the shores of Lake Burley Griffin.

National libraries in Asia

At least 25 nations in Asia have national libraries, some of them—including the National Library of Peking, the National Diet Library in Tokyo, and the Jewish National and University Library in Jerusalem—with holdings of more than 2,000,000 volumes. The National Library of India (formerly the Imperial Library) in Calcutta, founded 1903, has over 1,000,000 volumes. Each of the countries of Central and South America has its own national library: that of Argentina is located in Buenos Aires, that of Mexico in Mexico City, and that of Brazil in Rio de Janeiro, this being by far the largest. In the developing countries of Africa the creation of national libraries is still largely a dream, though many of them, by setting up national library services, have taken a useful step toward realizing it.

University libraries. Like the national libraries, university libraries are in the main general in scope. Some of them, notably Harvard University Library at Boston and the Bodleian Library at Oxford, compare with the greatest national libraries in their collections' quality and range. Harvard is by far the largest university library in the world, rivalling in size both the British Museum and the Bibliothkque Nationale. Its collections have been conscientiously built up, since its foundation in 1638, by a succession of eminent librarians, and it now has outstanding collections in English and foreign literature, law, economics, Slavonica, and Orientalia. The director of the library has under his control nearly 100 specialist and separate departmental libraries. The Bodleian Library owes its foundation, in 1602 on the site of a former library, to Sir Thomas Bodley, a successful diplomat and bibliophile. Bodley was able to attract many benefactions to the library and was successful in concluding an agreement with the Stationers' Company in London whereby they deposited a copy on request of every new book published, a privilege which has been confirmed up to the present by successive Copyright Deposit Acts. The Bodleian is in process of reorganization as a federation of dependent libraries, including the Radcliffe Science Library, the Indian Institute Library, the Law Library, and others.

Many other university libraries, especially in the United States, have very large general collections, in some cases numbering upwards of 3,000,000 volumes, notably at Berkeley, California, Chicago, Los Angeles, and Princeton; they often have fine special collections. The same is true of the libraries belonging to many academies in Eastern Europe, notably the Academy of Sciences in Leningrad, which was founded in 1714 and was greatly enlarged after the 1917 Revolution.

Public Libraries. Public libraries play an indispensable role in the life of the community: they promote the reading of literature among grown-ups, adolescents, and children; they provide familiar information and reference centres, where the modern citizen can secure reliable and disinterested advice on practical matters and where he is helped to get his bearings in the complexities of modern community life.

Role of public libraries

From modest beginnings in the middle years of the 19th century, public libraries now provide well-stocked reference libraries and wide-ranging loan services based on systems of branch libraries. They are further supplemented by travelling libraries, which serve outlying districts. Special facilities are provided for the old and the disabled, and in many cases library services are organized for local schools, hospitals, and jails. The services provided vary in proportion to the size of the municipality or the area covered. In the case of very large municipalities, library provision may be on a considerable scale, including a reference library, which has many of the features associated with large research libraries. The New York Public Library, for example, has rich collections in many research fields; and Boston Public Library, the first of the great city public libraries in the United States (and the first to be supported by direct public taxation), has had from the first a twofold character as a library for scholarly research as well as for general reading. In the United Kingdom, the first tax-supported public libraries were set up in 1850; 120 years later, they provided a highly significant part of the country's total national library service. The importance of public library activities has been recognized in many countries by legislation designed to ensure that good library services are available to all without charge.

In many cases, public libraries build up collections that relate to local interests, often providing information for local industry and commerce. It is becoming more and more usual for public libraries to provide, for home loan: music scores, phonograph records, and, in some countries--notably in Sweden and the United Kingdom—original works of art for use, against a deposit, in the home.

Not all countries provide public library services of an equally high standard, but there has everywhere been a tendency to recognize their value and to improve services where they exist or to introduce them where they do not. Public librarians work strenuously, through such organizations as the International Federation of Library Associations, for such developments.

Special libraries. The national, university, and public libraries mentioned above form the backbone of the general library system in a country, looking after the needs of scholars and the general public. They are supplemented by libraries established to meet the highly specialized requirements of professional or business groups. These special libraries spring from the need felt in science and industry for better control of bibliographical information and greater efficiency in its communication and use than was possible in libraries of the traditional kind. Special libraries form an important element in a country's planned library system, whether directly, as in the case of the National Library of Medicine and the Library of the Department of Agriculture in the United States, or indirectly, as in the National Reference Library of Science and Invention in the British Museum, and in other parts of large reference libraries (the map and music libraries in the national libraries, for example, or the specialist libraries in the Harvard University and the Bodleian complexes). The libraries of specialist colleges and universities are, naturally, themselves special libraries.

Functions and services of specialist libraries

Special libraries are often attached to official institutions such as government departments, hospitals, museums, and the like. For the most part, however, they come into being to meet specific needs in commercial and industrial organizations. They are planned on strictly practical lines, with activities and collections carefully controlled in size and scope, though they may be and often are large and wide-ranging in their activities; they cooperate widely with other libraries. They are largely concerned with communicating information to specialist users, in response to, or preferably in anticipation of, their needs. They have therefore been much concerned with the theoretical investigation of information techniques, including the use of computers. Their activities have led to a critical reappraisal of library practices and have thus greatly benefitted library work in general.

School libraries. In countries in which public library legislation exists, public library services are often available to school libraries. Technical procedures, such as book purchase and processing, are carried out by the local library; the books are then sent to the school ready to be put onto its library shelves. This is the case in Denmark, for example. There is similarly close cooperation in other European countries, notably Germany, Sweden, and the United Kingdom. The service thus made available is paralleled in the United States by commercial cataloging services; in the U.S.S.R. it is performed by the All-Union Book Chamber, which prepares bibliographical descriptions and distributes them to all the school libraries of the country. The importance of school libraries is widely accepted on both sides of the Atlantic; teachers have too little time to spare for running the libraries properly, and so it is becoming more common for full-time professional librarians to be appointed.

Private libraries. The libraries owned by private individuals are as varied in their range of interest as the individuals who collected them, and so they do not lend themselves to generalized treatment. The phrase private library is anyway unfortunate because it gives little idea of the public importance such libraries may have. The private collector is often able to collect in depth on a subject to a degree usually impossible for a public institution; being known to booksellers and other collectors, he is given early information about books that will be of interest to him and that will fit into his collection; he can give close attention to the condition of the books he buys. In these ways he adds greatly to the sum of bibliographical knowledge (especially if he makes his collection available to scholars).

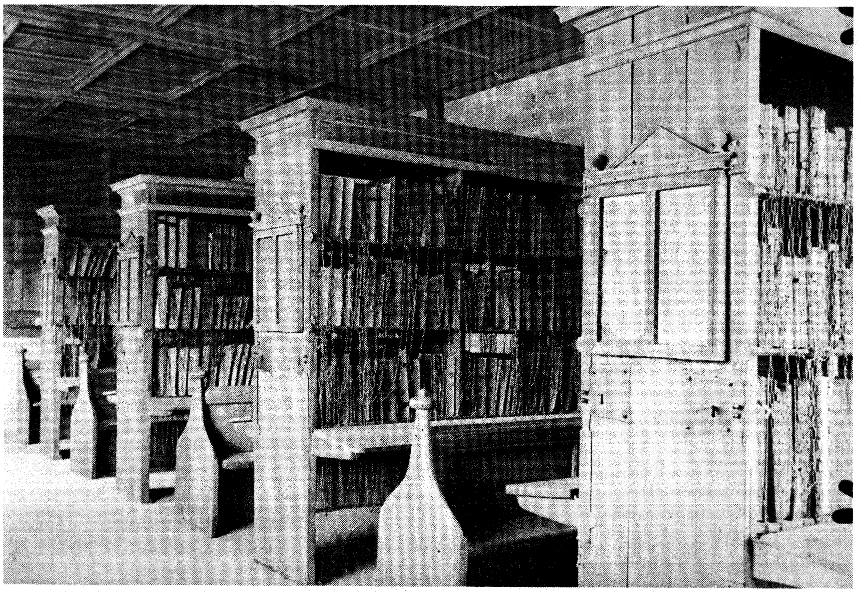
Henry Clay Folger, for example, collected no fewer than 70 copies of one book—the first collected edition of Shakespeare's plays. (In 1932 he opened the Folger Shakespeare Library in Washington, D.C., which had been built to house his collection.) As a result of the uninhibited nature of his collecting he added greatly to the sum of knowledge about the printing of Shakespeare's plays and about 17th-century printing in general. Collectors of private libraries may confer another inestimable benefit upon their fellows by leaving their collections to public institutions or by founding a library bearing their name. This practice has been common in recent years in the United States, where such names as Henry E. Huntington, John Carter Brown, William L. Clements, Pierpont Morgan, Henry Clay Folger, to mention but a few, are held in high esteem by scholars. They were following a tradition which had long been established in Europe, where notable private libraries (such as those of the Medici, Mazarin, and George III) have greatly enriched public institutions.

Subscription libraries. Part public, part private, these libraries enjoyed much popularity from the late 17th to the 19th centuries. Many of them were set up by associations of scholarly professional groups for the benefit of academies, colleges, and institutions; but their membership was also open to the general public. Some of them are still in existence: perhaps the most famous are the Library Company of Philadelphia, founded by Benjamin Franklin in 1731; the Boston Athenaeum, founded in 1807; and the London Library, opened largely at the instance of Thomas Carlyle in 1841, which today has a wide-ranging collection of over 500,000 volumes for loan to its members in their homes.

Famous subscription libraries

During the 19th century, the great size of many subscription libraries enabled them to wield much influence over publishers and authors: Mudie's Library, for instance, established in London in 1842, would account for the sale of as much as 75 percent of a popular novel's edition. Nevertheless, these libraries were for the most part unable to survive, and the service they gave is now largely provided by the free public libraries.

Archives. Archives are created in the course of conducting business activities of a public or private body; they are natural accumulations of documentary material. Originally, such records were not distinguished from library materials and were preserved in the same places as other manuscripts, right up until the mid-15th century and the invention of printing. The importance nowadays accorded to public records in an organized society has been recognized as one outcome of the French Revolution, when for the first time an independent national system of archive administration was set up, for whose preservation and maintenance the state was responsible and to which there was public access.



Chained library at Hereford Cathedral, England.
BY courtesy of the Dean and Chapter of Hereford Cathedral, England; photograph, F.C. Morgan

The science of archive administration embraces the study of records management, records appraisal, accessioning and arrangement, archival buildings and storage facilities, preservation and rehabilitation, and reference services, including exhibition and publication. Academic courses offering instruction in archive work are available in many parts of the world in universities and schools of library science.

While the administration of archives shares with libraries the basic obligation to collect, to preserve, and to make available, it has to employ different principles and management techniques. Libraries might be described as collecting agencies, whereas archival institutions are receiving agencies: they do not select—their function is to preserve documents as organic bodies of documentation. They must respect the integrity of these bodies of documents and maintain as far as possible the order in which they were created. And, of course, the documents need catalogs and finding aids, or guides.

Distinction between public and private archives

A distinction has to be drawn between public and private archives. Every state, broadly speaking, now recognizes the need to preserve its own official records and is expected to maintain a system of archive administration, which has the function of collecting them, preserving them, and making them publicly available after the appropriate lapse of time. Among the best known are the Archives Nationales in France, the U.S. National Archives, and the British Public Record Office. Nonofficial archives—the records of the day-to-day activities of an institution or a business—are now recognized as having great value for socio-economic history, and they are frequently sought by libraries for their historical value and preserved in manuscript and similar collections. It is the practice for many institutions such as universities, professional and commercial organizations, and ecclesiastical establishments, to set up their own archive departments.

LIBRARY ARCHITECTURE

Function and design. The basic function of a library building is to house the library's collections, to provide adequate space for staff administration and procedures, and to offer acceptable accommodation in which the collections can be used. The form given to the library is undoubtedly of great importance, but its success will in the long run be measured by its ability to meet the basic functional requirements. These requirements have to be interpreted differently according to the objectives and needs of different types of library, whether large research, university, public, or special.

In planning a research or university library, due regard to the following considerations is essential: (1) The space set aside for the collections must not only be adequate for immediate needs but also allow for planned growth over a reasonable period of time. (2) Adequate provision should be made to ensure the preservation and safety of the collections, such as proper temperature and humidity control, air conditioning, and a building plan calculated to provide security against misuse. (3) Provision should be made for the convenience and comfort of the readers and of the library staff, including the avoidance of unnecessary and time-wasting movement within the building.

The same strictures apply, on the whole, to the planning of other kinds of libraries. A good public library would pay more attention to providing an attractive exterior, a convenient layout, and pleasing conditions for users so that they can "browse" at leisure and discuss their reading and information requirements comfortably with the staff employed for that purpose. The functional emphasis in a special library would be on ease and speed of consultation.

Essential collaboration between architect and librarian

The final form to be aimed at in all these cases is a library building both functionally and architecturally distinguished. This a first-class architect can accomplish, but he is unlikely to do so unless he understands the institution he is working for and its objectives and has a satisfactory program to work to. Hence, the closest collaboration is essential between librarian and architect, the former supplying as complete a brief as possible (after careful discussion with library colleagues) and the latter endeavouring to understand the librarian's brief and interpret it in building terms. Difficulties and misconceptions are removed by close and constant discussion. Cooperation of this kind has become normal practice only in recent years; indeed, it is only in recent years that the concept of a fully documented brief prepared by the librarian has been accepted or even understood by architects and librarians alike. In the early days, libraries were often associated with temples; and it is still common to look upon a library as a central feature in a municipal or university complex, hence the temptation to erect a "prestigious" building, which may and often does subordinate function to appearance.

Historical developments. The clay-tablet rooms in the temple at Nippur in Babylonia may be looked upon as functionally designed for their purpose, as was an "archive room" excavated at Herculaneum; and, in the cloister of the early monasteries, the *armarium*, or book cupboard, was appropriately placed for its purpose in the well-lighted cloister. Monastery and cathedral libraries of the later medieval period employed the chained library stall system and were eminently functional in their form. Each consisted of a long room that ran the length of the building, usually on the first upper floor, lighted on both sides with rows of windows. Between the windows and at right angles to the walls were rows of back-to-back presses, with two or three fixed shelves on each side and with desks attached to the presses at the level of the lowest shelf; the books were chained to a long bar running the width of the presses by chains just long enough to enable them to lie open on the desk. When not in use, the books, usually at the time bulky folios and quartos, stood on the shelves, with their front edges facing outward so that the chains, which were normally attached to the front edge of the upper or lower cover, should not damage the binding of the other books on the shelves. A centre gangway ran the length of the room between the presses; and in the alcoves formed by the presses there were fixed benches for readers, who were thus able to get at the books on the desks and have enough light from the windows to read by.

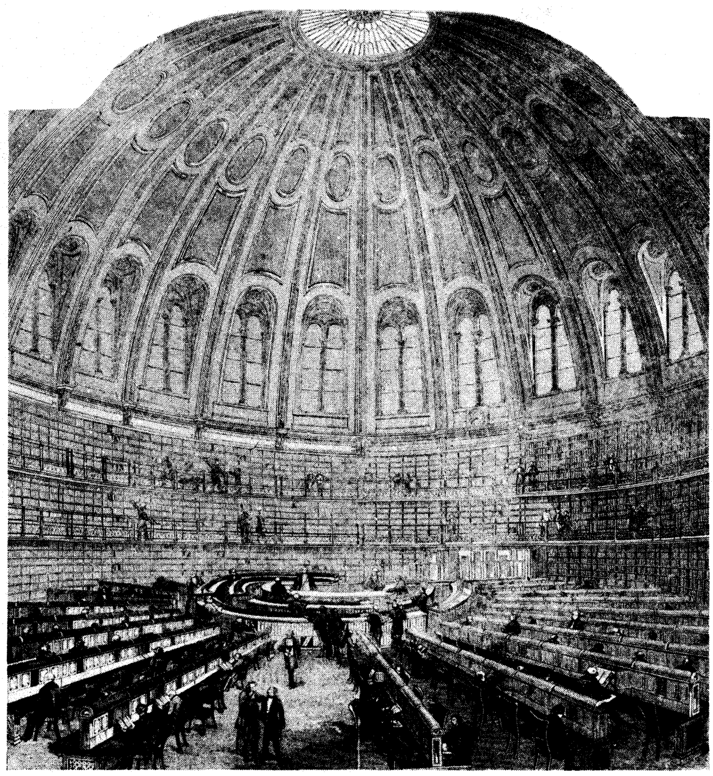
17th and 18th centuries. The "long room" was retained even in 17th- and 18th-century libraries, a reminder of the older system. The abandonment of chaining, however, coupled with a change in the pattern of publication—marked by increasing cheapness and greater numbers of publications, particularly in smaller sizes—made possible the adoption of wall shelving. Wall shelving, in turn, made possible the greater use of the height of a building—the higher the walls, the greater the number of books that could be shelved; besides this, the disappearance of chains made it possible for books to be taken away from the shelves for reading, and seats could be disposed as desired; the walls thus left unencumbered with desks and benches could be experimented with, and the main floor area could be left empty or provide an elegant setting for statuary or exhibits of other kinds, such as coin cabinets. Sometimes spaciousness and display were more highly prized than convenience for study. An example of this mood is to be found in the Prunksaal in the Austrian National Library in Vienna, erected in 1723–26 after a plan by Johann Fischer von Erlach, which consists of a magnificent Baroque hall, decorated with marble Corinthian columns; a statue of Charles VI stands in the middle and other statues of princes of the empire are placed along the walls. A number of the libraries in the Oxford and Cambridge colleges—the Wren Library at Trinity College, Cambridge, and that at Christ Church, Oxford—also provide excellent examples. Perhaps the last great flourish of library design of this kind was the King's Library in the British Museum, built between 1823 and 1826 to house the library of King George III, acquired by the museum in 1823. It consists of a room 300 feet (about 100 metres) in length with a gallery; the walls are lined with book shelves from floor to ceiling.

Wall shelving

British Museum reading room. A complete innovation in functional planning was made with the celebrated reading room in the British Museum. By the middle of the 19th century, readers were too numerous for the space provided for them in the existing reading rooms. In 1854 Antonio Panizzi, its principal librarian, by a brilliant application of the "engineering age" to library construction, planned a bookstack in cast iron with exterior enclosing walls and widely spaced brick piers, to hold some 1,500,000 volumes; it was designed to surround and support a huge, circular reading room whose radiating rows of reading desks with places for 450 readers naturally followed the circular plan, facilitating supervision, and whose walls provided space for an open access general reference library of 25,000 volumes. The "iron library," which provided an extensive library in close proximity to the reading area, reduced the risk of fire and demonstrated immense saving of space by eliminating brick supports and substituting metal. Modern stack construction has greatly improved on it, substituting steel for cast iron and generally economizing on space; but the basic plan of this stack was followed in subsequent stacks.

Bookstacks and reading rooms

The modern period. Panizzi's creation focussed attention once more on functionalism in library planning. From that time up to World War II, the design of libraries was conditioned largely by architects' interpretation of the functions to be performed by the library, and their buildings rarely gave any indication of what was actually performed within them. General style tended toward the monumental, as befitted (so it seemed to them) a building devoted to noble and scholarly pursuits. The librarian had little say—indeed, as a rule probably had little to say--on the design of the building, except to indicate the proportion of space to be allotted to bookstacks, reading rooms, and administrative quarters.



British Museum Reading Room. designed by Antonio (later Sir Anthony) Panizzi, 1854. Illustration by Sydney Smirke, from The Illustrated London News. 1857.

By courtesy of the trustees of the British Museum; photograph, J.R. Freeman & Co. Ltd.

A specific part of the building was set aside for bookstacks, and, not surprisingly in view of the constant growth in library collections, the stack frequently came to be designed as a tower. Reading rooms were treated as separate architectural units, in many cases, related neither in proportion nor access, to the bookstacks. Finally, the space provided for administrative purposes and the routine library procedures (such as control and recording of acquisitions, cataloging, binding, and so on) was frequently planned without properly appreciating the importance of a relationship between bookstacks, reading rooms, and public. Fixed function planning produced buildings with little possibility of the flexible use of space and as a rule with only limited, if any, possibility of expansion. Much frustration was caused to the librarian, who had to put up with (and provide) light and heat for unnecessarily high reading rooms, spacious corridors and, sometimes, monumental halls and staircases. The reader was expected to climb unnecessary stairs and to walk long distances from entrance to reading room and often from the catalog and the reference shelves back to his reading place.

Access to the stacks by readers was rarely provided in the large research libraries (nor is it always desirable that it should be). Public libraries, however, though they too were often housed in monumental buildings, found it possible to introduce open access to their stock of books very early on (in England during the 1890s). Open access made bookstacks with high shelves impracticable for readers, and it encouraged the use of classified arrangement of books so that readers could easily locate the subjects and the volumes of special interest to them.

A remarkable change of attitude toward the design and construction of library buildings took place following

World War II. It is of considerable interest for the history of libraries and their development to consider what the reasons for this change may have been. Scientists, technologists, and others had, during the war, needed precise information, speedily supplied; there was undoubtedly much dissatisfaction with the traditional library services and hence a more critical attitude toward libraries and their characteristic features. Afterward, the needs of users and their convenience were given greater consideration and more careful regard. There were also many more people using libraries—more students in a greater number of universities and colleges and more members of the general public looking to libraries for information, instruction, and recreation. Significant improvements in building techniques, moreover, made new building designs and methods possible, whereas very many new libraries were needed, partly because of the interruption of building during the war, partly because many libraries in Europe and other parts of the world had been destroyed or damaged. The old kinds of library, with their monumental and inflexible structure, were no longer acceptable.

The modular system of building

The new look at library design, thus prompted, has been further encouraged by development of the modular system of building. In the modular building, the basic floor area is divided into equal rectangles defined by structural columns at the corners, which are the only weight-bearing structures within the building. Subdivisions of these areas can be created by non-weight-bearing walls, bookstacks, and freestanding furniture; it follows, in theory, that nothing within the building is fixed and immovable except the columns (and of course stairways, elevators, heating facilities, ducts, and plumbing) so that the use made of an area can be extended or modified at will.

Erik Uluots

Interior of the library in Växjö, Sweden, building and furniture designed by Erik Uluots. completed 1985.

The value of such a building needs no emphasizing. The modular method may, however, have a restrictive effect on the architect. It is likely to be most effective if it is used in a basically rectangular building. There is thus a real danger that library buildings designed on this basis may be commonplace or even downright ugly.

An eye to the future. If one of the major requirements of library architecture, maximum flexibility, has been provided by the use of the modular system, there are nonetheless many other matters which still have to be seriously considered by the library planner. It is commonly held that, on present showing, research library collections are doubling in size every 16 to 20 years—even the largest national libraries are increasing at this rate. Such an increase can hardly be supported indefinitely, so librarians (and the bodies to which they are responsible) even now are being urged to plan their strategy to deal with the problem. Demands for reading places also continue to grow, and it is likely that society will require sympathetic responses to this and other demands, including longer hours of opening and more convenient and comfortable reading and study conditions.

CONTEMPORARY LIBRARY SCIENCE

The libraries mentioned above are, for the most part, splendidly maintained and a source of pride to the institutions or public bodies that administer them. But it would be idle to pretend that research libraries, in particular, do not face great problems as their number proliferates: how to maintain comprehensiveness and balance in the collections, given the enormous increase in book production both from traditional sources and from new publishing centres; how to cope with local, fluctuating demands for collections in depth to meet important projects; how to solve the difficulty of increased pressure on funds and space; even, at times, how to avoid internecine rivalry for the possession of certain items or groups of items as they become available.

Experimental solutions to library problems. *Interorganization cooperation.* The problems listed above are insistent problems—common to libraries everywhere—and they are not being ignored. If convincing solutions to them have not yet been found, many expedients are at least under investigation and experimentation. There is first the rationalization of collection, by means of acquisitions schemes such as the Farmington Plan in the United States and the Scandia Plan in the Scandinavian countries, whereby individual libraries agree voluntarily to cover, in as much depth as possible, certain subjects or subject fields without abrogating their right to continue collecting in any other subject for which they may have special responsibility (a similar system is in practice in West Germany, under the aegis of the Deutsche Forschungsgemeinschaft); there is also the use of cooperative centres for the acquisition and preservation of out-of-the-way books, as the Center for Research Libraries (formerly the Midwest Inter-library Center, Chicago); the use of joint storage for little used material; centralized or cooperative cataloging, with the ultimate aim of making cataloging an internationally shared activity, the catalog entries being provided from central sources in each country; the introduction of the practice of printing a catalog entry in every book as it is published; the formation of networks or consortia to establish computer-aided technical-processing centres. While these experiments hold out much promise, it is by no means clear whether they are sufficiently far-reaching or whether they get to the root of the problem. Neither is it certain that institutions will be willing to contemplate any diminution of their cherished independence of policy and action. Some much more drastic replanning—and perhaps some fundamental rethinking of library activities, especially in very large libraries—may be necessary before solutions to these difficult problems appear.

Substitution of microform copies

Microforms. It seems advisable to many that the expansion now typical of almost every library should be limited as far as possible so that only selected libraries, such as national collections, by reason of their special place in the community, should continue to collect and preserve comprehensively. There are, however, certain expedients that may provide temporary alleviation of the problems associated with growth. One of the most heavily canvassed is the substitution of microform copies of library material instead of original volumes. Such copies take the form of microfilm or the extension of that concept of the microfiche, or film card. This is a promising expedient but one that so far has suffered from many disadvantages: the cost of microfilming has been high, though the expense can properly be weighed against the saving in the cost of valuable library space; the devices for reading microforms for the most part have been clumsy and inconvenient and possibly trying to the reader forced to use them over long periods. In many cases, particularly when rare and valuable books are involved, when scholars may need to examine closely the details of printing, paper, layout, and the like, microforms may be unreliable or misleading. Much research is under way to remove or reduce these disadvantages, especially as many libraries are likely to be forced—either because copies of the originals can no longer be found or because of limitations of space—to use microforms.

Space-saving techniques. Another way of economizing on space is to use "compact shelving": briefly, this means that standard stack ranges, mounted on rails with ball-bearing wheels, can be pushed tightly together, thus adding very substantially to the amount of storage available in any given area. The disadvantages are high cost, the need for extremely strong floor construction, and diminished availability to the reader.

Projected developments. The introduction of devices such as these will inevitably have a major impact both on design and on library procedures, as will also the mechanical and data-processing devices that are more and more frequently being employed. All of these will inevitably play a major part in future library services, particularly the circulation systems, interlibrary cataloging, and information retrieval. Facilities for storing and using audio-visual aids, motion-picture films, tapes, and computer tapes will all have to be provided in the libraries of the future.

Librarianship is clearly at a time of great potential development. Above all, there is a need for reappraisal and fundamental rethinking regarding the place of libraries in the community. The traditional independence that libraries have always enjoyed may have to be modified because of a need for cooperation; for example, standard procedures may have to win a more general acceptance. The advent of automation, which may help the librarian to exploit the material under his care more effectively than is possible by traditional methods, might well have a decisive influence here.

It is important that libraries should hasten to make these modifications themselves rather than, by delaying, have more drastic changes forced upon them. For it is an indication of the contemporary reaction to the difficulty and complexity of library problems that they are beginning to be dealt with nationally rather than locally. Although many libraries and librarians place great confidence in local initiative and distrust the effects of enforced uniformity and centralism, the economics and the technology of the modern world suggest that local support and initiative cannot solve the problems alone. In many countries there is a national ministry with responsibility for libraries, with a system of inspectors to ensure uniformity of services and achievements. Such systems exist already in the countries of Eastern Europe, and France and Italy have long had their directorates of libraries. Norway has recently joined the ranks, but with a National Library Advisory Service (Biblioteks tjenesten) that can apparently allow scope for initiative and independent action. In the United Kingdom, steps are being taken to set up a National Library Authority but so far with control only over the strictly national libraries.

Library systems and procedures in advanced countries inevitably influence those in the developing countries, and in most of the contacts between representatives of the two there has been an assumption that what works well in the advanced countries will work equally well for the others. Many now recognize that this may be a mistaken assumption, and it may be more desirable that such public educational institutions as libraries should endeavour to embody and reflect local cultural patterns. International library associations, including the International Federation of Library Associations (IFLA) and the International Federation for Documentation (FID), with active members in all parts of the world, are playing an increasingly important role in providing a meeting place for nationals of many countries, including those in course of development, where such problems can be discussed.

Despite the attention paid to research in librarianship, little study has been devoted to the deeper implications of the current library situation, and it may well be that it is too early and the complexities too involved for it to be possible for a librarian to take a dispassionate view of his profession, however desirable an objective reappraisal may be. Library literature is plentiful and good on the practical problems of library work, but it would appear that librarians seem for the most part to be too heavily engaged in the day-to-day activities and in the routine tasks of administration to be able to devote much time to the study of their profession. Work on the history of libraries generally tends to be similarly descriptive but deficient in studies in depth—for which there is ample raw material. The best work has been done on the early history of libraries by, for example, John Willis Clark (***The Care of Books,*** 1900) and F. Wormald and C.E. Wright (***The English Library before 1700, 1958***). The library associations, of which there are very many throughout the world covering the work of libraries of different kinds, seem to be similarly preoccupied in the discussion of current professional problems.

BIBLIOGRAPHY. The literature of libraries is very large, but of uneven quality.

Encyclopaedias: There are numerous encyclopaedias; the most authoritative of these is probably by FRITZ MILKAU (in German), but this consists of a series of studies on individual aspects of libraries and librarianship and is not so convenient to consult as those alphabetically arranged by topics. A. KENT and H. LANCOUR (eds.), ***Encyclopedia of Library and Information Science,*** vol. 1–3 (1968–70); C. LOEFFLER and J. KIRCHNER (eds.), ***Lexikon des gesamten Buchwesens,*** 3 vol. (1935–37); ***Nordisk Leksikon for Bogvaesen,*** 2 vol. (1949–57); FRITZ MILKAU (ed.), ***Handbuch der Bibliothekswissenschaft,*** 2nd ed., 3 vol. (1950–57); J. KIRCHNER, ***Lexikon des Buchwesens,*** 4 vol. (1952–56), vol. 3 and 4 plates only; SVEND DAHL (ed.), ***Nordisk Håndbok i Bibliotekskundskap,*** 3 vol. (1957–60); THOMAS LANDAU (ed.), ***Encyclopaedia of Librarianship,*** 3rd ed. (1966); H. KUNZE and G. RUCKL, ***Lexikon des Bibliothekswesens*** (1969).

Periodicals: Among the wide range of library journals, the ***Journal of Documentation*** (quarterly) is authoritative in its field, and the following are widely respected as important sources of information on library matters in general: ***Journal of Librarianship, Library Quarterly, Library Trends, Libri, Unesco Bulletin for Libraries, Zentralblatt fur Bibliothekswesen, College and Research Libraries, School Librarian, Special Libraries. (Abstract journals)*: *Library Literature*** and ***Library Science Abstracts.***

Guides to libraries: Among guides to libraries, the ***World of Learning*** (annual), gives factual information about libraries of all kinds, in all parts of the world. The following are useful guides to the range of library and bibliographical services in the various countries with which they deal. ROBERT B. DOWNS, ***American Library Resources: A Bibliographical Guide*** (1951); GISELA VON BUSSE and H. ERNESTUS, ***Das Bibliothekswesen der Bundesrepublik Deutschland*** (1968); G. OTTERVIK and S. MOHLENBROCK, ***Bibliotek i Sverige*** (1964); PAUL L. HORECKY, ***Libraries and Bibliographical Centres in the Soviet Union*** (1959).

Techniques of librarianship: The emphasis formerly placed on the techniques of librarianship in such valuable works as MARGARET HUTCHINS, ***Introduction to Reference Work*** (1944); and MARGARET MANN, ***Introduction to Cataloguing and the Classification of Books,*** 2nd ed. (1943), is being supplemented by attention to the functions of libraries as in the following: R.R. SHAW (ed.), ***The State of the Library Art,*** 5 vol. (1960–61); RONALD STAVELEY ***et al., Introduction to Subject Study*** (1967); MAURICE F. TAUBER ***et al., Technical Services in Libraries*** (1954); E.E. WILLIAMS, ***Farmington Plan Handbook*** (1961); LOUIS R. WILSON and MAURICE F. TAUBER. ***The University Library: The Organization, Administration and Functions of Academic Libraries,*** 2nd ed. (1956).

Types of libraries: A.J.K. ESDAILE, ***National Libraries of the World,*** 2nd ed., 2 vol. (1957), is a purely descriptive account, but in general, books on types of libraries show a concern for function rather than description. UNESCO, ***National Libraries: Their Problems and Prospects*** (1960); HERMAN H. FUSSLER and J.L. SIMON, ***Patterns in the Use of Books in Large Research Libraries*** (1961); W.A. MUNFORD, ***Penny Rate: Aspects of British Public Library History, 1850–1950*** (1951); WILFRED ASHWORTH (ed.), ***Handbook of Special Librarianship and Information Work,*** 2nd ed. (1962); GUY R. LYLE, ***The Administration of the College Library,*** 3rd ed. (1961); C.A. STOTT, ***School Libraries: A Short Manual,*** 2nd ed. (1955); MARY P. DOUGLAS, ***The Teacher-Librarian's Handbook,*** 2nd ed. (1949).

History of libraries: The history of libraries, apart from descriptive histories of separate libraries, which is too bulky to mention here, has not in general attracted the scholarly attention it deserves. The following, however, demonstrate what can be done by scholarly treatment. J.W. CLARK, *The Care of Books*, 2nd ed. (1902); R. IRWIN, *The English Library* (1966); N.R. KER, *Medieval Libraries of Great Britain: A List of Surviving Books* (1941); C.B. OLDMAN, *et al.*, *English Libraries 1800–1850* (1958); JAMES W. THOMPSON, *The Medieval Library* (1939, reprinted 1957); FRANCIS WORMALD and C.E. WRIGHT, *The English Library Before 1700* (1958).

Library education: Library education has a wide literature. A useful survey of the subject is given in LARRY E. BONE (ed.), *Library Education: An International Survey* (1968).

Automation: The importance of automation in libraries is shown by a growing body of literature on the subject. ROBERT M. HAYES and J. BECKER, *Handbook of Data Processing for Libraries* (1970); and R.T. KIMBER, *Automation in Libraries* (1968), give an overall picture. HENRIETTE D. AVRAM, *The Marc Pilot Project* (1968), describes the important development by the Library of Congress in the use of automation to provide comprehensive bibliographic control.

Architecture and buildings: The universal acceptance of the importance of libraries in the community, and the growth in the size of research libraries, has resulted in great attention to library buildings and the creation of many new library buildings, and as a consequence in a number of outstanding contributions to the literature of the subject, notably: ANTHONY THOMPSON, *Library Buildings of Britain and Europe* (1963); RALPH E. ELLSWORTH, *Planning the College and University Library Building* (1960); KEYES D. METCALF, *Planning Academic and Research Library Buildings* (1965); H.R. GALVIN and M. VAN BUREN, *The Small Public Library Building* (1958); J.L. WHEELER and A.M. GITHENS, *The American Public Library Building* (1941).

Archives: The comparatively recent development of archive administration on a scientific basis is being supported by a growing literature, of which the following are significant: H. JENKINSON, *A Manual of Archive Administration*, 2nd ed. (1937); L.J. REDLAND and F.W. STEER, *Local Records: Their Nature and Care* (1953); T.R. SCHELLENBERG, *Modern Archives: Principles and Techniques* (1956) and *The Management of Archives* (1965); E. POSNER, *American State Archives* (1964).

(F.C.F.)

Library Science

Library science is the discipline that encompasses all aspects of library operation. It did not become an identifiable discipline, nor did librarianship begin to emerge as a profession, until the latter half of the 19th century. Although libraries have existed since ancient times, they were previously related to religious or educational institutions, and their direction had been merely one of the duties of members of the organization.

The first modern work on library science, the *Advis pour dresser une Bibliothèque* (1627; "Advice on Establishing a Library"), was prepared by Gabriel Naudé, who was appointed by Cardinal Mazarin to run the Bibliothèque Mazarine in Paris. Translated into English in 1661, this source must have had a considerable influence on the wealthy collectors and scholars of the Restoration period. Further identification of the principles of operation and creation of codes of practice began to occur with the creation of the royal libraries in the 18th century. A significant advance in librarianship occurred in the British Museum under the direction of Sir Anthony Panizzi, whose rules for creation of the catalog of that library have had a most important influence on library practices until today.

Library economy was the designation first used for the subject and continued to be used, particularly in Great Britain, into the 20th century. The term library science was current in the United States by the middle of the 19th century and has persisted despite objections from time to time that the field cannot qualify as a science.

More recently, two other fields of study related to library science have evolved: documentation and information science. The scope of these subjects has not been clearly defined; some use them interchangeably with each other, while others consider them two distinct subjects. Documentation deals with the creation, reproduction, analysis, and retrieval of documents. Information science, usually considered to be somewhat broader in conception, is concerned with the entire information-transfer process, whether through the medium of recorded documents or otherwise. To some, however, information science is identified with the electronic transmission of signals and is closely related to information theory. To others, information science is synonymous with computer science. In any event, library science is closely related to both information science and documentation and is making increasing use of the concepts being developed by them.

Related fields: documentation and information science

The librarian's work encompasses the collection, organization, and making available of the records of all human activity. To achieve these ends, thousands of libraries have developed to serve more or less homogeneous groups on the basis of subject, interest, age, or geographic distribution. The materials acquired may consist of books and booklike documents, art in all media, phonograph records, motion pictures, models, and even data in digital form. Although the book represents the most numerous category of documents, the proportion of other forms is increasing as a result of the technological changes of the 20th century.

Despite the fact that the totality of the library's contacts with users numbers high into the millions, it is not an institution of mass communication. With few exceptions, each loan or consultation of a document is in response to the individual needs and desires of one person. The uses to which such consultations are put may run from the most casual entertainment to the most serious research, but it is the emphasis on satisfying the needs of the individual at the time of his own choosing and in private that differentiates libraries from museums, theatres, or classrooms. (See also LIBRARY; INFORMATION PROCESSING.)

TECHNICAL SERVICES

Acquisition. *Selection policy.* The clientele of each library influences the materials to be acquired, the organization of that material, and the policies to be followed in making it available. The contents of libraries serving apparently similar constituencies are quite varied. Because there are tens of millions of books to choose from, it is obvious that many will not be found in even the largest libraries. In two university libraries of the same size, more than half of the contents of one library will be different from those of the other.

Because facilities and budgets do not permit acquisition of all materials published, each library must follow some selection policy. Authorities recommend that a formal selection policy be adopted in a library, but in practice the policy is more likely to be implicit than explicit. Each library limits its acquisitions on the basis of assumed or demonstrated needs of its users. A college or university library will usually exclude all elementary or secondary school textbooks; a public library will not buy material in an exotic foreign language; an elementary school library will not consider scientific-research reports. From these obvious limitations each library naturally progresses to finer definitions of categories of material to be acquired and the depth of coverage; thus, one library may collect only a representative group of materials on the archaeology of Crete, while another will seek to obtain every book, journal, or report on the subject in every language.

Limitations of acquisitions

Problems in development of a selection policy are relatively simple in libraries of institutions with clearly focussed and narrowly defined missions. As these limits are broadened and the focus becomes less clear, conflict between various interests and points of view appear and reach their peak in the public library. Because the public library serves the entire spectrum of interests of a large population, it has endured a continuing struggle between satisfaction of popular demands for recreational reading and the more diverse demands of intellectually elite minorities. This often results in a contest between entertainment and education or between best sellers and quality books.

Censorship. This same lack of agreement on the primary goals of the public library make it a fruitful ground for controversy or censorship in the selection of materials. Because of concern for children, the school library is an equally sensitive area.

Pressures on libraries to prohibit, restrict, or label certain materials arise from various sources: individuals; social, religious, political, or ad hoc organizations; public officials; governing boards; and even on occasion members of the library staff. Many examples of library censorship exist, among them the rigid government control of German libraries during Nazi rule, when books by Jewish and Communist writers were banned.

At times it becomes difficult to identify a practice as censorship or as selection policy. In recent years in the United States, there have been accusations of censorship against the library by minorities who feel that the selection policy has systematically excluded material of interest to or in support of the minority point of view; these objections have included claims that books on the Afro-American contribution to the culture have been ignored or that books with a conservative point of view have been rejected. Concern with the problem of censorship led the American Library Association to adopt a library bill of rights in 1939, which was last amended in 1967. This document stresses the right of access to all points of view, rejects the notion that books should be excluded because of the race, nationality, or beliefs of its author, and states that libraries should resist both private and official censors.

Censorship and library bill of rights

Evaluation procedures. Few libraries are in a position to examine any significant portion of books to be acquired. More commonly, dependence must be on selection and reviewing media. The American Library Association, for example, publishes the *Booklist and Subscription Books Bulletin,* which includes short reviews of books that may be appropriate for a small library. It also publishes *Choice,* which has short, unsigned reviews of books considered appropriate for college libraries. Several commercially published library periodicals contain reviews of current books.

In addition, libraries utilize literary periodicals and book-review sections of newspapers, such as the London *Times Literary Supplement* and the *New York Times Book Review.* Many scholarly journals review books in the scope of the journal, but these are far less current and are often delayed from six months to two years after the original publication date of the book reviewed.

Bibliographic verification. Many books are purchased without benefit of review, because of either the delay in reviewing or the inadequacy of reviewing media to cover all books. Often acquisition is based upon the reputation of the author or of the publisher. Newer books may be requested as a result of publisher announcements; older books may be requested in response to inclusion in a bibliography or citation in some other work.

In many cases it is necessary to verify the authorship, publisher, place of publication, edition, date of publication, or the price. At times it is even necessary to establish the existence of the book. Tools used in this process include national bibliographies published in many countries, such as the *British National Bibliography* and the *Japanese National Bibliography,* published at frequent intervals. There is no national bibliography in the United States, but the *National Union Catalog,* published by the Library of Congress, serves the same purpose fairly well. A number of commercial lists of current books, known as trade bibliographies, are published in various countries. In the United States these include the *Cumulative Book Index,* which lists each month new books reported by author, title, and subject.

National and trade bibliographies

Sources of purchase. Libraries obtain current books from the publisher, through a retail book store, or through a jobber or wholesaler who specializes in library business. Periodical subscriptions may be obtained either directly from the publisher or from subscription jobbers or agents.

Libraries acquiring foreign materials seldom attempt to deal directly with publishers. They have a choice of using a native jobber who specializes in importing books or of using a jobber in each country who specializes in export business. There are advantages in each type of arrangement, depending in part on the volume of purchasing being done in the country.

Since the 1960s there has been an expansion of approval-order plans, particularly among larger academic libraries. Various jobbers have developed plans by which a library will be supplied automatically with all current books that fit the specifications of a profile that the library has supplied; the books may be returned within a specified period of time if not wanted. Advocates of these programs point to the increased speed in receipt of material; opponents cite the tendency to take everything and thus to fill shelves with material of questionable value.

Every library must also acquire older material not available from the publisher or through normal trade channels, either as replacement or because it was missed at the time of publication. The antiquarian book trade is a complex of independent dealers who handle secondhand or rare books, often specializing by subject or some other characteristic. Finding the book needed often involves advertising in one of the periodicals devoted to the antiquarian book market.

Gifts and exchanges of books

Other sources of supply. Gifts and exchanges are supplementary sources of library materials. Gifts by individuals to libraries may on occasion be the only source of acquisition. Many of the outstanding scholarly collections resulted from the philanthropy of a book collector; numerous weak and struggling libraries could not exist without the gifts of ordinary books from numerous small donors. Exchange of materials between libraries further enhances their collections; arrangements may exist between libraries for the regular exchange of publications of the institutions of which they are a part. Universities and learned societies have extensive exchange relations throughout the world.

A number of organizations conduct programs for piece by piece exchange of unwanted duplicates. This traffic includes books but concentrates most on single volumes or issues of periodicals. Establishment of such operations was proposed in the 19th century. In the United States, for example, the American Library Association, the American Association of Law Libraries, and the Medical Library Association each conduct exchanges.

Cataloging. Cataloging and classification, although associated, are two distinct operations, and each can be utilized in a library without the other. Cataloging consists of listing the bibliographic items that make up a library, usually by author, by title, by subject, or by a combination thereof. Classification consists of assigning each item to a specific class, which in most systems is subdivided by subject. The call number is the notation of the specific class to which the book is assigned, with additional symbols to identify the specific book. Books are generally arranged by number when they are placed on shelves, although some libraries, especially in Germany and Austria, still arrange books by order of acquisition and by size (when access to bookshelves is restricted).

Book and card catalogs. Cataloging is far older than classification, and catalogs have been in existence for many centuries. The earliest catalogs were manuscripts, themselves in book form. Manuscript book catalogs exist today in some of the most important libraries of the world; *e.g.,* the Bodleian Library of Oxford University. Various methods were used to interpolate new entries as they were added to the catalog.

Early catalogs

The first important French catalog was made in 1622 for the collection that was to become the Bibliothèque Nationale. Another catalog was prepared in 1684, arranged into 23 classes, each designated by one or more letters (this is still in use). During the 19th century the practice of libraries printing their catalogs in book form became rather common. Pressure by the trustees of the British Museum in 1834 for a printed catalog was the impetus for creation of Panizzi's cataloging code of 1839. Although this project met nearly insuperable obstacles, two editions have been published: 1881–1900 and 1931–66. In the United States the Library of Congress Catalog was first printed in book form in 1942–46, and cumulative supplements have followed regularly.

The card catalog was the product of the small group of library pioneers in the 1870s; in the United States it existed almost to the exclusion of other forms, while in other countries with long library histories book-form catalogs continued in use despite the preponderance of card catalogs. The Library of Congress has printed cards for its catalogs since 1898 and has sold cards to other libraries since 1901. The impetus for the card catalog was the difficulty in incorporating new entries into the book-form catalogs and the near impossibility of updating in a single alphabet the various copies of the printed book catalog. But in time disadvantages of the card catalog became apparent; it became difficult, particularly in catalogs containing several million cards, for the user to find the materials needed in sections that have many subdivisions, such as "Shakespeare" or the "Union of Soviet Socialist Republics."

The electronic computer, with its capacity for manipulating large stores of data rapidly, has been studied as a possible tool to eliminate the difficulties of book catalogs. There have been numerous experimental uses of the computer to create book catalogs, but in only a few instances have they replaced card catalogs. Experimentation continues, and there is the possibility that both the book and card catalogs may be replaced eventually by an electronic catalog that can be projected, one item at a time as requested, on a cathode-ray-tube screen.

Arrangement of entries in library catalogs is usually alphabetical by word rather than letter by letter. When author, title, and subject entries are all filed in the same sequence, the result is a dictionary catalog. When the author and title entries are filed in one sequence and subject cards in another, the result is called a divided catalog. The subject section of the catalog may be arranged according to the classification scheme used in the library rather than alphabetically by the indexing terms used. Such an arrangement results in a classed catalog.

Types of catalogs

Cataloging codes. Whenever a library consists of more than a few thousand volumes, the complexities of cataloging reach such proportions that a formal set of cataloging rules must be established. Cataloging codes have existed since at least the 16th century. One code that has exerted strong influence on modern libraries was that developed for the British Museum by Sir Anthony Panizzi. A direct lineal descendant was the Anglo-American rules of 1908, which was the standard for the English-speaking world for 40 years. In 1949, in an effort to simplify cataloging, new American rules were published. These modifications were never completely satisfactory, and a concerted attempt was made to create a new code that would restore uniformity to cataloging practices in all English-speaking countries and at the same time approach agreement with the practices in Europe that were based upon the Prussian Instructions, a set of cataloging rules for the Preussische Staatsbibliothek (now the Deutsche Staatsbibliothek) that had long served as a model for libraries in German-speaking areas and influenced practices in countries using other Germanic languages. The new Anglo-American code adopted in 1967 achieved some but not all objectives of its promoters.

Subject cataloging is the procedure that bridges cataloging and classification. Traditional subject cataloging consists of assigning one or more subject headings to represent the content of the book. These headings are selected from an approved list of terms, such as the Library of Congress ***Subject Headings Used in the Dictionary Catalogs of the Library of Congress*** (eighth edition). Provision is made for cross-references from unused terms and from one term to a related one. A catalog containing these entries is known as a syndetic catalog.

Classification. Systematic classifications of knowledge, influenced by Aristotle, Bacon, and other philosophers, have been utilized in libraries for many years. In China, which has a library tradition dating back to the 6th century BC, a system of classification using four categories (classics, history, philosophy, belles lettres) was employed but has been modified under the influence of the modern Universal Decimal and Library of Congress systems. The pioneers of classification by subject were the booksellers and bibliographers, including a group in the 18th century known as "the Paris booksellers," and the Swiss doctor Conrad Gesner, who published a classified list of 20,000 scholarly books, the ***Pandectarum sive partitionurn universalium***, in 1548. France has produced many classified bibliographies, notably the ***Manuel du libraire*** of J.C. Brunet in 1810.

Widespread interest in classification developed in the latter half of the 19th century, a period of great growth of libraries and the expanding concept of open stacks. The Americans Melvil Dewey and Charles Cutter and other librarians developed systems soon after 1870. The organization of the American Library Association in 1876, followed very shortly by the Library Association in Great Britain, made possible formal discussion and propagation of interest in the subject.

Development of classificatory systems

Classification systems may be categorized as enumerative or as facetted. Enumerative systems set forth major subjects and various levels of subdivisions in a linear fashion. Facetted systems attempt to represent more than one logical relationship through the representation of a number of facets of a subject by combining segments of the system. The most well-known facetted systems are the Universal Decimal Classification, originally an adaptation of the Dewey Decimal Classification, and the Colon Classification (so-called for its use of the colon to represent facets), developed by the Indian librarian S.R. Ranganathan.

Dewey Decimal Classification. The most widely used classification system is that developed by Melvil Dewey in 1873 for Amherst College. It is used in far more libraries in the United States than all other systems combined and is employed throughout Europe, Asia, and Africa, as well as Latin America. Editions and modifications have been published in many languages; there have been 18 editions of the standard version and nine editions of an abridged edition used in schools and other small libraries. Taking its name from the decimal notation used, the system uses 999 major numbers, but additional classes can be created almost without limit by the generation of decimal subdivisions. An example from the schedule for chemistry shows how numbers are subdivided:

540	chemistry and allied sciences
541	physical and theoretical chemistry
541.2	theoretical chemistry
541.3	physical chemistry
541.34	solutions
541.35	photochemistry
542	laboratories, apparatus, equipment.

Another feature of the Dewey system is the mnemonics used for certain types of subdivisions. Thus, many subjects can be subdivided geographically by the use of the historical-geographical number as decimals:

900	general geography and history
970	history of North America
973	history of the United States.

Combining with the art schedule, the number for history of art in the United States is obtained:

700	the arts
709	history of art
709.73	history of art in the United States.

Library of Congress Classification. In the United States another widely used system is that developed by the Library of Congress. It was first published in outline in 1904 and since then has been continually expanded and revised. It is spreading rapidly in the United States, particularly in college and university libraries. By 1970 an estimated 400 libraries were using this system. The Library of Congress Classification owes much to Charles A. Cutter, whose Expansive Classification was originated almost simultaneously with Dewey's Decimal Classification. Pragmatic in approach, it is planned as a classification of books rather than of knowledge, and its allocations of space for expansion are more nearly in proportion to the publication rates than those of the Dewey system.

Popularity of Library of Congress system in the U.S.

Universal Decimal Classification. The Universal Decimal Classification was first published in French in 1905; it has since been expanded in great detail by the Fédération Internationale de Documentation and published in many languages. The wider subject divisions are similar to those of the Dewey Decimal Classification, but the detail has diverged considerably. Any two class numbers in the scheme can be linked by a colon to form a more specific subject. Thus, 621.785 "heat treatment" may be combined with 669.14 "steel" to form 621.785: 669.14 "heat treatment of steel."

The combination of classes in this way leads to long class numbers, inconvenient for the spines of books and catalog cards. The founders of the Universal Decimal Classification, Paul Otlet and Henri Lafontaine, realized that shorter numbers could be obtained if the subdivisions of a main class were sorted into homogeneous groups, such that combinations would often occur among groups but rarely within groups. Terms from each group can be combined with a shortened class number.

SERVICES TO USERS

Reference services. Reference service consists of providing personal assistance to individual library users in pursuit of information: it requires the recognition by the library that such assistance-is necessary in fulfilling its duties, and that a specific administrative organization of qualified personnel must be provided. The emergence of reference service as a major aspect of librarianship is rather recent in comparison to the acquisition, organization, and custody of library materials. The term did not appear until about 1890, although recognition of the need for aid to readers had been emerging for several decades.

As first envisaged in Europe and the United States, reference service was to be rendered to the uninitiated user largely through assistance with the bibliographic apparatus of library operation. The most austere or conservative approach to reader assistance was found in the college and university libraries. Scholars were presumed not to need help, and librarians were considered incapable of rendering assistance to them. Public libraries were frequently more generous in assistance, but, until the end of the 19th century, the reference librarian was considered essentially a guide or at best a teacher.

The liberal approach to reference service

The so-called liberal approach to reference service arose in special libraries, such as those serving industrial corporations or specialized government agencies. The point of view emphasized the information dispensing function of the library and recognized such services as literature searches, preparation of abstracts, evaluations of sources, and the dispensing of information orally. The advocates of this approach did not turn their backs upon the teaching function of reference service but included preparation of reading lists as appropriate activites. In most parts of the world, current practices in most libraries fall somewhere between these extremes.

Within the older conception of reader assistance, the services of libraries have been extended to the more sophisticated and scholarly users, as well as novices. Subject specialization in reference work, with departments devoted to such areas as business, art, science and technology, and history, staffed by librarians with advanced knowledge of the subject, have become commonplace in larger libraries of all types.

Preparation and dissemination of bibliographies on topics of timely interest are a standard facet of library service. Sometimes the local bibliographic endeavours are supplementary to published guides and become part of the apparatus that the reference librarian uses.

In the organization of libraries, certain books needed frequently for consultation are usually set apart and held for use within the library. At one time reference librarians tended to look upon these reference collections as their chief concern, but, as the conception of reference service has broadened, they have come to rely upon the entire resources of the library and even beyond. Currently, reference departments frequently administer extensive interlibrary-loan programs and photocopying services.

Instruction in use of the library. In schools and colleges there is need for more extensive instruction in use of the library than can economically be rendered through regular reference services. Various forms of classroom or group instruction are found.

Public libraries seldom have formal meetings for instruction, relying instead upon information desks and the regular reference desk. Like college and university libraries, they sometimes have handbooks of library information for free distribution to patrons.

Attempts to create an ideal system for recording loans

Charging systems. Because the number of loans made in a large library may run into millions each year, the system for recording them should be economical for both the library and the borrower, as well as effective in providing the essential information. For the last century, librarians have sought the ideal system that would answer the questions: where is a particular book? when is it due? what books does a borrower have? In addition, the system should provide statistical data for improvement of book acquisition and retention policies and similar management functions.

After much experimentation, two widely used systems had evolved by the early 20th century. The Newark System, developed by the Newark (New Jersey) Public Library, was adopted almost universally by U.S. public libraries and is also used in other countries. A card with call number, author, title, and other identifying data was kept in a pocket on the inside cover of a volume. Each time the book was issued, the identification of the borrower and the date the book was due were entered on the card. The cards were filed by call number or by author and title and placed behind a date guide. This file provided an effective control on the prompt return of books. At the time of registration, each borrower was issued a borrower's card, on which the essentials of each loan transaction were recorded; thus, the number of books allowed to a borrower could be controlled, fines could be collected, and delinquent borrowers could be refused further loans. Discovering the location of a book was more difficult, because at least 12 separate date files were to be consulted.

The second system that was widely used in college and university libraries in the United States and other countries was similar, except there were two cards. One was filed by date as in the Newark System; the other was interfiled in a master file by call number. College libraries seldom limited borrowers to a fixed number of books, so that the borrower card was rarely used.

Dissatisfaction with the Newark System began to be manifest about 1925. The first new development was the book-charging machine, which was used to imprint the borrower's registration number, using an embossed plate attached to the borrower's card. The next significant change in public-library systems was the use of the transaction card, of which there are two basic versions. In the Photocharger system, a camera placed at the circulation desk is used to photograph the borrower's identification card, the book identification (the book card or its equivalent pasted inside the cover), and a numbered transaction card as the record of a loan. The transaction card, on which the due date has been stamped, is inserted in the book pocket. Because the transaction cards are used consecutively, the file arrangement is purely chronological. This system has the advantages of accuracy and economy but makes impossible the location of a book on loan, imposes a single loan period, and makes renewal difficult. The audio-charge is similar, except that the elements of the record are recorded on magnetic tape or disk.

Computer-based circulation systems

The transaction system was not accepted by United States college libraries because of the loss of information considered essential to service. Instead, they moved in the direction of punched cards and, eventually, computers. The systems developed were considerably more expensive than those utilized in public libraries, but they approached the ideal system sought by librarians in the 19th century. When increased enrollments led colleges in the United States to issue identification cards, it was relatively simple to adapt them to data-collection devices used in conjunction with computers. The combined data of borrower's identification and book identification plus date and time of the transaction are recorded in a punched card or on a magnetic tape or disk. Although relatively expensive, the system is economical of borrower time, eliminates all filing and most of the work of removing records from the file, and provides answers to all of the basic questions asked of a charge file. Statistical data are generated automatically for more effective management.

Library extension. The public-library movement has been characterized in most countries by a missionary zeal to extend the benefits of library service to ever-increasing numbers of people. This extension has been achieved by the establishment of libraries to serve rural areas and smaller cities, by building branches in larger cities, by the use of bookmobiles, and by the use of the postal services to deliver books.

In metropolitan areas the public library may maintain widely dispersed branches, so that most potential users can have ready access to them. The branch undertakes to stock the most frequently used books for adults and for children, to provide basic reference services, and to bring the resources of the central library to the neighbourhood user by interbranch loans or by referral to it. In very large libraries there are regional branches that serve to augment the resources of the neighbourhood branches.

Bookmobiles serve to bring the library even closer to the entire constituency. In the United States they were first used in rural areas but more recently have been employed in cities as well. They are also used in many other areas of the world, including the British Isles and Australia. The bookmobile is a specially constructed van mounted on a truck chassis and equipped with book shelves. Stocked with a collection of materials selected to meet the various interests of users in the route to be served, it may make two or more stops in a day at shopping centres, schools, parks, or village squares. Only limited reference service can be rendered directly, but the bookmobile librarian may take questions or requests for specific books back to the central library. The information may be sent by mail directly to the requester's home or delivered on the next return of the bookmobile.

In several countries "outreach programs" have been developed that do not rely on the user's walking into the branch or bookmobile. Participating libraries have active programs of making specialized and personalized library service available to patients in hospitals, to the elderly, and to the handicapped. Programs aimed at segments of society that have not used books in the past are also being tried, particularly in urban areas.

Interlibrary cooperation. Cooperation between libraries in provision of better services assumes many forms. Basic to all 'programs is the sharing of resources, but there are also programs for cooperative acquisition, cataloging, and storage of materials at both the local and the international level. Interlibrary loan is an essential part of the cooperative effort.

Inter-library loans

By 1926 the practice of interlibrary lending had become common enough in the United States that a national interlibrary loan code was adopted to clarify and standardize policies and procedures. Revised several times, this code restricts lending to serious scholarship. Local and regional arrangements are more generous and may make any materials available to serve any user. Developments in photocopying have altered lending practice through the provision of facsimile copies of short articles in lieu of the loan of bound volumes of periodicals, but this practice is somewhat limited by copyright. At the local level, the sharing of resources may include mutual honouring of library cards issued by the participating libraries. Thus, the residents of neighbouring towns may visit and borrow freely from any local public library, or the students of one college may use the library of another.

In the Latin American countries, several important projects for library cooperation are in progress, including the use of cooperative cataloging and union catalogs, as well as interlibrary loans.

Union catalogs and cooperative processing. One of the chief obstacles to effective interlibrary loan service is knowledge of the location of material. Cooperative efforts have been devoted to creation of union catalogs of the cooperating libraries. The first national union catalog to begin publication was the ***Deutscher Gesamtkatalog*** (1931–39), which was to include the books held by the major German libraries. Only 14 volumes covering A through Beethoven had been published when the project was suspended in 1939.

The ***National Union Catalog*** at the United States Library of Congress includes reports from several hundred participating libraries in every state. The catalog was originally in card form and was consulted by letter or telephone call to the Library of Congress. From 1956, additions to the catalog have been printed in book form in annual and later in five-year editions. The pre-1956 catalog is in process of publication and is said to be the largest work ever published; when completed it will have 610 volumes. This national catalog does not represent the holdings of smaller libraries, however, so that there exist many local and regional union catalogs.

The ***Union List of Serials*** provides information on the specific volume holdings of periodicals and other serial publications in the major participating libraries of the United States. Likewise, numerous local and state lists augment the information in the national list.

Coopera-tive processing centres

Other nations also have or are preparing union catalogs, among them Brazil and Italy. The activities of libraries in creating bibliographic guides to library resources for interlibrary loans has corollaries in cooperative processing centres that purchase or catalog materials or both for the participating libraries. The earliest of these centres were established by groups of small public libraries in which each individual library was too small to support an efficient operation alone. With the development of computer technology, the concept has spread to larger libraries and to wider geographic areas. The Ohio College Library Center, for example, which serves most of the colleges of Ohio, provides various services, including the creation of sets of catalog cards by the computer, consultation of the common catalog via computer terminals in each library, and insertion of new catalog data through these same terminals.

Cooperative acquisition. Because libraries have recognized the impossibility of even the most affluent to be sufficient alone, efforts at cooperative acquisition have been made at the local and national levels. In most cases the participating libraries assume major responsibility for collecting material on specified subjects, leaving the other libraries free to ignore these areas or to acquire only those items specifically in demand. The best known cooperative-acquisition program in the United States is the Farmington Plan, which aims to assure that at least one copy of every book of research value published anywhere in the world is acquired by a library in the United States. These materials are recorded in the ***National Union Catalog*** and are available for loan to any library. The program began operation in 1948 for three countries and has been expanded to include most of the world's book production.

Yet another type of cooperative-acquisition project is the Center for Research Libraries in Chicago, which acquires and houses materials that are available to its member libraries. Originally conceived as a cooperative storage centre to which members might transfer little-used materials, recent emphasis has been on original acquisition by the centre of important material for which a single copy will meet the demands of a large clientele. Supported by membership fees, it has members throughout the United States and Canada.

BUILDINGS AND STORAGE

Library buildings in the 19th and early 20th centuries, even the smaller ones, tended to be monumental in design, following the pattern of the British Museum. They usually contained a high-ceilinged reading room, often on the second floor, and a multitier steel bookstack. After World War II, new library buildings tended to emphasize functional rather than aesthetic considerations.

Post-World War II library architecture

With the development of the fluorescent light and the adoption of air conditioning, the need for high ceilings was eliminated. Library buildings began to be created with ceilings from eight to nine feet (about 2½ metres) in height throughout, and the multitier bookstack was often abandoned in favour of movable shelving. Thus, it became possible to utilize any area as a bookstack, reading room, or office. The flexibility increased the potential for growth and change in libraries that had been limited by the architectural features of older buildings, but the buildings tended to become austere, rectangular boxes. After several years the value of architectural aesthetics was recognized, and more recent buildings have sought to retain the flexibility, while restoring architectural beauty.

The changes in buildings have been related to changes in the philosophy of library organization. At one time, the major portion of the book stock was closed to the public. In large libraries the multitier stack was an efficient method of storage. But over the years the concept of free and open access to the books by all users gained more acceptance. Closely related to the idea of open stacks was the subject-divisional organization of services to readers. As with open access to books, this movement to place reference and other assistance to readers near the books they needed began in the public library and spread into university libraries. Under these influences, the distinction between reading rooms and stacks tended to disappear, and the architectural differentiation was inappropriate.

Experimentation and development in increasing the efficiency of the library building led to the creation of compact stacks with a considerably greater shelving capacity than traditional stacks. There are three basic types, all of which reduce the amount of space devoted to aisles. One has sections of shelves mounted on hinges; a second system places the books on pullout drawer-type shelves, which again are made accessible by temporarily blocking the aisle. The third type has a number of ranges of stack on tracks with space for a single aisle between them.

Automatic shelving and reshelving

A recent development, not yet widely used, is a shelving system by which books may be called for and reshelved automatically; it is most efficient if installed in new buildings specifically constructed to its specifications. In this system books are placed in containers of uniform size, the number in each container determined by the size and shape of each book. Because there can be no browsing among the shelves, arrangement is most efficient if based on size rather than on subject.

DEVELOPMENTS IN DOCUMENTATION

Information centres. A proliferation of publications in many technical fields, particularly those affected by government research grants, has led to the need for assistance beyond that traditionally provided in library-reference services. While libraries, as a matter of both economic necessity and basic philosophy, attempted to guide the reader to the materials he desired without influencing his own evaluation of them, information centres undertake to analyze, evaluate, select, and even on occasion to synthesize answers to questions by users.

These centres may but need not be part of a regularly constituted library and may rely upon human sources as well as documents in providing their answers. Some are operated or supported by governments. Such centres are set up to serve a narrowly defined, highly specialized field. An example is the Japanese Information Centre of Science and Technology, established in Tokyo in 1957 to collect, store, and retrieve information from foreign and domestic sources. The information received is classified according to the Universal Decimal Classification, and there are mechanical devices for information retrieval. The centre undertakes on request bibliographical and documentation research work and translations from Japanese into English and from the foreign language into Japanese. It also services microfilms. In the United States the Defense Documentation Center (DDC), which originated as the Armed Services Technical Information Agency, the Educational Resources Information Center (ERIC), a loosely coordinated group of more specialized clearinghouses, and the National Technical Information Service are well-known governmental centres. The work of these agencies is coordinated by a presidential Committee on Scientific and Technical Information (COSATI).

Services of information centres

Information centres vary in the services rendered and the clientele served, but they all publish guides to report literature, provide microform reproductions of the more significant fugitive publications, and do computerized searches of the literature in response to user requests. They generally publish abstracts of significant literature and often publish exhaustive reviews of the entire literature of the subject. These publications are made available on a regular basis to all potential users of the centre. Some maintain systems for selective dissemination of information (SDI), based upon the use of computers in the indexing functions, that provide timely notification to Individual users of new publications appearing to fall in their particular fields of interest.

Information centres rely heavily upon newer techniques in documentation, including computers, telecommunication, and microform reproduction. Centres that provide specific answers to individual questions often utilize computerized search techniques.

A few centres have completely automated the inquiry and response to a limited number of frequently asked questions. In a centre devoted to heart disease and stroke, for example, several hundred questions have been recorded on audio tape and are kept up to date. The user can call the centre by telephone, specifying the question by dialing the appropriate numbers, and then listen to a response. If more information is needed, he merely gives his name and address, which is recorded, and the information is provided after a search of the appropriate literature; photocopies of appropriate articles are sent to the user.

Microphotography. Microphotography was originally used in libraries to reproduce unique or rare items not otherwise obtainable, such as manuscripts or early books; material on paper subject to deterioration, such as newspapers; and bulky, little-used materials. More recently, original publication has occurred in microform.

A microform may be defined as a document in reduced format, such that the text cannot be read with the naked eye. There are essentially four types of microforms. Microfilm is the oldest and most widely known. Consisting of a cellulose acetate base to which a silver-bromide emulsion has been applied, it is exposed, developed, and fixed using standard photographic processes.

Microcards were conceived by Fremont Rider, a U.S. editor and publisher, who in his concern for the inexorable expansion of libraries proposed to reproduce the text of a book on a card of the dimensions used in catalogs. A brief catalog entry appears at the top of the card in normal size; from 60 to 80 pages of an average-sized book are reproduced on the lower portion of the card, with additional cards used as needed. The Microcard is a contact photographic print from a master negative. The similar microprint card is produced by photolithography, using a high-gloss paper and special drying inks. The microfiche is similar to the Microcard in format, except that it is on a transparent cellulose base, created as a positive contact print from a master negative. Microfiche was used in Europe for some time before it took hold in the United States.

Uses of microforms for varying types of publications

Although there is competition between each of the forms, each remains dominant in specific areas. The ribbon microfilm is most frequently used for newspaper reproduction. Microfilm is economical for very small editions; in fact, editions of one copy are most common. Microfiche and Microcard are most convenient for small publications that can be reproduced on a single fiche or card.

Most commercial publication in microform consists of massive collections of material representing a subject, a period of time, an outstanding library collection, or a geographic area, rather than single bibliographic items. In many cases the library must acquire the entire collection, and, when individual pieces are for sale, the prices are appreciably higher. Frequently some form of bibliographic guide for a project is provided, so that the library does not need to catalog the individual items.

Care of and conditions for microforms

Microforms, if properly developed and treated, will last as long as rag paper, but the conditions (*i.e.*, temperature and humidity) under which they must be stored are somewhat more rigorous than those required for paper. They must be protected from dust and chemical pollutants in the air. Except for archival master negatives, ideal conditions usually exist in recently constructed air-conditioned libraries.

The successful use of microforms depends upon the availability of devices for reading them. Each requires a separate type of machine. Efforts to build a satisfactory multiple-form reader have had only limited success. Because of the range of reduction ratios used in original photographing and the varying size of the originals, readers capable of handling all microform reproductions are bulky and expensive. Models of readers have been developed by which the user needing a specific image for intensive use may make an enlarged print automatically. These readers are even more expensive, however, and by some processes the cost of the print is high. Most libraries with large microform collections have enlargement facilities that can provide the prints more economically.

There continues to be user resistance to microforms. Whether this results from physiological or psychological causes, users complain of eyestrain, inconvenience in locating and consulting a specific page, and limitation of availability. The user will accept microforms if he must use the material and there is no alternative, but because of this resistance there is little likelihood that microforms will replace the extensively used lending-book collections.

Library automation. The introduction of data-processing equipment and procedures into library operation began about 1935 in the United States. The University of Texas Library introduced a circulation system using punched cards; the Boston Public Library made use of punched-card methods to analyze certain aspects of acquisition at about the same time. An experimental, highly advanced circulation system was installed at the Montclair, New Jersey, Public Library in 1941. Although there were some expressions of interest elsewhere, adoption was slow.

Computer problems

Various factors impeded the adoption of computer-based techniques: ignorance and fear of machines; limitation of character sets to capital letters, numbers, and a few punctuation marks; costs of conversion and operation. By the 1960s, however, enthusiasm for library automation was worldwide; unfortunately, performance did not always approach expectations. Computer specialists not familiar with library problems underestimated the complexity of them and promised too much; interested librarians had too much faith in computer magic. The result often was systems that were inadequate or that cost more than estimated. By the end of the decade, reality began to approach expectation, but only a small fraction of library operation was computer based.

Data processing. Interest in computers in libraries has been at two distinct levels: at the data-processing level, which includes the automation of operating records, and at the level of retrieving the intellectual contents of the library. There is a basic connection between the two that was not recognized at first. It is essential to conceive of the library as an information system, the chief components of which involve acquisition procedures and attendant fiscal operations, the inventory-control functions, including circulation records, and the bibliographic subsystem that includes creation and maintenance of catalogs and indexes for effective utilization of the library's contents. There is also the management function itself; the information about the other functions and operations is used in making management decisions.

Use of computers for retrieval

Information retrieval. Most of the attempts to utilize computers that originated in libraries were concerned with implementing an acquisition system or a record system for lending books or the creation of catalog cards. The projects at the information retrieval level were generated primarily in research settings, funded by grants from government and private sources. Project MAC at Massachusetts Institute of Technology, one of the most acclaimed, is a general purpose information retrieval system in which a number of users can simultaneously make inquiries via remote terminals. Its importance lies in the demonstration of the potentialities of the computer as an information storage and retrieval device. The Medlars project of the National Library of Medicine in the United States was originally directed toward the mechanization of publishing of Index *Medicus,* but outside interest in it has focussed more on its potential for mechanized retrieval of medical information.

Although entering the field of library automation somewhat later, the United States Library of Congress has assumed national leadership in the field. A study by a panel of experts published in 1964 reported the feasibility of automation of its bibliographic processes and envisaged access to its contents through computer consoles rather than through extensive card and book-form catalogs. The Marc project (Machine-Readable Cataloging) that followed established and tested a standard format for exchange of bibliographic information in machine-readable form. Tapes containing a significant portion of the cataloging output of the library are available for purchase by subscription and form the basis of a number of regional cooperative processing centres.

Interest in mechanized information retrieval has led to research in new indexing techniques. New classification systems, thesaurus construction, statistical analysis of language to create abstracts automatically, and improvements in concordances are examples of the variety of subjects studied. The most widely used technique for indexing by computer, although not the most sophisticated, is the KWIC, or Key-Word-In-Context, index. In most cases it is simply a permuted title index, with the remainder of the title surrounding it—in context.

There have been predictions that the computer revolution in information handling will eventually result in magnetic, electronic, and microform stores of information that can be retrieved and viewed from cathode-ray-tube screens in the office or the home, and that the traditional library with its bookstacks and reading rooms will be obsolete. A more realistic appraisal is that computers will be extensively used to create tools for utilization of the library's collections, and that only a small segment of the informational content will be under complete computer control. The reasons will be more economic than technological; the cost of information storage in a computer probably will remain many times the cost of storage in book form on library shelves.

Attempts to automate classification

Automatic classification research. Advances in information-processing technology have brought attempts to automate the process of classification. As pointed out previously, a library classification system is a scheme for organizing a collection of documents into groups, so that related subject matter is brought together in a systematic fashion. Establishing clearly demarcated groups or categories can reduce to reasonable proportions the number of items to be scanned by a user. If the process is to be automated, the principle for determining class membership by subject area must be stated in terms that a computer can process. Computers cannot understand ideas per se. The ideas in documents, however, are expressed in words, and computers can process and compare words, as well as numbers. Documents on different subjects will use different sets of words to express ideas. It follows, therefore, that documents can be ordered into classes on the basis of similarity or differences in vocabulary. This is the basic principle on which automated-document classification rests. Research in automated classification includes activities in two areas: (1) automated methods of deriving a classification system and (2) automated methods of classifying documents. One can set out to automate the role of the classificationist—the one who designs the classification schedule and determines the number and kinds of categories that will be used.

There are a number of mathematical techniques that enable the analysis of a document collection into a number of subjects. One commonly begins by characterizing each document as a list of content-word terms and then continues by comparing each document with every other one in the collection on the basis of the words they share in common. Subsets are thus obtained. Each subset is a category within the classification scheme. Initial studies indicate that these automatically produced categories are reasonably descriptive of the document collection and similar to the manually derived categories.

Automatic determination of the subject of a document

The second aspect of classification research deals with the specification of mathematical procedures for computing the probability of a document belonging to a designated category—to automate the job of the classifier. The aim is to be able to determine automatically the subject content of a document and thus classify it into its proper, pre-established category. The procedures used are based upon word-content similarity.

Numerous techniques have enabled researchers to achieve a certain measure of success in automatically classifying documents. The diversity indicates that no one method is clearly superior to all the others, and no one method has achieved a full measure of success.

LIBRARY EDUCATION

Library associations. The development of library services and the rapid growth of public libraries in Britain and the United States in the second half of the 19th century led to a natural desire of librarians to unite in the furtherance of their aims. The American Library Association (ALA) was founded in Philadelphia in 1876 and the Library Association of the United Kingdom in London in 1877. The ALA is the largest of the national library associations. Its object is to promote library service and librarianship in the United States and throughout the world. The Library Association of the United Kingdom, renamed the Library Association (LA) in 1896, was granted a royal charter in 1898. In 1962 the LA was reorganized and became fully professional.

After the formation of these library associations, numerous others were set up in countries all over the world. Worldwide interchange of ideas occurs through the International Federation of Library Associations (IFLA), which was founded in 1929 and meets annually. It sponsors research projects of international interest aimed at unifying the principles of library science. Independent regional library organizations include the Asian Federation of Library Associations (1957) and the Inter-American Bibliographical and Library Association (1930).

The International Federation of Documentation (FID) was founded in 1895 as the Institut International de Bibliographie. Its chief project was a universal card catalog of the world's great libraries, of which about 14,500,000 cards had been compiled in Brussels by 1934; little progress, however, has been made since then. The Institut became a federation in 1938; after its first postwar congress (16th, Paris, 1946) it devoted itself to problems of documentation, classification (especially the Universal Decimal Classification), document reproduction, information services, and, later, mechanical and electronic storage and retrieval of information. In this specialized field, the International Federation for Information Processing (IFIP) was founded in 1959. IFLA and FID cooperate closely with the International Organization for Standardization (ISO) in promoting the international standardization of bibliographical references, abbreviations of titles of periodicals, and transliteration.

UNESCO's interests in library science

The work begun by the International Institute of Intellectual Co-operation set up by the League of Nations was in part revived by UNESCO after World War II. UNESCO began with four main objects: improvement of bibliographical services, promotion of public libraries, promotion of international conventions facilitating the free flow of all forms of information, and development of the international exchange of publications. Although the program developed to suit changing conditions, changes flowed logically from one or more of the original aims. Thus, the improvement of bibliographical services led to the establishment of national libraries and bibliographical centres; development of bibliography and documentation in education, the natural and social sciences, and specialized subject fields; cooperation with ISO and FID in international bibliographical standardization; encouragement of national copyright laws; and improvements in cataloging. In this special field the work is coordinated by the UNESCO International Advisory Committee on Bibliography, Documentation and Terminology, a group of experts meeting regularly under various titles since 1953.

Training. In Great Britain, the Library Association is the recognized professional examining body. Training is undertaken at ten full-time schools of librarianship (one year for graduates; two years for nongraduates). There are also postgraduate diploma courses (one year) at the University of London School of Librarianship and Archives and at Sheffield, Strathclyde (Glasgow), and Queen's (Belfast) universities. There are university schools in Canada, Australia, New Zealand, India, Pakistan, Ceylon, Ghana, Nigeria, and Malaysia, and several university schools in South Africa.

Training of librarians in the U.S.

In the United States, there are one- and two-year postgraduate courses at university library schools. These and the schools in Canada are accredited by the American Library Association. The master's degree indicates professional status. The first formal education for librarianship occurred with the establishment of the School of Library Service by Melvil Dewey at Columbia University in 1887. In 1926 the Graduate Library School was established at The University of Chicago to provide education beyond the first professional degree, offering a Ph.D. degree in library science. Since 1950, the number of doctoral programs in the United States has increased. In addition, the emergence of information science as an identifiable discipline has brought a new dimension into curricula of library schools, the newer ones being designated as Schools of Library and Information Science.

In France there is an advanced school at the Bibliotheque Nationale, and the École Nationale des Chartes issues a diploma in archive work. In Germany and Sweden academic librarians and public librarians are trained in separate schools. Switzerland has a library school at Geneva. In The Netherlands training is controlled by the Central Association for Public Libraries. Norway, Denmark, and Finland each have a library school and professional certification. In Spain the Department of Archives, Libraries, and Museums holds courses to train librarians; a school at Barcelona trains public librarians.

The Soviet Union has a highly developed training system of government-controlled schools. In Poland examinations are controlled by the Ministry of Education. There are library schools in Czechoslovakia and Hungary, where qualifications are regulated by law. Turkey, Israel, Iran, Thailand, the Arab Republic of Egypt, Senegal, the Philippines, Taiwan, and Japan have library schools. In Latin America there are schools in Argentina, Brazil, Chile, Colombia (including the Interamerican Library School at Medellin), Cuba, Mexico, Peru, Uruguay, and Venezuela. Successive regional seminars on library development held by UNESCO in Asia, Africa, and Latin America have consistently urged governments to increase training facilities.

APPLICATIONS OF LIBRARY SCIENCE TO DIFFERENT TYPES OF LIBRARIES

National libraries. Most countries have national libraries at the apex of their library systems. In some cases the library is part of the executive department of the government, but in others the library that serves the legislature is also the national library (*e.g.*, the National Diet Library of Japan). In some instances (*e.g.*, that of Norway), the library of the national university serves also as the national library.

Functions of national libraries

Despite wide variations in organization and control, certain functions are usually performed. In most cases the national library is the recipient of copies of all copyrighted publications. This fact facilitates the publication by the library of comprehensive bibliographies of current book production and allows it to serve as a centre for cataloging services for other libraries within the country. In addition, it provides the mechanism for more effective interlibrary cooperation. The national library assumes direction of, or at least a leadership role in, extension of library services to the entire population. Examples of this are the provision of materials for the blind and the use of the library's resources for extensive interlibrary loans.

In some countries, including the United States, there is no single national library. The Library of Congress, which serves many of the functions of a national library, is legally only the library for the Congress. Its national role is shared with the National Library of Medicine, which was originally the library of the Surgeon General of the Army, and the National Agricultural Library, which was formerly the library of the Department of Agriculture. The functions of the national library in the United States*are shared by the state libraries.

In the Soviet Union, four great libraries—the State V.I. Lenin Library of the U.S.S.R., the State Public Scientific and Technical Library, the Central Library of Agricultural Science, and the State Central Scientific Medical Library—act as administrative centres. Thus, they advise individual libraries on acquisitions, as well as on cataloging, technical processes, and readers' services.

University libraries. University libraries share with national libraries the responsibility for supporting most of the research in a country. They continue to support the archival functions that characterized libraries in ancient and medieval times while making their resources more widely available to users than in the past. Development of comprehensive collections in areas of primary interest rather than uniform coverage of all fields characterizes university library operation.

In the United States, the phenomenal growth of university enrollment in the mid-20th century made it necessary for libraries to alter their organization in order to render adequate service to both undergraduates and research-oriented faculty and graduate students. Two trends have been apparent: the creation of separate undergraduate libraries and the increase in specialized subject libraries.

The responsibility for selection of materials has shifted markedly from faculty to the library staff. In most cases there is a joint responsibility with the faculty library committee, of which the library director is usually a member and serves in an advisory capacity on policy. Reasons for the increased responsibility for the library staff include a significant improvement in the educational qualifications of librarians, the emphasis on research, with its consequent pressures on the faculty, and the great increase in rates of book publication.

Adoption of Library of Congress system in university libraries

There has also been a trend in the United States toward use of the Library of Congress Classification instead of the Dewey Decimal Classification. Among the reasons for this shift, which began about 1930, is the belief that the Library of Congress system better reflects the disciplinary distinctions in a university. There is also an economic advantage in using without change the cataloging and classification information available through the printed catalog cards that the Library of Congress distributes.

School libraries. When the newer media—motion pictures, slides, sound recordings-came into existence they were adopted for use in elementary and secondary education sooner than in colleges and universities. In the United States, there arose a new group of audiovisual specialists who operated almost entirely independently of the library. Recently, however, there has been a movement for reuniting the two aspects of essentially the same function under the name of Instructional Materials Center. These centres may include individual study booths for listening to records of all types and for watching motion pictures as well as for reading books.

Role of the public library

Public libraries. The objectives of the public library are not limited by those of the parent institution; its policies are determined largely by its governing board, and the staff has far greater responsibility for its direction and activities than is the case in other types of libraries. Yet there have been more difficulties in defining the role of the public library. Is its primary function to serve as an educational institution? Or is its chief purpose provision of recreational reading? The pursuit of meaningful social values without becoming an active participant in community conflict has been of continuing concern. Public libraries generally consider their role to be more than the passive lending of books to those who walk in. Many provide meeting rooms for use by community groups. Libraries frequently sponsor lectures and forums, show educational motion pictures, and may actively participate in adult education programs.

Learned society libraries and special libraries. Scholarly libraries maintained by private bodies such as professional and learned societies usually specialize in one subject, and often have the finest collections in their specialty. Examples include the Royal Society of London and the Zoological Society of London. An increasingly popular modern variation of this type of library is the establishment of an academy of sciences, a network of research institutes covering most fields of learning. Particularly in socialist countries, these academies may be created and at least partly supported by the government. The libraries of learned societies and academies in the United States do not have an importance comparable to those in Europe. Prototypes of learned society libraries in the United States are those of the National Academy of Sciences and the American Antiquarian Society, which have national impact. There are also many regional and state scientific and historical societies that maintain significant libraries. Some of the scholarly functions of European societies are served by privately endowed libraries, such as the Newberry and the John Crerar libraries in Chicago, the Henry E. Huntington Library in San Marino, California, the Folger Shakespeare Library in Washington, D.C., and the Linda Hall Library in Kansas City, Missouri. In general, these libraries do not lend materials to individuals, but make their resources and services available to other libraries and to scholars through various means, including photocopying.

The term special library is at best ambiguous and is applied to libraries undertaking to serve only a limited subject interest; it may be the library of a corporation, a government department, or a division of a university. Special libraries are closely related to, and are often indistinguishable from, information centres. They tend to be nonconventional in their practices, creating, for example, their own rules for cataloging and classification.

BIBLIOGRAPHY. Although originally published in 1933, PIERCE BUTLER, *An Introduction to Library Science* (1961), remains one of the best explications of library science as an intellectual discipline. Another book that delves more deeply into the philosophical aspects of libraries is JESSE H. SHERA, *Libraries and the Organization of Knowledge* (1965). PAUL S. DUNKIN, *Cataloging U.S.A.* (1969), discusses technical aspects of library operation, but undertakes to emphasize the rationale of library cataloging practice. The Dewey Decimal Classification system is widely known and generally understood by library users, but JOHN P. IMMROTH, *A Guide to Library of Congress Classification* (1968), provides one of the first discussions of that system in depth. A good textbook on library selection and acquisition procedures may be found in MARY DUNCAN CARTER and WALLACE J. BONK, *Building Library Collections,* 3rd ed. (1969). The conflicting philosophies of library reference service are brought into focus in D.J. FOSKETT, *Information Service in Libraries,* 2nd ed. (1967); and in the somewhat older historical study of SAMUEL ROTHSTEIN, *The Development of Reference Services Through Academic Traditions, Public Library Practice, and Special Librarianship* (1955). A non-technical treatment of libraries and the newer technology may be found in PAUL WASSERMAN, *The Librarian and the Machine* (1964). Two influential books on library administration are JOSEPH L. WHEELER and HERBERT GOLDHOR, *Practical Administration of Public Libraries* (1962); and Louis R. WILSON and MAURICE F. TAUBER, *The University Library,* 2nd ed. (1956).

(R.H.Pa.)

Libya

The Libyan Arab Republic (in Arabic, al-Jumhūrīyah al-'Arabiyah al-Libiyah) is an independent Socialist republic of North Africa. It is bounded by the Mediterranean Sea on the north, Egypt on the east, The Sudan on the southeast, Niger and Chad on the south, and Tunisia and Algeria on the west. Largely composed of the Sahara, it covers an area of 675,200 square miles (1,748,700 square kilometres) and is one-fourth the size of the conterminous United States and almost twice the size of

Egypt The early 1970s population of about 2,084,000 was concentrated along the coast, where the de facto capital, Tripoli (Ṭarābulus), and Benghazi (Benghāzī), the de jure capital, are located.

Before the discovery of oil in the 1950s, Libya was poor in natural resources and severely limited by the climatic conditions of the Sahara. The country was almost entirely dependent upon foreign aid and the import of commodities necessary to the maintenance of the economy. Petroleum dramatically changed the economy, and Libya became one of the richest countries of the Middle East and Africa. Both the monarchy and the military government that took power in 1969 have been sensitive to the problems of dependence upon a single natural resource. The government has attempted to develop agriculture and industry with its vast oil revenues. It has also established a welfare state, which provides free medical care and education to all. For history, see NORTH AFRICA, HISTORY OF.

THE LANDSCAPE

Relief features. Libya is composed of basement rocks of Precambrian age (from 570,000,000 to 4,600,000,000 years old) that are overlain with marine and windborne deposits. There is also evidence of ancient volcanic activity. The major physical features are the Jabal Nafiisah (Arabic *jabal*, "mountains") and the associated Gefara plain in the northwest, al-Jabal al-Akhdar in the northeast, and the Saharan plateau. The Gefara plain covers about 10,000 square miles of Libya's northwestern corner. It rises from sea level at the coast to about 1,000 feet at the foothills of the Jabal Nafiisah. Composed of coastal cliffs, sand dunes, salt marshes, and steppe, the plain contains most of Libya's population and its largest city—Tripoli. The Jabal Nafiisah is a limestone massif that stretches for about 212 miles from west to east between the Gefara and the Sahara; it reaches altitudes of between 2,000 and 3,000 feet.

In the country's northeastern corner, al-Jabal al-Akhdar stretches for about 100 miles along the coast between al-Marj and Darnah. The limestone mountains rise steeply and stretch about 20 miles inland, attaining altitudes between 2,000 and 3,000 feet.

The desert and the plateaus

The Saharan plateau covers about 99 percent of Libya, making it truly a desert land. The sand-covered undulating plateau surface is broken by several physical features, including the al-Harūj al-Aswad, the al-Ḥammādah al-Ḥamrā', and the Tibesti mountains. Al-Harūj al-Aswad is a hilly basaltic plateau, which covers about 15,400 square miles in central Libya. It rises to about 2,600 feet and is crowned by volcanic peaks. The region is covered with angular stone fragments and boulders.

Al-Ḥammādah al-Ḥamrā' is a rocky plateau that lies behind the Jabal Nafiisah and covers an area of about 19,000 square miles. It contains bare rock outcrops and rises to 2,700 feet. An arm of the Tibesti stretches northward from the main massif in Chad. Picco Bette rises to 7,500 feet (2,286 metres) on the Libya–Chad border.

Drainage and soils. There are no perennial rivers. The numerous wadis (dry riverbeds) are filled by flash floods during the rains and quickly dry up or are reduced to a trickle. There is, however, extensive underground percolation (seepage) of water, and the numerous oases are watered by wells and springs that are fed by the underground water table. Along the coastal strip there are several salt lakes, or *sabkhah*s, formed by the ponding and evaporation of water behind coastal dunes, the capillary movement of water from the underground water table, or the collection of water where percolation is restricted. Principal salt lakes are those of Tāwurghā', Zuwārah, and the Banghāzī Plain.

On the western coast the gray-brown soils are highly saline. In the east, the soils of the Barce plain are light and fertile. The result of decomposition of limestone deposits, they contain potash and phosphoric acid. The rest of the country is covered by wind-eroded sand or stony desert.

Climate. The climate over most of the country is that of the hot, arid Sahara, but it is moderated along the coastal littoral by the Mediterranean Sea. The maritime influence is greater in winter, while the Saharan influence is stronger in summer. The large daily range of temperatures is especially notable in the Sahara; it is modified along the littoral, where the sea cools during the day and warms by night. Coastal cloud cover moderates winter coastal temperatures by limiting insolation (the delivery of direct solar radiation) by day and by acting as an insulator at night.

The warmest months are July and August, when Benghazi experiences average monthly temperatures of 72° to 85° F (22° to 29° C) and Tripoli has average temperatures of 62° to 86° F (17° to 30° C). The coolest months are January and February; Benghazi has winter monthly temperatures of 50° to 63° F (10° to 17° C) and Tripoli has 47° to 61° F (8° to 16° C). Al-‘Azīzīyah on the Gefara has recorded temperatures of about 136° F (58° C).

The effect of the *ghibli* wind

The amount of annual precipitation declines, and its variability increases, inland from the coast. Benghazi receives an annual average rainfall of 10 inches, and Tripoli receives 15 inches. In the maritime zone, however, the periodic high incidences of rainless periods in a year are sufficient to cause drought. In the Sahara, 200 consecutive rainless days in a year have been recorded in various areas, and the world's highest degree of aridity occurs at Sabhā. The dry climate is exacerbated by the ghibli, a hot, arid wind that blows from the south several times a year. It is preceded by a short lull in the prevailing winds, which is followed by the full force of the *ghibli*. The wind carries large quantities of sand dust, which turn the sky red and reduce visibility to less than 20 yards. The heat of the wind is increased by a rapid drop of relative humidity, which can fall from 80 to 10 percent within hours.

Vegetation and animal life. In years of good rainfall the northern plains are covered with herbaceous vegetation and annual grasses; the most noticeable of plants are the asphodel (an herb of the lily family) and jubule. The northern area of al-Jabal al-Akhdar—where the influence of the Mediterranean is most apparent—supports low and relatively dense forest (or macchia), in which the main trees are the juniper and lentisk. Annual plants are abundant and include species such as brome grass, canary grass, bluegrass, and rye grass. The forest becomes more scattered and stunted south of the mountain crest, and annual plants are less frequent. The Jabal Nafūsah has less plant life, and the natural vegetation of grassland lies between barren hills.

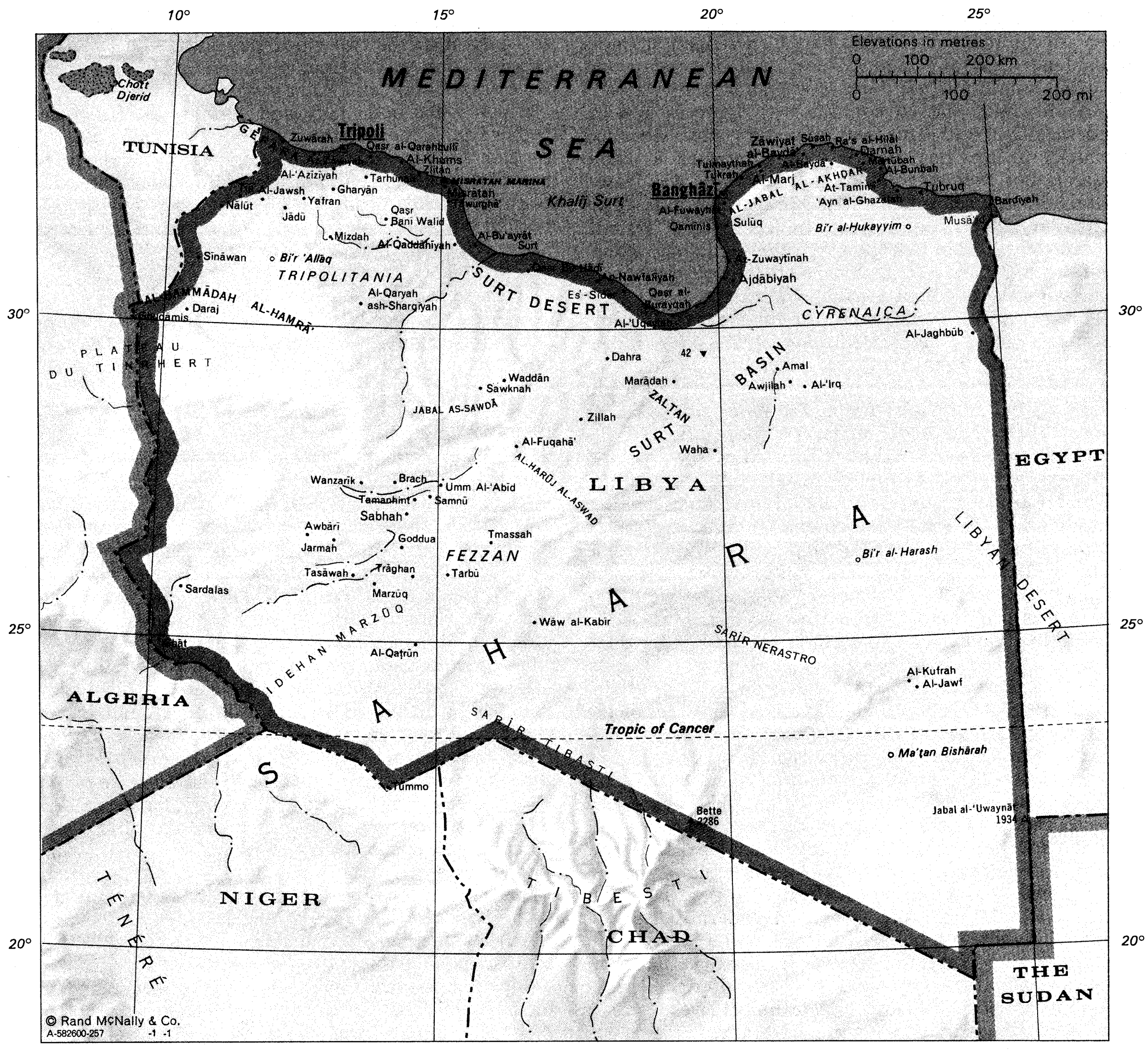
When rainfall is less than six inches, semidesert conditions prevail, and vegetation is sparse; isolated plants grow in generally barren pockets. The species most commonly found are saltwort (a plant used in making soda ash) and spurge flax (a shrubby plant), while goosefoot (an herb), wormwood (a woody herb), and asphodel also are widespread. Annual grasses grow in the rainy season, and leguminous plants appear in years of good rainfall. Although rainfall is extremely low in the true desert zone and the vegetation cover is scant, some of the plants of the semi-arid region penetrate into the occasional wadi (periodic river) valley, and date palms are grown in the southern oases.

Wild animals include desert rodents, such as the desert hare and the jerboa; hyenas; foxes, such as the fennec and the red fox; jackals; skunks; gazelles; and wildcats. The poisonous adder and krait are among the reptiles that inhabit the scattered oases and water holes. Insects are common; and the interior is crossed by migrating locusts, birds, and butterflies. Native birds include the wild ringdove, the partridge, the lark, and the prairie hen. Eagles, hawks, and vultures are common.

Tripolitania, Cyrenaica, and Fezzan

Traditional regions. The three former provinces of Tripolitania in the west, Cyrenaica in the east, and Fezzan in the southwest are coincident with Libya's traditional regions. There are deep geographical differences between the three areas, the centres of population of which are separated by hundreds of miles of barren and inhospitable country.

Economic and environmental factors accentuate strong feelings of provincial attachment. In Tripolitania (Arabic



LIBYA

MAP INDEX

Cities and towns

Ajddbiyah	30·48n	20·14e
al-'Azīzīyah	32·32n	13·01e
al-Bayḍā'	32·48n	21·45e
ai-Bu 'Ayrāt	31·24n	15·44e
ai-Bunbah	32·24n	23·08e
al-Fuqahā'	27·50n	16·22e
al-Fuwayhāt	32·04n	20·06e
al-'Irq	29·02n	21·33e
al-Jaghbūb	29·45n	24·31e
al-Jawf	24·11n	23·19e
al-Jawsh	32·00n	11·40e
al-Khums	32·39n	14·16e
al-Kufrah	24·10n	23·15e
al-Marj	32·30n	20·54e
ai-Qadddḥiyah	31·22n	15·14e
al-Qaryah ash-Sharqīyah	30·24n	13·36e
al-Qaṭrūn	24·56n	14·38e
al-'Uqaylah	30·16n	19·12e
Amal	29·25n	21·10e
an- Nawfalīyah	30·47n	17·50e
as-Sidar, see es'-Sider		
at-Tamīmī	32·20n	23·04e
Awbārī	26·35n	12·46e
Awjilah	29·09n	21·15e
'Ayn al-Ghazālah	32·10n	23·20e
az-Zāwiyah	32·45n	12·44e
az-Zuwaytīnah	30·58n	20·07e
Banghāzī (Benghazi)	32·07n	20·04e
Bardīyah	31·46n	25·06e
Bi'r al-Harash (oasis)	25·29n	22·11e
Bi'r 'Allāq (oasis)	31·02n	10·36e
Brach	27·32n	14·16e
Dahra	29·34n	17·50e
Daraj	30·09n	10·26e
Darnah	32·46n	22·39e
es'-Sider	30·38n	18·20e
Gharyān	32·10n	13.01 e
Ghāt	24·58n	10·11e
Ghuddmis	30·08n	9·30e
Goddua	26·26n	14·18e
Jādū	31·57n	12·01e
Jarmah	26·33n	13·04e
Marddah	29·14n	19·13e
Martūbah	32·35n	22·46e
Marzūq	25·55n	13·55e
Misrdtah	32·23n	15·06e
Mizdah	31·26n	12·59e
Musd'id	31·35n	25·03e
Nālūt	31·52n	10·59e
Qaminis	31·39n	20·03e
Qaṣr al-Burayqah	30·25n	19·34e
Qaṣr al-Qarahbullī	32·45n	13·43e
Qaṣr Banī Walīd	31·45n	14·01e
Qasr bū-Hādt	31·03n	16·40e
Ra's al-Hilāl	32·57n	22·10e
Ridotta Capuzzo, see Musā'id		
Sabhah (Sabhā)	27·03n	14·26e
Samnū	27·17n	14·53e
Sardalas	25·46n	10·34e
Sawknah	29·04n	15·47e
Sīnāwan	31·02n	10·36e
Sulūq	31·39n	20·15e
Surt	31·12n	16·35e
Sūsah	32·54n	21·58e
Tamanhint	27·13n	14·36e
Ṭarābulus, see Tripoli		
Tarbū	26·02n	15·10e
Tarhūnah	32·26n	13·38e
Tasāwah	25·58n	13·30e
Tāwurghā'	32·02n	15·09e
Tiji	32·01n	11·22e
Tmassah	26·22n	15·47e
Trāghan	25·59n	14·26e
Tripoli (Ṭarābulus)	32·54n	13·11e
Tubruq (Tobruk)	32·05n	23·58e
Tūkrah	32·32n	20·34e
Tulmaythah	32·42n	20·57e
Umm al-'Abid	27·31n	15·02e
Waddān	29·10n	16·08e
Waha	28·16n	19·54e
Wāw al-Kabir	25·20n	16·43e
Yafran	32·04n	12·31e
Zawiyat al-Baydd'	32·46n	21·43e
Zillah	28·33n	17·35e
Ẓlītan	32·28n	14·34e
Zuara, see Zuwārah		
Zuwārah	32·56n	12·06e

Physical features and points of interest

Akhdar, ai-Jabal al-, mountains	32·30n	21·30e
al-Harūj al-Aswad, hills	27·00n	17·10e
Barqah, see Cyrenaica		
Bette, peak	22·00n	19·12e
Bishārah, Ma'tan, oases	22·58n	22·39e
Bi'r al-Hukayyim, oases	31·36n	23·29e
Cyrenaica (Barqah), physical region	31·00n	22·30e
Fezzan (Fazzān), physical region	26·00n	14·00e
Gefara, *plain*	32·30n	11·45e
Hamrā', al-Ḥammādah al-, upland	30·22n	9·57e
Idehan Marzūq, dunes	24·30n	13·00e
Libyan Desert	26·00n	25·00e
Mediterranean Sea	33·00n	17·30e
Misrātah Marina, *port*	32·23n	15·13e
Nerastro, Sarir, desert	24·20n	20·37e
Sawdā', Jabal as-, hills	28·40n	15·30e
Surt, Khalij (Guif of Sidra)	31·30n	18·00e
Surt Basin	28·30n	19·30e
Surt Desert	31·00n	16·30e
Tibasti, Sarir, desert	24·15n	17·15e
Tinrhert, Plateau du	29·00n	9·00e
Tripolitania (Ṭarābulus), physical region	31·00n	15·00e
'Uwaynāt, Jabal al-, mountain	21·54n	24·58e
Zalṭan, hills	28·37n	19·46e

Tariibulus), the vast majority of the population is engaged in commercial farming based on irrigation; plots of land are generally held in individual ownership and are often small. On the Jabal Nafūsah, where water is less readily available, a sophisticated agrarian system, based on olive- and fruit-tree cultivation and associated livestock raising, has evolved. In Cyrenaica (Arabic Barqah), however, the traditional economy was based on nomadic and semi-nomadic pastoralism. Arable farming has largely been an adjunct of the pastoral system, with shifting dry-land cultivation rarely entailing sedentary farming. Land ownership is no longer exclusively tribal, but the system of tenancy contrasts sharply with that of ownership in the west. In the Fezzan (Arabic Fazzān), isolated irrigated farming at the oases represents a third traditional economic system.

Cultural differences between the provinces are also important. Tribal affiliations among the Arab groups are to smaller rather than larger units. The population of the west is far more cosmopolitan than that of the east and includes a higher proportion of people with Berber, Negro, and Turkish origins. Although Tripolitania came to uneasy terms with the Italian colonial regime, Cyrenaica was the scene of constant fighting and disruption; the commercial, administrative, and cultural impact of the Italians was therefore more appreciable in the west. Cyrenaica was profoundly affected by the teachings of the 19th-century Sanūsīyah, an Islāmic brotherhood, which had little influence in the west and south.

The landscape under human settlement. The most significant mode of life practiced in rural Libya is that of the sedentary cultivator. In the traditional oases most farmers rely on irrigation, and water is raised from shallow wells either by the animal-powered *dalū* (a goatskin bag drawn by rope over a pulley) or, increasingly, by electric or diesel pumps. Landholdings in the oases are small and fragmented; the average five or seven acres per farm are generally divided into three or four separate units. On the coastal lowlands farmers normally live on their land and often have rights to graze stock and undertake shifting grain cultivation. In both the east and the west, Arab farmers occupy large, formerly European estates, in which individual units range from 12 to 600 acres, and operations are conducted at a high technical level.

Nomadism is mainly a feature of the unirrigated region?, particularly al-Jabal al-Akhdar and its surrounding area in Cyrenaica. Nomadic groups subsist primarily on their holdings of sheep, goats, and camels but also participate in the shifting cultivation of cereals. In the east, the Bedouins move south as soon as pasture appears and remain there until the ephemeral grasslands die and necessitate the return to the northern hill lands.

The village was originally an alien institution to Libya's tribal organization. Since the first Turkish occupation in the 16th century, towns and villages were developed mainly as military posts or administrative centres by occupying powers. Many village sites have been occupied for centuries; smaller settlements often began as collecting centres for the nomadic tribes during their summer residence in the oases or hill pastures. Berbers in the west, however, are thought to have retained a more or less continuous thread of settlement in their fortified nucleated villages in the western Jabal Nafūsah. In the southern oases, the villages served both as defense posts for the scattered communities and as watering and provisioning points on the trans-Saharan caravan routes.

Urban settlement

Modern development has led to the expansion of villages into large towns and has fostered a movement of the rural population from the land to the centres of settlement. Villages now reflect a purely indigenous expression, and, except in isolated parts of the east, they are integrated into Libyan social and economic life.

The two main cities of Tripoli and Benghazi have expanded to the exclusion of virtually all others. The surrounding extensive garden suburbs, such as al-Fuwayhāt to the southeast of Benghazi and Giorgimpopoli to the west of Tripoli, are a collection of colourful and luxurious villas inhabited by the more wealthy nationals and aliens. Many old buildings have been replaced with new hotels, apartment blocks and offices. Shanty towns are also found near the two cities; and, although low-income housing is under construction, these areas present problems of sanitation and water supply.

Tripoli is the main port, and Tripoli and Benghazi together handle most of the country's maritime trade. Tripoli handles the bulk of the imports, particularly those associated with the oil industry and the booming trade in consumer goods. Tobruk (Tubruq), the third most important port, is located in a remote area of sparse population. There are many other lesser ports.

PEOPLE AND POPULATION

The tribe (qabilah) is the basis of the social order in Libya, and eight out of every nine persons reside in tribal domains. Tribal organization has persisted in spite of three centuries of successive occupation by Turkish, Karamanli, Italian, and British forces.

Population groups. The Berbers were the major original element in the population. Most Arab chroniclers divide them into two main groups, whose branches indigenous to Libya are the Luata, the Nefusa, and the Adassa. The Berbers lived in coastal oases and practiced sedentary agriculture. Most of them have been assimilated into Arab society except in the Jabal Nafūsah region, Awjilah, Hūn, Socra, and Zuwārah. The 50,000 Berbers speak their own Hamitic language but have adopted the Arabic alphabet. Many are bilingual in Berber and Arabic.

Origins of the Arab population

The first influence of the Arabs was felt during the invasions of the 7th century. The initial Arab incursion into Berber lands was essentially military and had little effect upon the composition of the population. The Banū Hilāl invasion of 1049 and succeeding attacks of the Banū Sulaym in the 11th century, however, were paramount factors in the ethnic character of the population. By the 20th century, about 97 percent of Libya's inhabitants were Arabic-speaking Muslims of mixed Arab and Berber descent.

The Banū Sulaym were composed of four main groups —the Banū Hebib, the Auf, the Debbab, and the Zegb. The Hebib settled in Cyrenaica, while the others went into Tripolitania. After the establishment of tribal groups, Libya underwent a period of disorder and tribal feuding, which was augmented by the incursion of other Arab adventurers from Egypt. Toward the close of the period of anarchy, the Debbab group took control of much of Tripolitania.

Several other social groups exist alongside the tribal unit. They are the Sharifs (holy tribes), who came originally from Fezzan; the marabouts (dervishes who are credited with supernatural powers), who infiltrated from Sakiet el-Hamra in Morocco; and the Koulouglis, who are descended from the Janissaries (elite Turkish soldiers). The Sharifs constitute a religious hierarchy that claims direct descent from Muhammad. Their blood relationship with the prophet gives them a powerful standing in Muslim society, where they are looked upon as holy men with divine powers of foresight. Extensive tracts of land are found under Sharif control in all the oases of western Libya.

Marabout tribes are descended from holy men who also claimed relation to Muhammad. Infiltration of marabouts from Algeria, Tunisia, and Spain began in the 14th century and continued intermittently for several centuries. The true marabouts founded their religious devotions upon an ascetic life manifested in their existence as hermits. In areas where their teachings and way of life made them acceptable to the local inhabitants, they settled and founded tribes pledged to the pure way of life.

Minority population groups

The 35,000 Koulougis are descendants of Janissaries, Berbers, and Christian slaves. Since Turkish times they have served as a secretarial class in several areas and are often concentrated in and around villages and towns. They speak Arabic and practice Islām.

The trans-Saharan trade in slaves, which continued during Turkish times, left a pronounced black element in many of the tribes, especially in the western littoral. There are about 35,000 blacks, most of whom live in Tripolitania and the Fezzan. Their languages are those of the central Sahara and the eastern Sudan; most also speak Arabic and have adopted Islām. Before World War II there were about 30,000 Jews in Libya, but their numbers have declined to little more than 6,000. It has been suggested that the Jewish peoples are descended from colonies that existed before the Hilālian invasion. They specialize in trade, commerce, and handicrafts and are concentrated in Tripoli and other urban centres.

Small groups of Tuareg tribesmen are found in the southwest, especially at Ghudiimis and Ghāt oases. Traditionally nomadic, they are gradually assuming a sedentary life style. Isolated Teda (Tebu) communities of the southeast are gravitating toward the north and al-Kufrah oasis in search of employment. The former Italian and Greek populations have declined to negligible proportions since 1969.

Demography. In 1972, Libya had a population estimated to be just over 2,084,000. The high rate of natural increase of 3.5 percent annually is well above the regional average. The crude birth rate is 43 births per 1,000 population, and the crude death late is 9 per 1,000. There are considerable regional variations, and isolated areas of the interior have high rates of infant mortality of about 300 deaths for 1,000 live births.

Libya, Area and Population

	area		population	
	sq mi	sq km	1964 census*	1972 estimate
Governorates (*muḥāfaẓāt*)				
al-Bayḍā'	6,600	17,100	...	121,000
al-Khums	9,700	25,200	...	182,000
Awbārī (al-Khalii)	278,000	720,000	...	107,000
az-Zāwiyah	2,800	7,200	...	219,000
Banghāzī	6,600	17,000	...	299,000
Darnah	39,600	102,600	...	112,000
Gharyān	58,000	150,300	...	241,000
Misrltah	57,000	147,600	...	173,000
Sabhā	215,800	558,900	...	89,000
Ṭarābulus	1,000	2,700	...	541,000
Total Libya	675,200†	1,748,700	1,564,000	2,084,000

*De jure. †Figures do not add to total given because of rounding.
Source: Official government figures.

Before the 1960s there was substantial emigration of Libyans to Tunisia and Egypt in search of employment. Recent economic changes have tended to reverse the permanent flow of Libyans abroad. Temporary emigration for health, commercial, and educational purposes has become common, but the major flow is that of expatriate Libyans returning from Tunisia and Chad. In the late 1960s there was a slightly higher rate of immigration than of emigration.

Libya is one of the most urbanized countries of the Near and Middle East. Over one-fourth of the population live in cities with a population of more than 100,000; more than one-half reside in centres of more than 5,000 inhabitants; and one-third is to be found in towns of more than 2,000. Possibly the outstanding feature of the urban situation is the concentration of two-thirds of the urban dwellers in Tripoli city proper (380,000) and Benghazi (279,000). Urbanization has been achieved at a rate of growth of 25 percent between 1954 and 1966 and at an accelerated rate since then. It has taken place within limited bounds; with the exceptions of al-Baydii' and al-Marj, no new settlements of stature have emerged. Most of the new urban settlers are of rural origin, and rural population densities are static or are in decline.

Besides Tripoli and Benghazi, there are 12 large towns. In the west Gharyiin, al-Khums, Misrātah, Tājūrā', Siiq al-Jum'ah, Zanziir, and Zāwiyah are the major centres of more than 5,000 inhabitants. In the east Ajdābiyah, al-Marj, al-Bayḍā', Darnah, and Tobruk are of the same rank.

Demographic trends

The population was expected to continue to grow at a rate of 3.5 percent or more through the 1970s as birth rates remain steady and infant survival is more assured. Mortality rates are also likely to continue to show rapid decrease. Movement from rural areas to the cities seems certain to be a feature of the 1970s. The major losses will probably occur in isolated areas such as the Fezzan and the western Jabal Nafiisah rather than in the more developed areas of northern Libya. Given political stability and positive governmental support for agriculture, farming will remain an important occupation for many Libyans. It is foreseen, however, that Tripoli and Benghazi will further enhance their primacy.

THE NATIONAL ECONOMY

Natural resources. Petroleum is Libya's most important mineral resource. First discovered in 1956 near the Algerian border, it has since been located mainly in the Surt Basin. The major oil fields are Zalṭan, Amal, and Intisar A in Banghiizi *muḥāfaẓah* (governorate); Dahra field in Misrātah *muḥāfaẓah;* and Sarir field in Darnah *muḥāfaẓah.* Deposits have been located near Ghudiimis on the western border, Marziiq in the southwest, and al-Kufrah oasis in the southeast. In 1971 proven oil reserves totalled 28,000,000,000 barrels. Libyan crude oil is low in sulfur content and therefore causes less corrosion and less pollution than most crude oils. The deposits are associated with natural gas.

There are important deposits of natron (hydrated sodium carbonate) in the Fezzan and of potash in the Surt Desert near Mariidah. The iron ore in the Fezzan is low in iron content, and development is hampered by the field's distance from the coast. Marine salt is produced in Tripolitania, where there are also small deposits of gypsum, manganese, and lignite coal. Sulfur has been found in the Surt Desert, and there are scattered deposits of chalk, limestone, and marble.

The arid climate supports few biological resources except for the grasslands of al-Jabal al-Akhdar and the Jabal Nafūsah, which are valuable for grazing. There are no hydroelectric resources, and oil represents the only domestic means of producing thermal electricity.

Sources of national income. Agricultural production has declined since the 1960s because of the exodus from rural areas to the cities. The oil industry has attracted a large amount of labor from the farms, and foodstuffs are increasingly imported. Cereals are the major crops throughout the country. Barley is the chief cereal grown because it adapts well to different climates and soils. Wheat is grown primarily on the eastern and western plateaus, and sorghum is raised in the Fezzan. Olive plantations were introduced by the Italians on the Gefara and on the Jabal Nafiisah, and there are smaller olive patches in the east. Orchards of almonds, citrus fruit, apricots, and figs occur on small and large farms and on small, crowded plots in the oases. Dates are the principal crop of the southern oases. Grapes, broad beans, and peanuts are also grown. Tobacco is raised in Tripolitania.

Animal husbandry is important in Cyrenaica, where the herds are raised on communal grazing lands. Livestock includes sheep, goats, cattle, camels, horses, mules, and donkeys. Animals are raised for their milk, meat, and hides or for their services as a means of transportation. Cattle often serve as draft animals. A small amount of milk is produced commercially, and commercial poultry farms are developing around the larger cities.

Before the 1950s, the only wooded area in Libya was the region of scrub brush on al-Jabal al-Akhdar. Since then, the government has launched a massive afforestation program. Between 1957 and 1964, 27,000,000 acacia, eucalyptus, cypress, cedar, and pine trees were planted in Tripolitania.

There is little demand in Libya for fish, and most fishing is carried out off the Tripolitanian coast by Libyan, Tunisian, Greek, and Maltese fishermen. The catch includes tuna, sardines, and red mullet. Sponge beds are also important. The sponges are harvested mainly by Greeks who are licensed by the Libyan government.

The oil industry

Libya constitutes the oil-bearing region most convenient to western Europe. By 1970 oil accounted for over 99 percent of the country's export earnings and over 80 percent of the government's planned expenditures. Petroleum, however, is a diminishing resource, and Libya has a dangerous dependence upon the industry.

Almost the entire oil industry is held by foreign companies. Unlike other oil-producing nations, Libya has not concentrated on one or two large companies but has licensed more than 40 concessionaries, mainly from the United States, the United Kingdom, and West Germany. Profits are divided equally between Libya and the oil companies, and taxes are paid on the posted price for oil. Competition is high and companies are encouraged to engage in development projects as part of their bids for concessions. If oil is discovered, some companies have offered to pay extra royalties, build refineries and petrochemical plants, sponsor agricultural development, and furnish scholarships to Libyan students.

The first pipeline was constructed from the Zalṭan field to Mars2 al-Burayqā' in 1961. Since then two lines have been built from Dahra to as-Sidar and to Ra's al-Uniif, and other pipelines connect the Tobruk field to Marsā al-Hariqah and the Intisar A field to az-Zuwaytinah. A natural-gas pipeline runs parallel to the oil pipeline from Zalṭan. The gas liquefaction plant at Marsā al-Burayqā' is the world's largest.

Libya produced more than 152,000,000 tons of oil in 1971 and is the seventh largest producer of oil after the United States, the Soviet Union, Saudi Arabia, Iran, Venezuela and Kuwait. Sales to Europe have been enhanced by the closure of the Suez Canal after the Arab-Israeli War of 1967. About 2 percent of the Libyan labour force is employed by the oil industry, along with about 4,000 aliens. It has been estimated that the industry will not be able to employ more than 5 percent of Libya's work force, even when fully developed.

Mining of other mineral resources is little developed. Natron is mined, and marine salt is produced in substantial quantities. Concessions have been granted for the exploitation of iron ore and potash, but mining had not begun by the early 1970s. Gypsum, limestone, chalk, and marble are quarried for the construction industry.

Industrial development

Industrial development is hampered by the lack of skilled labour, the limited domestic market, the poor transport system, and inadequate water and natural resources. Most factories are located in Tripoli and Benghazi and are managed by Arabs. The industrial work force is small: more than 90 percent of the factories employ fewer than 100 persons, and more than 50 percent employ fewer than 10. Most factories are engaged in the manufacture of processed food, beverages, and leather. The small textile industry is almost totally dependent upon imported raw materials. There are also oil-related industries, which produce steel drums, tanks, pipe fittings, ahd housing and office buildings.

The government also contributes to manufacturing. The Libyan Tobacco Monopoly processes all tobacco and controls its import and export. There are also government monopolies for salt and esparto grass. Other ventures include the production of processed food, leather, gypsum, rugs, and mineral water.

The production of electricity for public consumption is a government monopoly. There are also private plants, such as the 25,000-kilowatt facility built by an oil company at Marsā al-Burayqah. In 1971 the total installed capacity was about 300,000 kilowatts, more than half of which was located in the west. All electric plants are powered by oil.

Financial services are headed by the Bank of Libya, which supervises the banking system, regulates credit and interest policies, and promotes the transformation of foreign banks into Libyan institutions. The Industrial and Real Estate Bank extends credit to private and governmental operations and the National Agricultural Bank gives credit and guidance to cooperatives and farmers. There are also four foreign commercial banks from the United Kingdom, Italy, and Jordan.

Foreign trade

Since 1963 Libya has enjoyed a favourable balance of trade. Almost all of its exports are represented by oil, but agricultural products and hides and skins are also exported. Major customers for Libya's products are Italy, West Germany, the United Kingdom, France, and The Netherlands. Imports consist of equipment for the oil and construction industries, farm machinery, consumer goods, and agricultural products. Most imports are purchased from Italy, the United States, the United Kingdom, West Germany, and France.

Management of the economy. The Libyan economy may be characterized as one of foreign and private initiative and cautious government control. The administration has maintained control of the concessions granted to the foreign-dominated oil industry but, apart from British Petroleum (nationalized in 1972), has resisted nationalization. It has attempted to promote private industrial and agricultural development since 1966 by the granting of interest-free loans. The shortage of capital and technical knowledge in the private sector, however, has prompted the government to enter directly into industrial endeavours.

Rapidly growing oil revenues are the main source of the government's income. Other revenues are import duties, direct taxes, and profits from government monopolies. The income tax law promulgated in 1969 includes fixed-percentage qualitative taxes and progressive income taxes on individuals and corporations.

Trade unions are weak, and their membership is small. Labour regulations are put forth by the government. In practice, most labour disputes are handled directly between the employees and the owners, but larger problems are mediated by the government.

Unions are organized by craft and by industry; the strongest organizations are the two oil workers' unions. There are three labour federations—the National Federation of Trade Unions, the Professional Workers Federation. and the Federation of Trade Unions. Their 37,000 members are largely skilled and semiskilled workers.

Economic policies and problems

Economic policy emphasizes the development of the economic infrastructure and industry. There is increased government participation in such activities as the construction of roads, schools, housing, and ports. The development of Libyan-owned enterprises is promoted to combat the influence of the foreign-owned oil companies. It is possible that government trading corporations will be established in partnership with foreign capital.

The greatest economic problem is the shortage of skilled labour. Educational and training facilities must be expanded and the towns must be developed so as to stem the migration of workers from the rural areas to the cities. Because of the influx of money from oil, inflation has increased since the 1960s. Prices for food, housing, and services have been continually rising. To curb inflation, Libya needs more workers rather than more money.

Transportation. There is a total of more than 8,900 miles of roads, of which almost 3,800 miles are paved. The main artery is the 1,100-mile coastal highway between the borders of Tunisia and Egypt, which has branch roads to the coastal towns. The Sabhā road runs from the coastal highway at al-Qaddāḥīyah south and southwest to Ghāt near the Algerian border. There are several roads between the coast and the major oil fields, and a road is planned to run south from Sabhā to the borders of Chad and Niger. The two railroads that served Tripoli and Benghazi were closed in the 1960s, and no new lines are planned.

Tripoli and Benghazi are the main seaports; they handle all imports and exports except crude oil. Petroleum is shipped from as-Sidar and Marsā al-Burayqah. Misrātah, Zuwārah, and al-Khums are to be developed as fishing ports, and Darnah and Tobruk are also being improved. The two international airports are Tripoli International Airport and Banīnah outside of Benghazi. Domestic airfields include those at Sabhā, al-Bayḍā', Ghudāmis, and Ghāt. The Libyan Arab Airlines and foreign airlines operate domestic flights and services to Europe, Egypt, Lebanon, and Tunisia. There are also domestic flights operated by the oil companies.

ADMINISTRATION AND SOCIAL CONDITIONS

The government. In September 1969 the monarchy of Idris I was overthrown and the constitution suspended in a military coup d'etat. Since then, the country has been governed by the 12-member Revolutionary Command Council and by Col. Muammar al-Qaddafi, who is head of the Council, prime minister, minister of defense, and commander in chief of the armed forces. The Council appoints the 16-member Cabinet; there is no legislature.

Local government

The country is divided into the ten *muḥāfaẓāt* (governorates) of Ṭarābulus, Banghāzī, Sabhā, al-Bayḍā', Gharyān, Misrātah, az-Zāwiyah, al-Khums, Darnah, and al-Khalij (Surt, Ajdābiyah, and al-Kufrah oases). Each district is headed by a governor (*muḥāfiẓ*), who is appointed by the Revolutionary Command Council and is responsible for the administration of local services and utilities, security, and the supervision and control of municipalities in his district. The governor is assisted by a local council that is composed of the directors of the departments representing the 9 central ministries, 18 elected members, and 4 members chosen from the local communities of the Arab Socialist Union. The councils participate in development planning, which is conceived and supervised by the central ministries.

The *muḥāfaẓāt* are divided into about 30 *baladiynt,* which are further broken down into the smallest administrative units, or *maḥallāt.* Each tribe is headed by a sheikh or an imam (religious leader). One of the major functions of the regional administrative system is to increase national identification with Libya.

Both men and women have the right to vote, but there have been no elections since 1969. There were no political parties until the formation of the Arab Socialist Union in 1971.

The judiciary is officially independent from the government. It consists of two codes of justice and court systems for traditional Islāmic law (Shari'ah) and secular law. In practice the Shari'ah courts handle civil matters, whereas the courts of first instance and courts of appeal of the secular system deal with criminal matters. The Supreme Court serves as a final court of appeal for both systems and has the power of judicial review. Judges are appointed for life but must retire at 65 years of age.

The armed forces

The armed forces consist of a 20,000 man army, a 2,000 man navy, and a 3,000-man air force. Since 1969, the armed forces have been controlled by Colonel Qaddafi and a coterie of young officers and have played an increasingly political role in government.

The 11,000-man police force, the Security Force, is divided into ten regional divisions. They are responsible for the maintenance of law and order and assistance in criminal investigation.

Social conditions. Public education is free and primary education is compulsory for both boys and girls. Arabic is the language of instruction at all levels. The few private schools are attended largely by aliens. About 85 percent of the school-age population attends classes.

The school system is composed of a six-year primary level, a three-year intermediate and vocational level, and a three-year secondary and advanced vocational level. There are also about 100 Qur'ānic schools, which are financed by the government.

Higher education is offered by the secular state institutions of the University of Libya. Advanced religious training is obtained at a branch of the university at al-Bayḍā'. Libyan students also study abroad in Egypt, Europe, and the United States.

In order to increase the literacy rate from its present low level of 30 percent, the government has also sponsored an adult educational program.

The chief health problems are tuberculosis, trachoma, and schistosomiasis (a parasitic infestation of the liver or intestines). The incidence of malaria has declined, but gastroenteritis and tetanus remain major diseases.

Medical and hospital care and medicines are free. The country has a total capacity of one hospital bed for every 260 persons. Most care is available in hospitals as there are few outpatient or specialized-care facilities. Facilities are adequate only in Tripoli and Benghazi, but there are a number of health centres in rural areas.

Most medical personnel are aliens, and the training of Libyan women as nurses is hampered by religious custom. In 1971 there was only one doctor for every 692 persons.

Welfare programs

The National Social Insurance Institute operates social security programs. Workers covered by government insurance programs receive medical examinations and treatment, maternity benefits, and dental care. There are also old-age pensions and payments for incapacity or death as a result of work-related accidents.

Housing facilities range from the Bedouin tent to the luxury apartment. About half the population occupies Mediterranean-style row houses in the towns and cities, while another 43,000 persons live in single homes or apartments. The oil economy has exaggerated the growth of slums and shanty towns in the urban areas.

A minimum wage was established by the government in 1967 and revised in 1969, but wages vary greatly. Rates are generally lower in the rural areas than in the cities, and skilled workers may earn 600 percent more than unskilled labourers. The per capita income, however, has risen from an equivalent of U.S. $100 annually to $1,979 in 1971.

Life styles in both town and village are strongly influenced by tribal affiliations and the close identity with the family. The universal group in the tribal system is the *bayt* (tribal subgroup), which may include three or four generations within its organization. All of the families of a tribe claim direct descent from a single ancestor, usually the descendant of the founder of the larger tribe. Organization is patriarchal; the eldest or most influential male is invested with supreme authority in the conduct of all affairs that relate to other similar groups. The extended family remains important even in the largest urban areas, and it is not uncommon for the eldest male to collect all earnings of the family and to allocate funds irrespective of the individual contributions made. In this situation, social contacts take place within the family context, and marriage is often arranged by the senior male.

There are also economic divisions of the population. In 1969, 10 percent of the population received more than half of the national income. The elite controlled the country through its economic advantage and family influence.

CULTURAL LIFE

Libyan culture centres on folk art and traditions, which are highly influenced by Islām. The traditional arts of weaving, embroidery, metal engraving and leatherwork rarely depict people or animals because of the Islāmic prohibition against such representation. The dominant geometric and arabesque designs are best presented in the stucco and tiles of the Karamanli and Gurgi mosques of Tripoli. Surviving traditions are represented by festivals, horse races, and folk dances. Musical styles are basically Arabic and rely upon improvisation and embellishment of melody; major musical instruments are the cane pipe and drums. Nonreligious literature has developed largely since the 1960s; it is nationalistic in character but reveals Egyptian influences. The arts are supported by the government through the Ministry of Information, the Ministry of Education and National Guidance, and the al-Fikr Society, a group of intellectuals and professionals. Government subsidies are essential because national awareness of the fine arts and folklore is weak and needs encouragement.

Cultural institutions

Libraries include the Government Library and the Archives in Tripoli, the Public Library in Benghazi, the university libraries, and those of the United States and Egyptian cultural centres in Tripoli. The Department of Antiquities is responsible for the Archaeological Museum, the Leptis Magna Museum of Antiquities, the Natural History Museum, and the Sabratha Museum of Antiquities, all in former Tripolitania, and the archaeological sites of Ptolemais and Appolonia in Cyrenaica. The Sabhā Museum contains exhibits of ancient remains of the former Fezzan region.

Although the press is officially independent, there is considerable governmental control. The six daily newspapers, six weekly newspapers, and seven magazines published in 1971 were all printed in Arabic. All were printed in Tripoli or Benghazi; one newspaper and six magazines were owned by the government and the rest were independent.

Broadcasting is owned and controlled by the government. Radio Libya has two radio stations at Tripoli and one each at Benghazi and al-Bayḍā'. Programs are broadcast in Arabic, English, and French and include cultural programs and literacy lessons for adults. The television stations in Tripoli, Benghazi, and Sabhā broadcast in Arabic and English for three hours daily.

Prospects for the future. At the time of independence in 1951, Libya was considered to be one of the poorest countries in the world, with few known natural resources, a difficult natural environment, and a population that was small, poor, and illiterate. This unpromising picture was dramatically changed with the discovery of oil; since 1963 oil revenues have provided a new source of capital for economic and social development. The main problem of the oil boom is how best to use the large sums of money available to promote the well-being of the nation as a whole.

Only a minority of the population can expect to earn its living directly through the oil and related industries. Other sources of livelihood must be developed to provide employment and income for the majority, whose needs will not automatically be met by the growth of oil production. The government has devoted large sums of money to the development of agriculture and to the raising of livestock, which are the traditional mainstays of the Libyan economy. It is believed that—with good administration, better education, and increased capital investment—a much larger production can be obtained from the land, and that good opportunities exist for the development of fisheries and tourism.

BIBLIOGRAPHY. R.W. HILL, ***A Bibliography of Libya*** (1959), a reference guide.

Geography: W.B. FISHER, ***The Middle East: A Physical, Social and Regional Geography,*** 5th ed. (1963), includes a geographical account of Libya; R.J. HARRISON CHURCH ***et al., Africa and the Islands,*** rev. ed. (1965), contains a balanced presentation of Libyan geography.

History: O. BATES, ***The Eastern Libyans*** (1914), an excellent study of the ancient Libyans and their way of life; ETTORE ROSSI, ***Storia di Tripoli e della Tripolitania: dalla conquista araba al 1911*** (1968), the best historical record of Tripoli and Tripolitania from 642 to 1911; MAJID KHADDURI, ***Modern Libya: A Study in Political Development*** (1963), the first political study of Libya in the postwar era; J. DES-[illegible]IS, ***Le Djebel Nefousa, Tripolitaine*** (1935), an excellent study on the Berber society in the Tripolitanian mountains; E.E. EVANS-PRITCHARD, ***The Sanusi of Cyrenaica*** (1949), a study of the Sanūsīyah order and the Bedouin of Cyrenaica; UNIVERSITY OF LIBYA, FACULTY OF ARTS, ***Libya in History, Historical Conference 16–23 March, 1968*** (1970), the first original work on Libyan archaeology and history.

Economics: B.H. HIGGINS, ***Economic Development: Principles, Problems and Policies*** (1959), a discussion of Libya as a case study of an underdeveloped country before oil discovery; BANK OF LIBYA, RESEARCH DIVISION, ***Inflation in Libya*** (1961), covers the economic impact of oil company expenditure and aid; the BANK OF LIBYA, ***Economic Bulletin*** (monthly), offers good coverage of the overall economy; INTERNATIONAL BANK FOR RECONSTRUCTION AND DEVELOPMENT, ***The Economic Development of Libya*** (1960), appraises Libya's economic progress; LIBYA, MINISTRY OF PLANNING AND DEVELOPMENT, ***Housing in Libya, 2*** vol. (1964), discusses housing and transport problems; J.A. ALLAN, K.S. MCLACHLAN, and EDITH PENROSE, ***Agriculture and the Economic Development of Libya*** (1972).

Natural resources and geology: M. ANDRE BRICHANT, ***A Broad Outline of Geology and Mineral Possibilities in Libya,*** UN Report No. 7 (1952); F. FOWLAND and E. ROBB, "Survey of Land Resources in Tripolitaina" in ***British Military Administration of Occupied Territories During the Years 1941–43*** (1945); G.W. MURRAY, "The Water Beneath the Egyptian Western Desert," ***Geogrl.*** J., 118:443–452 (1952); A. FANTOLI, ***Le pioggie della Libia con particolare riguardo alle zone di avvaloramento*** (1952), a detailed study of climate, containing statistics, maps, and diagrams; LIBYA, MINISTRY OF PLANNING AND DEVELOPMENT, ***Agriculture in Libya and a Plan for Its Development*** (1966); UNIVERSITY OF LIBYA, FACULTY OF SCIENCE, ***Geology of Libya*** (1972).

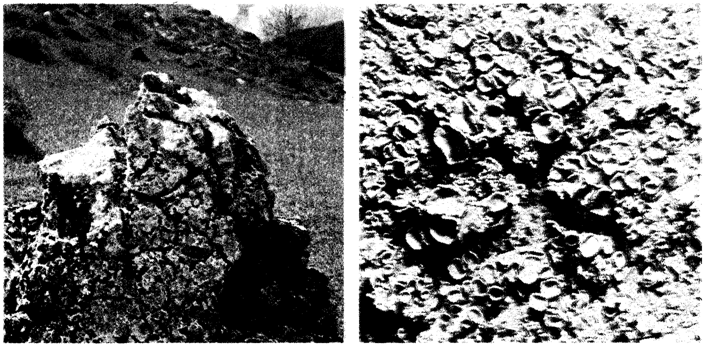
Statistics: The annual LIBYAN ARAB REPUBLIC, ***Statistical Abstract,*** published by the Census and Statistical Department, gives general statistical information.

(Mu.B.)

Lichen

Lichens, associations of fungi and algae, are among the most remarkable and successful examples of mutualism, a type of symbiosis, to be found in nature. Often confused with mosses, lichens appear to be single plants. Under a microscope, however, the association is seen to consist of millions of cells of algae (called the phycobiont) woven into a matrix formed of the filaments of the fungus (called the mycobiont). Approximately 15,000 different kinds of lichens, some of which provide forage for reindeer and products for man, have been described. Some lichens are leafy and form beautiful rosettes on rocks and tree trunks; others are filamentous and drape the branches of trees, sometimes reaching a length of nine feet. At the opposite extreme are those smaller than a pinhead and seen only with a magnifying lens. Lichens grow on almost any type of surface; bark, soil, rocks, and stones are common habitats. Glass, leaves, sun-bleached animal skulls, and the backs of certain weevils are unusual but not unknown habitats.

(Left) Vernon Ahmadjian, (right) Walter Dawn



Xanthoria, a foliose lichen: (Left) species on a limestone rock and (Right) closeup of X. *parietina.*

GENERAL FEATURES

Moisture requirements

Distribution. Almost all areas of the world contain lichens. They are especially prominent in bleak, harsh regions where few plants can survive. They grow farther north and farther south and higher on mountains than most plants. The body of a lichen, called a thallus, has one of several characteristic growth forms (crustose, foliose, or fruticose—see below). Each form has a different moisture requirement for growth and a favoured means for obtaining the moisture it needs. The relationship between growth and moisture requirement separates growth forms according to environmental habitats. Crustose thalli, which resemble a crust closely attached to a surface, are drought resistant and well adapted to dry climates. They prevail in deserts, Arctic and alpine regions, and ice-free parts of Antarctica. Foliose, or leafy, thalli grow best in areas of frequent rainfall; two foliose lichens, ***Hydrothyria venosa*** and ***Dermatocarpon fluviatile,*** grow on rocks in freshwater streams of North America. Fruticose (stalked) thalli and filamentous forms prefer to utilize water in vapour form and are prevalent in humid, foggy areas such as seacoasts and mountainous regions of the tropics. When growth conditions are optimal, lichens, despite their slow growth rates, cover extensive areas. Reindeer pastures in the far north consist of dense mats of ***Cladonia*** and ***Stereocaulon,*** and trees in the mountains or along seacoasts may become blanketed with ***Alectoria*** and ***Usnea.*** Few lichens grow near cities and industrial areas because of the extreme dryness and prevailing air pollution of these regions. Notable exceptions exist, however; and in England ***Lecanora conizaeoides*** actually is confined to areas of high pollution, apparently a product of the industrial age.

An example of the versatility of lichens is ***Verrucaria serpuloides,*** which is continuously submerged in the cold coastal waters of Antarctica. Other marine species of ***Verrucaria*** form black crusts on coastal boulders in intertidal zones. Some lichens inhabit the world's driest deserts, where they grow mostly underground, obtaining light and moisture through a window-like opening on the surface of the earth. A small group of desert lichens, not firmly attached, are distributed freely by wind. One such wandering lichen, *Lecanora esculenta,* is reputed to be the manna that in ancient days fell from the skies and served as a food source for man and his domestic animals.

Types. Although basidiomycetous fungi form only a few unions with algae, one well-known basidiolichen, *Cora pavonia,* is widely distributed in the tropics and has the blue-green alga *Scytonema* as the phycobiont. Several species of another basidiomycete (*Omphalina*) commonly form lichen unions with the green alga *Coccomyxa* in northern and alpine regions. The majority of mycobionts are placed in a single group of Ascomycetes called the Lecanorales, which are characterized by an open, often button-shaped fruit called an apothecium. The remaining mycobionts are distributed among various fungal groups —*e.g.*, Sphaeriales, Caliciales, Myrangiales, Pleosporales, and Hysteriales. Although there are various types of phycobionts, most of them also belong to a single group; *i.e.*, half the lichen associations contain species of *Trebouxia,* a single-celled green alga. Representatives of **26** genera of algae participate in lichen associations; 17 are green algae, eight are blue-green, and one is yellow-green. *Coccomyxa* and *Trentepohlia* are common green phycobionts; the latter, along with *Phycopeltis,* is especially prevalent in tropical lichens. Among the blue-greens, *Nostoc* and *Scytonema* are common phycobionts.

Hell Heikkila

Omphalina and *Coriscium.*

IMPORTANCE

Man has used lichens as food, as medicine, and in dyes. A versatile lichen of economic importance, *Cetraria islandica,* commonly called Iceland moss, is still being used widely by man more than **230** years after it was first prescribed for coughs and chest ailments. Nutritious and easily digested, Iceland moss sometimes is used either as an appetite stimulant or as a foodstuff in reducing diets; it has also been mixed with bread. Iceland moss also has been used to treat diabetes, nephritis, and catarrh. Lichens have little medical value. A compound (usnic acid) present in many lichens is processed in several countries as a drug for treating superficial wounds and infections. Although many lichen compounds have the ability to inhibit or destroy bacteria, molds, and viruses, lichens have not yet been used extensively to produce antibiotics.

Lichens as dye sources

Lichens are well known as dye sources. Dyes derived from them have an affinity for wool and silk and are formed by decomposition of certain lichen acids and conversion of the products. One of the best known lichen dyes is orchil, which has a purple or red-violet colour. Orchil-producing lichens include species of *Ochrolechia, Roccella,* and *Umbilicaria.* Litmus, formed from orchil, is widely used as an acid-base indicator. Synthetic coal-tar dyes, however, have replaced lichen dyes in the textile industry, and usage of orchil now is limited to its use as a food-colouring agent and an acid-base indicator. A few lichens (*e.g., Evernia prunastri*) are used in the manufacture of perfumes.

Caribou and reindeer depend on lichens for two-thirds of their food supply. In northern Canada, an acre of land undisturbed by animals for **120** years or more may contain **560** pounds of lichens; some forage lichens that form extensive mats on the ground are *Cladonia alpestris, C. mitis, C. rangiferina,* and *C. sylvatica.* Arboreal lichens such as *Alectoria, Evernia,* and *Usnea* also are valuable as forage. An acre of mature black spruce trees in northern Canada, for example, may contain over **600** pounds of lichens on branches within ten feet of the ground. Extensive consumption of lichens by animals in northern regions poses a problem, however, because lichens can absorb and retain many compounds. In addition to harmless minerals that accumulate in the thallus, radioactive derivatives of strontium and cesium also are retained in high concentrations. These isotopes, ingested by caribou and reindeer feeding on radioactive lichens, may reach dangerous levels in man when the animals are eaten by Eskimos and other human inhabitants of northern regions.

Although it has been said that lichens break down rocks and build soil on which other plants can grow, there is little evidence to support this idea. Lichen acids can extract metals from rocks, but the extent of this chemical weathering is limited. Windblown dust and soil that accumulate beneath lichen thalli growing on rocks probably have contributed to the idea that lichens are soil formers.

NATURAL HISTORY

Fungal–algal recombinations

Propagation and reproduction. Although the fungal symbionts of many lichens have fruiting structures on or within their thalli and may release numerous spores that develop into fungi, indirect evidence suggests that natural unions of fungi and algae occur only rarely among some lichen groups, if indeed they occur at all. Frequent occurrence of such unions should result in extensive genetic variability among lichen species, since fungal spores are genetically diverse, In addition, free-living algal populations, capable of serving as phycobionts, should be readily available. Evidence indicates, however, that neither condition exists in nature. Many variations that occur among lichens are induced by environmental factors affecting similar populations. In addition, free-living, potential phycobionts are not widely distributed; *e.g.*, despite repeated searches, free-living populations of *Trebouxia* have not been found. This paradox, an abundance of fungal spores and a lack of algae capable of forming associations, implies that the countless spores produced by lichen fungi are functionless, at least so far as propagation of the association is concerned. Some phycobionts —*i.e.*, species of *Nostoc* and *Trentpohlia*—*can* exist as free-living populations, so that natural reassociations could occur in a few lichens.

Some lichens have solved or bypassed the problem of recombination. In a few lichens (*e.g., Endocarpon, Staurothele*) algae grow among the tissues of a fruiting body and are discharged along with fungal spores; such phycobionts are called hymenial algae. When the spores germinate, the algal cells multiply and gradually form lichens with the fungus. Other lichens form structures, especially soredia, that are effective in distributing the association. A soredium, consisting of one or several algal cells enveloped by threadlike fungal filaments, or hyphae, may develop into a thallus under suitable conditions. Lichens without soredia may propagate by fragmentation of their thalli. Many lichens develop small thalloid extensions, called isidia, that also may serve in asexual propagation if broken off from the thallus.

In addition to these mechanisms for propagation, the individual symbionts have various methods of reproduction. Ascolichens—*i.e.*, lichens in which the mycobiont is an ascomycete—for example, form fruits (ascocarps) similar to those of free-living ascomycetes, except that the mycobiont's fruits are capable of producing spores for a longer period of time. The algal symbiont within the lichen thallus reproduces by the same methods as its free-living counterpart.

Symbiotic interactions. Most lichen phycobionts are penetrated to varying degrees by specialized fungal struc-

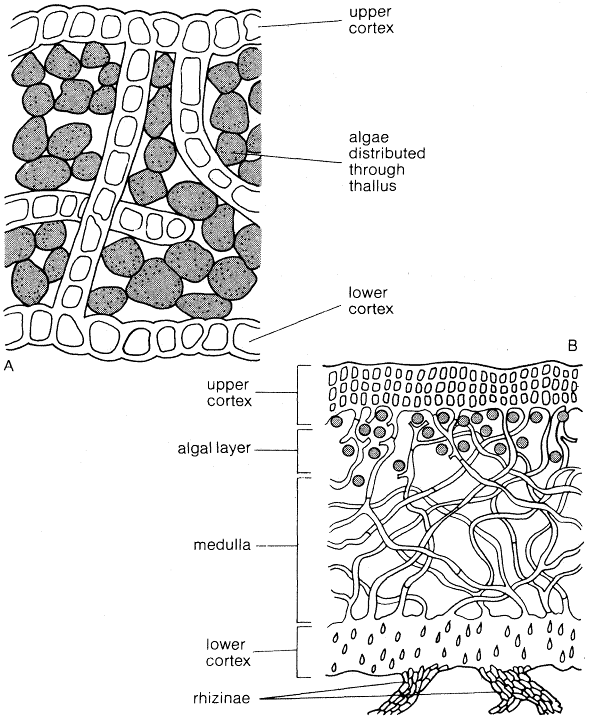


Figure 1: Distinctions based on distribution of algal cells: (A) homoiomerous thallus; (B) heteromerous thallus.

From *An Evolutionary Survey of the Plant Kingdom* by Robert F. Scagel, *et al.* © 1965 by Wadsworth Publishing Company, Inc., Belmont, California 94002. Reprinted by permission of the publisher

tures called haustoria. *Trebouxia* lichens have an interesting pattern, one in which deeply penetrating haustoria are prevalent in associations lacking a high degree of thalloid organization. On the other hand, superficial haustoria prevail among forms with highly developed thalli. *Lecanora* and *Lecidea,* for example, have individual algal cells with as many as five haustoria that may extend to the cell centre. *Alectoria* and *Cladonia* have haustoria that do not penetrate far beyond the algal cell wall. A few phycobionts, such as *Coccomyxa* and *Stichococcus,* which are not penetrated by haustoria, have thin-walled cells that are pressed close to fungal hyphae.

Although deeply embedded in an algal cell, a haustorium remains separated from the cell by walls and membranes; the haustorium wall is thinner than that of the usual fungal hypha. The contents of the algal cell appear to remain intact around the haustorium; in some lichens, however, a region containing many vesicle-like cavities occurs between the contents of the algal cell and the tip of the haustorium. Although the outer wall of *Trebouxia* is ruptured by the haustorium, the inner wall is more flexible, folding inward to accommodate the haustorium.

Carbohydrate flow

The flow of nutrients and metabolites between the symbionts is the basic foundation of the symbiotic system. A simple carbohydrate formed in the algal layer eventually is excreted, taken up by the mycobiont, and transformed into a different carbohydrate. Each phycobiont excretes a specific carbohydrate; *i.e.*, the mobile carbohydrate in *Nostoc* and other blue-green algae is a simple sugar, glucose. Green phycobionts excrete another type of carbohydrate known as sugar alcohols; *i.e., Trebouxia, Myrmecia,* and *Coccomyxa* excrete ribitol; *Hyalococcus,* sorbitol; and *Trentepohlia,* erythritol. The mobile carbohydrate formed by the phycobiont is taken up by the mycobiont and converted to mannitol, a sugar alcohol commonly found as a storage product in fungi. Smaller amounts of another sugar alcohol, arabitol, also may be produced. The release of carbohydrate by the phycobiont and its conversion by the mycobiont occur rapidly. In *Peltigera polydactyla,* for example, glucose forms and is released by the phycobiont within a minute, and mannitol is found in the mycobiont after two minutes. Whether or not the fungus influences the release of carbohydrate by the alga is not yet known with certainty, but carbohydrate excretion by the alga decreases rapidly if it is separated from the fungus.

Other symbiotic relationships

Carbohydrate transfer is only one aspect of the symbiotic interaction in lichens. The alga may provide the fungus with vitamins, especially biotin and thiamine, important because most lichen fungi that are grown in the absence of algae have vitamin deficiencies. The alga also may contribute a substance that causes structural changes in the fungus since it forms the typical lichen thallus only in association with an alga. One contribution of the fungus to the symbiosis concerns absorption of water vapour from the air; the process is so effective that, at high levels of air humidity, the phycobionts of some lichens photosynthesize at near-maximum rates. The upper region of a thallus provides shade for the underlying algae, some of which (*e.g., Trebouxia)* are sensitive to strong light. In addition, the upper region may contain pigments or crystals that further reduce light intensity and act as filters, absorbing certain wavelengths of light.

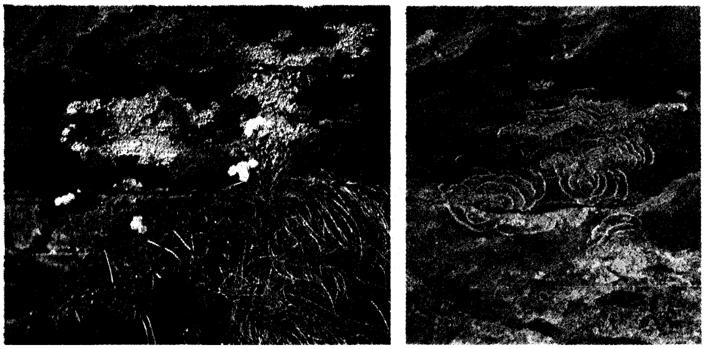
Lichens synthesize a variety of unique organic compounds that tend to accumulate within the thallus; many of these substances are coloured and are responsible for the red, yellow, or orange colour of lichens. Although these compounds have been assumed to be products of the symbiosis (*i.e.,* neither symbiont could produce them alone), certain mycobionts grown in the absence of algae form the same compounds that are found in the symbiosis. The role of lichen substances in the symbiosis has not yet been clarified. They might be waste products with no function, or, since some of the substances are effective chelating agents (*i.e.*, capable of binding metal atoms), they might extract minerals, thus making them available to the symbiosis. Other possible roles of lichen substances may be to increase the ability of compounds to penetrate the membranes of the algal cells, to stimulate algal metabolism, or to prevent the growth of decomposing organisms in the thallus.

FORM AND FUNCTION

Thalloid types. *Homoiomerous thallus.* A lichen thallus or composite body has one of two basic structures. In a homoiomerous thallus, the algal cells, distributed throughout the structure, are more numerous than those of the fungus. The phycobiont is a blue-green alga whose filaments are embedded in a matrix. In *Collema,* the mycobiont has a few of the threadlike filaments called hyphae growing between the algal filaments; in *Leptogium,* however, the mycobiont forms regions known as the upper and lower cortex.

Heteromerous thallus. The most common type of thallus, a heteromerous thallus, has four distinct layers, three of which are formed by the fungus and one by the alga. The fungal layers are called upper cortex, medulla, and lower cortex. The upper cortex consists of either a few layers of tightly packed cells or hyphae that may contain pigments. A cuticle may cover the cortex. The lower cortex, similar in structure to the upper cortex, participates in the formation of attachment structures called rhizines. The medulla, located below the algal layer, is the widest layer of a heteromerous thallus. It has a cottony appearance and consists of interlaced hyphae. The loose

Vernon Ahmadjian



Crustose Lichens.
(Left) Rust-coloured Caloplaca elegans. (Right) Concentric circle formation of a species of Lecanora.

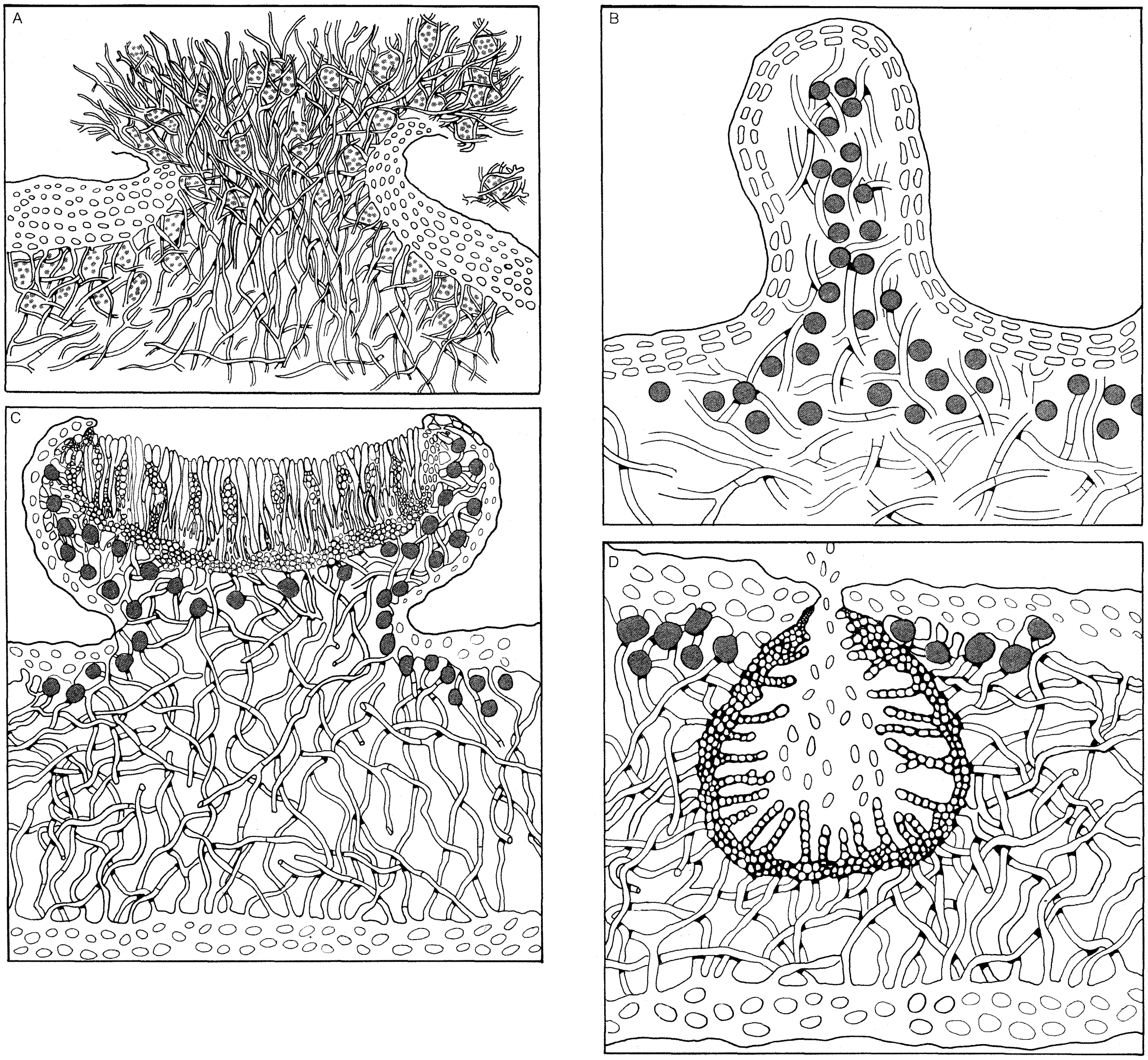


Figure 2: Lichen structure.
Section through (A) thallus of *Lobaria* verrucosa showing bursting mass of soredia; (B) isidium of *Lobaria pulmonaria;* (C) apothecium of Lecanora; (D) spermagonium (pycnidium) of Lecanora produces cells that act as the male gametes in sexual reproduction. Shaded areas are algal cells.

From An Evolutionary *Survey* of the Plant Kinodorn by Robert F. Scagel, et *al.* © 1965 by Wadsworth Publishing Company, Inc., Belmont, California 94002. Reprinted by permission of the publisher (C,D) Gebruder Borntraeger

nature of the medulla not only provides it with numerous air spaces but also allows it to hold large amounts of water.

The algal layer, about three times as wide as a cortex, consists of tightly packed algal cells enveloped by fungal hyphae from the medulla. The algal layer is not uniform in width; localized growth results in ridges or humps, some of which extend upward into the cortex or downward into the medulla. In addition to a primary phycobiont, some lichens also have a secondary one. *Lobaria*, Nephroma, and *Solorina* have a blue-green secondary phycobiont; the primary one is always green. The secondary phycobiont is found within the thallus either in clusters or as a distinct, uninterrupted layer directly below the primary alga. In some lichens, a gall-like structure called a cephalodium results from the association of the fungus with a secondary blue-green phycobiont.

Growth forms

A heteromerous thallus may have a stalked (fruticose), crustlike (crustose), or a leafy (foliose) form. Although these growth forms have considerable taxonomic significance, they are not inflexible structures, and many transitional types exist. It is not known, moreover, which growth form is primitive, which is advanced. Fruticose lichens, which usually arise from a primary thallus of a different growth form (*i.e.*, cmstose, foliose), may be shrubby, pendulous, or consist of upright stalks. The fruticose form usually consists of two thalloid types: the primary thallus is crustlike or lobed; the secondary thalli, which originate from the crust or lobes of the primary thallus, consists of stalks that may be simple, cup-shaped, intricately branched, and capped with brown or red fruiting bodies called apothecia. Fruticose forms such as *Usnea* may have elongated stalks with a central solid core that provides strength and elasticity to the thallus.

The crustose thallus is in such intimate contact with the surface to which it is attached that it usually cannot be removed intact. Some crustose lichens grow beneath the surface of bark or rock so that only their fruiting structures penetrate the surface. Crustose lichens may have a hypothallus—*i.e.*, an algal-free mat of hyphae extending beyond the margin of the regular thallus. Cmstose form varies: granular types such as Lepmria, for example, have no organized thalloid structure; but some Lecanora species have highly organized thalli, with lobes that resemble foliose lichens lacking a lower cortex.

The foliose forms are flat, leaflike, and loosely attached to a surface. The largest known lichens have a foliose form; species of Sticta may attain a diameter of several feet. Other common foliose genera include Cetraria, *Parmelia*, *Peltigera*, and Plzyscia. Umbilicaria, called the common rock tripe, differs from other foliose forms in its mode of attachment in that its platelike thallus attaches at the centre to a rock surface.

Specialized thalloid structures. *Rhizines, cilia, and pores.* Rhizines, hairlike growths that anchor the thallus to a surface, consist of bands of hyphae that originate in the medulla and lower cortex. Especially prevalent in foliose lichens, rhizines may be scattered over, or localized on, the under surface of the lichen. The tip of a rhizine may contain a sticky substance that increases its adhesive capability. Rhizines are not involved in the uptake of water or minerals to the lichen.

Cilia, long hairlike projections with no known function, originate from cortical hyphae on margins of a thallus or apothecium. Cyphellae are well-defined pores that facilitate gas exchange in the lower surface of some foliose lichens such as *Sticta.* Pseudocyphellae, less well-defined, irregular pores usually filled with a granular material, have a function similar to that of cyphellae.

Ceplzalodia. Cephalodia, gall-like structures connected to the main thallus, contain an algal type, usually a blue-green, different from that in the main thallus. A cephalodium develops when a foreign alga comes in contact with the thallus and stimulates the mycobiont to form a mass of tissue around it. Cephalodia have characteristic forms; *e.g., Peltigera aphthosa* has numerous disk-shaped cephalodia containing *Nostoc,* while *Stereocaulon* has saclike cephalodia containing either *Scytonema* or *Stigonerna.*

Soredia and isidia. Soredia (Figure 2A) are powdery masses that result from outgrowths of algae through cracks and pores in the upper cortex. A single soredium consists of one to several algal cells and the fungal hyphae accompanying them. Soredia, common in foliose and fruticose lichens but rare in crustose lichens, may be scattered over the thallus or limited to clumps called soralia. Soralia vary in size, shape, and thalloid position in different species—*e.g.,* elongated furrows on thalloid ridges of *Parmelia sulcata,* swollen knobs on the marginal lobes of *Hypogymnia physodes.* Powdery soredia are found on the stalks of many species of *Cladonia,* while the thallus of *Lepraria* consists entirely of soredia. Prolonged periods of moisture stimulate growth of the phycobiont in some lichens, resulting in increased soredial production.

Soralia

Isidia (Figure 2B) are similar to soredia in origin and reaction to moisture, but they differ from soredia in that the algae do not grow through the upper cortex. Instead, the algae extend the cortex into small knobs whose tissues are continuous with the tissues of the main thallus. Thalli of *Parrnelia furfuraccu* in moist localities are covered with isidia.

Vernon Ahmadjian

Fruticose lichen.
Alectoria sarmentosa, a northern forest lichen sensitive to warm temperatures.

Fruiting bodies and spores. The complex fruiting bodies (ascocarps) of lichen fungi are of several types. The factors that induce fruiting in lichens have not yet been established with certainty. Button-like fruiting bodies called apothecia (Figure 2C) are found on the upper surface or sides of Lecanorales thalli. Some thalli are covered with apothecia, while others may have few or none. An apothecium contains a spore-bearing tissue called the hymenium, which consists of saclike bodies (asci) containing spores (ascospores). Interspersed between the asci are sterile hyphae, called paraphyses, whose ends frequently are enlarged to form a hymenial surface layer (epithecium). The asci and paraphyses originate from a layer of fungal tissue called the hypothecium, which extends along the sides of the hymenium to form an apothecial margin—*e.g.,* as in *Lecidea.* Below the hypothecium may be another layer that is part of the thallus and contains algae. This layer may envelop the apothecial margin —*e.g.,* as in *Lecanora.*

Lichen fungi belonging to the Sphaeriales have a flask-shaped fruiting body called a perithecium, which has an outer margin, usually black in colour, and a tiny pore called an ostiole. Lichen fungi of the Hysteriales have a thick-walled narrow fruiting body. In the Caliciales, the asci disintegrate prematurely, and the mass of spores and paraphyses is called a mazaedium.

Ascospores of lichen fungi are of many sizes and shapes; *e.g., Pertusaria* has one or two large spores in one ascus; *Acarospora* may have several hundred small spores per ascus. An ascospore generally has one nucleus but may be single-celled or multicellular, brown or colourless; the *Pertusaria* spore, however, is a single cell containing 200 nuclei. Another type of fungal spore may be what are sometimes called spermatia (male fungal sex cells) or pycnidiospores; it is not yet certain that these structures can germinate and develop into a fungal colony. Few lichen fungi produce conidia, a type of asexual spore common among ascomycetes.

Separation and growth of symbionts. A phycobiont may be isolated from a lichen thallus by several techniques. If a thallus is fragmented in water, individual algal cells can be freed of visible debris and placed onto a nutrient agar. A visible colony will usually develop from such cells within two weeks. An easier isolation technique is to crumble a dry thallus over the surface of a mineral agar and to incubate the small fragments in the presence of light; the phycobiont grows faster than the fungus. Isolation of blue-green phycobionts is more complex because bacteria commonly are found in their gelatinous sheaths.

A mycobiont may be isolated either from spores or from pieces of hyphae. The hyphal isolation technique, best applied to foliose and fruticose lichens, involves separation and transfer of pieces of the medulla to a suitable growth medium. Both alga and fungus may be grown by isolating algal cells that have large fragments of fungus attached to them.

Culture methods

The growth or culture methods for lichen symbionts are similar to those used for other micro-organisms. The behaviour and requirements of many lichen fungi and algae, however, are unique. Most of the symbionts grow slowly, and the best growth is obtained on complex media rich in nutrients. Unlike other green algae, which require minerals alone for growth, *Trebouxia* grows well only if provided with a medium of sugar, organic nitrogen, and minerals. The phycobiont synthesizes nutrients both for itself and its fungal partner: *Trebouxia,* for example, can fix carbon by photosynthesis almost as well as other green algae; but ribitol, the photosynthetic product that accumulates, is incorporated so slowly into algal growth substances that it accumulates, begins to leak out of the cells, and is absorbed and converted to mannitol by the fungus.

A lichen develops only if the symbionts are forced into the union by stress conditions—*e.g.,* periodic drying or nutrient deficiency. Successful synthesis has been accomplished with *Cladonia cristatella* and *Endocarpon pusillum.* With *C. cristatella* the synthesis occurs in two phases, development of thalloid lobes and formation of fruiting bodies. The first occurs only after an interaction of the two symbionts; the second occurs in separate cultures of the mycobiont, but the fruiting bodies are immature and lack spores. The synthesis of *E. pusillurn,* on the other hand, represents the first complete synthesis of a lichen. The fungal and algal symbionts, isolated and grown separately, are mixed on a soil surface. Thallus lobes and mature fruits form within four months; in addition, spores from the fruiting bodies germinate.

Effects of drying

Factors that influence metabolic activity. *Water content.* The metabolic activity of lichens is greatly influenced by the water content of the thallus. The rate of photosynthesis may be greatest when the amount of water in the thallus is from 65 to 90 percent of the maximum. During drying conditions, the photosynthetic rate decreases and, below 30 percent is no longer measurable. Although respiration also decreases rapidly below 80 percent water content, it persists at low rates even when the thallus is air-dried. Since lichens have no mechanisms for water retention or uptake from the surface to which they are attached, they very quickly lose the water vapour they absorb from the air. The rapid drying of lichens is a protective device; *i.e.*, a moisture-free lichen is more resistant to temperature and light extremes than is a wet one. Frequent drying and wetting of a thallus is one of the reasons lichens have a slow growth rate. Studies have revealed that the phycobiont *Trebouxia* stores two products. If the lichen is dry, large numbers of fat-containing droplets are found within pyrenoid, or storage, structures in the algal cells. When the lichen is kept moist for several days, starch granules appear in addition to the fat droplets. Since starch is the usual storage product in green algae, fat may be a product that has evolved from the symbiosis with fungi. Both the natural habitats of lichens and the relatively short daily period of time during which adequate water balance exists indicate that fats may represent a product whose function is to provide a large amount of energy in a small period of time and possibly supply enough water to maintain the metabolic machinery of the symbionts.

Temperature and light. Maximum photosynthesis in lichens takes place at temperatures of 15° to 20° C. More light is needed in the spring and summer than in the winter. The photosynthetic apparatus of lichens is remarkably resistant to cold temperatures. Even at temperatures below 0° C, many lichens can absorb and fix considerable amounts of carbon dioxide. Respiration is much less at low temperatures so that, in nature, the winter months may be the most productive ones for lichens.

Crustose lichens grow less than one millimetre per year, while fruticose and foliose forms increase in size at a rate of a few centimetres a year. Thalli of some arctic crustose lichens such as *Rhizocarpon* and *Lecidea* may be 4,500 years old. In temperate regions, however, changes in a lichen community occur more rapidly. Lichens, which are sensitive to small changes in climate, appear or disappear from an area depending on climatic changes. Lichens may form thalli and fruits within two to eight years after growth on a new surface has begun.

EVOLUTION

Trebouxia lichens

The fossil record of lichens is inconclusive. It has not yet been established with certainty when and how these associations evolved, although lichens must have evolved more recently than their components and probably arose independently from different groups of fungi and algae. It seems, moreover, that the ability to form lichens can spread to new groups of fungi and algae. Interesting lichens, in terms of evolution, are *Trebouxia* lichens. It has been postulated that *Trebouxia* is a genus that originated as a result of an association with fungi. Its ancestor may have been a type of alga commonly found in soil. Two groups of *Trebouxia* are recognized; one group (I) occurs in *Cladonia* and related lichens, while the other group (II) is found in foliose and in many crustose forms. Certain features of group I *Trebouxia*, similar to features found in free-living soil algae, may reflect those of the ancestral type. On this basis, *Cladonia* and its related forms would have given rise to lichens with Group II *Trebouxia*. This hypothesis is contrary to the usually accepted idea that fruticose lichens represent the most advanced type.

Vernon Ahmadjian

Foliose lichen.
A species of Pseudocyphellaria showing fruiting bodies.

CLASSIFICATION

Distinguishing taxonomic features. The taxonomy of lichens is based mainly upon the character of the fungal symbiont. The primary taxonomic features involve details of the fruiting bodies such as type of ascocarp, size and nature of spores, and height of the hymenium. Secondary taxonomic characters include details of the external structure of the thallus—*i.e.*, colour; width of lobes; presence, absence, shape, size, and location of isidia, soredia, cyphellae, psuedocyphellae, cephalodia, cilia, and rhizines.

Other secondary characters include internal structural details of the thallus—*i.e.*, thickness of the cortex, type of algal symbiont, nature of the chemical substances found in the thallus. Geographical distribution and the type of growth surface also are factors in taxonomic evaluation. In some cases, a grouping of secondary features has been considered sufficient to give species rank to a lichen.

Annotated classification. The following abridged classification is adapted from one published by M.E. Hale in 1967.

PHYLUM ASCOMYCOTA (sac **fungi)**

Class Ascomycetes

Sexual spores contained within asci.

Subclass Ascomycetidae

Asci mainly unitunicate (one wall), arranged in a hymenium with unbranched paraphyses.

Order Lecanorales. Fruiting bodies of the apothecium type.

Family Acarosporaceae. Thallus crustose or scalelike; numerous minute spores in an ascus; on rocks in Arctic and temperate regions; *Acarospora* apothecia sunk in thallus; *Biatorella* thalli may be embedded in rock substrate.

Family Cladoniaceae. Two types of thalli; primary thallus crustose or scalelike; secondary thallus fruticose; on rocks, soil, and rotted wood in Arctic and temperate regions; *Baeomyces roseus*, pink apothecia on small stalks, common on roadside soil banks; *Cladonia,* a commonly collected type of lichen.

Family Collemataceae. Thallus gelatinous when wet; phycobiont blue-green; Arctic and temperate regions; on soil, rocks, and trees in moist locations; *Collema,* thallus black or dark brown, dull; *Leptogium,* thallus slate-blue, shiny.

Family Graphidaceae. Thallus crustose; apothecium elongated with narrow slitlike opening; on trees, predominantly in tropical and subtropical regions; *Graphis,* called "script lichen" because clusters of apothecia resemble hieroglyphic writing.

Family Lecanoraceae. Lichen family second only to Lecideceae in number; crustose, sometimes scalelike; apothecium with a thallus rim; on trees and rocks in Arctic and temperate regions; *Lecanora,* over 1,000 species.

Family Lecideaceae. Largest lichen family; thallus crustose; apothecia lack thallus rim; on rocks and trees in Arctic and temperate regions; *Lecidea,* over 1,000 species, simple spores; *Bacidia,* elongated, multicellular spores.

Family Parmeliaceae. Thallus foliose or subfruticose, well developed; on rocks and trees in Arctic, temperate, and tropical regions; *Parmelia,* largest genus of foliose lichens, common on deciduous trees in southeastern United States, apothecia on thallus surface; *Cetraria,* common on conifers in cold regions, apothecia on thallus margin.

Family Peltigeraceae. Thallus foliose, often with gall-like structures (cephalodia); on soil, mosses, and trees in Arctic and temperate regions; *Peltigera,* common in damp woods, large, thick thalli; *Hydrothyria venosa,* aquatic form in streams at high elevations.

Family Physciaceae. Thallus crustose or foliose; spores brown; on trees and rocks, Arctic to tropical; *Physcia,* small, ash-white thalli commonly rosette-shaped; *Buellia*, small, inconspicuous crustose thallus.

Family Stictaceae. Thallus foliose; includes some of the largest known lichens; both cyphellae and pseudocyphellae pore types common; on trees in temperate and tropical regions; highly developed in Southern Hemisphere; *Lobaria pulmonaria*, "lungwort lichen"; *Sticta*, with thalli up to several feet in diameter in Southern Hemisphere.

Family Teloschisfaceae. Thallus yellow or orange, crustose, foliose, or fruticose; on trees and rocks in Arctic and temperate regions; *Xanthoria parietina*, common along seashores; *Caloplaca elegans*, one of the most colourful lichens.

Family Unzbilicariaceae. Thallus foliose, attached to rocks at a single point (umbilicus); Arctic and temperate regions; *Umbilicaria*, "rock tripes."

Family Usneaceae. Thallus fruticose, sometimes pendulous to great lengths; on bark and rocks in Arctic and temperate regions; *Usnea*, "old-man's beard"; *Letharia vulpina*, covers branches of trees in mountains of western U.S.

Order Sphaeriales. Fruiting bodies of perithecium type.

Family Pyrenulaceae. Thallus crustose; mostly on trees in tropical areas; *Pyrenula*, thallus inconspicuous, often immersed within growth surface.

Family Strigulaceae. Thallus crustose; tropical; *Strigula*, often parasitic on leaves.

Family Verrucariaceae. Thallus crustose or scalelike; on rocks, trees, and soil; Arctic and temperate regions; many freshwater and marine species; *Verrucaria*, forms black crusts on rocks along seashores; *Endocarpon*, fruits contain algal cells that are discharged along with spores.

Order Caliciales. Fruiting bodies of the mazaedium type (asci disintegrate at maturity).

Family Caliciaceae. Thallus crustose, poorly developed; fruits on slender stalks, resembling tiny nails; on bark and soil in temperate regions; *Calicium, Coniocybe.*

Family Opegraphaceae. Thallus crustose; on trees and soil in temperate regions; highly developed in Southern Hemisphere; *Sphaerophorus*, shrubby thallus with branched, intertangled stalks.

Subclass Loculoascomycetidae

Asci bitunicate (two walls) an arranged with branched paraphyses in an ascostroma.

Order Myrangiales. Ascostroma poorly differentiated; asci distributed irregularly.

Family Arthoniaceae. Thallus crustose, poorly formed, embedded in trees in temperate and tropical regions; *Arthonia*, widely distributed.

Order Pleosporales. Ascostroma well differentiated, resemble perithecia; asci distributed in regular fashion.

Family Arthopyreniaceae. Thallus crustose; on trees and rocks; primarily tropical; *Arthopyrerzia*, smooth, whitish thalli partly within growth surface.

Order Hysteriales. Ascostroma well differentiated, resemble apothecia or are elongated and irregular.

Family Opegraphaceae. Thallus crustose; on trees and rocks in temperate and tropical regions; *Opegrapha*, "script lichen" similar to *Graphis* (family Graphidaceae).

Family Roccellaceae. Thallus crustose or fruticose; on rocks and trees in tropical and subtropical regions; *Roccella*, fruticose thalli, common in dry, coastal areas of eastern Pacific, Africa, and the Mediterranean.

Class Basidiomycetes

Sexual spores borne on basidia; *Cora*, on soil and trees in tropical regions; *Omphalina*, on soil in Arctic and mountainous regions.

Class Fungi Imperfecti

Sexual spores absent; *Lepraria*, granular, mealy thallus of irregular form; common on rocks and soil in moist, shaded areas.

Critical appraisal. Lichens are a biological group lacking formal status in the taxonomic framework of living organisms. Although the mycobiont and phycobiont have Latin names, the product of their interaction, a lichen, does not. Earlier names given to lichens as a whole now are considered names for the fungus alone. Classification of lichens is difficult and remains controversial. Part of the problem is that the taxonomy of lichens was established before their dual nature was recognized; *i.e.*, the association was treated as a single entity. Another difficulty encountered in lichen classification is inherent in the field of taxonomy—*i.e.*, whether or not taxonomic systems should reflect the natural phylogeny of living organisms or conveniently identify them. In a system designed for convenience, it is advantageous to consider the association as one unit.

BIBLIOGRAPHY. V. AHMADJIAN, *The Lichen Symbiosis* (1967), a well-illustrated, modern reference source in experimental lichenology, with an extensive annotated list of references; Y. ASAHINA and S. SHIBATA, *Chemistry of Lichen Substances* (1954; orig. pub. in Japanese, 1949), an excellent, classic reference book on lichen chemistry; H. DES ABBAYES, *Traité de lichénologie* (1951), a general, well-illustrated reference book for most aspects of lichenology; B. FINK, *Lichen Flora of the United States* (1935, reissued 1960), an advanced source with key, descriptions, and photographs of lichens found in the U.S.; M.E. HALE, *The Biology of Lichens* (1967), an excellent, intermediate-level introduction to all aspects of lichenology, and *How to Know the Lichens* (1969), an authoritative, readable guide; G.G. NEARING, *The Lichen Book* (1947, reissued 1962), still the best popular guide to the identification of lichens, containing drawings of lichens, their popular names, and descriptions; G.D. SCOTT, *Plant Symbiosis* (1969), an excellent, brief work that compares the different types of plant symbioses at a morphological and physiological level and discusses the concept of symbiosis and its origin and development; A.L. SMITH, *Lichens* (1921), the classical book on all aspects of lichenology, which is still a useful reference source; C.J. TAYLOR, *Lichens of Ohio*, 2 vol. (1967–68), an introduction to the identification of lichens that contains keys and excellent photographs of lichens.

(V.A.)

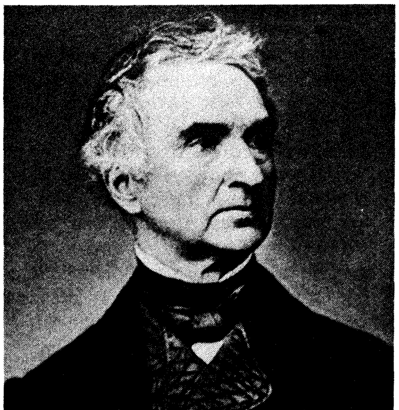
Liebig, Justus von

Justus von Liebig, German chemist, made many important contributions to the early systematization of organic chemistry, to the application of chemistry to biology (biochemistry), to chemical education, and to the basic principles of agricultural chemistry. He was born at Darmstadt on May 12, 1803. After studying pharmacy for a short time he followed his ambition to become a chemist by entering the University of Bonn to study with a leading chemist of the day, Karl Wilhelm Gottlob Kastner.

Education

When Kastner moved to the University of Erlangen, Liebig accompanied him and received his doctorate there in 1822. He then received a grant from the Hessian government to study in Paris. Through the influence of Alexander von Humboldt he was able to work in the private laboratory of Joseph-Louis Gay-Lussac.

By courtesy of the Gesellschaft Liebig-Museum, Giessen, West Germany



Liebig, photograph by F. Hanfstaengl, 1868.

In 1824 he joined the faculty of the University of Giessen, where he became a full professor in 1826. At Giessen he established the first laboratory in which the methods of chemical research were taught systematically to young chemists. The laboratory soon became world famous, and students from all over Europe came to study with him. Among them were many of the notable chemists of the following generation, including August Wilhelm von Hofmann, Sir Edward Frankland, F.A. Kekule von Stradonitz, and Charles-Adolphe Wurtz. Liebig's laboratory set the pattern for chemical education that came to prevail in Germany and was in large part responsible for the great development of German chemistry later in the 19th century. In 1845 he was made a baron, and in 1852 he became professor of chemistry at the University of Munich, where he remained until his death in Munich on April 18, 1873.

Development of radical theory

Liebig was an active worker in many fields of inorganic and organic chemistry, but certain of his studies stand out because of their great significance in the later development of chemistry. His early studies on the isomerism of cyanic and fulminic acids made a great impression on his contemporaries and led him into the field of organic chemistry. Through his work in organic chemistry, he became acquainted with another outstanding chemist, Friedrich Wohler. His friendship with Wohler lasted a lifetime, and the two men collaborated in several investigations. The most important of these was the study of bitter almond oil (benzaldehyde), in which the same chemical group was found to pass unchanged through a great variety of reactions. Such a group was called a radical, and the radical theory that Liebig did much to develop was the first major attempt at systematization in organic chemistry.

Liebig's studies in organic chemistry were greatly aided by the simple method he developed for the analytical determination of carbon and hydrogen. He also developed a method for the analytical determination of halogens, published important work on polybasic organic acids, and did much to support the hydrogen theory of acids. He popularized, but did not invent, the Liebig condenser, still used in many laboratory distillations.

After 1838 Liebig's interest shifted from pure organic chemistry to the chemistry of plants and animals. He made a large number of analyses of tissues and body fluids and carried out a study of the nitrogenous products of the animal organism. A by-product of this investigation was the Liebig extract of beef. He then became greatly interested in agricultural problems and in 1840 published *Organic Chemistry in Its Application to Agriculture and Physiology*. This book exerted very great influence on practical agriculturalists. He rejected the old theory that humus supplied plants with food and showed that plants took carbon dioxide, water, and ammonia from the air and soil. He advocated the use of mineral fertilizers to supply other elements that might become depleted in soils.

In his later years Liebig's reputation became so great that he was regarded as the final authority in chemical matters. He never hesitated to express his sometimes rather rigid views. As a result he was often involved in scientific controversies in which he was not always in the right. Much of his work was published in the journal that he founded in 1832, the *Annalen der Pharmacie*, which ultimately became the *Annalen der Chemie*, one of the major chemical journals.

After he accepted the professorship at Munich, Liebig gradually lost his enthusiasm for laboratory work. He soon refused to accept new students and devoted himself more and more to literary activities. In these, aside from his frequent polemics, he was occupied in stressing the broader applications of chemistry to human life.

BIBLIOGRAPHY. Most of the biographies of Liebig have been written in German. The earliest English account is A.W. VON HOFMANN, *The Life Work of Liebig in Experimental and Philosophic Chemistry* (1875). A good account is by F. HABER in E. FARBER, *Great Chemists* (1961). The work of Liebig and its results are treated in *Liebig and After Liebig*, ed. by F.R. MOULTON (1942). An excellent picture of Liebig's relations with his contemporaries is H. VON DECHEND (ed.), *Justus von Liebig in eigenen Zeugnissen und Solchen seiner Zeitgenossen*, 2nd ed. (1963).

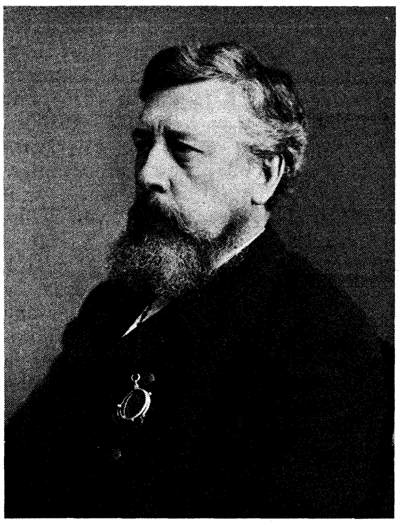
(H.M.L.)

Liebknecht, Wilhelm and Karl

Between the German revolutions of 1848–49 and 1918–19 Wilhelm and Karl Liebknecht, father and son and prominent German Socialists, successively provided much of the leadership to the Marxian wing of the German Socialist movement, which was characterized by a collectivist, internationalist, antimilitary, and anti-war tradition. The revolutions of 1848–49 in Germany and elsewhere in Europe marked the beginning of Wilhelm Liebknecht's active espousal of Socialism. The German Revolution of 1918, which claimed the life of Karl Liebknecht, closed a period in which father and son significantly shaped the nature and destiny of German Socialism.

Wilhelm. Wilhelm Liebknecht was born at Giessen in Hesse, on March 29, 1826. He was still a child when his father died, but he was brought up comfortably. He attended the universities of Giessen, Marburg, and Berlin and developed an interest in French Socialist thinking, formulated in the early 1830s by Henri Saint-Simon, Charles Fourier, and Pierre-Joseph Proudhon. Dissatisfied with his life in Germany, he planned to emigrate to the United States when he was in his 20s. Instead, he accepted an invitation to teach at a Swiss elementary school, where for the first time he came in close contact with the working classes. He then decided to study law and be called to the bar in Switzerland.

Archiv fur Kunst und Geschichte



Wilhelm Liebknecht, c. 1890.

This was in 1847. On February 23, 1848, revolution erupted in Paris. He arrived too late to become involved and returned to Germany, where he participated in several revolutionary insurrections that failed. During an attempt to fan the fading revolutionary embers in Baden, he was captured and held prisoner for eight months. In 1849, after his release, he returned to Switzerland.

Years with Marx

Liebknecht's stay in Switzerland was short, for the Austrian and Prussian governments, fearful of his growing influence among the Swiss workers, succeeded in having him expelled from Geneva. In 1849 he went to England, where he remained for 13 years. In London he joined the Communist League, working closely with Karl Marx and Friedrich Engels and supporting himself as London correspondent for the *Augsburger allgemeine Zeitung* ("Augsburg Gazette"). In 1862 the Prussian government granted him amnesty; he returned to Berlin and became a writer for the *Norddeutsche allgemeine Zeitung* ("North German Gazette"), soon becoming an influential Socialist. But Otto von Bismarck, who had become minister president (prime minister) in 1862, resented Liebknecht's influence among the working classes and, failing to gain his support, had him expelled from Prussia in 1865.

In Leipzig, where he moved, Liebknecht joined the floundering Allgemeiner Deutscher Arbeiterverein (General German Workers' Association), founded by the Socialist leader Ferdinand Lassalle in 1863. He also formed a friendship with August Bebel, a woodturner, who on his travels as a journeyman had become familiar with the poverty of the masses throughout Germany. Liebknecht, the writer, and Bebel, the orator and practical politician, complemented one another and together they provided the leadership for German Socialism for the remainder of the century. In Leipzig, Liebknecht worked hard to win new recruits for the cause and continued his efforts to educate the masses through the *Demokratisches Wochenblatt.* In 1867 the workers elected Liebknecht to the North German Reichstag, where he opposed Lassalle's advocacy of a "paternalistic" state socialism. In 1869, at a congress at Eisenach, Liebknecht and Bebel organized the Sozialdemokratische Arbeiterpartei (Social Democratic Labour Party) and affiliated it with the First International (International Workingmen's Association) headquartered in London.

The outbreak of the Franco-Prussian War in 1870 put Liebknecht's devotion to international Socialism to a practical test. His failure to vote for war credits and his writings against the war and the government resulted in his conviction on charges of "treasonable intentions" in 1872. He was sentenced to two years' confinement in the fortress of Hubertusburg, along with Bebel, who was similarly charged.

The Prussian military victory in 1871 did nothing to abate the Socialists' growing strength in the Reichstag, and Liebknecht continued to be a thorn in Bismarck's side. Bismarck's determination to repress the Socialists brought about the merger of the Lassalleans and Liebknechtians as the Sozialistische Arbeiterpartei Deutschlands (Socialist Labour Party) at Gotha in 1875. The Gotha Program, a compromise between the positions of the two parties—although criticized by Marx for its call for government-aided productive organizations—remained the charter of German Socialism until the adoption of the Erfurt Program in 1891, which discarded the state-aid provisions of the Gotha Congress and pledged the party to a Marxist program. Bismarck won his battle to repress the Socialists in 1878 when the Reichstag adopted the Anti-Socialist Law that, among other things, forbade the publication of Socialist literature. Until the legislation against the party expired in 1890, Liebknecht did double duty by establishing an office in Leipzig to aid victims of the law and devoting his writing talents to the preparation of scholarly books and pamphlets.

Socialist leader

Notwithstanding a dozen years of repression, the party continued to grow significantly. When the law expired in 1890, it was obvious that Liebknecht's tactic of education, not conspiracy, had been productive. When the liberated party met at Erfurt in 1891, it adopted a charter embodying the 19th century's fullest expression of Social Democratic ideas. Thereafter, the party was known as the German Social Democratic Party (SPD). Liebknecht became one of its leading spokesmen, primarily as a writer for *Vorwarts,* the party's most prominent newspaper, founded in 1890. It was his gift as a writer, giving expression to his belief in the education of the masses, that made him a great leader of German Social Democracy, which was perhaps the most important political development in Europe between the Franco-Prussian War and World War I. Liebknecht died in Berlin on August 7,1900.

Karl. On August 13, 1871, while Wilhelm was in prison awaiting trial, his son, Karl, was born in Leipzig. Karl grew up during the years when the Anti-Socialist Law was in force against his father's party. Lacking the means for professional study, Karl independently studied nature, poetry, art, and music. Later, with financial help from the party, he studied law and political economy, first at Leipzig and then at Berlin, where he earned his doctor's degree. He planned to devote his career to the defense of Marxism.

After serving with the Imperial Pioneer Guards in Potsdam during 1893–94 and subsequently as a junior barrister in Westphalia, he returned to Berlin in 1898. In 1900, the year of his father's death, he married his first wife, Julie Paradies, by whom he had three children. She died ten years later, and in 1912 he married Sophia Ryss, a woman of Russian birth who had graduated from the University of Heidelberg.

Lawyer and revolutionary

In 1904, at a trial at Konigsberg, he defended propertyless peasants accused of infiltrating Socialist propaganda from East Prussia into tsarist Russia. His defense of the accused was primarily an apology for Social Democracy and provided him with a platform for his attacks against militarism. The key to the fight against militarism, he maintained, was the education of youth. Encouraged by the Russian Revolution of 1905, he worked among the youth, criticizing the "Moloch of Militarism," which he promised to fight "to the last drop of blood." In 1907 he played a principal role in the establishment of the International Union of Socialist Youth Organizations in Stuttgart. His publication of *Militarismus und Antimilitarismus* in the same year earned him a jail sentence of 18 months in Glatz, Silesia. While still in prison, he won a seat in the Prussian Landtag, and in 1912 he entered the Reichstag as the chief spokesman against the government and against the growing movement within the Social Democratic Party to revise its Marxist doctrine.

Karl **Liebknecht,** 1913.
Interfoto-Friedrich Rauch, Munich

During World War I Karl Liebknecht became a leading figure in the development of opposition movements to the wartime government. He was the first in the Reichstag to vote against the war credits and spoke out publicly, as early as January 1915, for the transformation of the national war into a civil or class war. The government conscripted him as a noncombatant but furloughed him to fulfill his duties as a deputy in the Reichstag and the Prussian Assembly. He served on the Düna sector of the Russian front, felling trees, peeling potatoes, and burying the rotting corpses of the dead, until he suffered physical collapse in October 1915. In 1916 he was expelled from the SPD for opposing the party's leadership. The ouster brought him into close alliance with another revolutionary personality, Rosa Luxemburg. Together, they provided the leadership for illegal opposition to the war through the subversive Spartacusbund, which disseminated through its network of confidential underground agents various kinds of revolutionary propaganda. Liebknecht edited the famous illegal "Spartacus Letters," the "official" organ of the Spartakusbund.

Ouster from Social Democratic Party

On May 1, 1916, Liebknecht participated in a May Day demonstration in Berlin and called for the overthrow of the government and an end to the war. The Spartakusbund continued its propaganda activity in the spirit of Liebknecht after the government tried and imprisoned him for his May Day activities. In prison he worked as a shoemaker, heartened by evidence of German disintegration on the military and domestic fronts. In October 1918 the climate in Germany had become more revolutionary and Liebknecht was granted an amnesty by the government of Prince Max of Baden.

Liebknecht entered the maelstrom of the German revolutionary period with great expectations. The Russian Soviet government celebrated his release from prison by a dinner for him at its embassy in Berlin. He planned to develop, through the Spartakusbund, a German revolution after the Soviet pattern. While the SPD, under the leadership of Friedrich Ebert, channelled the revolution along moderate lines, Liebknecht harangued the masses to win support for a "real" revolution. He played a leading role in the formation of the German Communist Party (KPD), which attempted without success to organize the radical elements. A series of bloody clashes between the provisional government formed by Ebert after the downfall of the monarchy and the extreme radicals culminated in the January 1919 *Putsch* in which Liebknecht resorted to force, a tactic both he and his father had strongly opposed. His use of force stimulated the growth of the counterrevolution, and both he and Rosa Luxemburg were among its first victims. On January 15, 1919, they were shot to death by counterrevolutionary volunteers on the pretext of attempted escape while under arrest. Their deaths paved the way for Soviet influence in the KPD, which, in turn, provided a stimulant for the growth of German political reaction.

BIBLIOGRAPHY. Sympathetic biographies of the Liebknechts were written in German by their contemporaries. KURT EISNER, *Wilhelm Liebknecht, sein Leben und Wirken* (1900), is the most available reference, written by a left-wing Socialist and journalist. HARRY SCHUMANN, *Karl Liebknecht*, rev. ed. (1923), renders a good profile of Liebknecht as a person; K.W. MEYER, *Karl Liebknecht: Man Without a Country* (1957), the only recent and English-language biography, stresses his political activities.

(K.W.M.)

Liechtenstein

The tiny Principality of Liechtenstein is one of the smallest states of Europe. Nestled between Switzerland and Austria, it covers 62 square miles (160 square kilometres), an area just smaller than Washington, D.C. The total population of about 22,000 is slightly smaller than that of Lugano, Switzerland. The national capital is at Vaduz.

Throughout most of its history, Liechtenstein has been a quiet, rural corner of the world that was largely unaffected by its European neighbours. After World War II, however, the country underwent a remarkably rapid period of industrialization, which has transformed the economy but which has not as yet been attended by environmental pollution or social dislocation.

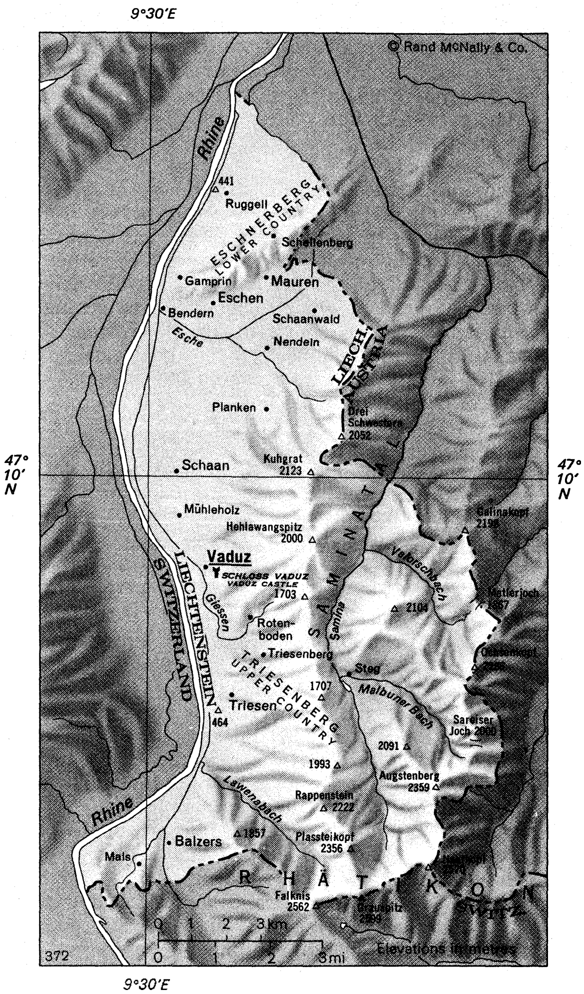
The landscape. *Relief and climate.* The eastern two-thirds of the country is composed of the rugged foothills of the Rhatikon Massif, part of the central Alps. The mountains rise to altitudes between 5,900 and 8,600 feet. Their lower slopes are clothed by evergreen forests and alpine flowers, while their bare peaks are blanketed by snow. The mountains contain three major valleys and are drained by the Samina River. The western section of the country is occupied by the Rhine River floodplain. This, together with the valley of the Ill River, forms a triangular lowland widening northward. The river valley was once marshy, but a drainage channel built in the 1930s has made its rich soils highly suitable for agriculture.

The climate is mild and is greatly affected by the warm southerly wind known as the foehn. The annual totals of rainfall and snowfall range, according to location, between 36 and 45 inches. In winter the temperature rarely falls below 5° F (−15° C), while in summer the average temperature varies from 68° to 82° F (20° to 28° C). These conditions allow for the cultivation of grapes and maize (corn), unusual in a mountainous area.

Vegetation and animal life. Small though it is, Liechtenstein has a remarkable variety of vegetation. Water milfoil and mare's-tail as well as reeds, bulrush, bird's eye primrose, and orchids can be found. The forests comprise a mixed woodland with copper beeches, common and Norway maple, sycamore, lime, elm, and ash. As to the animal life, Liechtenstein is rich in red deer, roe deer, chamois, hare, marmot, blackcock, pheasant, hazel grouse, partridge, fox, badger, marten, polecat, stoat, weasel, and others.

Ancient lordships

Traditional regions. The Rhine plain has always been the focus of settlement. For centuries, the valley was occupied by two independent lordships of the Holy Roman Empire, Vaduz and Schellenberg. Although they were united under the princes of Liechtenstein in 1719, they are still recognized as unique regions—the Upper Country and the Lower Country—and they form separate electoral districts.



LIECHTENSTEIN

The population. *Ethnic origins.* The Liechtensteiners are descended from the Alemanni tribe that came into the region after AD 500. Although the official language is German, the population still speaks an Alemanni dialect containing local variations in pronunciation and vocabulary. The Walsers, descendants of immigrants from the Swiss canton of Valais, settled in Triesenberg at the end of the 13th century and continue to speak a particularly distinctive form of the language. Over 90 percent of the population is Roman Catholic.

Demography. The total population numbered 21,350 at the census of 1970. To the south, the more industrial Upper Country contains the communes of Vaduz, Balzers, Triesen, Triesenberg, Schaan, and Planken. The Lower Country, on the north, is divided into the communes of Eschen, Mauren, Gamprin, Ruggell, and Schel-

Liechtenstein, Area and Population

	area		population	
	sq mi	sq km	1960 census	1970 census
Gemeinden (communes)				
Balzers	7.58	19.62	2,115	2,704
Eschen	3.99	10.33	1,571	2,114
Gamprin	2.37	6.14	434	660
Mauren	2.88	7.45	1,536	2,055
Planken	2.05	5.30	122	177
Ruggell	2.85	7.37	783	866
Schaan	10.37	26.85	3,022	3,890
Schellenberg	1.37	3.55	444	513
Triesen	10.17	26.35	1,789	2,637
Triesenberg	11.49	29.77	1,414	1,813
Vaduz	6.67	17.28	3,398	3,921
Total Liechtenstein	61.78*	160.01	16,628	21,350

*Converted area figures do not add to total given because of rounding.
Source: Official government figures.

lenberg. As a result of postwar industrialization, there has been a movement of people to the larger communes *(Gemeinden)*. The most populous communes are Vaduz, the administrative and commercial centre, and Schaan, the principal industrial community. The birthrate is 1.95 percent of the population and the mortality rate a low 0.78 percent.

The economy. Liechtenstein has no natural resources, and virtually all raw materials, including wood, have to be imported. Industrialization, therefore, has as yet been free of the pollution problems suffered by other developed areas. All of the principality's 10,400 acres of forest are protected in order to maintain the ecology of the mountain slopes and to guard against erosion. There is no heavy industry, and the small manufacturing concerns—the largest employs about 1,000 workers—are spread throughout the country.

Of the working population, about 46 percent are employed in industry and trades; 15 percent in construction; 6 percent in agriculture; 25 percent in commerce, banking, transportation, and the hotel industry; and 8 percent in other occupations.

Growth of industry

The burgeoning industrial economy has been borne on a larger European economic expansion. Total exports rose from 15,200,000 Swiss francs (SFr. 4.373=$1 U.S.; SFr. 10.495=£1 sterling, on December 1, 1970) in 1950 to more than SFr. 400,000,000 in 1972. Production includes metalworking, pharmaceuticals, food processing, and the manufacture of consumer goods.

In 1921 Liechtenstein adopted Swiss currency, and in 1924 it joined the Swiss customs union. By the early 1970s, over half of the exports went to the European Free Trade Area (including Switzerland and Finland) and a third to the European Economic Community.

Despite the small and declining percentage of the population engaged in agriculture, production rose by the 1970s to exceed SFr. 5,000,000. Farming units have been increasing in size, and the largest concerns concentrate on stock breeding and dairying. Maize (corn) and potatoes are traditional crops, but the production of cereals and vegetables is steadily increasing. Vineyards are few and are split into small units. The Alpine slopes are used for grazing during the three summer months.

Tourism

Tourism is sponsored by the government and generated an annual income in the 1970s exceeding SFr. 15,000,000. Most visitors come from the surrounding European countries and centre their activities around Vaduz. The registration of foreign firms in Liechtenstein provides a source of tax income.

There is a network of excellent roads connecting the country with its neighbours. The railway, part of the Paris–Vienna express route, passes through the northern sections of the country. There is no airport.

Administration and social conditions. *Government structure.* Liechtenstein is a constitutional principality in which governmental power derives from both the prince and the populace. Succession to the throne is hereditary through the male line and is determined by the regulations of the princely house. The constitution of 1921 provides for a unicameral Landtag, or parliament, composed of 15 members who are elected for four-year terms. The government consists of a chief and deputy chief of government and three councillors who are appointed by the prince, on proposition of the Landtag, for four-year terms. Through a civil-service staff of about 250, the government is responsible for all administrative matters. The 11 *Gemeinden* (communes) are governed autonomously—but under government supervision—by mayors and city councils, elected every three years.

Politics. The franchise, of secret and direct ballot, extends to all male citizens over 20 years of age. The citizenry is also guaranteed the right to referendum (submission of legislation to a popular vote), plebiscite (popular choice of a type of government or a ruler), and initiative (popular proposal of legislation). Political activity is channelled through a multi-party system. In 1970 the Patriotic Union (formerly People's) Party won its first majority in 32 years.

Justice. The judiciary is fully independent. There are two administrative courts—the Administrative Tribunal and the State Tribunal—that hear appeals against governmental decisions and decrees, as well as complaints concerning the violation of civil rights. In matters of civil law, original jurisdiction is exercised by the Lower Courts. Criminal cases originate in the Lower Courts for petty offenses, the assize court for misdemeanours, the criminal court for felonies, or the juvenile court. Both civil and criminal cases can be appealed to the High Court and the Supreme Court of Justice. The government maintains a 33-man police force, but the standing army was abolished in 1868.

Social services. Matters of public health are in the hands of a committee of public health, the standing executive watchdog authority, which is headed by a state medical officer. Liechtenstein's small, 30-bed hospital and two maternity wards are supplemented by the excellent neighbouring Swiss facilities, to which the principality contributes support. Social security is sustained by a variety of compulsory insurance schemes to which citizens contribute between the ages of 20 and 65. The financing of these comprehensive plans is shared by employers, employees, and the government, the last also assuming any deficits.

Education is supervised by the National Board of Education. The school system consists of eight-year primary schools, three-year secondary schools, a vocational school, an eight-year grammar school, a five-year commercial high school, a music school, and a technical college offering evening courses. There is also a five-year commerical high school for girls, run by a religious order. All citizens over the age of seven are literate.

Cultural life and institutions. The five-member Advisory Council for Culture and Youth was established by the government in 1964. Its major functions include the encouragement of artistic activity and the promotion of works by Liechtenstein artists.

Art collections and museums

The world-famous art collections of the princes of Liechtenstein, exhibited in the newly enlarged Englanderhaus in the centre of Vaduz, include outstanding works of the Flemish painters Rubens, Van Dyck, Brueghel, and Teniers, as well as of the Dutch painters Ostade, Steen, Rembrandt, Ruisdael, Cuyp, Hobbema, and van de Velde. The State Art Collection, opened in 1969, includes paintings by Frans Hals, Van Dyck, and Jan van Goyen and a modern-art section with graphic art by Picasso, Braque, Barlach, Max Beckmann, Goya, Gustave Klimt, Kokoschka, and Alberto Giacometti.

The Liechtenstein Postal Museum (founded in 1930) exhibits a large stock of stamps, including national issues since 1912 and stamps received since 1921 as a result of Liechtenstein's membership in the Universal Postal Union. The Liechtenstein National Museum in Vaduz houses primarily early and Roman artifacts, but it also contains a collection tracing the history of the arts in Liechtenstein and another collection devoted to the principality's folklore. The Liechtenstein National Library was established in 1961 as a public foundation.

Future prospects. As Liechtenstein continues to develop its industries, it is increasingly faced with certain attendant problems. Its citizens are moving to urban areas, and the rural calm that was so highly prized is beginning to be threatened. The consolidation of agricultural holdings and the expansion of towns increased the value of land, and speculation has become a serious problem. Although industrial pollution is not yet a real threat, the government is alert to such dangers and is guarding against them.

BIBLIOGRAPHY. WALTER KRANZ (ed.), *The Principality of Liechtenstein: A Documentary Handbook* (1967); RUSSELL FRANKLIN MOORE, *The Principality of Liechtenstein: A Brief History* (1960); PRESS AND INFORMATION OFFICE OF THE LIECHTENSTEIN GOVERNMENT, *The Economy of the Principality of Liechtenstein* (n.d.); PIERRE RATON, *Le Liechtenstein: Histoire et institutions*, 2nd ed. (1967; Eng. trans., *History and Institutions of the Principality*, 1970). See also *Jahrbücher des Historischen Vereins für das Fürstentum Liechtenstein*, which is the yearbook of the historical society of the principality of Liechtenstein, published annually since 1900.

(W.K.)

Life

The profusion of life on earth has been studied in great detail, and a number of general principles have been revealed. Foremost among them is the principle of evolution by natural selection — the stepwise adaptation of organisms to their environment with increasing precision by small random mutations, or changes, in their hereditary material — which is the feature that distinguishes living from non-living matter. This article treats first the varieties of definitions of life and then covers, in some detail, the similarities and differences among organisms on earth. It deals with the problem of the origin of life on earth and concludes with a consideration of the possibility of life beyond the earth.

This article is divided into the following sections:

DEFINITIONS OF LIFE

What is known about life

A great deal is known about life. Anatomists and taxonomists have studied the forms and relations of more than a million separate species of plants and animals. Physiologists have investigated the gross functioning of organisms. Biochemists have probed the biological interactions of the organic molecules that make up life on our planet. Molecular biologists have uncovered the very molecules responsible for reproduction and for the passage of hereditary information from generation to generation, a subject that geneticists had previously studied without going to the molecular level. Ecologists have inquired into the relations between organisms and their environments, ethologists the behaviour of animals and plants, embryologists the development of complex organisms from a single cell, evolutionary biologists the emergence of organisms from pre-existing forms over geological time. Yet despite the enormous fund of information that each of these biological specialties has provided, it is a remarkable fact that no general agreement exists on what it is that is being studied. There is no generally accepted definition of life. In fact, there is a certain clearly discernible tendency for each biological specialty to define life in its own terms. The average person also tends to think of life in his own terms. For example, the man in the street, if asked about life on other planets, will often picture life of a distinctly human sort. Many individuals believe that insects are not animals, because by "animals" they mean "mammals." Man tends to define in terms of the familiar. But the fundamental truths may not be familiar. Of the following definitions, the first two are in terms familiar in everyday life; the next three are based on more abstract concepts and theoretical frameworks.

Physiological. For many years a physiological definition of life was popular. Life was defined as any system capable of performing a number of such functions as eating, metabolizing, excreting, breathing, moving, growing, reproducing, and being responsive to external stimuli. But many such properties are either present in machines that nobody is willing to call alive, or absent from organisms that everybody is willing to call alive. An automobile, for example, can be said to eat, metabolize, excrete, breathe, move, and be responsive to external stimuli. And a visitor from another planet, judging from the enormous numbers of automobiles on the earth and the way in which cities and landscapes have been designed for the special benefit of motorcars, might well believe that automobiles are not only alive but are the dominant life form on the planet. Man, however, professes to know better. On the other hand, some bacteria do not breathe at all but instead live out their days by altering the oxidation state of sulfur.

Metabolic. The metabolic definition is still popular with many biologists. It describes a living system as an object with a definite boundary, continually exchanging some of its materials with its surroundings, but without altering its general properties, at least over some period of time. But again there are exceptions. There are seeds and spores that remain, so far as is known, perfectly dormant and totally without metabolic activity at low temperatures for hundreds, perhaps thousands, of years but that can revive perfectly well upon being subjected to more clement conditions. A flame, such as that of a candle in a closed room, will have a perfectly defined shape with fixed boundary and will be maintained by the combination of its organic waxes with molecular oxygen, producing carbon dioxide and water. A similar chemical reaction, incidentally, is fundamental to most animal life on earth. Flames also have a well-known capacity for growth.

Biochemical. A biochemical or molecular biological definition sees living organisms as systems that contain reproducible hereditary information coded in nucleic acid molecules and that metabolize by controlling the rate of chemical reactions using proteinaceous catalysts known as enzymes (*q.v.*). In many respects, this is more satisfying than the physiological or metabolic definitions of life. There are, however, even here, the hints of counterexamples. There seems to be some evidence that a virus-like agent called scrapie contains no nucleic acids at all, although it has been hypothesized that the nucleic acids of the host animal may nevertheless be involved in the reproduction of scrapie. Furthermore, a definition strictly in chemical terms seems peculiarly vulnerable. It implies that were man able to construct a system that had all the functional properties of life, it would still not be alive if it lacked the molecules that earthly biologists are fond of — and made of.

Instructions for living systems

Genetic. All organisms on earth, from the simplest cell to man himself, are machines of extraordinary powers, effortlessly performing complex transformations of organic molecules, exhibiting elaborate behaviour patterns, and indefinitely constructing from raw materials in the environment more or less identical copies of themselves. How could machines of such staggering complexity and such stunning beauty ever arise? The answer, for which today there is excellent scientific evidenne, was first discerned by the evolutionist Charles Darwin in the years before the publication in 1859 of his epoch-making work, the *Origin* of Species. A modem rephrasing of his theory of natural selection goes something like this: Hereditary information is carried by large molecules known as genes, comprised of nucleic acids. Different genes are responsible for the expression of different characteristics of the organism. During the reproduction of the organism the genes also reproduce, or replicate, passing the instructions for various characteristics on to the next generation. Occasionally, there are imperfections, called mutations, in gene replication. A mutation alters the instructions for a particular characteristic or characteristics. It also breeds true, in the sense that its capability for determining a given characteristic of the organism remains unimpaired for generations until the mutated gene is itself mutated. Some mutations, when expressed, will produce characteristics favourable for the organism; organisms with such favourable genes will reproduce preferentially over those without such genes. Most mutations, however, turn out to be deleterious and often lead to some impairment or to death of the organism. To illustrate, it is unlikely that one can improve the functioning of a finely crafted watch by dropping it from a tall building. The watch may run better, but this is highly improbable. Organisms are so much more finely crafted than the finest watch that any random change is even more likely to be deleterious. The accidental beneficial and inheritable change, however, does on occasion occur; it results in an organism better adapted to its environment. In this way organisms slowly evolve toward better adaptation, and, in most cases, toward greater complexity. This evolution occurs, however, only at enormous cost: man exists today, complex and reasonably well adapted, only because of billions of deaths of organisms slightly less adapted and somewhat less complex. In short, Darwin's theory of natural selection states that complex organisms developed, or evolved, through time because of replication, mutation, and replication of mutations. A genetic definition of life therefore would be: a system capable of evolution by natural selection.

Genetic definition of life

This definition places great emphasis on the importance of replication. Indeed, in any organism enormous biological effort is directed toward replication, although it confers no obvious benefit on the replicating organism. Some organisms, many hybrids for example, do not replicate at all. But their individual cells do. It is also true that life defined in this way does not rule out synthetic duplication. It should be possible to construct a machine that is capable of producing identical copies of itself from preformed building blocks littering the landscape but that arranges its descendants in a slightly different manner if there is a random change in its instructions. Such a machine would, of course, replicate its instructions as well. But the fact that such a machine would satisfy the genetic definition of life is not an argument against such a definition; in fact, if the building blocks were simple enough, such a machine would have the capability of evolving into very complex systems that would probably have all the other properties attributed to living systems. The genetic definition has the additional advantage of being expressed purely in functional terms: it does not depend on any particular choice of constituent molecules. The improbability of contemporary organisms—dealt with more fully below—is so great that these organisms could not possibly have arisen by purely random processes and without historical continuity. Fundamental to the genetic definition of life then is the belief that a certain level of complexity cannot be achieved without natural selection.

Life as an ordered system

Thermodynamic. Thermodynamics distinguishes between open and closed systems. A closed system is isolated from the rest of the environment and exchanges neither light, heat, nor matter with its surroundings. An open system is one in which such exchanges do occur. The second law of thermodynamics states that, in a closed system, no processes can occur that increase the net order (or decrease the net entropy) of the system (see THERMODYNAMICS). Thus the universe taken as a whole is steadily moving toward a state of complete randomness, lacking any order, pattern, or beauty. This fate has been known since the 19th century as the heat death of the universe. Yet living organisms are manifestly ordered and at first sight seem to represent a contradiction to the second law of thermodynamics. Living systems might then be defined as localized regions where there is a continuous increase in order. Living systems, however, are not really in contradiction to the second law. They increase their order at the expense of a larger decrease in order of the universe outside. Living systems are not closed but rather open. Most life on earth, for example, is dependent on the flow of sunlight, which is utilized by plants to construct complex molecules from simpler ones. But the order that results here on earth is more than compensated by the decrease in order on the sun, through the thermonuclear processes responsible for the sun's radiation.

Some scientists argue on grounds of quite general open-system thermodynamics that the order of a system increases as energy flows through it, and moreover that this occurs through the development of cycles. A simple biological cycle on the earth is the carbon cycle. Carbon from atmospheric carbon dioxide is incorporated by plants and converted into carbohydrates through the process of photosynthesis. These carbohydrates are ultimately oxidized by both plants and animals to extract useful energy locked in their chemical bonds. In the oxidation of carbohydrates, carbon dioxide is returned to the atmosphere, completing the cycle. It has been shown that similar cycles develop spontaneously and in the absence of life by the flow of energy through a chemical system. In this view, biological cycles are merely an exploitation by living systems of those thermodynamic cycles that pre-exist in the absence of life. It is not known whether open-system thermodynamic processes in the absence of replication are capable of leading to the sorts of complexity that characterize biological systems. It is clear, however, that the complexity of life on earth has arisen through replication, although thermodynamically favoured pathways have certainly been used.

The existence of diverse definitions of life surely means that life is something complicated. A fundamental understanding of biological systems has existed since the second half of the 19th century. But the number and diversity of definitions suggest something else as well. As detailed below, all the organisms on the earth are extremely closely related, despite superficial differences. The fundamental ground pattern, both in form and in matter, of all life on earth is essentially identical. As will emerge below, this identity probably implies that all organisms on earth are evolved from a single instance of the origin of life. It is difficult to generalize from a single example, and in this respect the biologist is fundamentally handicapped as compared, say, to the chemist or physicist or geologist or meteorologist, who now can study aspects of his discipline beyond the earth. If there is truly only one sort of life on earth, then perspective is lacking in the most fundamental way.

LIFE ON EARTH

Mechanism and vitalism. Human beings are ambulatory collections of some 10^{14} cells. Human cells are in many fundamental respects similar to those that make up all the other animals and plants on the earth. Each cell typically consists of a central, usually spherical, nucleus and an outer more heterogeneous region, termed the cytoplasm. The substance of nucleus and cytoplasm together has for many decades been called protoplasm. Use of this term implied that there was some special substance underlying living organisms. In the use of the word protoplasm there is occasionally an implication that life cannot be explained solely by physics and chemistry, that some mysterious "vital force" must be invoked. A living cell is a marvel of detailed and complex architecture. Seen through a microscope there is an appearance of almost frenetic activity. On a deeper level it is known that molecules are being synthesized at an enormous rate. Almost any enzyme catalyzes the synthesis of more than 100 other molecules per second. In ten minutes, a sizable fraction of the total mass of a metabolizing bacterial cell has been synthesized. The information content of a simple cell has been estimated as around 10'' bits, comparable to about a hundred million pages of the *Encyclopædia Britannica.* Faced with all this or its equivalent, it is not surprising that early biologists felt despair at ever being able to understand the detailed workings of life.

A Stone Age man, confronted for the first time with a watch, might also deduce that there was some special watch substance in nature, or perhaps even a god of the watch. In ancient times, the most common of biological activities, such as the hatching of an egg or the blooming of a flower, were attributed to the intercession of a deity. After the epochal work of Sir Isaac Newton, when the motion of the planets and comets of the solar system was predictable to some very great precision and understood on the basis of an underlying principle, the idea developed that organisms were also nothing more than a particularly intricate kind of clockwork. But when early investigations failed to unveil the clockwork, a kind of ghostly mainspring was invented—the "vital force." This force was a

"Vital force"

rebellion from mechanistic biology, an explanation of all that mechanism could not explain or for which mechanism could not be found. It also appealed to those who felt debased by the implication that they were "nothing more" than a collection of atoms, that their urges and apparent free wills arose merely from the interaction of an enormously large number of molecules in a way that, although too complex to use predictably, was in principle determined.

Not only is there no evidence for a vital force but the idea itself is hardly thought out; it is a sort of catchall concept, covering anything otherwise inexplicable. The alternative approach, that all organisms are made of atoms and nothing else, has proven especially useful and has led to a fundamental new understanding of biological systems. This situation does not imply, of course, that atoms cannot be put together in so complex a way that their collective behaviour is too difficult to understand in terms of the individual atoms; in this sense there may be particular laws of biology not readily derivable from the elementary interaction of atoms. But this is a very different thing from a vital force. Indeed, there is nothing debasing in the thought that man is made of atoms alone; it means he is intimately connected with the matter that comprises the inanimate universe. What a wonder that atoms can be put together in so complex a pattern as to produce man. Man is a tribute to the subtlety of matter. As the American anthropologist Loren Eiseley has written, ". . . if 'dead' matter has reared up this curious landscape of fiddling crickets, song sparrows, and wondering men, it must be plain even to the most devoted materialist that the matter of which he speaks contains amazing, if not dreadful powers. . . ." *(The Immense* Journey, Random House, New York, 1957.)

Nucleic acids. It is now known that many if not all of the fundamental properties of cells are a function of their nucleic acids, their proteins, and the interactions among these molecules. Within the nuclear regions of cells is a mélange of twisted and interwoven fine threads, the chromosomes. During cell division, in all but the simplest organisms, the chromosomes display an elegantly choreographed movement, separating so that each daughter cell of the original cell receives an equal complement of chromosomal material. This pattern of segregation corresponds in all details to the theoretically predicted pattern of segregation of the genetic material implied by the fundamental genetic laws (see HEREDITY). The chromosomes are composed of nucleic acids and proteins in a combination called nucleoprotein. The nucleic acid stripped of its protein is known to carry genetic information and to regulate cellular metabolism; the protein in nucleoprotein undoubtedly plays some secondary, probably regulatory, role.

The role of nucleic acids

The specific carrier of the genetic information in higher organisms is a nucleic acid known as DNA, short for deoxyribonucleic acid. DNA is a double helix, two molecular coils wrapped around each other and chemically bound one to another by bonds connecting adjacent bases. Each helix has a backbone that consists of a long sequence of alternating sugars and phosphates. Attached to each sugar is a base. Each sugar–phosphate–base combination is called a nucleotide; a nucleic acid strand can be thought of as a sequence of nucleotides. There is a very significant one-to-one base pairing in the connection of adjacent helices, in the sense that once the sequence of bases along one helix is specified, the sequence along the other is also specified. The specificity of base pairing plays a key role in the replication of the DNA molecule, where each helix makes an identical copy of the other from molecular building blocks in the cell. These nucleic acid replication events are mediated by enzymes, and with the aid of enzymes have been produced in the laboratory.

Ribonucleic acid (RNA) differs from DNA in having a slightly different five-carbon sugar, and in replacing one of the four bases that make up DNA by a slightly different base. RNA does not appear to exist in a double-stranded form. Now DNA, RNA, and the enzymes have a curiously interconnected relation, which appears ubiquitous in all organisms on earth today.

Commonalities among organisms on earth. The genetic code was broken in the 1960s. It was found that three consecutive nucleotides code for one amino acid of a protein molecule; *e.g.,* an enzyme. By controlling the synthesis of enzymes, the nucleic acids control the functioning of the cell. Of the four different bases taken three at a time, there are 43 = 64 possible combinations. The meaning of each of these combinations, or codons, is known. Most of them represent a particular amino acid. A few of them represent punctuation marks; for example, instructions to start or stop a synthesis. Some of the code is degenerate in the sense that more than one nucleotide triplet may specify a given amino acid. These interactions among nucleic acids and proteins seem absolutely central to living processes on earth today. Not only are these processes apparently the same in all organisms on earth but even the particular dictionary that is used for the transcription of nucleic acid information into protein information seems to be essentially the same in all organisms. Moreover, this code has various chemical advantages over other conceivable codes. The complexity, ubiquity, and advantages of these processes clearly argue that the present interactions among proteins and nucleic acids are themselves the product of a long evolutionary history. At the time of the origin of life this very complex replication and transcription apparatus could of course not have been in operation. A fundamental problem in the origin of life is the question of the origin and early evolution of the genetic code.

Energy-storing molecules

There are many other commonalities among organisms on earth. For example, there is only one class of molecules that store energy for biological processes until the cell has use for it, and these molecules are all nucleotide phosphates. The most common example is ATP (adenosine triphosphate). For this very different function, a molecule identical to the building blocks of the nucleic acids is employed. There are metabolically important molecules —*e.g.*, molecules known as FAD (flavin adenine dinucleotide) and coenzyme A—which include subunits similar to the nucleotide phosphates. Porphyrins represent another category of ubiquitous molecules. Porphyrins are the chemical basis of hemoglobin, which carries oxygen molecules through the bloodstreams of animals; of chlorophyll, which is the fundamental molecule mediating photosynthesis in plants; and of the colours that many animals display. The left- or right-handedness of many biological molecules--discussed more fully below—runs identically through all organisms on earth. In fact, of the billions of possible organic compounds, less than 1,500 are employed by contemporary life on earth, and these 1,500 are constructed from less than 50 simple molecular building blocks. Similarly, organisms as diverse as paramecia and human sperm cells have little whiplike appendages called cilia or flagella used to propel themselves through their liquid environments. The cross-sectional structure of the cilia and flagella is almost always nine pairs of peripheral and one pair of internal fibres. There is no immediately obvious selective advantage of the 9:1 ratio. These commonalities indicate that a few basic chemical and functional patterns are being used over and over again, reflecting the extremely close relations all organisms on earth have to one another. Many biologists believe that these commonalities—particularly where no obvious selective advantage exists—imply that all organisms on the earth are descendants of a single common ancestor. But it is possible that there are more subtle selective advantages. The issue may have to await the first detailed study of an extraterrestrial organism.

The number of possible ways of putting nucleotides together in a chromosome is enormous. The renowned geneticist H. J. Muller estimated that in a human chromosome there are about 4 x 10^9 base pairs. Each base pair position could be filled by any one of four possible bases; accordingly, the number of possible varieties of human chromosomes is $44^{\times 10^9}$ ($102.4^{\times 10^9}$), an inconceivably large number. By contrast, the number of elementary particles (electrons and protons) in the entire physical universe is only about 10^{80}. Thus a human being is an extraordinarily improbable object. Most of the $10^{2.4 \times 10^9}$

possible sequences of nucleotides would lead to complete biological malfunction. Our nucleotides work because natural selection, over a 4,000,000,000-year history of life, has destroyed enormous numbers of combinations that did not work. But there still may be combinations that work far better than any now present, and the future holds the promise that man will be able to assemble nucleotides in any desired sequence to produce whatever characteristics of human beings are thought desirable, an awesome and disquieting prospect.

Metabolism. The chemical bonds that make up living organisms have a certain probability of spontaneous breakage. Accordingly, mechanisms must exist to repair this damage, or to replace the broken molecules. In addition, the meticulous control that cells exercise over their internal activities requires the continued synthesis of new molecules. These processes of synthesis and breakdown of the organic molecules of the cell are collectively termed metabolism, and for synthesis to keep ahead of the thermodynamic tendencies toward breakdown, energy must be supplied to the living system. Organisms acquire this energy by two general methods. Some organisms are heterotrophs, acquiring their energy by the controlled breakdown of pre-existing organic molecules (food)—generally those supplied by other organisms. Human beings and most other animals are heterotrophs. Alternatively, organisms may be autotrophs, acquiring their useful free energy from some other source, either from the energy of sunlight, in which case the organism is called a photoautotroph, or from the controlled chemical reaction of inorganic materials, in which case the organism is known as a chemoautotroph. Organisms that use both modes are called photochemoautotrophs.

Methods of acquiring energy

A green plant is a typical example of a photoautotroph. It uses sunlight to break water into oxygen and hydrogen. Hydrogen is then combined with carbon dioxide to produce such energy-rich organic molecules as ATP and carbohydrates, and the oxygen is released back into the atmosphere. Many animals, on the other hand, utilize the atmospheric oxygen to combine chemically with organic materials they have eaten and release carbon dioxide and water as waste products in extracting energy from the organic materials. This is an example of an ecological cycle in which a material (here carbon) is pumped through two different organisms.

More generally, such metabolic cycles—used by the organism to extract useful energy from the environment—can be described in terms of oxidation-reduction reactions. In the case of respiration, molecular oxygen accepts electrons from glucose or other sugars. The oxygen is said to be an electron acceptor (it has a great affinity for electrons), the glucose an electron donor. This is the prototype of oxidation-reduction reactions, but not all such reactions necessarily involve oxygen. Biological electron acceptors other than oxygen include nitrates, sulfates, carbonates, nitrogen, and methanol. Biological electron donors other than sugars include nitrogen, sulfides, methane, ammonia, and methanol. For acceptor-donor transformations to occur over any period of time, biological cycles are necessary. It is possible that, for geologically short periods of time, organisms have lived off a finite supply of material, but for any long-term continuance of life, a dynamic cycling of matter, involving at least two different varieties of organisms, is necessary. If there is life on other planets, a similar cycling must exist. A search for such molecular transformations is one method of detecting extraterrestrial life.

On the earth, all such useful biological electron transfer reactions lead to the net production of one or more molecules of ATP. Two of the three phosphates of this molecule are held by "energy-rich" bonds sufficiently stable to survive for long periods of time in the cell, but not so strong that the cell cannot tap these bonds for energy when needed. ATP and very similar molecules, all of them having a base, a five-carbon sugar, and three phosphates, are, so far as is known, the general and unique energy currency of living systems on earth.

Metabolic processes do not occur in one step. The ordinary sugar, glucose, is not oxidized to carbon dioxide and water by living cells in the same way that occurs if a flame is applied to glucose in air. The resulting release of energy would be much too sudden, and concentrated in too small a volume, for such a process to be utilized safely by the cell. Instead, the glucose is broken down by a series of successive and coordinated steps, each mediated by a particular and specific enzyme. In almost all organisms that metabolize glucose, the sugar is first broken down in a set of anaerobic steps (that is, in the absence of oxygen). The total number of such steps is about 11. Some organisms are anaerobes; that is, they do not utilize molecular oxygen. In them the anaerobic steps are as far as the glucose metabolism is carried. Other organisms, including man, carry the oxidation of glucose further, gingerly combining glucose breakdown products with molecular oxygen. Such aerobic oxidation of glucose requires about 60 more enzymatically catalyzed steps. Another indication of the relative simplicity of the anaerobic breakdown of sugar is that all the enzymes used are free in solution in the cell; the aerobic steps use enzymes that are localized in specific regions of the cell. The complete aerobic breakdown of sugar to carbon dioxide and water is about ten times more efficient than the breakdown accomplished by anaerobes; ten times as many ATP molecules are produced. Similar themes and variations exist for the metabolism of other molecules (see METABOLISM).

Efficiency of aerobic oxidation

The energy made available in this way to ATP is used in a variety of ways by the cell; for example, for motility. When an amoeba extends pseudopods, or a man walks, ATP molecules are being tapped for their energy-rich phosphate bonds. In addition, ATP molecules are used for the synthesis of molecules that the organism needs and does not have available. Among such molecules may be amino acids, the particular five-carbon sugars involved in nucleic acids, the nucleic acid bases, and so on. Each of these synthetic processes is again controlled and enzymatically mediated and may start from a variety of building blocks available to the organism, some simple, some more complex. For example, the amino acid L-leucine is produced from pyruvic acid, itself the product of the anaerobic breakdown of glucose. Synthesis of L-leucine from pyruvic acid involves eight enzyme-mediated steps and the addition of acetic acid and water.

These exquisitely interlocked and controlled metabolic steps are not usually performed in a diffuse manner all over the cell. Instead there is, at least in all higher organisms, a marvellously architectured cellular interior with particular specialized regions where particular chemical reactions are performed. Those oxidation-reduction reactions that involve molecular oxygen occur in an inclusion within the cytoplasm called the mitochondrion. The mitochondrion itself has an intricate substructure, and particular enzymes are thought to reside in particular sites within it; the molecule being metabolized may be passed on from one enzyme to another as through a conveyor belt in a factory. Similarly, photosynthesis occurs in a cytoplasmic inclusion called a chloroplast, which contains the chlorophyll and other pigments that absorb visible light, as well as the detailed enzymatic apparatus for the photosynthetic process. Chloroplasts and mitochondria, as well as other cytoplasmic inclusions at the base of flagella and cilia, all contain DNA. Moreover, this DNA has a somewhat different distribution of bases from that of the nucleus. It has been suggested that the cytoplasmic inclusions are the remnants of once free-living forms that, because of the favourable conditions found there, have taken up residence in the insides of other organisms.

The work of the mitochondrion

Nucleic acids are known to pass from cell to cell and to perform their replication and coding functions efficiently in the new cell. In fact, viruses are essentially strands of nucleic acid, with a protein coat, that operate in just this way. It is also known that pieces of the genetic material from one cell may migrate into another cell of the same species and produce genetic and permanently heritable changes there. Alternatively, part of the virus nucleic acid may be permanently bound to the nuclear DNA of a host cell. It is likely that a virus is a degenerate form, now highly specialized to live off a specific host, of an organism once free-living and much more generally capable of performing a wide range of metabolic tasks. A virus must use the genetic transcription apparatus of its host cell. Many viruses do this extremely efficiently, turning a bacterium from a factory for making other bacteria into a factory for making viruses. In some cases it takes no more than ten minutes for a bacterium infected by a single virus to produce a hundred new virus particles, which then burst forth from the host bacterium, destroying it. The line between benign or useful cytoplasmic inclusions and infective agents is not a very sharp one (see VIRUS).

Differences in nuclear structure

Eucaryotes and procaryotes. In the very simplest one-celled organisms one may distinguish between eucaryotic and procaryotic cells. Many familiar one-celled organisms, such as paramecia and amoebas, as well as the cells of all higher organisms including man, are eucaryotic. Such cells undergo mitosis, a fundamental sequence of events that occurs after DNA replication and that ensures that the DNA is precisely and equally distributed to the daughter cells. Eucaryotic cells have nucleoprotein in their nuclei, There is a membrane that separates the nucleus from the cytoplasm; Mitochondria are generally present in the cytoplasm, as is a very intricately convoluted structure, called the endoplasmic reticulum, that probably serves as the anchoring point for many cytoplasmic enzymes not contained in such inclusions as mitochondria or chloroplasts.

On the other hand there are procaryotic cells, which are most generally typified by the bacteria and the blue-green algae. In these cells nuclear division is nonmitotic, there is no nucleoprotein, and a nuclear membrane is absent. While eucaryotic cells may have more than one chromosome, procaryotic cells have one chromosome only, and that one is dispersed in the cytoplasm. Mitochondria, chloroplasts, and the endoplasmic reticulum are always absent. It is clear from this description that the procaryotes are in many respects more primitive than the eucaryotes. A basic unsolved evolutionary question concerns the evolution of procaryotes into eucaryotes.

An interesting subject of biological speculation concerns what the smallest and simplest contemporary free-living organism might be. The smallest free-living cells now known are the pleuropneumonia-like organisms (PPLO). While an amoeba has a mass of 5×10^{-7} grams, a PPLO weighs 5×10^{-16} grams and is only about $^1/_{10}$ of a micrometre across. It can be seen only in the electron microscope. Such organisms grow very slowly. There may be smaller organisms that grow even more slowly, but they would be extremely difficult to detect. Even an organism of the size of PPLO has room for only about a hundred enzymes. A much smaller organism would have room for many fewer enzymes, and its ability to accomplish the functions that contemporary living systems must accomplish would be severely compromised. Were there, however, an environment in which all the necessary organic building blocks and such energy sources as ATP were provided "free," a functioning organism could be substantially smaller than PPLO. In fact the inside of a cell provides just such an environment; this is why infectious agents, such as viruses, can be substantially smaller than PPLO. But it must be emphasized that such agents are not free-living organisms.

Metazoa, embryology, and sex. The distinction between single-celled and many-celled organisms (in animals, between protozoa and metazoa) is far from a sharp one. An interesting illustration is the slime molds, which undergo an extraordinary sequence of events during their life cycle. The cycle begins with single cells, somewhat like amoebas, which swarm, or combine, into a slimy mass with many nuclei called a plasmodium. The plasmodium in turn forms a sluglike mass that is certainly a multicelled organism. The slug develops into a stalked, fruitlike sporangium, still multicellular. The sporangium produces spores with cellulose cell walls similar to those of plants. The spores in turn germinate into small cells bearing flagella. The flagella are lost and the life cycle is completed with the production of an amoeboid form (for details, see SLIME MOLD).

Biology is replete with life cycles of comparable complexity. The swarming of individual cells to form a plasmodium may in fact be an example of the events that led to the production of metazoa in the early history of the earth. Such life cycles, while apparently very exotic, are shared by many organisms, including man, where a one-celled, free-swimming sperm stage is part of the life cycle.

The problem of development

The life cycle of slime molds, or men, or any other multicellular organism brings up a fundamental and still largely unsolved problem. These organisms develop from a single cell that has a single complement of the genetic material. These cells then divide, forming many identical cells. The very early embryology of man goes through stages with 2, 4, 8, 16, etc., cells. Since the genetic information is identical in each cell, how does it ever happen that cells become specialized, forming hair cells, teeth cells, liver cells, blood cells, or bone cells? How can any given cell "know" what sort of specialized cell it must become, since all cells contain identical nucleic acids? Possibly the answer to this question has to do with geometry. After the 16- or 32-cell stage, there is a distinct difference between a cell on the inside of the embryo, which is entirely surrounded by cells, and a cell on the outside of the embryo, which is not entirely surrounded by other cells. One of the earliest major steps in embryonic development is a distinction in function between interior cells (the endoderm) and exterior cells (the ectoderm). There are physical and chemical interactions among adjacent cells. Perhaps any cell then has the capability of becoming any specialized cell, but cells are, as a result of their external cellular environment, called upon to develop in different ways. Occasional embryonic anomalies or cysts occur in which, for example, hair or teeth develop in totally inappropriate portions of the body. Similarly, eyes have been caused to develop on the limbs of frogs. Such incidents demonstrate the capability of the "wrong" cells to produce particular cellular specializations (see GROWTH).

Much of the beauty and diversity of contemporary life on earth is due to sex. A totally asexual organism will be genetically identical to its (single) parent except for occasional mutations. The development of any major new adaptation would then require the acquisition of large numbers of appropriate mutations. Consider, for example, how the ability of an organism to metabolize a given molecule depends on the interaction of many enzymes, each produced by the transcription of the genetic information in hundreds of nucleotides, each nucleotide being the product of a single mutation. Thus, the chance development of any advantageous adaptation in an asexual organism requires the mutations to wait in line for a fortuitous juxtaposition.

Sex: an added dimension

Sex solves this problem in an elegant way. The genetic material of the parents is reassorted so that totally new combinations of genes are produced. In this way mutations acquired by any member of the population are rather quickly distributed to other members, and mutations arising in separate organisms can be combined. The likelihood of producing a useful sequence of mutations is thereby greatly enhanced. The advantages of sexual reproduction are so great that even many simple forms, such as bacteria or protozoa, which largely reproduce asexually, have occasional sexual encounters. While two sexes are clearly adequate for such a random assortment of genetic material, some organisms have developed more sexes: paramecia, for example, have somewhere between five and ten sexes, defined in terms of the elaborate taboos about which organisms can combine their genetic material. In the process of genetic reassortment, some organisms make a very large number of attempts; for example, frogs lay millions of eggs at a time, and the number of sperm cells in a single human ejaculation is about 3×10^8 (see SEX AND SEXUALITY).

The varieties of organisms and environments. The environment of the earth is heterogeneous. There are mountains, oceans, and deserts, extremes of temperature and humidity. In addition, there are diverse microenvironments: oxygen-depleted oceanic oozes, ammonia-rich soils, mineral deposits with a high radioactivity content, and so on. The environment of an organism also includes the other organisms in its surroundings. For each of these environmental situations there are corresponding ecological niches, and the variety of ecological niches populated on the earth is quite remarkable. Furthermore, ecological niches can be filled independently several times. For example, quite analogous to the ordinary mammalian wolf is the marsupial wolf that lives in Australia; the two have striking similarities in physical appearance and in predation behaviour. As another example, the same streamlined shape for high-speed marine motion has evolved independently at least three times: in *Stenopterygius* and other Mesozoic reptiles; in the tuna, which are fish; and in the dolphins, which are mammals. This case of convergent evolution must arise from the fact that hydrodynamics admits a narrow range of solutions to the problem of high-speed marine motion by large animals. Similarly, the eye has independently evolved several times among animals on the earth; apparently such a structure is the best solution to the problem of visual recording. In those cases where physics or chemistry establishes one most efficient solution to a given ecological problem, natural selection will often tend to reach the solution, but not always. Some adaptations of undoubted utility, such as tractor treads in swampy environments, have never been evolved by natural selection on the earth.

Life in extreme environments

There is an extraordinarily wide range of ecological niches to which organisms have adapted through the operation of natural selection. The same basic fabric of life has been used to produce very diverse organisms. The alga *Cyanidium caldarium* can grow in concentrated solutions of hot sulfuric acid. Other bacteria, algae, and fungi can live in extremely acidic (pH of 0) or extremely alkaline (pH near 13) environments. Procaryotic bacteria live in pools at Yellowstone National Park at temperatures above 90° C (194" F), almost at the boiling point of water. Sulfate-reducing bacteria are reported to grow and reproduce at 104" C (219" F) under very high pressures. Many organisms employ organic or inorganic antifreezes to lower the freezing point of their internal liquids, so that they can live at several tens of degrees below 0° C (32" F). Some insects use dimethyl sulfoxide as an antifreeze. Other organisms live in briny pools in which dissolved salts lower the freezing point. For example, Don Juan Pond in Antarctica has about one molecule of calcium chloride for every two water molecules and does not freeze until −45° C (−49° F). It contains a possibly unique microflora that continues to metabolize at least down to −23" C (−9" F). Biological activity does not cease at the freezing point of water; in fact some enzymes are actually more active in ice than in water. Many single-celled organisms can be frozen indefinitely to extremely low temperatures—the temperature of liquid air for example—and then be thawed with no decrease in activity. The primary damage that freezing causes is apparently due to the unavailability of liquid water and to the expansion and contraction attendant to freezing and thawing. Some arthropods can be severely dehydrated and then revived simply by adding water. In the dehydrated state they can be brought to any temperature from close to absolute zero to above the boiling point of water without apparent damage. When encysted in response to dehydration, some such organisms seem indistinguishable from a weathered grain of sand.

The great majority of familiar organisms on the earth, however, are much more sensitive to the temperature of their surroundings. Warm-blooded animals internally regulate their temperatures for this reason. A human being whose body temperature drops below 30" C (86° F) or rises above 40° C (104" F) soon dies. Organisms that inhabit cold climates have special insulating layers of fat and fur. Other organisms adapt to seasonal temperature changes by producing dormant forms, such as spores or eggs, to survive the low temperatures. In all cases dormancy appears to be accompanied by dehydration.

Importance of water

Since organisms are composed largely of water, the availability of water is clearly a limiting factor. Here also, however, remarkable adaptations exist. Certain micro-organisms can live on the water adsorbed on a single crystal of salt. Other organisms, such as the kangaroo rat and the flour beetle, obtain no water at all in the liquid state, relying entirely on metabolic water; that is, on water released from chemical bonds through the metabolism of food. A variety of plants, including Spanish moss, live in environments where they have no contact with groundwater—for example, on telephone wires—apparently extracting water directly from the air, although such plants require a relatively high humidity. Plants that live in deserts and other very dry environments have evolved wide-spreading root systems that adsorb subsurface water from a great volume of adjacent soil.

Organisms have been found from the stratosphere to the ocean depths. Bacteria and fungal spores have been discovered near the base of the stratosphere by balloons, and searches for organisms at much greater altitudes (up to 100,000 feet) have been attempted with ambiguous results. Birds have been observed flying at altitudes as great as 27,000 feet, and jumping spiders have been found at 22,000 feet on Mt. Everest. At the opposite extreme, micro-organisms, fish, and a variety of other metazoa have been recovered from the ocean depths down to thousands of feet, where the corresponding pressures are hundreds of times that at sea level. At these depths no light can penetrate and the organisms, some of which are quite large and include unique phosphorescent adaptations to the dark, ultimately live off particles of organic matter raining down from the upper reaches of the oceans.

There is a range of adaptations to the radiation environment of the earth. Some micro-organisms are readily killed by the small amount of solar ultraviolet light that filters through the earth's atmosphere at wavelengths near 3000 angstrom units (A; 1Å = one ten billionth of a metre). On the other hand, the bacterium *Pseudomonas radiodurans* thrives in the large neutron flux at the cores of swimming-pool reactors, to the continuing annoyance of nuclear physicists. Organisms can avoid radiation by shielding. For example, some algae and some desert plants live under a superficial coating of soil or rocks that are more transparent to visible light than to ultraviolet light. In addition, organisms have active methods of undoing the damage produced by radiation. Some of these repair mechanisms work in the dark; others require visible light. The usual reason for the ultraviolet sensitivity of organisms is that their nucleic acids absorb ultraviolet light very effectively at a wavelength near 2600 A. Generally speaking, there is an upper limit to the amount of ionizing radiation (such as gamma rays, X-rays, electrons or protons) that an organism can receive without being killed: in the vicinity of 1,000,000 roentgens. Such a lethal dose applies only to extremely radiation-resistant micro-organisms; mammals, for example, are killed by much lower doses because there is more that can go wrong with such complex organisms. A lethal dose of ionizing radiation for human beings is a few hundred roentgens applied to the whole body. A thermonuclear weapon dropped on a populated area may deliver, through direct radiation and fallout, doses of a few hundred roentgens or more to people within a radius of some tens of miles of the target. Much smaller doses can produce a variety of diseases and predominantly deleterious mutations in the hereditary material. Moreover, the effect of small doses is cumulative. But until very recently human beings have not lived in environments characterized by large doses of ionizing radiation (see RADIATION, BIOLOGICAL EFFECTS OF).

Size range of living things

The sizes of organisms on the earth vary greatly. As discussed above, the smallest free-living organisms on the earth, PPLO, are about 1000 A in diameter; a limitation on the size of the smallest free-living organism is its volume: it must contain all the molecules necessary for metabolism. A variety of influences place an upper limit to the size of organisms. One is the strength of biological materials. Galileo calculated in 1638 that a tree taller than roughly 300 feet would, when displaced slightly from the vertical (for example, by a breeze), buckle under its own weight. (Sequoias, some of which exceed 300 feet, are apparently near the upper limit of height for an organism.) Because of the buoyancy of water, large whales are not presented with such stability problems, but other difficulties arise. For a fixed shape, the volume of tissues to be nourished increases as the cube of the characteristic length of the organism, but the surface of the gut, which adsorbs the ingested food, increases only as the square of the length. As the length is increased, a point of diminishing returns is ultimately reached.

The range of organic molecules that organisms on earth can metabolize is very wide and occasionally includes such foods as formaldehyde or petroleum, which seem unlikely from a human point of view. *Pseudomonas* bacteria are capable of using almost any organic molecule as a source of carbon and of energy, provided only that the molecule is at least slightly soluble in water. Micro-organisms cannot metabolize plastics, not because of any fundamental chemical prohibitions but probably because plastics have not been part of the environment of micro-organisms for very long. Man tends to think of oxygen as extremely important for life, but there are facultative anaerobes that can take their oxygen or leave it, and obligate anaerobes that are actually poisoned by oxygen. Such organisms use a variety of alternative electron acceptors, as previously discussed.

Chemical constituents

The water content of organisms usually represents between 50 and 90 percent of the live weight. Unless there is a massive mineral skeleton, the dry matter of organisms constitutes about one-half carbon by weight, reflecting the fact that organic molecules are based upon carbon. A wide variety of other chemical elements are used for diverse functions. Amino acids are made of nitrogen and sulfur in addition to carbon, hydrogen, and oxygen. Nucleic acids, as has been seen, employ phosphorus in addition to hydrogen, nitrogen, oxygen, and carbon. Sodium and potassium are used in maintaining the electrolyte balance, and calcium and silicon as structural materials. Iron plays a fundamental role in the transport of molecular oxygen as part of the hemoglobin molecule. In some ascidians (sea squirts), however, vanadium replaces iron. Ascidian blood also contains unusually large amounts of niobium, tantalum, titanium, chromium, manganese, molybdenum, and tungsten. The vanadium and niobium compounds in ascidian blood may be adaptations to low oxygen levels. Occasional organisms use selenium or tellurium as electron acceptors; others may produce the fully saturated gas hydrides of arsenic, phosphorus, or silicon, as metabolic wastes. Still others form compounds of carbon with such halogens as chlorine or iodine. Many of the foregoing elements, plus copper, zinc, cobalt, and possibly gallium, boron, and scandium, perform particular functions in the enzymatic apparatus of cells. Many of these elements, both the uncommon ones and those as common as phosphorus, are very highly concentrated in organisms over their general availability in the environment. This concentration must indicate that such chemicals play unique functional roles where other more abundant elements will not serve.

Behaviour and sensory capabilities. Analogous to the wide range of physiological adaptations and the great variety of elements used by organisms on earth, there is an enormous range of behaviour patterns and sensory capabilities. Coded into its nucleic acids is the information that allows a bird raised from the egg in the absence of other birds to migrate when migration time arrives, to build a nest characteristic of its species, or to engage in elaborate courtship rituals. Those birds that do not perform acceptably do not leave descendants. Such behavioral information must itself have evolved. Rats that pass through mazes easily can be interbred, as can rats that pass through with difficulty; eventually two populations with inherited characteristics called "maze-smart" and "maze-dumb" will be produced. Fruit fly populations attracted to the light can be separated from those that avoid light. Classical genetic crossing experiments reveal that the two populations differ largely in a small number of genes for phototropism. Similar genetic determinants of behaviour exist in man. Possession of a supernumerary Y-chromosome in males is strikingly correlated with aggressive tendencies—which may, however, have been a selective advantage in more primitive societies. Myopia may have had strong survival value in earlier times: near-sighted males, useless in the hunt, stayed home and painted the walls of the cave. As technology develops, natural selection enters new behavioral arenas; for example, in an age of artificial contraception, the clumsy and forgetful preferentially reproduce.

Response to the electromagnetic spectrum

Human beings use only a small part of the total electromagnetic spectrum, the part called visible light, which extends from about 4000 to about 7000 A in wavelength. While many plants and animals are sensitive to this same range of wavelengths, many of them are sensitive to other wavelengths as well. Most insects are sensitive to ultraviolet light at wavelengths below 4000 A, and many flowering plants take advantage of this fact and present patterns visible only in the ultraviolet range. Honeybees use polarized light, which the unaided human eye is quite unable to detect, for direction finding on partly cloudy days. The "pit" of such pit vipers as the rattlesnake is an infrared receptor and direction finder. These reptiles can sense the thermal radiation emitted by warm-blooded prey, radiation to which human beings are completely insensitive.

It is common knowledge that some animals (for example, dogs) are sensitive to sounds that the human ear cannot detect. Bats emit and detect sound waves at ultrahigh frequencies, in the vicinity of 100,000 cycles per second, about five times the highest frequency to which the human ear is sensitive. Bats use these sounds not so much to communicate, however, as to echolocate their prey and were doing this for millions of years before radar and sonar were invented. The audio receptors of many moths that are prey to bats are responsive only to the frequencies emitted by the bats. When the bat sounds are heard, the moths take evasive action. Dolphins have a very wide frequency range and several communication channels, as well as a "click" echolocator. Dolphins and whales use their blowholes rather than their mouths to utter these sounds. Sharks and other marine predators are said to locate their prey by the low-frequency sounds the prey makes when in distress. Some animals develop highly specialized and exotic organs for the detection or transmission of sound—for example, a European grasshopper has a relatively large parabolic antenna on its back that looks very much like a small radio telescope. This antenna is used for producing noises evidently thought attractive by the female of the species.

Many organisms are capable of smell and taste; that is, the detection of specific chemical molecules. According to one theory of smell, there are particular olfactory sensors, each receptive only to a specific chemical group on airborne molecules. The ultimate in olfactory specialization is probably the male silkworm moth: with its feathery antennae it is able to smell essentially nothing except the chemical sex attractant discharged by the female of the species. But it can detect this molecule very well, needing an impact of only 40 molecules per second on its antennae to produce a marked response. One female silkworm moth need release only 10^{-8} grams of sex attractant per second in order to attract every male silkworm moth in a volume hundreds of metres to kilometres on a side.

Extraordinary senses

Besides the senses of sight, hearing, smell, taste, and touch, various animals have a wide variety of other senses (see SENSORY RECEPTION). Man has an inertial orientation system and accelerometer in the cochlear canal of the ear. The water scorpion (*Nepa*) has a fathometer sensitive to hydrostatic pressure gradients. Most higher plants have chemically amplified gravity sensors. Fireflies and squids communicate with their own kind by producing time sequences or patterns of light on their bodies. The African freshwater fish *Gymnarchus niloticus* operates a dipole electrostatic field generator and a sensor to detect the amplitude and frequency of disturbances in the impressed field, an adaptation well suited for its nocturnal activities in turbulent waters. Other organisms have salinity sensors, or humidity sensors. There may be sensors involved in homing instincts of animals that have not yet been discovered. All of these senses confer upon their possessors an awareness of the environment that may be very different from that of such other organisms as man. Man, however, has the remarkable ability to extend his sensory and intellectual capabilities artificially, through the use of instrumentation.

THE ORIGIN OF LIFE

Hypotheses of origins. Perhaps the most fundamental and at the same time the least understood biological problem is the origin of life. It is central to many scientific and philosophical problems and to any consideration of extraterrestrial life. Most of the hypotheses of the origin of life will fall into one of four categories:

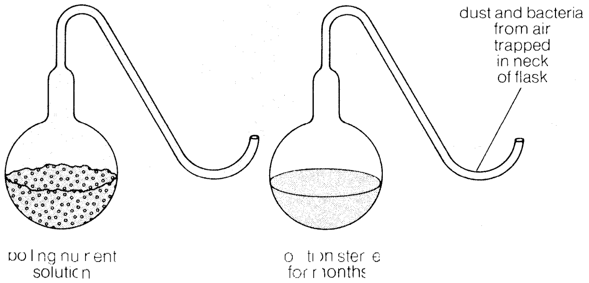
1. The origin of life is a result of a supernatural event; that is, one permanently beyond the descriptive powers of physics and chemistry.
2. Life—particularly simple forms—spontaneously and readily arises from nonliving matter in short periods of time, today as in the past.
3. Life is coeternal with matter and has no beginning; life arrived on the earth at the time of the origin of the earth or shortly thereafter.
4. Life arose on the early earth by a series of progressive chemical reactions. Such reactions may have been likely or may have required one or more highly improbable chemical events.

The early theological view

Hypothesis 1, the traditional contention of theology and some philosophy, is in its most general form not inconsistent with contemporary scientific knowledge, although this knowledge is inconsistent with a literal interpretation of the biblical accounts given in chapters 1 and 2 of Genesis and in other religious writings. Hypothesis 2 (not of course inconsistent with 1) was the prevailing opinion for centuries. A typical 17th-century view follows:

> [May one] doubt whether, in cheese and timber, worms are generated, or, if beetles and wasps, in cowdung, or if butterflies, locusts, shellfish, snails, eels, and such life be procreated of putrefied matter, which is to receive the form of that creature to which it is by formative power disposed[?] To question this is to question reason, sense, and experience. If he doubts this, let him go to Egypt, and there he will find the fields swarming with mice begot of the mud of the Nylus [Nile], to the great calamity of the inhabitants.

It was only in the Renaissance, with its burgeoning interest in anatomy, that such transformations were realized to be impossible. A British physiologist, William Harvey, during the mid-17th century, in the course of his studies on the reproduction and development of the king's deer, made the basic discovery that every animal comes from an egg. An Italian biologist, Francesco Redi, in the latter part of the 17th century, established that the maggots in meat came from flies' eggs, deposited on the meat. And an Italian priest, Lazzaro Spallanzani, in the 18th century, showed that spermatozoa were necessary for the reproduction of mammals. But the idea of spontaneous generation died hard. Even though it was proved that the larger animals always came from eggs, there was still hope for the smaller ones, the micro-organisms. It seemed obvious that, because of their ubiquity, these microscopic creatures must be generated continually from inorganic matter.


dust and bacteria from air trapped in neck of flask
boiling nutrient solution
solution sterile for months

Pasteur's swan-necked flask experiment (see text).

Meat could be kept from going maggoty by covering it with a flyproof net, but grape juice could not be kept from fermenting by putting over it any netting whatever. This was the subject of a great controversy between the famous French bacteriologists Louis Pasteur and F.A. Pouchet in the 1850s, in which Pasteur triumphantly showed that even the minutest creatures came from germs floating in the air, but that they could be guarded against by suitable filtration. Actually, Pouchet was arguing that life must somehow arise from nonliving matter; if not, how had life come about in the first place?

Life from outer space

Toward the end of the 19th century Hypothesis 3 gained currency, particularly with the suggestion by a Swedish chemist, S.A. Arrhenius, that life on earth arose from panspermia, micro-organisms or spores wafted through space by radiation pressure from planet to planet or solar system to solar system. Such an idea of course avoids rather than solves the problem of the origin of life. In addition, it is extremely unlikely that any micro-organism could be transported by radiation pressure to the earth over interstellar distances without being killed by the combined effects of cold, vacuum, and radiation.

Pasteur's work discouraged many scientists from discussing the origin of life at all. Moreover they were anxious not to offend religious feeling by probing too deeply into the subject. Although Darwin would not commit himself on the origin of life, others subscribed to Hypothesis 4 more resolutely, notably the famous British biologist T.H. Huxley in his *Protoplasm, the Physical Basis of Life* (1869), and the British physicist John Tyndall in his "Belfast Address" of 1874. Although Huxley and Tyndall asserted that life could be generated from inorganic chemicals, they had extremely vague ideas about how this might be accomplished. The very phrase "organic molecule" implies that there exists a special class of chemicals uniquely of biological origin, despite the fact that organic molecules have been routinely produced from inorganic chemicals since 1828. In the following discussion the word organic carries no imputation of biological origin. In fact the problem largely reduces to finding an abiological source of appropriate organic molecules.

The primitive atmosphere. Darwin's attitude was: "It is mere rubbish thinking at present of the origin of life; one might as well think of the origin of matter." The two problems are, in fact, curiously connected, and modern scientists are thinking about the origin of matter. There is convincing evidence that thermonuclear reactions and subsequent explosions in the interiors of stars generate all the chemical elements more massive than hydrogen and helium and then distribute them into the interstellar medium from which subsequent generations of stars and planets form. Because of the commonality of these thermonuclear processes, and because some thermonuclear reactions are more probable than others, there exists a cosmic distribution of the major elements, so far as is known, throughout the universe. Table 1 compares, for some atoms of interest, the relative numerical abundances in the universe as a whole, on the earth, and in living organisms.

Table 1: Relative Abundances of the Elements (percent)

atom	universe	life (terrestrial vegetation)	earth (crust)
Hydrogen	87	16	3
Helium	12	0*	0
Carbon	0.03	21	0.1
Nitrogen	0.008	3	0.0001
Oxygen	0.06	59	49
Neon	0.02	0	0
Sodium	0.0001	0.01	0.7
Magnesium	0.0003	0.04	8
Aluminum	0.0002	0.001	2
Silicon	0.003	0.1	14
Sulfur	0.002	0.02	0.7
Phosphorus	0.00003	0.03	0.07
Potassium	0.000007	0.1	0.1
Argon	0.0004	0	0
Calcium	0.0001	0.1	2
Iron	0.002	0.005	18

*0% here stands for any quantity less than 10^{-6}%.

Composition of life

There is of course some variation in composition from star to star, from place to place on the earth, and from organism to organism, but such comparisons are nevertheless very instructive. The composition of life is intermediate between the average composition of the universe and the average composition of the earth. Ninety-nine percent both of the universe and of life is made of the six atoms, hydrogen (H), helium (He), carbon (C), nitrogen (N), oxygen (O), and neon (Ne). Can it be that life on earth arose when the chemical composition of the earth was much closer to the average cosmic composition, and that some subsequent events have changed the gross chemical composition of the earth?

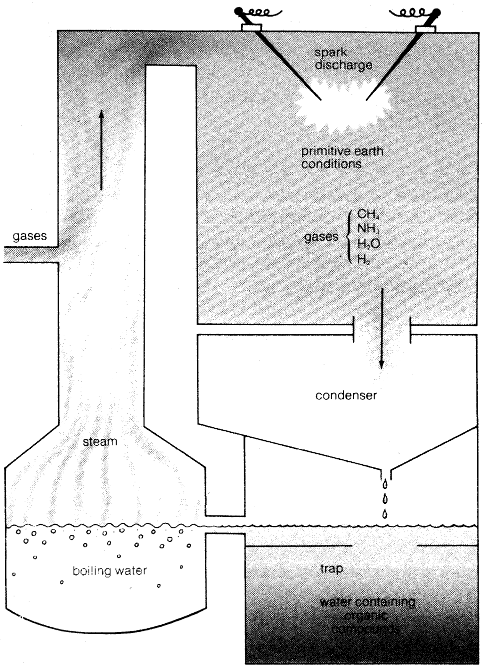
The Jovian planets (Jupiter, Saturn, Uranus, and Neptune) are much closer to cosmic composition than is the earth. They are largely gaseous, with atmospheres composed principally of hydrogen and helium. Methane (CH,) and ammonia (NH_3) have been detected in smaller quantities, and neon and water are suspected. This circumstance very strongly suggests that the Jovian planets were formed out of material of typical cosmic composition. They have very large masses, and because they are so far from the sun their upper atmospheres are very cold. Therefore it is impossible for atoms in the upper atmospheres of the Jovian planets to escape from their gravitational fields; escape was probably very difficult even during planetary formation. The earth and the other planets of the inner solar system, however, are much less massive and most have hotter upper atmospheres. It is possible for hydrogen and helium to escape from the earth today, and it may well have been possible for much heavier gases to have escaped during the formation of the earth. It is reasonable to expect that in the very early history of the earth a much larger abundance of hydrogen prevailed, which has subsequently been lost to space. Thus the atoms carbon, nitrogen, and oxygen were present on the primitive earth, not as CO_2 (carbon dixoide), N_2, and O_2 as they are today but rather in the form of their fully saturated hydrides, CH_4 (methane), NH_3 (ammonia), and H_2O. In the geological record, the presence of such reduced minerals as uraninite (UO_2) and pyrite (FeS_2) in sediments formed several billions of years ago implies that conditions then were considerably less oxidizing than they are today.

In the 1920s J.B.S. Haldane in Britain and A.I. Oparin in the Soviet Union recognized that the abiological production of organic molecules in the present oxidizing atmosphere of the earth is highly unlikely; but that, if the earth once had more reducing (in this context, hydrogen-rich) conditions, the possible abiogenic production of organic molecules would have been much more likely. If large numbers of organic molecules were somehow synthesized on the primitive earth, there would not necessarily be much trace of them today. In the present oxygen atmosphere, largely produced by green-plant photosynthesis, such molecules would tend, over geological time, to be oxidized to carbon dioxide, nitrogen, and water. In addition, as Darwin recognized, the first micro-organisms would consume prebiological organic matter produced prior to the origin of life.

Synthesis of amino acids

Production of simple organic molecules. The first deliberate experimental simulation of these primitive conditions was carried out in 1953 by a U.S. graduate student, S.L. Miller, under the guidance of the eminent chemist H.C. Urey. A mixture of methane, ammonia, water vapour, and hydrogen was circulated through a liquid water solution and continuously sparked by a corona discharge elsewhere in the apparatus. The discharge may be thought to represent lightning flashes on the early earth. After several days of exposure to sparking, the solution changed colour. Subsequent analysis indicated that several amino and hydroxy acids, intimately involved in contemporary life, had been produced by this simple procedure. The experiment is in fact so elementary, and the amino acids can so readily be detected by paper chromatography, that the experiment has been repeated many times by high school students. Subsequent experiments have substituted ultraviolet light or heat as the energy source or have altered the initial abundances of gases. In all such experiments amino acids have been formed in large yield. On the early earth there was much more energy available in ultraviolet light than in lightning discharges. At long ultraviolet wavelengths, in which methane, ammonia, water, and hydrogen are all transparent, but in which the bulk of the solar ultraviolet energy lies, the gas hydrogen sulfide (H_2S) is a likely ultraviolet absorber.

Following such reasoning, a U.S. astrophysicist, Carl Sagan, and his colleagues made amino acids by long wavelength ultraviolet irradiation of a mixture of methane, ammonia, water, and HS. The amino acid syntheses, at least in many cases, involve hydrogen cyanide and aldehydes (*e.g.*, formaldehyde) as gaseous intermediaries formed from the initial gases. It is quite remarkable that amino acids, particularly biologically abundant amino acids, can be made so readily under simulated primitive conditions. When laboratory conditions become oxidizing, however, no amino acids are formed, suggesting that reducing conditions were necessary for prebiological organic synthesis.



Miller–Urey spark-discharge apparatus.

Under alkaline conditions, and in the presence of inorganic catalysts, formaldehyde spontaneously reacts to form a variety of sugars, including the five-carbon sugars fundamental to the formation of nucleic acids and such six-carbon sugars as glucose and fructose, which are extremely common metabolites and structural building blocks in contemporary organisms. Furthermore, the nucleotide bases as well as porphyrins have been produced in the laboratory under simulated primitive earth conditions by several investigators. While there is still debate on the generality of the experimental synthetic pathways and on the stability of the molecules produced, most if not all of the essential building blocks of proteins, carbohydrates, and nucleic acids can be readily produced under quite general primitive reducing conditions, plus probably, ATP as well.

Polymerization as dehydration reactions

Production of polymers. The construction of polymers, long-chain molecules made of repeating units of these essential building blocks,·however, is a much more difficult experimental problem. Polymerization reactions are generally dehydrations, in which a molecule of water is lost in the formation of a two-unit polymer. Dehydrating agents must be used to initiate polymerization. The polymerization of amino acids to form long protein-like molecules has been accomplished through dry heating by a U.S. investigator, S.W. Fox. The polyamino acids that are formed are not random polymers and have some distinct catalytic activities. The geophysical generality of dry heating and return to solution, however, has been questioned. Long polymers of amino acids can also be produced from hydrogen cyanide and anhydrous liquid ammonia. Some evidence exists that nucleotide bases and sugars can be combined in the presence of phosphates or cyanides under ultraviolet irradiation. Some condensing agents such as cyanamide are efficiently made under simulated primitive conditions. Despite the breakdown by water of molecular intermediates, condensing agents are often quite effective in inducing polymerization, and polymers of amino acids, sugars, and nucleotides have all been made this way.

A famous British scientist, J.D. Bernal, has suggested that adsorption of molecular intermediates on clays or other minerals may have concentrated these intermediates. Such concentration could offset the tendency for water to break down polymers of biological significance. Of special interest is the possibility that such concentration matrices included phosphates, for this would help explain how phosphorus could have been incorporated preferentially into prebiological organic molecules at a time when biological concentration mechanisms did not yet exist. Mineral catalysis implies that organic synthesis could also occur in deep water where ultraviolet light had been filtered out.

Quite apart from concentration mechanisms, the primitive waters themselves may have been a not very dilute solution of organic molecules. If all the surface carbon on the earth were present as organic molecules in the contemporary oceans, or if many known ultraviolet synthetic reactions producing organic molecules were permitted to continue for a billion years with products dissolved in the oceans, a 1 percent solution of organic molecules would result. For similar reasons, Haldane suggested that the origin of life occurred in a "hot dilute soup." In addition, concentration mechanisms do exist, such as evaporation or freezing of pools or adsorption on interfaces or the generation of colloidal enclosures called coacervates.

The origin of the code. It has been shown that all the essential building blocks for life and their polymers may have been produced in some fair concentration on the primitive earth. This possibility is certainly relevant to the origin of life, but it is not the same thing as the origin of life. By the genetic definition of life discussed in *Definitions of life,* above, a self-replicating, mutable molecular system, capable of interacting with the environment, is required. In contemporary cells the nucleic acids are the sites of self-replication and mutation. Laboratory experiments have already shown that polynucleotides can be produced from nucleotide phosphates in the presence of a specific enzyme of biological origin and a pre-existing "primer" nucleic acid molecule. If the primer molecule is absent, polynucleotides are still formed, but they of course contain no genetic information. Once such a polynucleotide spontaneously forms it then acts as primer for subsequent syntheses.

Production of polynucleotides

Imagine a primitive ocean filled with nucleotides and their phosphates and appropriate mineral surfaces serving as catalysts. Even in the absence of the appropriate enzyme it seems likely, although not yet proved, that spontaneous assembly of nucleotide phosphates into polynucleotides occurred. Once the first such polynucleotide was produced it may have served as a template for its own reproduction, still of course in the absence of enzymes. As time went on there were bound to be errors in replication. These would be inherited. A self-replicating and mutable molecular system of polynucleotides, eventually leading to a diverse population of such molecules, may have arisen in this way. Alternatively, the primitive hereditary material may have involved some other molecule altogether, but no concrete suggestion for such a molecule has ever been proposed.

In any case, a population of replicating polynucleotides cannot quite be considered alive because it does not significantly influence its environment. Eventually, all the nucleotides in the ocean would have been tied in polynucleotides and the entire synthetic process would then have ground to a halt. So far as is known, polynucleotides have no catalytic properties, and proteins have no reproductive properties. It is only the partnership of the two molecules that makes contemporary life on earth possible. Accordingly, a critical and unsolved problem in the origin of life is the first functional relation between these two molecules, or, equivalently, the origin of the genetic code. The molecular apparatus ancillary to the operation of the code —the activating enzymes, adapter RNAs, messenger RNAS, ribosomes, and so on—are themselves each the product of a long evolutionary history and are produced according to instructions contained within the code. At the time of the origin of the code such an elaborate molecular apparatus was of course absent.

It has been proposed that a weak but selective chemical bonding does exist, even in the absence of any of this apparatus, between amino acids and nucleotides. There need not be a very great selectivity; a given nucleotide sequence might in primitive times have coded for many different amino acids or, conversely, the same amino acid may have been coded for by several different nucleotide sequences. All that is required is that a particular linear sequence of nucleotides must code for some nonrandom sequence of amino acids. The active sites largely responsible for the catalytic activity of contemporary enzymes are generally only five or six amino acids long; the remainder of the enzyme is devoted to more sophisticated functions, such as arranging for the enzyme to be turned on and off by the machinery of the cell. With, say, 20 different varieties of amino acids available in the primitive environment, the chance of any given active site being produced by a random sequence of nucleotides is one in 20^5, or one in about 3,000,000. But 3,000,000 combinations to form units five amino acids long is not a very large number for the chemistry and time periods in question. To conclude this speculation, then, if polynucleotides were initially capable of crude, nonenzymatic replication, and if a crude primitive genetic code existed, then any one of a very large number of catalytic properties was available to some self-replicating polynucleotides on the primitive earth. This situation is all that would be necessary for the origin of life; those polynucleotides that could code for a primitive protein having catalytic properties furthering the replication of the polynucleotide would preferentially replicate. Other polynucleotides coding for less effective proteins would have replicated more slowly. The foregoing is one of several possibilities for the origin of the first living systems. Many separate and rather diverse instances of the origin of life may have occurred on the primitive earth, but competition eventually eliminated all but one line. Every organism on earth today would be a descendant of that line.

The earliest living systems. One curious feature of biological organic molecules is their optical activity: they rotate the plane of a beam of plane-polarized light. Organic molecules produced abiologically do not show optical activity. Molecules made of the same units can be put together in complementary ways like a left- and right-handed glove. The same building blocks can be used to produce molecules that are three-dimensional mirror images of each other. This asymmetry is responsible for optical activity. At the time of the origin of life, organic molecules, corresponding both to left- and right-handed forms, were produced. The laboratory simulation experiments always produce both types. But the first living systems could have been made only of one type, for the same reason that carpenters do not use random mixtures of screws with left- and right-handed threads. Whether left- or right-handed activity was adopted was probably purely a matter of chance, but once a particular asymmetry was established in the first living systems, it maintained itself. This belief implies that optical activity should be a feature of life on any planet, and also that the chances should be equal of finding a given terrestrial organic molecule or its mirror image molecule in extraterrestrial life forms.

Optical activity of biological molecule

The first living systems probably resided in a molecular garden of Eden, where all the building blocks that contemporary organisms must work hard at synthesizing were available free. Under such conditions the numbers of organisms must have increased very rapidly. But such increases cannot go on indefinitely. In time the supply of some molecular building block must have become short. Those primitive organisms that had the ability to synthesize the scarce building block, say A, from a more abundant one, say B, clearly had a competitive advantage over those organisms that could not perform such a synthesis. In time, however, the secondary source of supply, B, would have also become depleted and those organisms that could produce it from a third building block, *C,* would have preferentially replicated. A U.S. biochemist, N.H. Horowitz, has proposed that in this way the enzymatic reaction chains of contemporary organisms—each step catalyzed by a particular enzyme—originally evolved.

Even the evolution of enzymatic reaction chains may have occurred in free nucleic acids before the origin of the cell. The cell may have arisen in response to the need for maintaining a high concentration of scarce building blocks or enzymes, or as protection against the gradually increasing abundance of oxygen on the primitive earth. Oxygen is a well-known poison to many biological processes, and in contemporary higher organisms the mitochondria that handle molecular oxygen are kept in the cytoplasm, far from contact with the nuclear material. Even today processes are known whereby polyamino acids form small spherical objects, microns to tens of microns across, with some of the properties of cells. These objects, called proteinoid microspheres by Fox, are certainly not cells, but they may indicate processes by which the ancestors of cells arose. Procaryotic cells almost certainly preceded eucaryotic cells and the evolution of so extremely complex an apparatus as the mitotic spindle (which ensures equal segregation of replicated chromosomes) must have taken very long periods of time to evolve.

Proteinoid micro-spheres

The development of mitochondria and chloroplasts (each of which contains its own DNA) in the eucaryotic cell may have been the result of a symbiosis, a cooperative arrangement entered into at first tentatively by originally free-living cells.

As the competition for building blocks increased among early life forms, and also perhaps as the abiological production of organic molecules dwindled because of the increasing oxygen abundance, the strictly heterotrophic way of life became more and more costly. The utilization of porphyrins, which are also made abiologically, by primitive photoautotrophs would have had great selective advantage. Many of the intermediates and enzymes in photosynthesis and in the anaerobic breakdown of carbon compounds are similar, but there is no generally accepted view of the origin of the photosynthetic process. Photosynthesis in procaryotes is more primitive than in such eucaryotes as green plants. In bacteria, water is not the ultimate source of hydrogen atoms for reducing carbon dioxide, and therefore oxygen is not produced. In addition, when a chlorophyll-containing cell is exposed both to light and to oxygen, it is killed unless it also contains an accessory carotenoid pigment. Thus green-plant photosynthesis had to wait until the appearance of carotenoids while bacterial photosynthesis, which does not produce oxygen, could function without carotenoids.

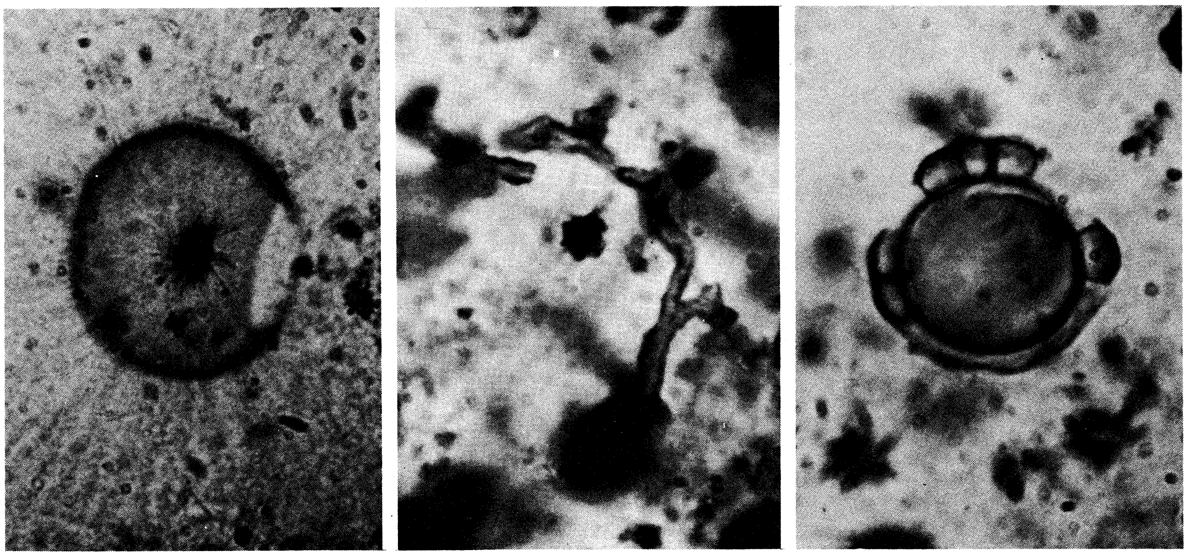
The antiquity of life. Among the oldest known fossils are those found in the Fig Tree chert from the Transvaal, dated at 3,100,000,000 years old. These organisms have been identified as bacteria and blue-green algae. It is very reasonable that the oldest fossils should be procaryotes rather than eucaryotes. Even procaryotes, however, are exceedingly complicated organisms and very highly evolved. Since the earth is reliably known to be itself only about 4,500,000,000 years old, this suggests that the origin of life must have occurred within a few hundred million years after the origin of the earth.

By performing chemical analyses on the oldest sediments, it is possible to say something about the sorts of organic molecules produced, either biologically or abiologically, in primitive times. Thus, amino acids and porphyrins have been identified in the oldest sediments, as have pristane and phytane, typical breakdown products of chlorophyll. There are several indications that these organic molecules, dating from 2,000,000,000 to more than 3,000,000,000 years ago, are of biological origin. For one thing their long-chain hydrocarbons show a preference for a straight chain geometry, whereas known abiological processes tend to produce a much larger proportion of branched chain and cyclic hydrocarbon molecular geometries than have been found in these sediments. Abiological processes tend to produce equal amounts of long-chain carbon compounds with odd and with even numbers of carbon atoms. But the oldest sediments show a distinct preference for odd numbers of carbon atoms per molecule, as do products of undoubted biological origin. Finally, a C^{12} enrichment, for which no abiological process seems able to account, has been discovered in the oldest sediments, evidence that suggests that plantlike life, which concentrates the carbon isotope C^{12} preferentially to C^{13}, was present very early. These departures from thermodynamic equilibrium are often considered to be compelling signs of biological activity. Such evidence again points to the great antiquity of life on earth.

Oceanic origin of life

The fossil record, in any complete sense, goes back only about 600,000,000 years. In the layers of sedimentary rock known by geological methods and by radioactive dating to be that old, most of the major groups of invertebrates appear for the first time. All these organisms appear adapted to life in the water, and there is no sign yet of organisms adapted to the land. For this reason, and because of a rough similarity between the salt contents of blood and of seawater, it is believed that early forms of life developed in oceans or pools. With no evidence for widespread oxygen-producing photosynthesis before this time, and for cosmic abundance reasons described above, the oxygen content of the earth's atmosphere in Precambrian times was very likely less than today. Accordingly, in Precambrian times, solar ultraviolet radiation, especially near the wavelength of 2600 A, which is particularly

By courtesy of Elso S. Barghoorn



Photomicrographs of 2,000,000,000-year-old organisms from the geological stratum called the Gunflint Chert. These organisms are about ten microns across. Organisms more than 1,000,000,000 years older are known.

destructive to nucleic acids, may have penetrated to the surface of the earth, rather than being totally absorbed in the upper atmosphere by ozone as it is today. In the absence of ozone, the ultraviolet solar flux is so high that a lethal dose for most organisms would be delivered in less than an hour. Unless extraordinary defense mechanisms existed in Precambrian times, life near the earth's surface would have been impossible. Sagan has suggested that life at this time was generally restricted to some tens of metres and deeper in the oceans, at which depths all the ultraviolet light would have been absorbed, although visible light would still filter through. As the amount of atmospheric oxygen and ozone increased, due both to plant photosynthesis and to the photodissociation of water vapour and the escape to space of hydrogen from the upper atmosphere, life increasingly close to the earth's surface would have been possible. It has been suggested that the colonization of the land, about 425,000,000 years ago, was possible only because enough ozone was then produced to shield the surface from ultraviolet light for the first time.

Life then had insinuated itself between the sun and the earth. It diverted solar energy to its own uses and contrived more and more ways of exploiting more and more environments. Some experiments were faulty and the lines became extinct; others were more successful and the lines filled the earth. Evolution through natural selection directed the proliferation of a growing array of life forms throughout the biosphere (see EVOLUTION; PHYLOGENY).

EXTRATERRESTRIAL LIFE

It is not known what aspects of living systems are necessary in the sense that living systems everywhere must have them; it is not known what aspects of living systems are contingent in the sense that they are the result of evolutionary accident, so that somewhere else a different sequence of events might have led to different characteristics. In this respect the possession of even a single example of extraterrestrial life, no matter how seemingly elementary in form or substance, would represent a fundamental revolution in biology. It is not known whether there is a vast array of biological themes and counterpoints in the universe, whether there are places that have fugues, compared with which our one tune is a bit thin and reedy. Or it may be that our tune is the only tune around. Accordingly the prospects for life on other planets must be considered in any general discussion of life.

The chemistry of extraterrestrial life. The significance of the search for life beyond the earth has already been discussed. What are the methods and prospects for such a search? Each of the definitions of life described in *Definitions* of *life,* above, implies a method of searching for life. Particular physiological functions, particular metabolic activities, such specific molecules as proteins and nucleic acids, self-replication and mutation, processes not in closed-system thermodynamic equilibrium—all these might be sought. All the search methods significantly depend upon chemistry.

Life on earth is structurally based on carbon and utilizes water as an interaction medium. Hydrogen and nitrogen have significant accessory structural roles; phosphorus is important for energy storage and transport, sulfur for three-dimensional configuration of protein molecules, and so on. But must these particular atoms be the atoms of life everywhere, or might there be a wide range of atomic possibilities in extraterrestrial organisms? What are the general physical constraints on extraterrestrial life?

Prerequisites for life

In approaching these questions several criteria can be used. The major atoms should tend to have a high cosmic abundance. A structural molecule for making an organism at the temperature of the planet in question should not be extremely stable, because then no chemical reactions would be possible; but it should not be extremely unstable, because then the organism would fall to pieces. There should be some medium for molecular interaction. Solids are not appropriate because the diffusion times are very long. Such a medium is most likely a liquid (but could possibly be a very dense gas) that is stable in a number of respects. It should have a large temperature range (for a liquid, the temperature difference between freezing point and boiling point should be large). The liquid should be difficult to vaporize and to freeze; in fact, it should be very difficult to change its temperature at all. In addition it should be an excellent solvent. There should also be some gas on the planet in question that could be used in various biologically mediated cycles, as CO_2 is in the carbon cycle on earth.

The planet, therefore, should have an atmosphere and some near-surface liquid, although not necessarily an ocean. If the intensity of ultraviolet light or charged particles from the sun is intense at the planetary surface, there must be some place, perhaps below the surface, that is shielded from this radiation but that nevertheless permits useful chemical reactions to occur. Since after a certain period of evolution, lives of unabashed heterotrophy lead to malnutrition and death, autotrophs must exist. Chemoautotrophs are, of course, a possibility but the inorganic reactions that they drive usually require a great deal of energy; at some stage in the cycle, this energy must probably be provided by sunlight. Photoautotrophs, therefore, seem required. Organisms that live very far subsurface will be in the dark, making photoautotrophy impossible. Organisms that live slightly subsurface, however, may avoid ultraviolet and charged particle radiation and at the same time acquire sufficient amounts of visible light for photosynthesis.

Thermodynamically, photosynthesis is possible because the plant and the radiation it receives are not in thermodynamic equilibrium; for example, on the earth a green plant may have a temperature of about 300" K while the sun has a temperature of about 6,000" K. (K = Kelvin temperature scale, in which 0° K is absolute zero; 273" K, the freezing point of water; and 373" K, the boiling point of water at one atmosphere pressure.) Photosynthetic processes are possible in this case because energy is transported from a hotter to a cooler object. Were the source of radiation at the same (or at a colder) temperature as the plant, however, photosynthesis would be impossible. For this reason the idea of a subterranean plant photosynthesizing with the thermal infrared radiation emitted by its surroundings is untenable, as is the idea that a cold star, with a surface temperature similar to that of the earth, would harbour photosynthetic organisms.

Chemical requirements

It is possible to approach some of the foregoing chemical requirements and see just which atoms are implied. When atoms enter into chemical combination, the energy necessary to separate them is called the bond energy, a measure of how tightly the two atoms are bound to each other. Table 2 gives the bond energies of a number of chemical bonds, mostly involving abundant atoms. The energies are in electron volts (eV; 1 eV = 1.6×10^{-12} ergs). The symbols are as follows: H, hydrogen; C, carbon; N, nitrogen; 0, oxygen; S, sulfur; F, fluorine; Si, silicon; Bi, bismuth (very underabundant, biologically uninteresting, and present only as an illustration of the relatively weak chemical bonds in some metals). Bond energies generally vary between 10 eV and about 0.03 eV; double and triple bonds where two or three electrons are shared between two atoms tend to be more energetic than single bonds, single

Table 2: Energies of Representative Chemical Bonds

(Hydrogen bonds 0.08–0.45 eV; van der Waals bonds 0.04 eV)

bond	energy (eV)	bond	energy (eV)
N≡N	9.8	Si—0	3.8
C≡N	9.4	C—0	3.7
CEC	8.4	C—C	3.6
C—0	7.4	S—H	3.5
C=C	6.4	Si—H	3.1
H—F	5.4	C—N	3.0
O—H	4.8	Si—Si	1.8
N=N	4.4	N—N	1.7
C—H	4.3	Bi—Bi	1.1
N—H	4.1	O_2N—NO_2	0.57

bonds more energetic than hydrogen bonds where a hydrogen atom is shared between two other atoms, and hydrogen bonds more energetic than the very weak (van der Waals) forces that arise from the attraction of the electrons of one atom for the nucleus of another. At room temperature, atoms, free or bound, move with an average kinetic energy corresponding to about 0.02 eV. Some of the atoms have greater energies, some lesser. At any temperature a few will have energies greater than any given bond energy; hence bonds occasionally will break. The higher the temperature, the more atoms there are moving with sufficient energy to spontaneously break a given bond.

Suppose it is decided arbitrarily (although the decision will not critically affect the conclusions) that for life to exist at any time the fraction of bonds broken by random thermal motions must be no larger than 0.0001 percent. It then turns out that any hypothetical life where the structural bonds are based upon van der Waals forces can only exist where the temperature is below 40" K, for hydrogen bonds below about 400° K, for bonds of 2 eV below 2,000" K, and for bonds of 5 eV below 5,000" K. Now, 2,000" to 5,000" K are typical surface temperatures of stars; 400" K is somewhat above the highest surface temperature found on earth; and 40" K is about the cloud-top temperature of distant Neptune. Thus, over the entire range of temperatures, from cold stars to cold planets, there seem to exist chemical bonds of appropriate structural stability for life, and it would appear premature to exclude the possibility of life on any planet on grounds of temperature.

Life on earth lies within a rather narrow range of temperature. Above the normal boiling point of water, much loss of configurational structure or three-dimensional geometry occurs. At these temperatures proteins become denatured, in part because above the boiling point of water the hydrogen bonding and van der Waals forces between water and the protein disappear. Also, similar bonds within the protein molecule tend to break down. Proteins then change their shapes, their ability to participate in lock-and-key enzymatic reactions is gravely compromised, and the organism dies. Similar structural changes, some of them connected with the stacking forces between adjacent nucleotide bases, occur in the heating of nucleic acids. But it is significant that these changes are not fragmentations of the relevant molecules but rather changes in the ways they fold. There appears to be no reason that configurational bonds should not have been evolved that are stable at higher temperatures than terrestrial organisms experience. On planets hotter than the earth there seems to be no reason that slightly more stable configurational forces should not be operative in the local biochemistry.

Molecular factors

While the bonds that characterize life on earth are too weak at high temperatures, they are too strong at low temperatures, tending to slow down the rates of chemical reactions generally. There are less stable bonds (*e.g.*, hydrogen bonds, silicon–silicon bonds, and nitrogen-nitrogen bonds), however, that might play structural roles at significantly lower temperatures. At higher temperatures, multiple bonds (*e.g.*, in aromatic, or ring-shaped, hydrocarbons) might be utilized for life. There clearly is a rich variety of little-studied chemical reactions that proceed at reasonable rates either at much lower or at much higher temperatures than those on earth.

Except for bismuth and fluorine, all the atoms in Table 2 have relatively high cosmic abundances. At terrestrial temperatures, carbon is the unique atom for biological structure. Not only does it have high abundance but it forms a staggering variety of compounds of great stability, it lends itself to compounds that are configured by weaker bonds, and it enters into multiple bonds. These double- and triple-bonded molecules, among other useful properties, absorb long wavelength ultraviolet light, a process leading to the synthesis of a variety of more complex molecules. A photon of ultraviolet light at a wavelength of 2000 A has an energy of 6.2 eV, capable of breaking many bonds, and permitting more complex reactions among the resulting molecular fragments. Photons of blue light have energies of about 3 eV, and of red light about 2 eV.

Silicon compounds do not form double bonds. Silicon–oxygen bonds are slightly more stable than carbon-carbon bonds, but they tend to produce molecules like the silicates, which are crystals of the same unit repeated over and over again, rather than molecules with aperiodic side chains with potential information content. On low-temperature planets, silicon–silicon bonds are more promising than carbon bonds in terms of reaction times, but they do not form double bonds and the carbon abundance is likely to be greater. Nevertheless, silicon compounds may be of limited biological importance both on high-temperature and low-temperature worlds.

Hydrogen bonding confers on liquids the stability properties necessary for life. There seem to be very few reasonable candidates for liquid interaction media. By all odds water is the most suitable. The other candidates, all to some extent hydrogen bonded, are ammonia, hydrogen fluoride, hydrogen cyanide, and mixtures of liquid hydrocarbons. Hydrogen fluoride can be excluded because it is too scarce cosmically. The hydrocarbons are not good solvents of salts, but life elsewhere may not be based on the same acid–base chemistry as life on earth. The liquid range of water is larger than commonly thought, ranging from about 210" K in saturated salt solutions to 647" K at enormous atmospheric pressures. Water is the biological liquid medium of choice above 200" K, particularly in view of its extremely high cosmic abundance. At lower temperatures ammonia or hydrogen cyanide could serve as a liquid medium.

There are functional roles for specific atoms in biology, but except for considerations of structure and a liquid interaction medium they do not seem fundamental. For example, the energy-rich phosphate bonds in ATP are in fact of relatively low energy; they are about as energetic as the hydrogen bonds (see Table 2). The cell must store up large numbers of these bonds to drive a molecular degradation or synthesis. On high-temperature worlds the energy currency may be much more energetic per bond, and on low-temperature worlds much less energetic per bond.

It may be concluded that, in our present state of ignorance, it is premature to exclude life on grounds of temperature on any other planet, particularly when account is taken of the temperature heterogeneity of the other planets. But life does require an interaction medium, an atmosphere, and some protection from ultraviolet light and from charged particles of solar origin.

The conclusion that for the earth, carbon-based aqueous life is the most appropriate may be slightly suspect, since terrestrial life is manifestly carbon based and aqueous. In 1913 a U.S. biochemist, L.J. Henderson, published The Fitness of the Environment in which the biological advantages of carbon and water were stressed for the first time in terms of comparative chemistry. He was struck by the fact that those very atoms that are needed are just those atoms that are around; it remains a remarkable fact that atoms most useful for life do have very high cosmic abundances.

Exobiology

The search for extraterrestrial life. Exobiology, a term coined by a U.S. biologist, J. Lederberg, for the study of extraterrestrial life, has been called a science without a subject matter. It is certainly true that, as of the early 1970s, no strong evidence for life beyond the earth has been adduced. Exobiology, however, has deep significance even if extraterrestrial life is never found. The mere design of exobiological experiments forces man to examine critically the generality of his assumptions about life on earth. In addition, a lifeless neighbouring planet presents a very interesting quandary: How is it that life has originated and evolved on earth, but not on the planet in question? There is an entire spectrum of possibilities. A given planet may be lifeless and have no vestiges of primitive organic matter and no fossils of extinct life. It may be lifeless but may have either organic chemical or fossil relics. It may possess life of a simple sort or life of a quite complex biochemistry, physiology, and behaviour. It may possess intelligent life and a technical civilization. Establishment of any one of these five possibilities would be of fundamental biological importance.

The difficulties and opportunities inherent in exobiological exploration, in determining which of these five possibilities applies to a given planet, is most clearly grasped by imagining the situation reversed, with man on some neighbouring planet, say Mars, examining the earth for life with the full armoury of contemporary scientific instrumentation and knowledge. First a distinction must be made between remote and *in situ* testing. In remote testing light of any wavelength reflected from or emitted by the target planet can be examined, but with *in situ* studies samples of the planet must be acquired by visiting them or by sending instruments that land on the planet, perform experiments, and radio back their findings. Since biological exploration involves the detailed characterization of any life found, rather than its mere detection, *in situ* experiments are necessary even if remote experiments are successful.

Sensing methods

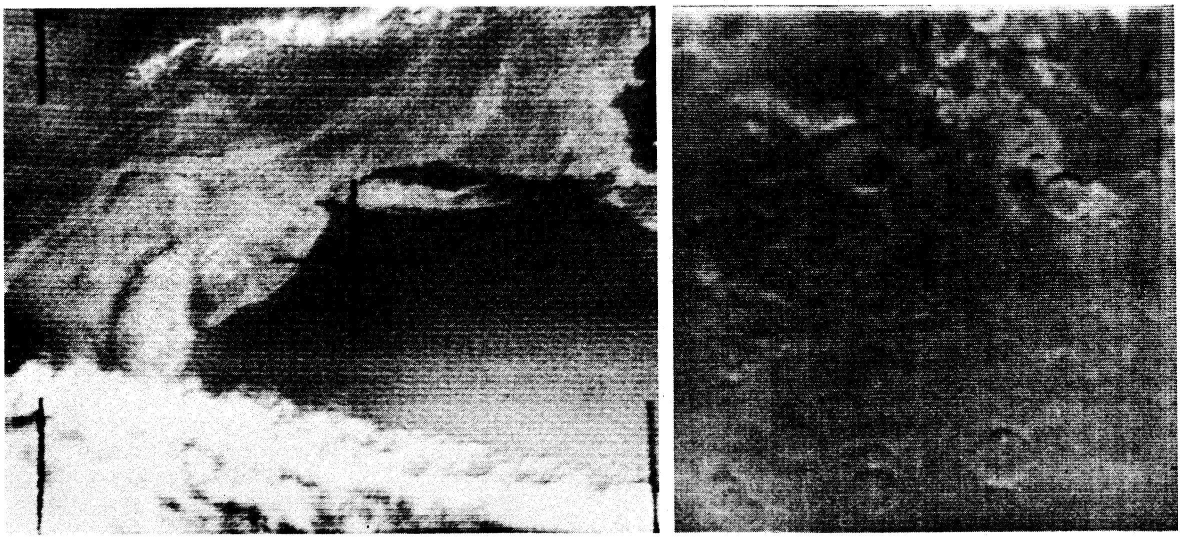
The bulk of the remote sensing methods are directed toward finding some thermodynamic disequilibrium on the planet. This may be a chemical disequilibrium, a mechanical disequilibrium, or *a* spectral disequilibrium. For example, it would be quite easy to determine spectroscopically from Mars that the earth's atmosphere contains large amounts of molecular oxygen and about one part per million (10^6) of methane. It would also be possible to calculate that, at thermodynamic equilibrium, the abundance of methane should be less than one part in 10^{35}. This huge discrepancy implies the existence of some process continuously generating methane on the earth so rapidly that methane increases to a very large steady-state abundance before it can be oxidized by oxygen. Now such a methane-production mechanism need not be biological. It is conceivable that relatively stable aromatic hydrocarbons were produced abiologically in the early history of the earth and that their slow thermal degradation leads to *a* continuous loss of methane from the planetary subsurface. But this and similar nonbiological explanations of the observed disequilibrium are unlikely. From Mars this thermodynamic discrepancy would be considered not as proof of life on earth but as *a* significant hint of life on earth. In fact the methane abundance on the earth is produced by bacteria that, in the course of the reduction of a more oxidized form of carbon, release methane. Some methane bacteria live in swamps (hence, the term marsh gas for methane), and others — a significant fraction — live in the intestinal tracts of cows and other ruminants. The methane abundance over India is probably larger than over most other areas of the world, and if an extraterrestrial observer knew how to interpret the methane disequilibrium accurately (which is unlikely) it would be possible for him to deduce cows on earth by spectrochemical analysis. The existence of relatively large quantities of methane in the presence of an excess of oxygen would remain a tantalizing but enigmatic hint of life on earth. Similarly, the large amount of oxygen might itself be a sign of life if one could reliably exclude the possibility that the photodissociation of water and the escape to space of hydrogen were the source of oxygen. Also such relatively complex reduced organic molecules as terpenes, a hydrocarbon given off by plants, might conceivably be detected spectroscopically, perhaps by a spectrometer in orbit about the earth. Not only would the chemical disequilibrium of terpenes in an excess of oxygen be suggestive of life, but equally suggestive would be the fact that terpenes are much more abundant over forested areas than over deserts.

Photographic observation

Photographic observations of the daytime earth from Mars would give equivocal results. Even with a resolution of 100 metres (that is, an ability to discriminate fine detail at high contrast only if its components are more than 100 metres apart) — better than the best Mariner photos of Mars at the end of the 1960s, and about a thousand times better than the best ground-based photographs of Mars — it would be extremely difficult to discern cities, canals, bridges, the Great Wall of China, highways, and other large-scale accoutrements of the earth's technical civilization. In satellite photographs with 100-metres resolution only about one in a thousand random photographs of the earth yields features even suggestive of life. As the ground resolution is progressively improved, it becomes increasingly easy to make out the regular geometrical patterns of cultivated fields, highways, airports, and so on. But these are only the products of a civilization recently developed on earth, and even photographs of the earth with a ground resolution of 10 metres, but taken 100,000 years ago, would still have shown no clear sign of life. The lights of the largest cities might be just marginally detectable from Mars at night. Seasonal changes in the colour or darkness of plants would be detectable from Mars, but such cycles might easily have nonbiological explanations, and their accurate interpretation would be difficult.

To detect individual animals a ground resolution of a few metres is required, and even here a low sun and long shadows are generally necessary. This detection could be accomplished with a large telescope in earth orbit. It would then be possible to determine, for example, that objects with the general shape of cows are frequent on the earth. But suppose that members of the civilization examining the earth thus remotely are not even approximately quadrupedal and do not immediately associate the shape of cows with life. They would nevertheless be able to

By courtesy of National Aeronautics and Space Administration



(Left) Eastern seaboard of the United States photographed by a TIROS weather satellite; Cape Cod (right), Long Island (centre), and Delaware Bay (centre left) can be seen. (Right) Surface of Mars photographed by Mariner 4 spacecraft. Both photographs at one-kilometre resolution. No sign of life, intelligent or otherwise, can be discerned on either planet.

deduce life. They would observe that certain locales on earth have a quantity of raised lumps connected to the ground by four stilts. It would be possible to calculate that wind and water erosion would cause the lumps to topple to the ground in geologically short periods of time. Such stilted lumps are mechanically unstable; they are not in equilibrium; if pushed hard, they fall. Accordingly, there must be a process for generating stilted lumps on the earth, and in short periods of time. It would be very difficult to avoid the implication that this generation process is biological.

Radio emission

A third detection technique arises upon scanning the radio spectrum of the earth. Because of domestic television transmission, the high-frequency end of the AM broadcast band, and the radar defense networks of the United States and of the Soviet Union, the amount of energy put out by the earth to space at certain radio frequencies is enormous. At some frequencies, if this radiation were interpreted as ordinary thermal emission, the temperature of the earth would have to be hundreds of millions of degrees, according to an estimate made by a Soviet astrophysicist, I.S. Shklovskii. Moreover, it would be possible to determine that this radio "brightness temperature" of the earth had been steadily increasing with time over the last several decades. Finally, it would be possible to analyze the frequency and time variation of these signals and deduce that they were not purely random noise.

Now imagine *in situ* studies by vehicles that enter the earth's atmosphere and land at some predetermined locale. There are many places on the earth (the ocean surface, the Gobi Desert, Antarctica) where large organisms are infrequent and a life-detection attempt based solely on television searches for large life forms would be a risky investment. On the other hand, if such an experiment were successful (the camera records a dolphin cavorting, a camel chewing its cud, a penguin waddling) it would provide quite convincing evidence of life.

Although the oceans, the Gobi Desert, and Antarctica are relatively devoid of large life forms, they are in many places replete with minute life forms. Therefore, micro-organism detectors would be a good investment. A television camera coupled to a microscope (optical or electron) would be a promising life detector if the sample acquisition problem could be solved: the early Dutch microscopist Anton van Leeuwenhoek had no difficulty at all in identifying as alive the little "animalcules" that he found in a drop of water, although nothing similar had previously been seen in the history of man.

In addition to morphological criteria for the detection of micro-organisms, there are metabolic and chemical criteria. For example, a sample of terrestrial soil, or seawater, say, might be acquired and introduced into a chamber containing food the investigators guess the earthlings might find tasty. Such food might be an abundant product of prebiological organic synthetic experiments. It could then be determined whether any characteristic molecules, such as carbon dioxide or ethanol, are produced metabolically, or whether the medium containing food and terrestrial sample changes its acidity or becomes cloudy because of the growth of micro-organisms, or it might be investigated whether there is heat given off in the chamber containing sample and food. Alternatively, photosynthesis could be tested by measuring the fixation of some gas, say carbon dioxide, as a function of illumination provided artificially to the sample by the instrument. Along chemical lines a direct test of terrestrial soil or seawater for optical activity might be made. Organic molecules could certainly be searched for with a combined gas chromatograph and mass spectrometer, or by a remote analytic chemistry laboratory. The detection of any organic matter would of course be interesting and relevant, whether or not it was biological in origin. Such criteria as have been used in the analysis of Precambrian sediments (described in *The antiquity of life,* above) might be used to test for biological origin.

Ambiguities of tests for life

It is remarkable, however, that many of these tests are ambiguous. It would be possible, for example, for the Martian investigator to guess wrong about what terrestrial organisms eat and to make incorrect assumptions about their structural chemistry or their interaction medium. If forms of regular geometry that do not move were detected microscopically, there might be serious questions of biological versus mineralogical origin. Chemical criteria (such as the expectation that if odd-numbered carbon chains are more prominent than even-numbered carbon chains, then life is detected) might not be valid unless it was certain which processes actually occurred in the prebiological organic chemistry of the planet in question. In addition, there might be the galling problem of contamination. The Martians' spacecraft might carry living organisms from their own planet and report them as detected on the planet earth. For this reason great care would have to be taken that spacecraft were rigorously sterilized.

In fact, many of these problems have already arisen in an analysis of a variety of meteorite called carbonaceous chondrites. These meteorites, which fall on earth probably from the asteroid belt contain about 1 percent organic matter by mass, far too much to be largely the result of terrestrial contamination. The most abundant organic molecules, however, are not clearly of biological origin, and some of the biologically more interesting molecules may be contaminants. Reports of optical activity have been contested and might alternatively be due to contamination. Geometrically interesting microscopic inclusions have been detected in these bodies. The most abundant inclusions, however, are probably mineralogical in origin, while the most highly structured and lifelike are very rare and, at least in some cases, are obviously due to contamination (in one case by ragweed pollen). Finally, claims have been made of the extraction of viable micro-organisms from the interiors of carbonaceous chondrites. These meteorites are porous, however, and "breathe" air in and out during their entry into the atmosphere. There also have been significant opportunities for their contamination after arrival on the earth. Moreover, one of the organisms extracted was a facultative aerobe. Since, as of the early 1970s, no planet in the solar system besides the earth was known to contain significant quantities of molecular oxygen, it seems quite curious that the complex electron transfer apparatus required for oxygen metabolism would be evolved out on the asteroid belt in expectation of ultimate arrival on the earth. Here, again, contamination has proved a serious hazard. The large amounts of organic matter found in carbonaceous chondrites, however, suggest that the production of organic molecules occurred with very great efficiency in the early history of the solar system.

From such a hypothetical exercise as the instrumental detection of life on earth by an extraterrestrial observer, and from the actual experience acquired in the analysis of carbonaceous chondrites, the following conclusions can be drawn: There is no single and unambiguous "life detector." There are instruments of great generality that make few ambiguous assumptions about the nature of extraterrestrial organisms, particularly their chemistry. These systems, however, require a fair degree of luck (an animal must walk by during the operating lifetime of the instrument), or they require the solution of difficult instrumental problems (such as the acquisition and preparation of samples for remote microscopic examination). Other instruments, such as metabolism detectors, have great sensitivity and are directed at the more abundant micro-organisms. They are quite specific, however, and are critically dependent upon certain assumptions (for example, that extraterrestrial organisms eat sugars) that are in the early 1970s no better than informed guesses. Therefore, an array of instruments, both very general and very specific, seems required. Stringent sterilization of such spacecraft appears necessary, both to avoid confusion of the life-detection experiments, and to prevent interaction of contaminants with the indigenous ecology. Many of the instruments and strategies discussed in the preceding paragraphs continua to be adapted by the United States and the Soviet Union in attempts to search for life on the moon and the nearby planets (see SPACE EXPLORATION).

An exobiological survey of the solar system. A brief survey of the physical environments and biological pros-

Biological prospects of the moon

pects of the moons and planets of the solar system, so far as is known by the early 1970s, follows. The moon's surface seems inhospitable to life of any sort. The diurnal temperatures range from about 100" to about 400" K. In the absence of any significant atmosphere or magnetic field, ultraviolet light and charged particles from the sun penetrate unimpeded to the lunar surface, delivering in less than an hour a dose lethal to the most radiation-resistant micro-organism known. For other reasons already mentioned, the absence of an atmosphere and of any liquid medium on the surface also argues against life. The subsurface environment of the moon is not nearly so inclement. About a metre or so subsurface there is no penetration of ultraviolet light or solar protons, and the temperature is maintained at a relatively constant value about 230" K. Even there, however, the absence of an atmosphere and the probable absence of abundant liquids make the biological prospects rather dim.

It is not out of the question, however, that prebiological organic matter, produced in the early history of the moon, might be found sequestered beneath the lunar surface. Such organic matter may have been produced either in an original lunar atmosphere that has subsequently been lost to space, or in a secondary lunar atmosphere produced by release of gases after the formation of the moon, and also subsequently lost to space. The depth at which such organic matter may be found depends upon the unknown history of the early lunar atmosphere, if any, and upon whether the moon has, on the whole, gained or lost matter due to meteoritic impact. An apparent gaseous emission near the lunar crater Alphonsus was recorded in 1958 and a spectral identification was made of the molecule C_2, a likely organic fragment, but this identification subsequently has been disputed.

Because of contamination by unmanned spacecraft, the lunar surface had accumulated a microbial load estimated by the late 1960s at some 100,000,000 micro-organisms. Since such organisms will be immediately killed unless shielded from radiation, and since the likelihood of their growth seems remote, such contamination may not be a serious problem in subsequent microbial analysis of returned lunar samples. A much more serious contamination problem occurs during the acquisition of such samples by astronauts. Samples obtained during the historic Apollo 11 moon landing in July 1969 were tested for possible organic molecules, but results were inconclusive. Such a finding might shed significant light on the early history of organic molecules in the solar system.

The environment of Mercury is rather like that of the moon. Its surface temperatures range from about 100" to about 620" K, but about a metre subsurface the temperature is constant, very roughly at comfortable room temperature on earth. But the absence of any significant atmosphere, the unlikelihood of bodies of liquid, and the intense solar radiation make life unlikely.

The possibility of life on Mars

Mars has long been considered a likely abode for life. It has a thin atmosphere comprised largely of carbon dioxide. The surface pressure is equivalent to the pressure at an altitude about 100,000 feet above the earth's surface. Small amounts of carbon monoxide and water also have been detected, and molecular nitrogen may be present. Ultraviolet spectroscopy of Mars from rockets and from the space probes Mariners 6 and 7 indicates that ultraviolet light at wavelengths longer than 2000 A is falling unimpeded upon the Martian surface. The temperature at an equatorial noontime locale on Mars is about 300" K (about 85° F), comfortable by terrestrial standards. The variation from day to night at the equator, however, or from equator to pole at noontime are both over 100" K. The nighttime and polar temperatures on Mars are very cold—so cold that atmospheric water vapour probably condenses out each night on Mars, and atmospheric water vapour and carbon dioxide probably condense out at the winter pole, accounting for the polar caps. While large open pools of pure liquid water are not expected on Mars because of the low atmospheric pressures, briny pools in salt deposits are possible, and there is some reason to believe that liquid water may form temporarily each morning on Mars in soil interstices where the vaporization of water is limited by diffusion.

Experiments have been performed in which terrestrial micro-organisms are introduced into simulated Martian environments. In one set of experiments every sample of terrestrial soil from a wide variety of locales contained some micro-organisms that survived Martian conditions indefinitely. These were micro-organisms capable of living in the absence of oxygen; they were shielded from ultraviolet light by small grains of surface material and had some degree of tolerance to cold. The Martian surface contains a reddish material (probably ferric oxide) that strongly absorbs ultraviolet light. When small quantities of water were introduced into simulated Martian environments in experiments, growth of terrestrial micro-organisms occurred. Even when liquid water was available for only about 15 minutes a day, significant growth occurred. These experiments suggest that there are no insuperable physical or chemical problems even for terrestrial life on Mars and emphasize the necessity for sterilization of Mars-bound space vehicles. Since Martian organisms, if any, should be even better adapted to the Martian environment than terrestrial organisms, such experiments indicate, although they certainly do not prove, that Mars may be an abode of life. Hypothetical and speculative ecological schemes have been designed in which, for example, photosynthetic organisms live slightly subsurface where ultraviolet light does not penetrate but where visible light does, or in which organisms have evolved shields opaque to ultraviolet light but transparent to visible light.

Direct evidence for life on Mars has been claimed for many decades. The prevailing opinion in the early 1970s, however, was that such remote clues to life on Mars must be interpreted with great caution. The corresponding difficulty in detecting life on earth from a vantage point on Mars has already been mentioned. The first such argument was posed by a French astronomer, E.L. Trouvelot in 1884: "Judging from the changes that I have seen to occur from year to year in these spots, one could believe that these changing grayish areas are due to Martian vegetation undergoing seasonal changes." The seasonal changes on Mars have been reliably observed, not only visually but also photometrically. There is a conspicuous springtime increase in the contrast between the bright and dark areas of Mars. Accompanying colour changes have been reported, but their reality has been disputed. While such changes have been attributed to the growth of vegetation, seasonally variable dust storms are an equally convincing possibility.

The most famous case, historically, for life on Mars is the discovery of the "canals," a set of apparent thin straight lines that cross the Martian bright areas and extend for hundreds and sometimes thousands of kilometres. They change seasonally as do the Martian dark areas. These lines, first systematically observed by an Italian astronomer, G.V. Schiaparelli, in 1877, were further cataloged and popularized by a U.S. astronomer, Percival Lowell, around the turn of the century. Lowell argued from the unerring straightness of the lines that they could not be of geological origin but must instead be the artificial constructs of a race of intelligent Martians. He suggested that they might be channels carrying water from the melting polar caps to the parched equatorial cities of Mars. While considerable skepticism has been expressed about these straight lines, there is no doubt that approximately rectilinear features do exist on the Martian surface. Less exciting explanations, however, are more probable, including crater chains, terrain contour boundaries, faults, mountain chains, and ridges analogous to the suboceanic ridge systems that are features of the earth.

Several spectroscopic claims have been advanced for the existence of organic molecules on Mars in quantities representing a marked thermodynamic disequilibrium. The difficulties in unambiguous detection of such molecules from the earth have proved to be very great, however, and no such claims have stood the test of time. Thus, while life on Mars seems possible, there is no direct evidence for its

Photograph, from a TIROS weather satellite, of a region near Cochrane, Ontario. The crisscross pattern in white (top left) shows logging swaths, a sign of intelligent life on earth.
By courtesy of National Aeronautics and Space Administration

presence. Mars stands lonely and enigmatic in the night sky, awaiting more direct explorations.

Features of Venus

According to both ground-based and spaceborne observations, the average surface temperatures of Venus are around 750° K. It does not seem likely, either at the poles or on the tops of the highest Venus mountains, that the surface temperature will be below 400" K, and noontime temperatures are probably significantly hotter than 700" K. Thus, quite apart from the other surface conditions, the temperatures on Venus seem too hot for terrestrial life. It is still not possible to exclude a Venus surface life with a rather different chemistry, although hydrogen bonding would be much less suitable for the geometrical configuration of polymers on Venus than it is on earth. The clouds of Venus, however, are another matter. There, carbon dioxide, sunlight, and (according to the results of the Venera space vehicles) water are to be found. These are the prerequisites for photosynthesis. Some molecular nitrogen also is expected at the cloud level, and some supply of minerals can be expected from dust convectively raised from the surface. The cloud pressures are about the same as on the surface of the earth, and the temperatures in the lower clouds also are quite earthlike. Despite the fact that there is little oxygen, the lower clouds of Venus are the most earthlike extraterrestrial environment known. While there are no recorded cases of organisms on earth that lead a completely airborne existence throughout their life cycle, it is not impossible that such organisms could exist in the vicinity of the Venus clouds, perhaps buoyed, as is a fish by its swim bladder, to avoid downdrafts carrying them to the hotter lower atmosphere.

A similar speculation can be entertained with regard to the lower clouds of Jupiter. On Jupiter the atmosphere is composed of hydrogen, helium, methane, ammonia, and probably neon and water vapour. But these are exactly those gases used in primitive earth-simulation experiments directed toward the origin of life. Laboratory and computer experiments have been performed on the application of energy to simulated Jovian atmospheres. In addition to the immediate gas-phase products, such as hydrogen cyanide and acetylene, more complex organic molecules, including aromatic hydrocarbons, are formed in lower yield. The visible clouds of Jupiter are vividly coloured, and it is possible that their hue is due to such coloured organic compounds. There is also an apparent absorption feature near 2600 A, in the ultraviolet spectrum of Jupiter, which has been attributed both to aromatic hydrocarbons and to nucleotide bases. In any event it is likely that organic molecules are being produced in significant yield on Jupiter; it is possible that Jupiter is a vast planetary laboratory that has been operating for 5,000,000,000 years on prebiological organic chemistry.

Satellites and comets of the solar system

The other Jovian planets, Saturn, Uranus, and Neptune, are similar in many respects to Jupiter, although much less is known about them. Their cloud-top temperatures progressively decrease with distance from the sun. In the case of Saturn, microwave studies have indicated that the atmospheric temperature increases with depth below the clouds; similar situations are expected on Jupiter, Uranus, and Neptune. Thus, it is by no means clear that the low temperatures of the upper clouds of the Jovian planets apply to the lower clouds, or to the underlying atmosphere. The environment of Pluto is almost completely unknown. In addition to these planets, the solar system contains 32 natural satellites, some of which, such as Titan, a satellite of Saturn, and Io, a satellite of Jupiter, appear to have atmospheres. There are also tens of thousands of comets, which, judging from their spectra, contain organic molecules, as well as some thousands of asteroids and asteroidal fragments revolving about the sun between the orbits of Mars and Jupiter. These are the presumed sources of the carbonaceous chondrites, which contain organic matter.

In short, there is a wide range of environments of biological interest within the solar system. There is no direct evidence for extraterrestrial life on these planets, but, on the other hand, there is no strong evidence against life on many of these worlds. Beyond this is the near certainty that biologically interesting organic molecules will be found throughout the solar system.

Intelligent life beyond the solar system. For thousands of years man has wondered whether he is alone in the universe or whether there might be other worlds populated by creatures more or less like himself. The common view, both in early times and through the Middle Ages, was that the earth was the only "world" in the universe. Nevertheless, many mythologies populated the sky with divine beings, certainly a kind of extraterrestrial life. Many early philosophers held that life was not unique to the earth. Metrodorus, an Epicurean philosopher in the 3rd and 4th centuries BC, argued that "to consider the earth the only populated world in infinite space is as absurd as to assert that in an entire field sown with millet, only one grain will grow." Since the Renaissance there have been several fluctuations in the fashion of belief. In the late 18th century, for example, practically all informed opinion held that each of the planets was populated by more or less intelligent beings; in the early 20th century, by contrast, the prevailing informed opinion (except for the Lowellians) held that the chances for extraterrestrial intelligent life were insignificant. In fact the subject of intelligent extraterrestrial life is for many people a touchstone of their beliefs and desires, some individuals very urgently wanting there to be extraterrestrial intelligence, and others wanting equally fervently for there to be no such life. For this reason it is important to approach the subject in as unbiased a frame of mind as possible. A respectable modern scientific examination of extraterrestrial intelligence is no older than the 1950s. The probability of advanced technical civilizations in our galaxy depends on many controversial issues.

The Green Bank formula

A simple way of approaching the problem, which illuminates the parameters and uncertainties involved, has been devised by a U.S. astrophysicist, F.D. Drake. The number N of extant technical civilizations in the galaxy can be expressed by the following equation (the so-called Green Bank formula):

$$N = R_* f_p n_e f_l f_i f_c L$$

where R_* is the average rate of star formation over the lifetime of the galaxy; f_p is the fraction of stars with planetary systems; n_e is the mean number of planets per star that are ecologically suitable for the origin and evolution of life; f_l is the fraction of such planets on which life in fact arises; f_i is the fraction of such planets on which intelligent life evolves; f_c is the fraction of such planets on which a technical civilization develops; and L is the mean lifetime of a technical civilization. What follows is a brief consideration of the factors involved in choosing numerical values for each of these parameters, and an indication of some currently popular choices. In several cases these estimates are no better than informed guesses and no very great reliability should be pretended for them.

There are about 2×10^{11} stars in the galaxy. The age of the galaxy is about 10^{10} years. A value of $R_* = 10$ stars per year is probably fairly reliable. While most contemporary theories of star formation imply that the origin of planets is a usual accompaniment of the origin of stars, such theories are not well enough developed to merit much confidence. Through the painstaking measurement of slight gravitational perturbations in the proper motions of stars, it has been found that about half of the very nearest stars have dark companions with masses ranging from about the mass of Jupiter to about 30 times the mass of Jupiter. The nearest of these dark companions orbit Barnard's star, which is only six light years from the sun and is the second nearest star system. The most direct indication that planetary formation is a general process throughout the universe is the existence of satellite systems of the major planets of our own solar system. Jupiter, with 12 satellites, Saturn with 10, and Uranus with 5, each closely resemble miniature solar systems. It is not known what the distribution of distances of planets from their central star are in other solar systems and whether they tend to vary systematically with the luminosity of the parent star. But considering the wide range of temperatures that seem to be compatible with life, it can be tentatively concluded that $f_p n_e$ is about one.

Because of the apparent rapidity of the origin of life on earth, as implied by the fossil record, and because of the ease with which relevant organic molecules are produced in primitive earth simulation experiments, the likelihood of the origin of life over a period of billions of years seems high, and some scientists believe that the appropriate value of f_l is also about one. For the quantities of f_i and f_c the parameters are even more uncertain. The vagaries of the evolutionary path leading to the mammals, and the unlikelihood of such a path ever being repeated has already been mentioned. On the other hand, intelligence need not necessarily be restricted to the same evolutionary path that occurred on the earth; intelligence clearly has great selective advantage, both for predators and for prey.

Similar arguments can be made for the adaptive value of technical civilizations. Intelligence and technical civilization, however, are clearly not the same thing. For example, dolphins appear to be very intelligent, but the lack of manipulative organs on their bodies has apparently limited their technological advance. Both intelligence and technical civilization hove evolved about halfway through the relevant lifetime of the earth and sun. Some, but by no means all, evolutionary biologists would conclude that the product $f_i f_c$ taken as 10^{-2} is a fairly conservative estimate.

Still more uncertain is the value of the final parameter, L, the lifetime of a technical civilization. Here, fortunately for man, but unfortunate for the discussion, there is not even one example. Contemporary world events do not provide a very convincing counterargument to the contention that technical civilizations tend, through the use of weapons of mass destruction, to destroy themselves shortly after they come into being. If we define a technical civilization as one capable of interstellar radio communication, our technical civilization is only a few decades old. If then L is about ten years, multiplication of all of the factors assumed above leads to the conclusion that there is in the second half of the 20th century only about one technical civilization in the galaxy—our own. But if technical civilizations tend to control the use of such weapons and avoid self-annihilation, then the lifetimes of technical civilizations may be very long, comparable to geological or stellar evolutionary time scales; the number of technical civilizations in the galaxy would then be immense. If it is believed that about 1 percent of developing civilizations make peace with themselves in this way, then there are about 1,000,000 technical civilizations extant in the galaxy. If they are randomly distributed in space, the distance from the earth to the nearest such civilization will be several hundred light-years. These conclusions are, of course, very uncertain.

Communications with extraterrestrial civilizations

How is it possible to enter into communication with another technical civilization? Independent of the value of L, the above formulation implies that there is about one technical civilization arising every decade in the galaxy. Accordingly, it will be extraordinarily unlikely for man soon to find a technical civilization as backward as his. From the rate of technical advance that has occurred on the earth in the last few hundred years, it seems clear that man is in no position to project what future scientific and technical advances will be made even on earth in the next few hundred years. Very advanced civilizations will have techniques and sciences totally unknown to 20th-century man. Nevertheless man already has a technique capable of communication over large interstellar distances. This technique, already encountered in the discussion of life on earth, is radio transmission. Imagine that we employ the largest radio telescope available on earth, the 1,000-foot-diameter dish of Cornell University, the Arecibo Observatory in Puerto Rico, and existing receivers. and that the identical equipment is employed on some transmitting planet. How distant could the transmitting and receiving planets be for intelligible signals to be transmitted and received? The answer is a rather astonishing 1,000 light-years. Within a volume centred on the earth, with a radius of 1,000 light-years, there are over 10,000,000 stars.

There would of course be problems in establishing such radio communication. The choices of frequency, of target star, of time constant, and of the character of the message would all have to be selected by the transmitting planet so that the receiving planet would, without too much effort, be able to deduce the choices. But none of these problems seem insuperable. It has been suggested that there are certain natural radio frequencies (such as the 1420-megacycle line of neutral hydrogen) that might be tuned to; the first choice might be to listen to stars of approximately solar spectral type: in the absence of a common language there nevertheless are messages whose intelligent origin and intellectual content could be made very clear without making many anthropocentric assumptions.

Because of the expectation that the earth is relatively very backward, it does not make very much sense to transmit messages to hypothetical planets of other stars. But it may very well make sense to listen for radio transmissions from planets of other stars. Project Ozma, a very brief program of this sort, oriented to two nearby stars, Epsilon Eridani and Tau Ceti, was organized in 1960 by Drake. On the basis of the Green Bank formula, it would be very unlikely that success would greet an effort aimed at two stars only 12 light-years away, and Project Ozma was unsuccessful. It remains, however, the first pioneering attempt at interstellar communication. Related programs were organized on a larger scale and with great enthusiasm in the 1960s in the U.S.S.R., where a state scientific commission devoted to such an effort was organized. Other communication techniques including laser transmission and interstellar spaceflight have been discussed seriously and may not be infeasible, but if the measure of effectiveness is the amount of information communicated per unit cost, then radio is the method of choice.

The search for extraterrestrial intelligence is an extraordinary pursuit, in part because of the enormous significance of possible success, but in part because of the unity it brings to a wide range of disciplines: studies of the origins of stars, planets, and life; of the evolution of intelligence and of technical civilizations; and of the political problem of avoiding man's self-annihilation. But at least one point is clear. In the words of Loren Eiseley (also from *The Immense* Journey),

> Lights come and go in the night sky. Men, troubled at last by the things they build, may toss in their sleep and dream bad dreams, or lie awake while the meteors whisper greenly overhead. But nowhere in all space or on a thousand worlds will there be men to share our loneliness. There may be wisdom; there may be power; somewhere across space great instruments, handled by strange, manipulative organs, may stare vainly at our floating cloud wrack, their owners yearning as we yearn. Nevertheless, in the nature of life and in

the principles of evolution we have had our answer. Of men [as are known on earth] elsewhere, and beyond, there will be none forever.

BIBLIOGRAPHY

The definition of life and life on Earth: H.J. MOROWITZ, *Energy Flow in Biology (1968)*; A.L. LEKNINGER, *Bioenergetics: The Molecular Basis of Biological Energy Transformations* (1965); P. HANDLER (ed.), *Biology and the Future of Man* (1970); G. and M. BEADLE, *The Language of Life: An Introduction to the Science of Genetics* (1966); V.M. INGRAM, *The Biosynthesis of Macromolecules* (1965); G.N. COHEN, *Biosynthesis of Small Molecules* (1967); R.B. SETLOW and E.C. POLLARD, *Molecular Biophysics* (1962); "Explanation *in* Biology," J. *Hist. Biol.*, vol. 2, no. 1 (1969); L.J. HENDERSON, *The Fitness of the Environment: An Inquiry into the Biological Significance of the Properties of Matter* (1958); J. KEOSIAN, *The Origin of Life*, 2nd ed. (1968).

Origin of life and extraterrestrial life: I.S. SHKLOVSKII and C. SAGAN, *Intelligent Life in the Universe* (1966); C.S. PITTENDRIGH, W. VISHNIAC, and J.P.T. PEARMAN (eds.), *Biology and the Exploration of Mars* (1966); E.A. SHNEOUR and E.A. OTTESEN (comps.), *Extraterrestrial Life: An Anthology and Bibliography* (1966); S. GLASSTONE, *The Book of Mars* (1968); S.W. FOX (ed.), *The Origins of Prebiological Systems and of Their Molecular Matrices* (1965); A.G.W. CAMERON (ed.), *Interstellar Communication* (1963); W. SULLIVAN, *We Are Not Alone: The Search for Intelligent Life on Other Worlds*, rev. ed. (1966); C. SAGAN, *The Cosmic Connection* (1973).

Evolution of life: G.G. SIMPSON, *The Meaning of Evolution* (1949) and *This View of Life* (1964); L. EISELEY, *The Immense Journey* (1957); G.J. HARDIN, *Nature and Man's Fate* (1959).

(C.Sn.)

Life-Span

It is a commonplace that organisms die, some after only a brief existence, like that of the mayfly whose adult life burns out in a day, and others whose term seems almost endless, like that of the gnarled bristlecone pines that have lived thousands of years. The limits of the life-span of each species appear to be determined ultimately by heredity. Locked within the code of the genetic material are instructions that specify the age beyond which a species cannot live given even the most favourable conditions. But many environmental factors act to diminish that upper age limir, for conditions are never altogether favourable.

GENERAL BACKGROUND

Maximum and average life-spans

The maximum life-span is a theoretical number whose exact value cannot be determined from existing knowledge about an organism; it is often given as a rough estimate based on the longest lived organism known to date. A mole meaningful measure is the average life-span, a statistical concept derived by the analysis of mortality data for populations of each species. A related term is the expectation of life, a hypothetical number computed for humans from mortality tables drawn up by insurance companies. It represents the average number of years that a group of persons, all born at the same time, might be expected to live, and it is based on the changing death rate over many past years.

The concept of life-span implies that there is an individual whose existence has a definite beginning and end. What constitutes the individual in most cases presents no problem: among organisms that reproduce sexually the individual is a certain amount of living substance capable of maintaining itself alive and endowed with hereditary features that are in some measure unique. In some organisms, however, extensive and apparently indefinite growth takes place and reproduction may occur by division of a single parent organism, as in many protists, including bacteria, algae, and protozoans. If these divisions are incomplete, a colony results; if the parts separate, genetically identical organisms are formed. In order to consider life-span in such organisms, the individual must be defined arbitrarily since the organisms are continually dividing. In a strict sense, the life-spans in such instances are not comparable to those forms that are sexually produced.

The beginning of an organism can be defined by the formation of the fertilized egg in sexual forms; or by the physical separation of the new organism in asexual forms (many invertebrate animals and many plants). In animals generally, birth is considered to be the beginning of the life-span. The timing of birth, however, is so different in various animals that it is only a poor criterion. In many marine invertebrates the hatchling larva consists of relatively few cells, not nearly so far along toward adulthood as a newborn mammal. For even among mammals, variations are considerable. A kangaroo at birth is about an inch long and must develop further in the pouch, hardly comparable to a newborn deer, who within minutes is walking about. If life-spans of different kinds of organisms are to be compared, it is essential that these variations be accounted for. The end of an organism's existence results when irreversible changes have occurred to such an extent that the individual no longer actively retains its organization. There is thus a brief period during which it is impossible to say whether the organism is still alive, but this time is so short relative to the total length of life that it creates no great problem in determining life-span.

Potential immortality

Some organisms seem to be potentially immortal. Unless an accident puts an end to life, they appear to be fully capable of surviving indefinitely. This faculty has been attributed to certain fishes and reptiles, which appear to be capable of unlimited growth. Without examining the various causes of death in detail (see DEATH) a distinction can be made between death as a result of internal changes (*i.e.*, aging) and death as a result of some purely external factor, such as an accident. It is notable that the absence of aging processes is correlated with the absence of individuality. In other words, organisms in which the individual is difficult to define, as in colonial forms, appear not to age.

HUMAN BEINGS

The exact duration of human life is unknown, although there is presumably a maximum life-span for the human race established in the genetic material. At first thought, this statement seems irrational. Surely no human being can live 1,000 years. Even though all may agree that the likelihood of an individual living 1,000 years is infinitesimal, there is no scientific proof that this statement is or is not true. The indeterminacy of the maximum limit of human life is made more comprehensible if one chooses a number that may appear to be a more reasonable limit.

Since there is no verified instance of a person having lived 150 years, this number may, for purposes of illustration, be arbitrarily accepted as the maximum limit of the span of human life. But if the possibility is admitted that an individual may live exactly 150 years, there is no valid reason for rejecting the possibility that some other individual may live 150 years and one minute. And if 150 years and one minute is accepted, why not 150 years and two minutes, and so on? Thus, based on existing knowledge of longevity, a precise figure for the span of human life cannot be given.

Studies on longevity. Much information concerning the inheritance of longevity has come from the study of genealogical records of nobility and landed gentry.

The early genealogical studies were criticized on the grounds that the downward trend in the death rate (attributable generally to scientific advancements) introduced a spurious correlation in statistics derived from records extending over long periods of time. It was argued that in some instances records were included of persons who, at the time of the study, had not had the opportunity of living out their possible life-span. The general finding of such investigations was that the expectation of life of sons of long-lived parents (*i.e.*, those living to age 70 years or older) was greater than that of sons of shorter-lived parents (*i.e.*, those having attained less than age 50 at the time of death).

An American biostatistician attempted to avoid the defects of genealogical studies by collecting records of the family histories of 365 nonagenarians (90-year-old persons) and of a comparison group of 143 individuals of

The TIAL concept

varying ages, selected because all of their six immediate ancestors were dead. The study introduced the concept of "total immediate ancestral longevity," or TIAL—the sum of the ages at death of the two parents and the four grandparents of a given person—as a measure of longevity. This number is unlikely to be greater than 600 or less than 90. The average TIAL of the nonagenarians and centenarians definitely exceeded that of the comparison group. This held true not only for the six immediate ancestors as a group but also for each category—father, mother, paternal and maternal grandparents. In the same study, investigators also computed the expectation of life for sons of fathers as classified in three groups by age at death: (1) under age 50, (2) from age 50 to age 79, and (3) age 80 or over. The expectation of life for the three groups at birth was 47.0, 50.5, and 57.2 years, respectively. The same relative ranking continued through the lifetime of the sons, their expectation of life at age 40 being 27.3, 28.9, and 32.0 years, respectively.

While certain doubts have been raised about the validity of these as well as earlier studies, taken at their face value, these data show clearly that long-lived persons had parents and grandparents who lived longer than the parents and grandparents of shorter lived persons.

Since longevity is important in life insurance underwriting, several studies have been made of the relationship between heredity and the life-span by an analysis of life insurance records. Such analyses showed that policyholders both of whose parents were living when the policy was written live longer than those whose parents were dead when the policy was written. These results are in conformity with those obtained from genealogical records and family histories.

Each of the various types of studies of the inheritance of longevity—genealogical records, life insurance records, and family histories of the general population—has limitations that restrict the applicability of the findings. The principal studies indicate, nevertheless, that the children of long-lived parents are more likely to be long-lived than are the children of short-lived parents. Conversely, the immediate ancestors—parents and grandparents—of long-lived persons on the average are older at death than are the immediate ancestors of persons who die at a relatively young age. These studies support the conclusion, mentioned earlier, that longevity is determined in part by heredity.

Actual versus possible life-span. It should be observed that this conclusion relates to the inheritance of longevity—the observed expression of the span of life—and not to the span of life itself. The actual length of life itself is shorter than the possible life-span, since the former reflects the effect of unfavourable environmental factors. In the absence of any biological data from which the maximum limit of the span of life can be determined precisely, an estimate of the limit must be obtained from observation of the actual length of life of persons who already have died. But such observations cannot establish a fixed limit for the span of life.

The estimation of the length of the span of life from observed data is a form of sampling from a large but incomplete population. The tabulation of the ages at death of a large number of persons from a large general population of the United States will give an asymmetrical frequency distribution with two modes, or peaks, of highest frequency: the first at age less than one year and the second between ages 75 and 80 years. The frequency distribution is bounded by age zero at the lower limit but there is no boundary at the upper limit. The number of deaths of persons whose length of life is near the upper limit of this frequency distribution (*e.g.*, 100 years or more) varies from year to year. The age of the oldest person dying also varies from year to year.

The number of deaths of centenarians (100-year-old persons) depends in part upon the number of deaths counted. Ages at death are frequently unverified, so that the true numbers of centenarians almost certainly deviate from those given in official vital statistics. Moreover, only a very small proportion of the deaths that have occurred throughout the history of the human race have been registered. The potential number of future deaths greatly exceeds the number that already has occurred. Statistical theory supports the expectation that as the total number of deaths continues to increase, the death of a person whose length of life will be longer than that of any person previously known will be recorded.

Observation of the length of life of persons who have died can show that it is possible for a human being to live to the oldest age recorded as of any specified date and can provide an estimate of the relative frequency or probability of that event. But such observations do not provide a logical basis for fixing any age as the maximum possible limit of the life-span.

The continuation of the worldwide decline in the death rate will naturally result in an increase in the number of persons who live until age 100 years or more. Since the number of persons who may live to an advanced age, such as 110 or 115 years, is directly related to the number of persons who live to age 100, an increase in the latter number will increase the probability that the death of an individual attaining some greater age (*e.g.*, 115 years) will be recorded at some future date.

Famous age records

Many instances of persons alleged to have died at an age considerably greater than 100 years have been recorded. Statements concerning the age at death of biblical characters such as Methuselah can be dismissed, since scientific verification is impossible. Three of the most frequently cited cases of more recent times are: Thomas Parr, who died in November 1635 at the alleged age of 152 years; Henry Jenkins, who died in December 1670 at the alleged age of 169 years; and Catherine, countess of Desmond, who died in 1604 at the alleged age of 140 years. William Harvey, a famous English physician, performed an autopsy on Thomas Parr and the account of the autopsy was cited for many years as evidence that Harvey—in his paper—had confirmed Parr's age. Quite apart from the fact that it is impossible accurately to determine the age of a person by an autopsy, Harvey made no attempt to verify Parr's age but merely referred to the current estimates. Subsequent investigations have revealed that no proof exists of the age at death of any of these three individuals and that their reported ages were based solely upon hearsay.

An example with more definite documentation is that of Christian Jacobsen Drakenberg, stated to have been born on November 18, 1626, and to have died on October 9, 1772, aged 145 years and 325 days. Although the authenticity of his age was attested to by many persons, including two celebrated Scandinavian actuaries, later investigations cast doubt upon the record. It is difficult to accept the statements concerning Drakenberg's age at death, since this age is more than 30 years greater than the next oldest verified age at death—a difference that in itself casts doubt on its authenticity.

Of eight individuals for whom records substantiate the fact that each had lived more than 108 years, seven were females. Six of the eight were more than 110 years old at death. The oldest was Pierre Joubert, who was born July 15, 1701, and died November 16, 1814, aged 113 years and 124 days. Discounting the Drakenberg record, this is the oldest age at death that has been generally accepted as authentic.

It may be concluded that the span of human life is at least 114 years, but that this is not the maximum upper limit. This does not mean the span of life of each individual now living or to be born in the future is at least 114 years. The span of life, since it is determined by heredity, varies from one individual to another as do other genetically determined traits.

Factors influencing life-span of man

A significant proportion of human embryos and fetuses die before birth. Other infants at birth have defects that limit their span of life to a few years. Some malformations (*e.g.*, certain cardiovascular defects) are developmental rather than genetic in the strict sense of the word and can be corrected so that the length of life of such persons is extended.

In the past the length of life of most individuals has been considerably shorter than their possible span of life because of unhealthful environmental factors. As these factors are increasingly brought under control or eliminated, the actual length of life will approach more closely the span of life. At the end of the 18th century the expectation of life at birth in North America and northwestern Europe was about 35 or 40 years. By 1970 it exceeded 70 years, and at some future date the death of a person at an authenticated age of more than 114 years can be expected.

There is no evidence that the span of human life has increased since the beginning of recorded history. Neither is there any evidence that the death rate of centenarians has decreased. The expected increase in the number of centenarians results from a decrease in the death rate at ages under 100 years and not from any demonstrable increase in the maximum length of the span of life. The remarkable increase in the average length of life during the past 2,000 years—from 20–25 years to 70 years under favourable conditions—has increased the likelihood that a person may live to the maximum limit of his span of life.

SUBHUMAN ANIMALS

Much of what is known of the length of life of animals other than man derives from observations of domesticated species in laboratories and zoos. One has only to consider how few animals reveal their age to appreciate the difficulties involved in answering the apparently simple question of how long they live in nature. In many fishes, a few kinds of clams, and an occasional species of other groups, growth is seasonal, so that annual zones of growth, much like tree rings, are produced in some part of the organism. Among game species, methods of determining relative age by indicators such as the amount of tooth wear or changes in bone structure have yielded valuable information. Bird bands and other identifying marks also make age estimation possible. But one of the consequences of the fact that animals move is that very little is known about the life-span of most species as they exist in nature.

Maximum and average longevity. The extreme claims of longevity that are occasionally made for one species or another have consistently been proven false when subjected to critical scrutiny. Although the maximum life-span that has been observed for a particular species cannot be considered absolute, since a limited number of individuals at best has been studied, this datum probably provides a fair approximation of the greatest age attainable for this kind of animal under favourable conditions. Animals in captivity, which provide most of the records of extreme age, are exposed to far fewer hazards than those in the wild. In the accompanying table of maximum longevity, particular species have been so selected as to encompass the known range of longevity of other members of the taxonomic group to which they belong.

Table 1: Maximum Longevity of Animals in Captivity*

animal	life-span in years	animal	life-span in years
Mammals		Amphibians	
Bat (*Eptesicus fuscus*)	2	European black salamander (*Salamandra atra*)	
Grizzly bear (*Ursus horribilis*)	31	Spotted salamander (*Ambystoma maculatum*)	25
Cat (*Felis catus*)	21	Frog (*Rana* species)	5–15
Chimpanzee (*Pan troglodytes*)	37	Fishes	
Dog (*Canis familiaris*)	34	Eel (*Anguilla rostrata*)	6
Elephant, Indian (*Elephas maximus*)	57	Goldfish (*Carassius auratus*)	25
Goat (*Capra hircus*)	18	Sturgeon (*Acipenser transmontanus*)	50
Golden hamster (*Mesocricetus auratus*)	1.8	Insects	
Horse (*Equus caballus*)	62	Ant (*Lasius* species)	15
Lion (*Panthera leo*)	29	Buprestid beetle (*Buprestis splendens*)	30
Mouse (*Mus musculus*)	3	Fruit fly (*Drosophila melanogaster*)	0.1
Ox (*Bos taurus*)	30	**Arachnids**	
Squirrel, gray (*Sciurus carolinensis*)	15	Bird spider (*Avicularis avicularis*)	15
Wild boar (*Sus scrofa*)	27	Rocky Mountain wood tick (*Dermacentor andersoni*)	3-4
Birds		Crustaceans	
Blue jay (*Cyanocitta cristata*)	4	Crayfish (*Astacus fluviatilis*)	30
Canary (*Serinus canaria*)	24	Water flea (*Daphnia magna*)	0.2
Macaw (*Ara macao*)	64	Mollusks	
Nightingale (*Luscinia luscinia*)	3.8	Clams, various species	1–10
Pigeon (*Columba livia domestica*)	35	Snails, various species	1–30
Titmouse (*Parus major*)	9	Annelid worms	
Reptiles		Earthworm (*Lumbricus terrestris*)	10
Alligator (*Alligator mississipiensis*)	56	Medicinal leech (*Hirudo medicinalis*)	27
Garter snake (*Thamnophis sirtalis*)	6	Rotifers	
Box turtle (*Terrapene carolina*)	123	Various species	0.03–0.1
Giant tortoise (*Testudo elephantopus*)	177		
Water turtle (*Pseudemys scripta*)	7		

*Condensed and adapted from W.S. Spector (ed.), *Handbook of Biological Data*, 1956.

Environmental influences. Life-span usually is measured in units of time. Although this may seem eminently logical, certain difficulties may arise. In cold-blooded animals in general, the rate of metabolism that determines the various life processes varies with the temperatures to which they are exposed. If aging depends on the expenditure of a fixed amount of vital energy, an idea first proposed in 1908, life-span will vary tremendously depending on temperature or other external variables that influence life-span. There is considerable evidence attesting at least to the partial cogency of this argument. So long as a certain range is not exceeded, cold-blooded invertebrates do live longer at low than at high temperatures. Rats in the laboratory live longest on a somewhat restricted diet that does not permit maximum metabolic rate. Of perhaps even greater significance is the fact that many animals undergo dormant periods. Many small mammals hibernate; a number of arthropods have life cycles that include periods during which development is arrested. Under both conditions the metabolic rate becomes very low. It is questionable whether such periods should be included in computing the life-span of a particular organism. Comparisons between species, some of which have such inactive periods while others do not, are dangerous. It is possible that life-span could be measured more adequately by total metabolism; however, the data that are necessary for this purpose are almost entirely lacking.

Temperature effects

Length of life is controlled by a multitude of factors, which collectively may be termed environment, operating on a genetic system that determines how the individual will respond. It is impossible to list all the environmental factors that may lead to death. For analytical purposes it is, however, useful to make certain formal separations. Every animal is exposed to (1) a pattern of numerous events, each with a certain probability of killing the individual at any moment and, in the aggregate, causing a total probability of death or survival; (2) climatic and other changes in the habitat, modifying the frequency with which the various potentially fatal events occur; and (3) progressive systemic change, inasmuch as growth, reproduction, development, and senescence are characteristics intrinsic in the organism and capable of modifying the effects of various environmental factors.

Patterns of survival. Consider a group of similar animals of the same age. Although no two individuals can have precisely the same environment, let it be assumed that the environment of the group remains effectively constant. If the animals undergo no progressive physiological changes, the factors causing death will produce a death rate that will remain constant in time. Under these conditions, it will take the same amount of time for the population to become reduced to one-half its former number, no matter how many animals remain at the beginning of the period considered. The animals therefore survive according to the pattern of an accident curve. This is the sense in which many of the lower animals are immortal. Although they die, they do not age; how long they have already lived has no influence on their further life expectation.

Another group of animals may consist of individuals that differ markedly in their responses to the constant environment. They may be genetically different, or their previous development may have caused variations to arise. Those individuals that are most poorly suited to the new environment will die, leaving survivors that are better adapted. The same result can also be achieved in other ways. If the environment varies geographically, those individuals that happen to find areas in which existence can be maintained will survive, while the remainder will die. Or, as a result of their own properties, animals in a constant environment may acclimate in a variety of ways, thus adjusting to the existing conditions. The pattern of survival that results in each of these cases is one in which the death rate declines with time, as illustrated by the selection–acclimation curve.

In the absence of death from other causes, all members of a population may exist in their environment until the onset of senescence, which will cause a decline in the ability of individuals to survive. In a sense they can be considered to wear out as does a machine. Their survival is best described by individual differences among members of the population that determine the curvature of the survival line (wearing-out curve). The more the population varies, the less abrupt is the transition from total survival to total death.

Under the actual conditions of existence of animals the three types of survival (accident pattern, selection–acclimation pattern, wearing-out pattern) above all enter as components of the realized survival pattern. Thus in animals that are carefully maintained in the laboratory, survival is approximately that of the wearing-out pattern. Environmental accidents can be kept to a minimum under these conditions, and survival is almost complete during the major part of the life-span. In all known cases, however, the early stages of the life-span are characterized by a noticeable contribution of the selection–acclimation pattern. This must be interpreted as a result of developmental changes that accompany the early life of the individuals and of selective processes that operate on those organisms whose genetic constitutions are ill fitted for that environment.

Highest mortality in aquatic animals

In some of the larger mammals in nature, the existing evidence points to a similar survival pattern. In a variety of other animals, however, and including fishes and invertebrates, mortality in the young stages is so high that the selection–acclimation curve predominates. One estimate places the mortality of the Atlantic mackerel during its first 90 days of life as high as 99.9996 percent. Since some mackerel do live for several years, a mortality rate that decreases with age is indicated. Similar considerations probably apply to all those animals that have larval stages that serve as dispersal mechanisms.

When the postjuvenile portion of the life-span is considered by itself, a number of animals for which such information has been gathered — including primarily fishes and birds — have survivorship curves that are dominated by the accident pattern. In these species in nature, death from old age apparently is rare. Their chance of surviving to an advanced age is so small that it may be statistically negligible. In modem times, human predation is a large factor in the mortality of these species in many cases. Since deaths from fishing and hunting are largely independent of age, once an animal has reached a certain minimum size, such a factor only makes the survival curve steeper but does not change its shape. One consequence of such increased mortality is that fewer old and large individuals are noticed in a population.

More complex survival patterns, such as the hypothetical one illustrated, undoubtedly exist. They should be looked for in those species in which extensive reorganization of the animal is part of the normal life cycle. In effect, these animals change their environment radically, in some cases several times during a lifetime. The frog offers a familiar example. During its period of early development and until shortly after hatching, the animal is subject to major internal, and some external, change. As a tadpole it is adjusted to an aquatic, herbivorous life. The metamorphosis to the terrestrial, carnivorous adult form is accompanied by varied physiological stresses that must be expected to produce a temporary increase in mortality rate. In some insects the eggs, larvae, pupae, and adults are exposed lo and respond to quite different environments, and a survivorship pattern even more complex than that described by the composite curve may exist.

The same species will exhibit changed survival in different environments. In captivity an animal population may approach the wearing-out pattern; in its natural habitat survivorship may vary with age in a quite different way. Although one can assign a maximum potential life-span to an individual — while realizing that this maximum may not be attained — it is impossible to specify the survivorship pattern unless the environment is also specified. This is another way of saying that life-span is the joint property of the animal and the environment in which it lives. (P.W.F.)

PLANTS

Plants grow old as surely as do animals; however, a generally accepted definition of age in plants has not yet been realized. If the age of an individual plant is that time interval between the reproductive process that gave rise to the individual and the death of the individual, the age attained may be given readily for some kinds of plants but not for others. Table 2 lists maximum ages, both estimated and verified, for some seed plants.

Problem of defining age. An English oak that has 1,000 annual rings in the trunk is 1,000 years old. But age is less certain in the case of an arctic lupine that germinated from a seed that, containing the embryo, had been lying in a lemming's burrow in the arctic permafrost for 10,000 years.

Ageless embryonic tissue

The mushroom caps that appear overnight last for only a few days, but the network of fungus filaments in the soil (the mycelia) may be as old as 400 years. Because of important differences in structure, the life-span of higher plants cannot be compared with that of higher animals. Normally, embryonic cells (that is, cells capable of changing in form or becoming specialized) cease to exist very early in the life of an animal. In plants, however, embryonic tissue — the plant meristems — may contribute to growth and tissue formation for a much longer time, in

Table 2: Maximum Ages for Some Seed Plants

plant	maximum age in years		locale of verified specimen
	estimated	verified	
Conifers			
Common juniper (*Juniperus communis*)	2,000	544	Kola Peninsula, north-eastern Russia
Norway spruce (*Picea abies*)	1,200	350–400	Eichstatt, Bavaria
European larch (*Larix decidua*)	700	417	Riffel Alp, Switz.
Scotch pine (*Pinus sylvestris*)		584	
Swiss stone pine (*Pinus cetnbra*)	1,200	750	Riffel Alp, Switz.
White pine (*Pinus strobus*)		400–450	
Bristlecone pine (*Pinus aristata*)		4.900	Wheeler Peak, Humboldt National Forest, Nevada
Sierra redwood (*Sequoiadenflron giganteum*)	4,000	2,200–2,300	Northern California
Flowering plants			
Monocots			
Dragon tree (*Dracaena draco*)	200*		Tenerife, one of the Canary Islands
Solomon's-seal (*Polygonatum*)	17†		
Dicots			
Dwarf birch (*Betula nana*)		80	Eastern Greenland
European beech (*Fagus sylvatica*)	900	250	Montigny, Normandy, France
English oak (*Quercus robur*)	2,000	1,500	Hasbruch Forest, Lower Saxony
Bo tree (*Ficus religiosa*)	2,000–3,000‡		Buddh Gaya, India; Anuradhapura, Ceylon
Linden (*Tilia*)		815	Lithuania
English ivy (*Hedera helix*)	440		Ginac, near Montpellier, France

'Exaggerated estimates for this historic specimen reach 6,000 years. †Scars on root-stock counted. ‡According to Buddhist and Roman history.

some cases throughout the life of the plant. Thus the oldest known trees, bristlecone pines of California and Nevada, have one meristem (the cambium) that has been adding cells to the diameter of these trees for, in many cases, more than 4,000 years and another meristem (the apical) that has been adding cells to the length of these trees for the same period. These meristematic tissues are as old as the plant itself; they were formed in the embryo. 'The wood, bark, leaves and cones, however, live for only a few years. The wood of the trunk and roots, although dead, remains a part of the tree indefinitely, but the bark, leaves, and cones are continually in the process of dying and sloughing off.

Among the lower plants only a few mosses possess structures that enable an estimate of their age to be made. The haircap moss (*Polytrichum*) grows through its own stem tip each year, leaving a ring of scales that marks the annual growth. Three to five years' growth in this moss is common, but life-spans of ten years have been recorded. The lower portions of such a moss are dead, though intact. Peat moss (*Sphagnum*) forms extensive growths that fill acid bogs with a peaty turf consisting of the dead lower portions of mosses whose living tops continue growing. Mosses that become encrusted with lime (calcium carbonate) and form "tufa" beds several metres thick also have living tips and dead lower portions. On the basis of their observed annual growth, some tufa mosses are estimated to have been growing for as long as 2,800 years.

No reliable method for determining the age of ferns exists, but on the basis of size attained and growth rate, some tree ferns are thought to be several decades old. Some club mosses, or lycopsids, have a "storied" growth pattern similar to that of the haircap moss. Under favourable conditions some specimens live five to seven years.

The woody seed plants, such as conifers and broadleaf trees, are the most amenable to determination of age. In temperate regions, where each year's growth is brought to an end by cold or dryness, every growth period is limited by an annual ring—a new layer of wood added to the diameter of the tree. These rings may be counted on the cut ends of a tree that has been felled or, using a special instrument, a cylinder of wood can be cut out and the growth rings counted and studied. In the far north growth rings are so close together that they are difficult to count. In the moist tropics growth is more or less continuous, so that clearly defined rings are difficult to find.

Errors in estimating **age**

Often the age of a tree is estimated on the basis of its diameter, especially when the average annual increase in diameter is known. The source of greatest error in this method is the not infrequent fusing of the trunks of more than one tree, as, for example, occurred in a Montezuma cypress in Santa Maria del Tule, a little Mexican village near Oaxaca. This tree, described by the Spanish explorer Hernan Cortés in the early 1500s, was earlier estimated on the basis of its great thickness to be 6,000 years old; later studies, however, proved it to be actually three trees grown together. Estimates of the age of some English yews have been as high as 3,000 years, but these figures, too, have turned out to be based on the fusion of close-growing trunks, none of which is more than 250 years old. Increment borings of single bristlecone pines have shown several specimens in Inyo National Forest, California, to be 4,600 years old; one, about 4,900 years old, was felled in the summer of 1964 in Nevada's Humboldt National Forest.

Growing season of seed plants. *Annuals.* Plants, usually herbaceous, that live for only one growing season and produce flowers and seeds in that time are called annuals. They may be represented by such plants as corn and marigolds, which spend a period of a few weeks to a few months rapidly accumulating food materials. As a result of hormonal changes—brought about in many plants by changes in environmental factors such as day length and temperature—leaf-producing tissues change abruptly to flower-producing ones. The formation of flowers, fruits, and seeds rapidly depletes food reserves and the vegetative portion of the plant usually dies. Although the exhaustion of food reserves often accompanies death of the plant, it is not necessarily the cause of death.

Biennials. These plants, too, are usually herbaceous. They live for two growing seasons. During the first season, food is accumulated, usually in a thickened root (beets, carrots); flowering occurs in the second season. As in annuals, flowering exhausts the food reserves, and the plants die after the seeds mature.

Perennials. These plants have a life-span of several to many years. Some are herbaceous (iris, delphinium), others are shrubs or trees. The perennials differ from the above-mentioned groups in that the storage structures are either permanent or are renewed each year. Perennials require from one to many years growth before flowering. The preflowering (juvenile) period is usually shorter in trees and shrubs with shorter life-spans than in those with longer life-spans. The long-lived beech tree (*Fagus sylvatica*), for example, passes 30–40 years in the juvenile stage, during which time there is rapid growth but no flowering.

Some plants—cotton and tomatoes, for example—are perennials in their native tropical regions but are capable of blooming and producing fruits, seeds, or other useful parts in their first year. Such plants are often grown as annuals in the temperate zones.

Longevity of seeds. Although there is great variety in the longevity of seeds, the dormant embryo plant contained within the seed will lose its viability (ability to grow) if germination fails to occur within a certain time. Reports of the sprouting of wheat taken from Egyptian tombs are unfounded, but some seeds do retain their viability a long time. Indian lotus seeds (actually fruits) have the longest known retention of viability (see above). On the other hand, seeds of some willows lose their ability to germinate within a week after they have reached maturity.

The oldest living seeds

The loss of viability of seeds in storage, although hastened or retarded by environmental factors, is the result of changes that take place within the seed itself. The changes that have been investigated are: exhaustion of food supply; gradual denaturing or loss of vital structure by protoplasmic proteins; breakdown of enzymes; accumulation of toxins resulting from the metabolism of the seed. Some self-produced toxins may cause mutations that hamper seed germination. Since seeds of different species vary greatly in structure, physiology, and life history, no single set of age factors can apply to all seeds.

(L.K.)

BIBLIOGRAPHY

Human beings: WARREN ANDREWS, *Anatomy of Aging in Man and Animals* (1970); SEYMOUR BAKERMAN (ed.), *Aging Life Processes* (1969); AUSTIN *M.* BRUES and G.A. SACHER (eds.), *Aging and Levels of Biological Organization* (1965); ALEXANDER COMFORT, *The Biology of Senescence* (1956); L.I. DUBLIN, A.J. LOTKA, and M. SPIEGELMAN, *Length of Life*, rev. ed. (1949): E.S. DEEVEY, JR., "The Probability of Death," *Scient. Am.*, 182:58–60 (1950); NATHAN W. SHOCK (ed.), *Biological Aspects of Aging*, 2nd ed. (1962).

Subhuman animals: W.C. ALLEE *et al.*, *Principles of Animal Ecology* (1949); F.S. BODENHEIMER, *Problems of Animal Ecology* (1938); CIBA FOUNDATION, *The Lifesparz of Animals* (1959); ALEX COMFORT, "The Life Span of Animals," *Scient. Am.*, 205:108–119 (1961); E.S. DEEVEY, JR., "Life Tables for Natural Populations of Animals." Q. *Rev. Biol.*, 22:283–314 (1947); D.L. LACK; *The Natural Regulation of Animal Numbers* (1954).

Plants: A.E. PORSILD, C.R. HARINGTON, and G.A. MULLIGAN, "Lupinus arcticus Wats. Grown from Seeds of Pleistocene Age," *Science*, 158:113–114 (1967); W.J. ROBBINS, "Physiological Aspects of Aging in Plants," in W.C. STEERE (ed.), ***Fifty*** *Years of Botany* (1958); B.L. STREHLER *et al.* (eds.), *Biology of Aging* (1960); FRITZ W. WENT. "Plant Growth and Plant Hormones," in W.H. JOHNSON and W.C. STEERE (eds.), *This Is Life* (1962).

(L.K./P.W.F.)

Life-Support Systems

A life-support system is any mechanical device that enables man to live, and usually work, in an environment in which he could not otherwise survive. A key factor in this definition, though unstated, is time. Just as a man might work underwater for a minute or two merely by holding his breath, he might possibly live in a confined space without supplementary support for quite some time. Eventually, however, he must exchange gases, nutrients, water, waste, and energy with the environment in order to survive. A life-support system provides an artificial mechanism for making these exchanges.

Life-support priorities

In a small, closed, isolated space, such as a spacecraft or submarine, the maintenance of a suitable atmosphere for breathing is generally the most immediate problem, followed by concern for water and food and a method of waste disposal. Temperature, comfort, the danger of contaminants, and psychological factors must also be considered.

Man is best adapted to an environment temperature in the vicinity of 20" C (68" F) and air pressure equivalent to that on the earth at sea level (one atmosphere); he does best breathing a gas containing about 20 percent oxygen; and he has certain nutritional requirements. Two matters of degree must be considered here: the first is time, as mentioned above, and the second is comfort. Just as it is obvious that a man can live longer without food than without breathing, it is obvious that he can survive conditions that, though not fatal, make it extremely difficult for him to do creative work. Life-support systems, therefore, aim to provide conditions of comfort, efficiency, and safety in a working environment.

This article is outlined as follows:

I. Basic principles of life support in unfavourable environments

NEEDS OF MAN

The life-support engineer is concerned with providing the best environment for man to work in and with devising ways in which he can exchange materials and energy with his surroundings.

Oxygen. Man's most vital point of contact with the environment is his lung; his most urgent needs relate to respiration. Any satisfactory life-support system must supply oxygen at pressures sufficient to provide adequate amounts in the blood. It is normally sufficient to provide a breathing mixture of 20 percent oxygen at sea-level pressure.

A normal man consumes about 250 millilitres, or about one-half pint, of oxygen per minute when resting; he consumes 400 to 500 millilitres per minute when moderately active, perhaps one litre per minute when moving around briskly, and up to three litres per minute when exercising at peak capacity. As important as the dangers of insufficient oxygen are the risks of oxygen poisoning due to high oxygen pressure. Man can tolerate oxygen at two or three times normal atmospheric pressure for short periods. For survival over many hours or days, however, oxygen pressure should be below one-third normal atmospheric pressure.

Carbon dioxide. Man is accustomed to breathing gas that is essentially free of carbon dioxide, although small amounts may be inhaled without ill effects and large amounts can be tolerated for brief periods. Normally a man breathes out carbon dioxide at an average rate about 0.82 times his oxygen consumption. This means that a resting man would produce 200 millilitres of carbon dioxide per minute and a man moving around briskly about 800 millilitres per minute.

Pressure. Although accustomed to a so-called normal atmospheric pressure, man can tolerate a wide range of pressures, the primary limiting factor being the partial pressure of oxygen; that is, the total pressure of the gas multiplied by the proportion of oxygen in the gas. In the normal atmosphere, since the air is about 20 percent oxygen, the partial pressure of oxygen is 0.2 atmosphere. The minimum pressure for normal body function is determined by that point at which a normal blood oxygen concentration can be maintained by breathing pure oxygen. Pressures should not be reduced below this level. The limit to man's tolerance of high pressures is not yet known, but men have tolerated pressures up to 53 times the normal atmospheric level, and animals have survived pressures several times this great.

Limits of tolerance to high pressure

Temperature. Most individuals know the temperatures at which they are most comfortable; this of course is a function of individual preference, dress, level of activity, etc. In the same way, relative humidity affects comfort at a particular temperature. Most people are comfortable between a range of 60° and 75° F (16°–24° C) and a relative humidity range between 35 and 70 percent.

Energy. Virtually the same number of calories of heat must be removed from the body each day as are consumed as food. On a short-term basis, heat production can be related directly to oxygen consumption at a level of 4,800 calories per litre of oxygen used. The problem is to prevent loss of heat (and to increase heat production); but in a closed space from which heat can escape only with difficulty, some means must be provided to remove this heat and to maintain a suitable temperature.

Food. A balanced diet of about 2,000 to 4,000 kilocalories per day is needed by a moderately active man. Experience in spacecraft and other confined spaces has shown that 2,000 to 2,250 kilocalories daily is usually suitable. Virtually all experiences with men confined in life-support systems have involved complaints about food. Subjective choice of food to be used is apparently highly desirable.

Water. Water balance is critical to good health and good performance. Two to three litres of water are lost each day by normal man and must be replaced. Water is needed to keep the kidneys functioning to remove waste from the body and as perspiration to help control the body temperature. Normally a man takes in about 1,500 millilitres per day as liquid water and another 600 millilitres as part of his food and converts 300 millilitres of the food eaten into metabolic water. This water is normally removed from the body in the following quantities: 1,400 millilitres as urine, 100 millilitres as fecal water, and 900 millilitres as insensible water loss—that is, water loss through the respiratory tract and the skin but not as visible perspiration. Both sides of this equation will be increased accordingly if the man is in a situation that causes him to perspire.

Waste. On a normal diet a man produces about 150 grams per day of solid waste as feces and another 1,000 millilitres of intestinal gas. These values are increased on a vegetable diet. On an all-meat diet, the solid waste can be reduced to less than one-third of the above value.

Circadian (day–night cycle) stimuli. In addition to his ordinary needs as mentioned above, man also requires periodic stimuli to help maintain his normal 24-hour physiological periodicity. A great many of the hormonal, metabolic, and biochemical functions of the body vary throughout the 24-hour solar day, and it is felt by some authorities that metabolic disorders will result without some kind of "synchronizer" in situations in which man is removed from the normal day–night cycle.

Ionization of a closed atmosphere

Ions. Although ions are difficult to measure, and the effects of ions in the atmosphere are subjective and similarly difficult to quantify, the ionization of a closed atmosphere seems to influence the well-being of those living in and breathing it. Tests have shown that positive ions are associated with asthma, head colds, headache, and fatigue and that negative ions are associated with a general feeling of well-being. Since ions are generated by radiation from many electronic devices in spaceships and submarines, a well-planned environmental control system considers their presence.

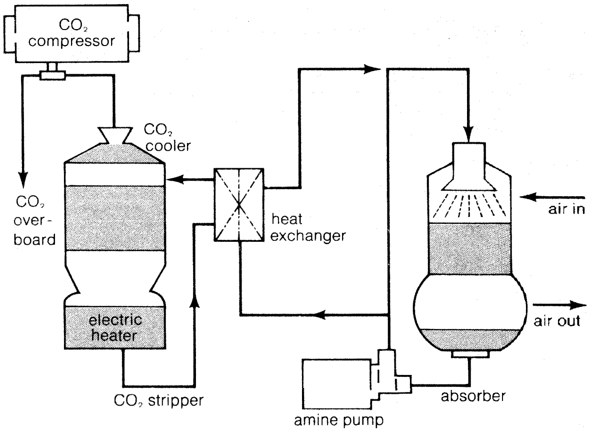
Time. Time is important in determining the characteristics of any life-support system. If only minutes are involved, provisions for oxygen supply and carbon dioxide removal may be all that will be necessary. After many minutes, other aspects of atmosphere composition, such as oxygen toxicity, gas density, temperature, and contaminants take on importance. After several hours food and water are needed, waste must be removed, and temperature and humidity need further control.

BASIC METHODS OF ATMOSPHERE CONTROL

Oxygen supply. For most uses the most expedient source of oxygen is as a compressed gas. If the rate of use is high, however, a low-temperature liquid oxygen supply may be employed, and in some systems now available oxygen can be stored for many weeks without appreciable loss. In applications in which light weight and long storage periods are criteria, a solid chemical source of oxygen may be best. There are two general types of such a source. A chlorate candle (a mixture of sodium chlorate and a small quantity of iron powder), if ignited, will burn the iron, causing decomposition of the chlorate and releasing oxygen. Second, alkaline superoxides, either sodium dioxide or potassium dioxide or both, will react with water or carbon dioxide or both to give off oxygen. There are several chemical methods by which oxygen can be removed from water or carbon dioxide; for example, by means of an electrochemical reaction. These are discussed below, under *Aviation and* space systems.

The molecular sieve

Carbon dioxide removal. The most expedient method for removing carbon dioxide from a closed atmosphere is with a so-called chemical scrubber, of which the most common are soda lime, Baralyme (a mixture of barium and calcium hydroxides), and lithium hydroxide. All three of these compounds are useful, and a choice will depend on factors such as weight, cost, and heat and water generation in the reactions. Lithium hydroxide has a disadvantage in that it produces an irritating dust. All three compounds can be used in refillable canisters through which the atmosphere is passed by means of small blowers; they cannot be regenerated and reused. One type of carbon dioxide scrubber that can be reused is a molecular sieve that absorbs carbon dioxide and from which carbon dioxide can later be removed by heating. Molecular sieves can also be used for other types of contaminant removal, including water; in fact, a limitation to their use as a scrubber is their tendency to absorb water in preference to carbon dioxide. The gas must be dried before carbon dioxide can be effectively removed by a molecular sieve. Another regenerative type of carbon dioxide removal system is an amine scrubber (such as is used on submarines; Figure 1), in which a carbon dioxide absorbing agent is heated to separate the carbon dioxide and reused in the same machine. As with oxygen, there are electrochemical methods of converting carbon dioxide to oxygen, but most of these systems require that the carbon dioxide be concentrated first.



$$2\,MEA + H_2O + CO_2 \rightleftharpoons (MEA)_2 \cdot H_2CO_3$$

Figure 1: Scrubber for removing carbon dioxide from air by absorbing the CO_2 in an amine (see text).

Reprinted with permission of the Macmillan Company from *Man's Dependence on the Earthly Atmosphere* by Karl E. Schaefer. Copyright © 1362 by American Institute of Biological Sciences

Contaminant removal and odour control. The most common means of removing organic molecules from a gaseous environment is to adsorb them on the surface of activated charcoal. Charcoal can also remove odours from liquids, as is necessary, for example, in recycling urine for water conservation. The effectiveness of activated charcoal depends on its immense surface area—so large that one pound of it may have as much as 9,000,000 square feet of surface. Trace organic compounds such as hydrogen, methane, and carbon monoxide can also be removed from an atmosphere with a catalytic burner in which a heated gas is passed over a catalyst such as platinum black or hopcalite.

Temperature and humidity control. Temperature and humidity are controlled in life-support systems by application of principles established in conventional engineering practice. Heat, obtained as a by-product from an engine or other source, is removed by radiation, conduction, convection, or evaporation. Humidity is usually supplied in sufficient quantity by the occupants of a system, it can be removed by chemical absorption or by condensation on cooling coils.

II. Aviation and space systems

THE ENVIRONMENT: ALTITUDE AND SPACE

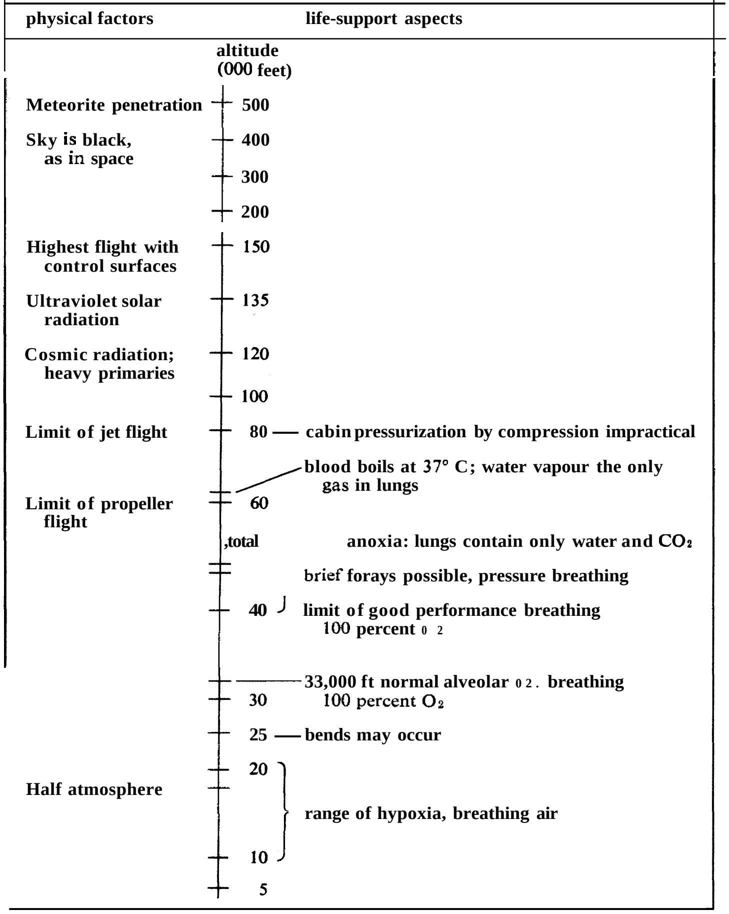
The lethal altitude

If a man not acclimatized to high altitude ascends in an unpressurized airplane, he will exhibit various symptoms that indicate the effects of altitude. Evidences of lack of oxygen begin at about 10,000 feet (3,000 metres). At 18,000 feet (about 5,500 metres) the pressure is half that at sea level, and man may feel definite symptoms; he will probably lose consciousness between 25,000 and 30,000 feet (7,500–9,000 metres). If pure oxygen is breathed, a man functions normally at up to 34,000 feet (10,000 metres); here his blood has the same oxygen pressure as it would if he were breathing air at sea level. A man may continue operating to 40,000 feet (12,000 metres) without serious problems; above that altitude he cannot function well without supplementary pressure. At 52,000 feet (15,800 metres) he will have reached the point at which there is no oxygen in his lungs and will lose consciousness in about 15 seconds. At 63,000 feet (19,200 metres) the atmospheric pressure is the same as the vapour pressure of water at body temperature; thus, there is no gas other than water in a man's lungs; at this pressure the blood boils at body temperature.

Other aspects of altitude and the limits of various phenomena are shown in the table. Though other items shown have physiological significance, pressure is the most easily controlled and is thus the most important factor in life-support systems.

Mountains. Man's earliest ventures to altitudes high enough to tax his adaptation were of course made on foot, climbing mountains. La Paz, Bolivia, is nearly 12,000 feet (3,600 metres) high, and in other parts of the

Changes in Physical and Life-Support Factors at Increasing Altitudes


physical factors
life-support aspects
altitude (000 feet)
Meteorite penetration
500
Sky is black, as in space
400
300
200
Highest flight with control surfaces
150
Ultraviolet solar radiation
135
Cosmic radiation; heavy primaries
120
100
Limit of jet flight
80
cabin pressurization by compression impractical
blood boils at 37° C; water vapour the only gas in lungs
Limit of propeller flight
60
,total
anoxia: lungs contain only water and CO_2
brief forays possible, pressure breathing
40
limit of good performance breathing 100 percent O_2
33,000 ft normal alveolar O_2. breathing 100 percent O_2
30
25
bends may occur
20
Half atmosphere
range of hypoxia, breathing air
10
5

Andes lifetime adaptation makes it possible for people to live at the 16,000-foot (4,800-metre) level and work at 18,000 feet (5,400 metres), without life-support equipment. In attempts to climb Himalayan peaks, courageous and well-acclimated climbers have reached altitudes as high as 28,000 feet (8,534 metres) without supplementary oxygen. In the first successful assault of the 29,028-foot (8,839-metre) Mt. Everest, however, life-support equipment was the deciding factor. In addition to boots designed especially for this mission, the successful climb of 1953 required two types of oxygen-breathing apparatus. An open-circuit type metered oxygen into the stream of air inhaled by the climber. This device could be set to deliver from one litre of pure oxygen per minute when requirements were low, as in sleeping, up to five litres per minute under difficult climbing conditions. This simple apparatus weighed only 28 pounds (12.7 kilograms). The second type, a closed-circuit apparatus, provided approximately the same duration of breathing and weighed 35 pounds (15.9 kilograms) and consisted of a breathing bag and valves arranged so that the gas exhaled by the climbers was recirculated through a soda lime canister that absorbed the carbon dioxide. The entire apparatus was filled with pure oxygen, so the user breathed oxygen undiluted by ambient air. This closed system was more likely to freeze up than the open system; it was heavier, and it generated heat. Furthermore, the user of the closed system became more dependent on the 100 percent oxygen supply; if it was lost for some reason he was likely to lose consciousness at high altitudes. Although both systems were used by members of the Everest expedition, the team that reached the top used the open-circuit type.

The Everest experience

Balloons. A 19th-century French physiologist, engineer, and physician, Paul Bert, conceptualized the basic principle of using supplementary oxygen to supply balloonists and mountain climbers who had ascended beyond the levels at which the oxygen in air is sufficient. Two of Bert's colleagues took a supply of about 150 litres of 70 percent oxygen on a balloon flight in 1875; but they failed to use it soon enough, and only one of the two survived. On this flight the oxygen was stored at ambient pressure in "goldbeater's bags" (made from cow's Intestines) and was to be inhaled by mouth tube through a humidifier containing an aromatic liquid whose purpose was both to humidify the gas and to counteract the odour of the bags. The maximum height attained by the balloon was over 28,000 feet (8,534 metres). These explorers had used oxygen on previous flights with success. Bert also designed a tank and regulator system with a capacity of 330 litres whereby mountain climbers could breathe oxygen near the peak on their ascent.

On November 11, 1935, on a mission sponsored by the National Geographic Society and the U.S. Army Air Corps, two men rode a hydrogen-filled balloon to a height of 72,395 feet (22,067 metres); this feat was particularly significant since the two had parachuted to safety 16 months earlier from a similar vehicle that had ripped apart at 60,000 feet (18,000 metres). Their magnesium alloy gondola was supplied with 45 percent oxygen and 55 percent nitrogen from a 25-litre liquefied gas container. Helium was used to drive the liquid air through the coils of an evaporator, and a battery-powered fan circulated cabin air over this coil for the removal of moisture and subsequently through a sodium hydroxide scrubber for the removal of carbon dioxide. To aid in stabilizing the interior temperature, the gondola was painted white on top and black below, to reflect and absorb heat.

Cosmic ray studies

Groups of mice and guinea pigs flown by balloon to altitudes above 100,000 feet (30,000 metres) for cosmic ray studies in the early 1950s were maintained in thin aluminum spheres with pressurized atmospheres. Oxygen was supplied from a high-pressure source. When only a few animals were involved, carbon dioxide was absorbed with soda lime in containers; with greater loads, forced air was used. Temperature was controlled by the evaporation of water into the near-vacuum at that altitude, and the cool surface of the water container condensed moisture out of the atmosphere of the capsule. Sponges were used to absorb this moisture.

A record for balloon flights was set on May 4, 1961, when a helium-filled polyethylene balloon ascended to about 113,740 feet (34,747 metres) over the Gulf of Mexico. Though the gondola was nothing more than a framework covered with venetian blinds, the balloonists were protected by special suits pressured by oxygen from a liquefied gas source and provided with electric heaters to keep the faceplates from fogging. The balloonists could reverse the blinds on their gondola from silver to black to help control temperature.

SYSTEMS USED IN HIGH-ALTITUDE AIRCRAFT

Cabin pressurization. For aircraft flying at altitudes high enough to cause illness among unprotected passengers, the most obvious and effective life-support system is a normal pressure environment, accomplished by pressurizing the occupied portions of the aircraft. The first commercial aircraft to provide cabin pressurization was the Boeing 307B, which began to carry passengers in 1940. Though some modern propeller-driven aircraft are not pressurized, pressurization is a necessity for jets that fly at high altitudes.

Aircraft cabin-pressurization systems are active systems; pressure is continuously provided to the cabin, and some is continuously leaked overboard. Because of this continuous leakage, no further provisions for carbon dioxide removal equipment are required. In reciprocating-engine aircraft, pressure is provided by an engine-driven turbine, while in jet aircraft the pressurization system is usually provided by gas drawn from the compressor section of the jet engine or from an engine-driven blower. Because of the heat of compression, the air that is drawn off the engine is very hot; even though some heat is required in the cockpit or cabin, it is generally necessary to cool the air, usually with a turbine refrigeration system, making this method of pressurization impractical at altitudes over 80,000 feet (24,000 metres).

Commercial airliners are pressurized to the equivalent of 4,000 to 8,000 feet (1,200–2,400 metres) during the climb, which may take the plane to an altitude of well over 40,000 feet (12,000 metres). All jets that operate above 20,000 feet (6,000 metres) are required to provide supplemental oxygen, through masks, in the event of cabin pressure failure. A small plastic bag attached to the mask receives piped oxygen when the passenger is not inhaling, and he inhales the next breath from this bag. Oxygen for such masks is usually supplied from pressurized cylinders, although systems now being developed hold the oxygen combined in chemical form, such as a chlorate candle. Such systems do not develop leaks, may be stored indefinitely, weigh less, and are less of a fire hazard than pressurized cylinders.

Oxygen systems. During World War I, when 20,000 feet (6,000 metres) was a significant aircraft altitude, a tube of oxygen continuously flowing and held in the lips of the pilot was generally quite adequate. By the beginning of World War II this crude system had been refined so that a continuous flow of oxygen was breathed by mask from a small balloon, which was allowed to fill between breaths. Soon this system was improved by the introduction of a demand regulator, which ensured that no gas flowed unless the pilot was inhaling. A refinement of the demand system is the pressure-demand system, which delivers oxygen to the pilot's mask at a pressure slightly higher than that of the cockpit. Increasing pressure in the mask increases pressure in the lungs and results in a higher pressure of oxygen in the blood. This system has undesirable side effects, however; the increased pressure in the mask is transmitted to the walls of the blood vessels in the chest, the diaphragm, the chest wall, and certain areas in the head and can be tolerated for only a limited time. The apparatus is, thus, generally used only in the event of cabin pressure failure.

World War II oxygen masks

The next step in the process of providing a higher oxygen pressure to the lungs of a pilot at a high altitude is a partial-pressure suit, a tightly fitting garment that covers the entire body and to which are attached bladders filled with air from the pressurization system of the aircraft. The pilot's head is enclosed in a helmet; pressurized oxygen is supplied. Because the bladders apply a counterpressure to the body, it is possible to supply a much higher pressure to the breathing gas than with the pressure-demand system alone. A pressure-demand system provides normal function at an altitude of 43,000 feet (13,000 metres); a partial-pressure suit, on the other hand, permits normal function at an altitude of 60,000 feet (18,200 metres).

The most recent step in the development of personal equipment for delivery of oxygen to a man exposed to reduced pressure is the full-pressure suit. Despite many refinements, the basic design is similar to that used by spacemen and lunar astronauts. The body and head are entirely enclosed in a sealed, pressurized envelope through which gas is supplied at the same pressure as the rest of the body. Full-pressure suits keep the body at a pressure equivalent to 33,000 feet (10,000 metres) above sea level; if supplied with pure oxygen, these suits effectively provide a normal oxygen level to the user's body. Since the full-pressure suit prevents effective heat exchange between the user's body and the environment, a cooling system of some sort must be provided. Pressure suits designed for use in aircraft generally have a perforated inner garment through which air can be circulated at an adjustable temperature and flow rate.

Systems for supplying oxygen have evolved along with breathing equipment. Originally, oxygen was supplied by low-pressure cylinders; in the early jets, pressures were increased, but later jets began to use low-temperature (cryogenic) systems supplied with liquid oxygen. This development was accelerated by the invention of in-flight refuelling, which meant that a small fighter plane might stay aloft for hours, its cruising time limited by its oxygen rather than by its fuel supply. Work is now in progress on an electrochemical method to separate oxygen from the air, trading energy for consumable components and providing an essentially unlimited supply of oxygen.

Research aircraft. Rocket-powered research aircraft have presented special problems in the design of life-support systems. With no engine and very little surrounding atmosphere from which to draw compressed air for cabin pressurization, such aircraft have used hermetically sealed cabins filled with pure nitrogen and equipped with a high-capacity refrigeration system. The pilot does not breathe this atmosphere but has his own breathing atmosphere supplied in his suit.

In the American X-I plane, the aircraft that first surpassed the speed of sound (1947), the life-support system was primitive. Pilots used World War II leather flying helmets over which they fitted the top of a football helmet for impact protection; they breathed through a standard demand-type oxygen mask. Later flights in later planes were flown in partial-pressure suits, which provided protection at altitudes above 50,000 feet (15,000 metres) in case of loss of cabin pressurization; on occasion this protection was needed and used. Because of the extreme cold outside the cockpit, the pilots placed a dehydrating canister on the expiratory outlets of their suits so that moisture from their breath would not escape into the cockpit and frost the windows.

The X-15 life-support system

The X-15, which exceeded altitudes of 60 miles (96.6 kilometres) and speeds six times that of sound, had a nitrogen environment cooled by a powerful refrigeration system. The suit worn by the pilot had an undergarment ventilated by oxygen; a pressure garment, and an external aluminized protective layer.

In view of the above, the question of how a supersonic transport can fly safely at altitudes of 60,000 to 70,000 feet (18,000–21.000 metres) must be dealt with. To begin with, cabin structure must be less subject to failure; this is accomplished in part by the use of armour plate either on the engine housings or on the fuselage in the areas where a lost engine turbine blade might penetrate. Pressure is supplied to the cabin by engine-driven compressors that can be turned up to a much higher speed in the event of cabin pressure loss. supplying air to the cabin faster than it can be lost through a broken window or hole.

SYSTEMS USED IN EARLY SPACE FLIGHTS WITH ANIMALS

During the years preceding manned space flights a number of significant flights were first made with animals aboard. The earliest of these were short rocket or balloon flights made to study weightlessness and cosmic rays; these were so short that life-support requirements were minimal, and the sealed capsule often was large enough to sustain the animals without special equipment.

Before the first manned space flights, Soviet investigators sent a number of dogs and other small animals on suborbital rocket missions from which they were usually recovered by parachute. Two basic life-support approaches were employed, one a sealed, insulated cabin and the other an unsealed type in which the animals were protected by space suits. The dogs used in Soviet space work were always extensively trained and conditioned to the noise, vibration, acceleration; and confinement of the flight, in order to reduce the number of variables that might affect biomedical measurements. The "sealed" cabin actually contained a calibrated leak, and a mixture of 40 percent oxygen and 60 percent nitrogen was allowed to bleed in. As the oxygen flowed in, it passed through a tube that circulated the cabin gas through carbon dioxide absorbers at a rate of about 120 litres per minute. Water vapour was removed by silica gel. Other dogs, which rode in unpressurized capsules, were trained and fitted with rubberized suits equipped with Plexiglas helmets. Oxygen flowed through the suit at six litres per minute, and pressure was maintained inside; no other steps were necessary to control carbon dioxide and water vapour. A Soviet dog rode in the second vehicle to be launched into earth orbit. On November 3, 1957, the 1,200-pound (540-kilogram) Sputnik 2 carried the 13.3-pound (6-kilogram) dog around the earth for nearly a week before she died from lack of oxygen. The spacecraft life-support system was essentially the same as that of the Vostok series that later carried the first men into orbit. Conditions in the capsule were maintained close to those found on earth,

The Soviet space dogs

with respect to air pressure, temperature, and oxygen and carbon dioxide content. Oxygen was supplied by release from a dry chemical, in conjunction with the absorption of carbon dioxide. The dog was fed a gelatinous suspension of ground meat and bread crumbs by means of an automatic feeder. Waste was collected in a rubber bag fitted to the pelvic region of the dog. Liquids were absorbed and deodorized by activated charcoal. Later (1960), two other dogs and 40 mice were recovered after 24 hours in orbit in Sputnik 5. They were the first living things to be recovered from orbit.

In 1959 two monkeys, a rhesus and a squirrel monkey, were launched to an altitude of 300 miles (480 kilometres) and a distance of 1,500 miles (2,400 kilometres) by a Jupiter rocket and were successfully recovered. Sealed in separate capsules, each was placed on a restraint couch and fitted with a helmet. The life-support system of one animal continuously received oxygen through a needle valve from a pressurized cylinder. Excess pressure was vented overboard. Carbon dioxide was absorbed by lithium hydroxide distributed in bags of small-pore plastic fabric, and water and ammonia were absorbed by similar bags of charcoal and a fused silica water absorber. Diapers helped control the amount of water and ammonia that reached the atmosphere. An electric heater and liquid nitrogen cooler provided temperature control. The unit of the other animal had an oxygen system that added oxygen to make up for that consumed by the animal, so as to maintain a constant cabin pressure. The atmosphere was nearly normal in pressure and gas composition.

SYSTEMS USED IN U.S. AND SOVIET MANNED SPACE FLIGHTS

Vostok. Manned space flight began on April 12, 1961, with a one-orbit flight in the Russian spacecraft Vostok 1. Pressure and oxygen concentration in the Vostok spacecraft were maintained at approximately those of sea level on earth. Carbon dioxide levels were kept at less than 0.5 percent throughout most of the flight. Temperatures varied from about 12° to 25° C (54"—77" F); relative humidities ranged comfortably between about 45 and 65 percent. As part of the experimental plan, conditions were maintained as close to normal as possible so that observed changes could be attributed to space flight and not to such things as high oxygen levels.

In the Vostok spacecraft a ventilating fan forced cabin air to a gas analyzer, a regenerator unit, and a dehumidifier. The gas analyzer provided readings of oxygen and carbon dioxide partial pressure. which along with cabin pressure, temperature, and humidity were displayed to the pilot and relayed to ground stations. When the cabin pressure dropped below the desired level, a regulator increased the flow through the regenerator, which absorbed carbon dioxide and generated oxygen by means of a chemical system probably based on alkali superoxides. Flow through the dehumidifier was controlled by a humidistat, which maintained humidity at desired levels much as a thermostat controls temperature. A separate temperature control system maintained cabin temperature by means of a regulator and pipes that conveyed excess heat to a radiator, where it was radiated into space.

Mercury. The first manned orbital flight in the United States space program, a three-orbit flight, was made on February 20, 1962, in a Mercury-Atlas 6 spacecraft. The life-support system for the Mercury spacecraft (Figure 2) was dual, with one subsystem to control the environment within the pilot's suit and another for the spacecraft cabin. A secondary but nonetheless important function of the cabin atmosphere control was to provide a suitable environment for operation of electronic and electrical apparatus. In the vacuum of space the insulating effect of the earth's air is missing, and artificial pressure is needed to prevent short-circuiting of high-voltage apparatus. Oxygen was circulated through the suit, entering near the waist, passing to the hands and feet, returning and exiting by way of the helmet. The system allowed the pilot to select a temperature between 50" and 90" F (10"–32" C) even if cabin temperature climbed as high as 180" F (82° C) during re-entry. As the oxygen left the helmet it entered a debris trap, then passed through a charcoal bed for odour removal and through a lithium hydroxide-carbon dioxide scrubber. A filter protected the pilot from lithium hydroxide dust. The gas was then cooled and dehumidified before passing back to the pilot. Heat was removed from the system by evaporating water into the vacuum of space. When necessary to compensate for leakage or consumption by the pilot, oxygen was bled into the system from a pair of high-pressure supply bottles. An additional cooling system worked on the cabin atmosphere independently of the suit. The pilot had instrument panel readouts of cabin pressure, temperature, oxygen partial pressure, and relative humidity.

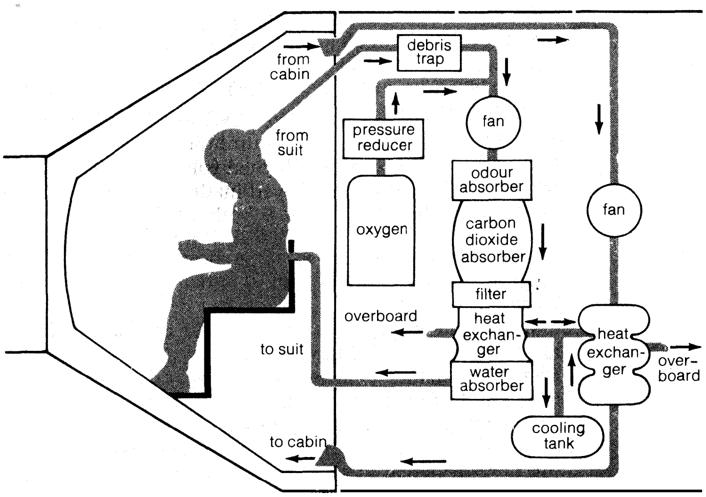


Figure 2: The Project Mercury environmental-control system.
By courtesy of the National Aeronautics and Space Administration

Defects of the Mercury system

Although the Mercury system performed its function adequately, it left something to be desired in terms of temperature control and particularly of control of water vapour and perspiration. All the astronauts who flew in this series returned from their flights dehydrated, some markedly so.

Gemini. The life-support and environmental-control systems for the Gemini spacecraft included several improvements over the Mercury; among these was a capability to sustain a longer (14-day) mission duration, an extravehicular-activity capability, and the use of fuel cells for crew water supplies. These cells, which produce water as a by-product while generating electricity, are used as a water source on spaceships. Like Mercury, Gemini used a pure oxygen atmosphere. The main life-support system was divided between the pressurized cabin and a removable "adapter" section; the cabin subsystem could operate without the adapter, as on re-entry. There was also a separate suit subsystem, and the crew could choose between using the suit or the cabin system.

The primary oxygen supply is a low-temperature cryogenic container, much lighter in weight than pressurized bottles. Bottles, however, are used for the cabin system and as a backup. The cooling system uses a liquid coolant and radiates unwanted heat into space, cooling cabin, astronauts, fuel cells, and electronic equipment. Water is removed from the exhaled gas and returned to the water storage tank. The primary source of water, as mentioned above, is from the fuel cells, as a product of the electrochemical reaction.

The suit is connected to a subsystem that contains a chemical canister of lithium hydroxide and charcoal for purification. This circuit also provides for cooling, water removal, and pressurization. Oxygen enters the suit at the waist, is directed at the face, arms, and legs, and leaves the suit at the waist. The suit system is designed to handle a heat load of 125 kilocalories (500 British thermal units) per hour per man and to maintain the suit at 70" F (21° C). Though the suit is an integral part of the system, it need not be worn by the crew at all times; on the longer Gemini missions the suits were removed and the crew carried on in their long underwear.

Extra-vehicular life-support system

Another integral part of the Gemini system is the extravehicular life-support system used when one of the pilots performs activity outside the craft. This is a semi-open loop, bleeding enough oxygen overboard to maintain an acceptable carbon dioxide level in the pressurized space suit. This gas is supplied by an umbilical hose, which also contains electrical leads. Designed for up to 500 kilocalories per hour, the system proved inadequate on two Gemini flights, when the astronauts both worked so hard during their extravehicular activities that the system could not keep up; as a result the faceplates fogged, partially blinding the astronauts. During extravehicular activity, the extravehicular life-support system contains a supply of oxygen good for 10 to 22 minutes if the supply that is fed to the astronaut from the ship through a hose should fail. The suit itself is a significant improvement over the Mercury suit, having greater mobility, a micrometeorite protective outer layer, and multiple visors. A special port is provided for drinking with the suit under pressure.

Apollo. The atmosphere chosen for the Apollo space cabin was pure oxygen, though the spacecraft was fiiled initially with a mixture of oxygen and nitrogen. The system sealed well enough that nitrogen remained a part of the cabin atmosphere for some days after launch, unless opened for extravehicular activity. Actually three systems, the Apollo life-support system provided essentially the same environment as Gemini, employing somewhat similar technique. The command module operated normally from equipment installed in the service module but was switched to a self-contained system for re-entry. In addition, there was a separate system in the lunar excursion module. Lithium hydroxide was used to remove carbon dioxide, and odour control and water removal were provided. Heat was radiated into space from the service module. Water, provided by fuel cells, was recovered from the spacecraft atmosphere. Oxygen was carried in low-temperature canisters, in a semiliquid, semigas form.

The crew wore sealed space suits during launch and docking operations to protect themselves against the possibility of sudden pressure loss. At other times they wore the main suit with gloves and helmet removed for more comfort, and at still other times a "constant-wear garment," which amounted to little more than a set of flameproof long underwear. In the event of meteorite penetration causing loss of the cabin pressure seal, the oxygen system was set to provide a gush of oxygen into the cabin, sufficient to maintain pressure even with a large leak for the five minutes it takes to don the suits.

Apollo 13: disaster and survival

The life-support capabilities of the entire Apollo complex became a matter of great concern throughout the flight of Apollo 13, during the week of April 11–18, 1970. Three-quarters of the way to the moon, the crew felt an explosion and found that the oxygen had been lost from the service module. As a consequence, the fuel cells soon became inoperative. Although the astronauts were able to store some water and charge their batteries during the last few hours of fuel cell operation, the loss left the mission definitely short of "consumables"—water, electric power, and oxygen. The crew cut power in the command module and transferred to the lunar module for the necessary trip around the moon and back to earth. The life-support system of the lunar module was not adequate for the three crewmen, so it was necessary to rig up a hose from the blower inside it (where there were batteries) to the command module (where there were unexpired canisters of lithium hydroxide for carbon dioxide removal). The lunar module had adequate oxygen; but there was a shortage of water, since the guidance system in the lunar module had to be kept cool (by evaporation), and no water was being produced by the fuel cells. There was enough for the flight, however, and the crew reached earth safely. Since there was only a minimal usage of electrical equipment in the lunar module, the entire spaceship was cold and damp and uncomfortable for the crew.

Lunar suit. Because the demands are different, two types of pressure suits were designed for the Apollo mission, one for use in the spacecraft and the other for exploring the lunar surface. Next to his skin the moon explorer wore a liquid-cooling garment, essentially a loosely woven set of long underwear with a network of flexible tubing through which cooling water from the backpack, or portable life-support system, circulated. The intravehicular suit included a knitted constant-wear garment next to the skin. Moving outward from the skin, both suits had (1) a lining layer of high-temperature-resistant nylon; (2) a ventilating and pressurization system consisting of ducts for circulating pressurizing oxygen to all parts of the suit and helmet, bringing fresh oxygen in and returning it, with its water, carbon dioxide, and heat, to the backpack; (3) a bladder made of neoprene-impregnated nylon, which acted to prevent gas escape from the suit; (4) a layer of nylon cloth; and (5) a "torso-limb suit," which contained convoluted bellows joints and disconnect mechanisms for removal of gloves and helmet. For the intravehicular suit, finally, there was a nylon-lined outer layer of a special fireproof cloth made of fibre glass with each fibre coated with Teflon, a material that does not burn even in pure oxygen. For the extravehicular suit there was next a thermai meteoroid barrier of neoprene and nylon; insulation provided by several layers of aluminized flexible plastic; and an outer fireproof layer protected at points of wear with woven metal fabric. The lunar boots, worn over the regular boots like galoshes, were insulated in the same way and had layers of fireproof felt cushion and soles of silicon rubber.

Supporting the function of the suit was the portable life-support system, or backpack. Weighing 190 earth pounds (86.2 kilograms), but only one-sixth as much as that on the moon, this system provided cooling water, oxygen flow and recirculation through purification and dehumidification equipment, electric power, and communications. The oxygen supply in the pack lasted four hours and was backed up by a short-duration purge system in case of failure. Waste containment units were also built into the suit.

Eating in space. A great deal of anxiety preceded the first few space flights with regard to the problem of eating in weightless space. In early Soviet and American flights, the astronauts ate pureed food such as applesauce, squeezed from flexible tubes, and bite-size chunks of meat, dried fruit, etc. No problems arose in eating any of the foods tried, including even a bologna sandwich smuggled on board a Mercury flight. Crumbly foods such as cookies were found to be a minor problem, since the crumbs were difficult to recover from the cabin.

The problem of eating under conditions of weightlessness is only one consideration. Space food must also have minimal weight and volume; must be nutritionally adequate and supply appropriate energy; must be pathogen-free and stable in the space environment, which may include repeated decompressions, vibration, heat, radiation, and a pure oxygen atmosphere; and must produce low residue and gas. And a primary factor in mission effectiveness is that it have good taste and texture.

Weight is conserved and stability assured by the process of freeze-drying and reconstituting with water before eating. The use of hot water provides a way of serving a warm meal under difficult conditions. In addition to the dried foods there are compressed, energy-dense, bite-sized types that do not require rehydration. Thermal stabilization is used on natural foods for improved flavour and texture; this amounts to "canning" the food in plastic bags. The problem of crumbs has been solved by coating cookies, etc. with a thin, edible layer of methyl cellulose. Food on Gemini and early Apollo flights was served in plastic bags with a one-way rehydration valve and a reclosable mouthpiece; a gun similar to a water pistol was used for adding water. In later flights, when consistencies were suitable, a spoon was used. Menus included roast beef, turkey, sausage, gravy, fruit, cheese, pudding, and cereal with fruit juice, coffee, and tea.

For long-term orbiting laboratory flights, the astronauts will have flat trays with compartments for various foods and will be able to choose which of several canned dishes they wish, put them on a warmer tray, and have a meal with maximum convenience.

Energy consumption in space

Astronauts in space have eaten less than expected, probably because of the lower energy requirements of living under weightlessness and being confined to a small cabin. Energy consumption has been about 2,000–2,100 kilocalories per day in U.S. flights, while the Soviets report using about 2,500 kilocalories. In early U.S. flights extra calcium was added to the diet to help prevent decalcification of bones, but this has been discontinued as not worthwhile.

Waste. The problem of managing body wastes and urine is, like many other problems, dependent on mission duration. In the Mercury program no waste removal was necessary, the need being obviated by use of a low-residue diet for a few days before the mission. Urine was collected by a constant-wear device and retained in the spacecraft. For the longer Gemini flights a zippered opening in the crotch of the suit permitted defecation in a device called a "fecal mitten," consisting of a plastic bag with pressure-sensitive adhesive for attaching to the buttocks and a glove insert into the bag to permit use of toilet tissue before removal of the bag. Tissue and a disinfectant are contained in the bag, which is sealed after use. These devices were also used in the Apollo program, but more easily, since a formal lavatory facility was built into the cabin. The bags were stored after use.

Urine was collected during the Gemini program for analysis, the balance being vented overboard after a sample had been stored. It was collected by means of a metal receiver having an inflatable cuff and drawn into a reservoir by means of a bellows, from which a sample bag (75 cubic centimetres) was filled.

CLOSED LIFE-SUPPORT SYSTEMS

As the length of space missions increases, the approach to life support will necessarily change. Initial lunar-exploration missions can be accomplished with durations of a few weeks, and early orbiting stations or laboratories are planned for a few months but will have a resupply capability. The same applies to lunar stations. But travel to Mars and Venus with propulsion systems presently conceivable will take hundreds of days, weight will be at a premium, and resupply is not feasible. It is necessary, therefore, to develop a system that is closed (*i.e.*, cannot be resupplied from outside) for as many requirements as possible (Figure 3). To seek a completely closed system will perhaps never be feasible, and the usual trade-offs will continue to be made.

From *Astronautics & Aeronautics Magazine* (1970), a publication of the American Institute of Aeronautics and Astronautics

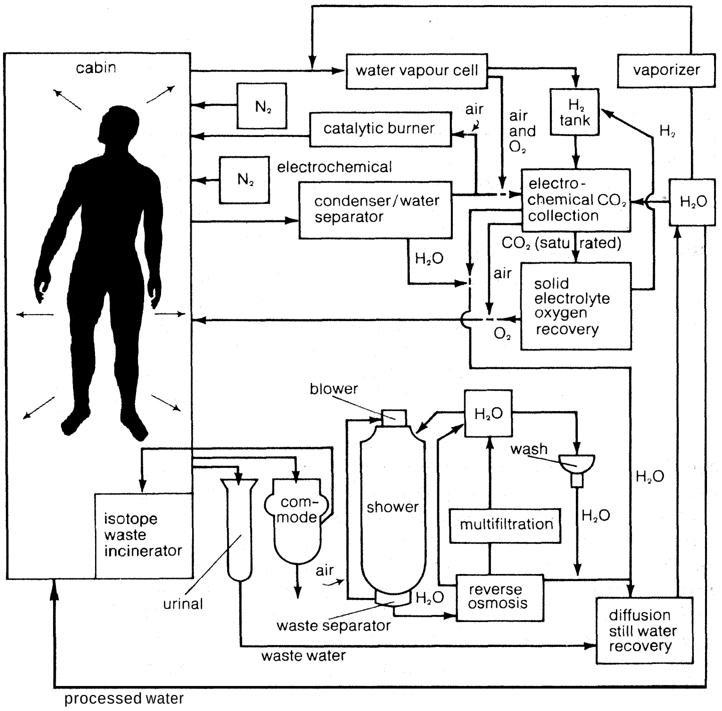


Figure 3: Advanced life-support system proposed for long U.S space flights and for U.S. space stations in the late 1970s and early 1980s.

Requirements for long--term missions

The problem reduces to one of supplying pressure, oxygen, water, food, perhaps heat, artificial gravity, protection against radiation, particles, and micro-organisms, and perhaps some psychological support such as day–night cycle stimuli; and at the same time removing carbon dioxide, gaseous toxins, humidity, solid and liquid wastes, and excess heat.

Operational spacecraft have so far involved only minor steps toward closure of any of these systems. The recirculation of an oxygen atmosphere through a chemical scrubber to remove carbon dioxide may be referred to (though incorrectly) as a "closed" system. Such a process is not really closed because it is only a means of using and reusing oxygen as a vehicle for carbon dioxide removal, and the oxygen actually consumed by the body must be replaced. Presently the only recycling that is done is the recovery and reuse of water, by condensing humidity from the atmosphere.

On long flights the savings to be gained in the use of regenerative systems are most apparent in the recovery of water, chiefly wash water but also the drinking-water–urine cycle, though it is less efficient because of the greater energy and weight required for urine reduction. Oxygen recovery will become practical in flights that exceed the efficient storage capability of low-temperature methods. Reduction of carbon dioxide to pure carbon will be practical at about the same time. Recycling metabolic products to food will effect a great saving, but this appears to be farthest in the future.

In addition to the supply of reactants and the removal of wastes there is the basic question of atmosphere composition and pressure. Oxygen tolerance limits are fairly well established, for short- and long-term exposures. The choice of total pressure and the best gas to use as the inert component are not fully determined, though it has been established that an inert diluting agent is highly desirable if not essential physiologically. Inert gas considerations include storage and handling, flow and heat transfer characteristics, fire, decompression risk, and crew comfort. Nitrogen is well understood, though helium and neon both have advantages, one of which is a reduced risk of caisson disease ("the bends") in case of decompression. This risk is increased by increased pressure, an argument for a cabin pressure of less than that on the earth at sea level. A fair compromise seems to be a pressure about one-half that on the earth at sea level, and roughly equal proportions of oxygen and inert gas.

In considering a short-term system one may select somewhat independently the control methods for each component involved. In a closed system many of these are interdependent.

Potential long-term energy

Another consideration is energy. A system that depends on consumable supplies (such as oxygen and hydrogen powering a fuel cell) to furnish its energy will have different requirements from one whose energy source is adequate and continually available. Nuclear fuels and solar cells both seem capable of fulfilling the latter requirement.

Oxygen production and carbon dioxide removal. There are a number of chemical methods of producing oxygen and removing carbon dioxide. The oxygen production methods include electrolysis of water, in which water is separated into hydrogen and oxygen by use of electric current; the Sabatier reactor, which combines carbon dioxide and hydrogen into water and methane; the Bosch reactor, which converts hydrogen and carbon dioxide into water and solid carbon; the solid electrolyte, which electrochemically separates oxygen directly from carbon dioxide; and the fused salt system, in which dehydrated air is passed through an electrolyte cell containing molten alkali carbonate, yielding oxygen and carbon. Most of the more promising methods of producing oxygen also reduce carbon dioxide, but first the carbon dioxide must be concentrated. These methods, most of which have been discussed above, include molecular sieves, solid amines, ion exchange resins, electrochemical carbonation cells, and membrane diffusion.

Waste recycling. Water may be recovered by several methods. At present urine must be treated with non-reusable chemicals before distillation. Electrochemical pretreatment that would not involve such waste is under development. Another method is vapour pyrolysis, in which urine, which is a mixture of water and various organic compounds, is placed in a distillation column along with a catalyst. The complex organic molecules in the urine are broken down by heat (pyrolized) into water and carbon dioxide; this process is assisted by the catalyst. In another recovery technique, water vapour passes through a membrane through which bacteria cannot pass, supplementing the distillation process. Wash water does not present the same problems as urine and is available in greater quantities. It is filtered through several types of filter beds, then sterilized by heat.

It does not appear likely at this time that chemical treatment methods will be found to make possible the reuse of feces. Biological methods offer some possibilities, however. One suggestion is to remove water and convert the remaining solids to fuel. In any case, feces can be stored by removing water and incinerating..

The methods described above might well be used to provide life support on year-long flights to neighbouring planets. These systems are not closed, however, and require a substantial supply of stored food while leaving solid waste and carbon. At present, it would appear that biological methods are needed to close the system. For several years space medicine investigators have been working on several possible photosynthetic systems that might eventually be expected to accomplish this closure. These include algae, of which the most often mentioned types are *Chlorella*, duckweed (Spirodela *polyrhiza*), and various higher forms such as sweet potatoes. Among biological systems, however, the most promising seems to involve *Hydrogenomonas*, a nonphotosynthetic bacterium. Cultures of these bacteria convert hydrogen, carbon dioxide, and urea from human waste to potentially useful food products.

Some authorities have suggested the possibility of raising animals on long space flights and using them to convert plant products into palatable and useful food. Any one of these small algae must be supplemented with other foods before it can form an adequate diet, for small animals and presumably for man also.

If one adds to the difficulties over and above the problems of bulk, of low-energy efficiency, and the possibility of viral or genetic disasters putting an entire system out of commission, it appears that considerable research remains to be done before a completely closed system becomes a reality. Yet the use of a biological edjunct to enrich an electrochemical routine remains inviting.

The Soviet simulator experiment

To illustrate how near to reality some of the above processes are, three Soviet scientists lived in a hermetically sealed simulator from November 1967 to November 1968. During the last part of the run the atmosphere was recycled and water was reclaimed. Carbon dioxide was concentrated, then used to make water. Special catalytic furnaces were used for atmosphere purification. Moisture was absorbed by chemical beds. Urine was reconstituted into water in several stages, including filtration and distillation. Atmospheric moisture was regenerated by filtration and ultraviolet exposure, followed by more filtration. Wash water received the same treatment plus another stage of filtration. Solid wastes were removed rather than regenerated. Food was mainly vacuum dehydrated, although a greenhouse (using a synthetic resin soil) provided cabbage, cucumbers, and watercress.

III. Underwater systems

Life-support systems that enable man to work underwater are primarily concerned with the gaseous atmosphere he must breathe. Another major problem is protection from cold water.

Undersea gaseous atmospheres can be divided into those that maintain sea-level atmospheric pressure and those that maintain ambient pressure (equivalent to that of the water outside the habitat). The sea-level atmospheric pressure category includes military submarines, research and industrial submarines, bathyscaphes, and certain underwater observation structures. In the ambient pressure category are included undersea habitats and caissons. Diving bells as currently used may be placed in either category.

SYSTEMS THAT MAINTAIN SEA-LEVEL PRESSURE

For vessels strong enough to hold out the pressure of the sea wherever they may operate, life support is intimately related to the time spent there. The problems include maintenance of a suitable atmosphere and regulation of temperature and humidity levels to within the range of comfort and good performance.

Military submarines. Life-support systems in the military submarines used before World War II require no description—they were essentially nonexistent. Every voyage was a test of endurance. There were no provisions for exhausting carbon dioxide or for adding oxygen. Ventilation was poor, humidity was always high, and the temperature was always either too hot or too cold. When the vessel was submerged for more than a few hours, oxygen could be added by bleeding air from the compressed air bank. Soda lime or lithium hydroxide granules were spread on bunks to absorb carbon dioxide, but this was not very effective.

It is generally recognized that the German submarine fleet in World War II suffered more from the inadequacies of its life-support systems than it did at the hands of the enemy. The volume of air available per man was adequate to run submerged during the hours of daylight without any problems; but when the Allies learned to detect surfaced submarines at night with radar, the duration of submergence soon exceeded the time the crew could easily endure. The problem was somewhat alleviated in the later submarines, which were equipped with snorkels that could be used to ventilate the crew quarters as well as furnish air to the diesel engines.

Postwar battery-powered submarines, with greatly improved air-conditioning systems, permitted longer periods of submergence made possible by the installation of lithium hydroxide scrubbers. These are quite simple devices, consisting of several canisters containing a bed of the absorbing agent and a fan for passing the air to be scrubbed through the bed. In addition to simplicity and modest power requirements, these scrubbers have a weight advantage over continuous systems, if they are not needed for longer than ten days. Finely divided lithium hydroxide is irritating, however, and wherever it is used it somehow manages to get into the atmosphere.

The nuclear submarines

A whole new situation presented itself when nuclear propulsion enabled the submarine to run submerged for months without surfacing. A completely controlled atmosphere had to be provided. In these crafts volume is at a premium, but weight and power requirements are not much of a problem. Chemical scrubbers, consequently, are a reasonable solution to the problem of carbon dioxide removal. In these devices, air from the crew compartment is placed in contact with a cool solution containing an organic agent, usually monoethanolamine, which readily takes up carbon dioxide, and the air is then returned to the cabin. The absorbing solution is next transferred by pump through a heat exchanger (a device that transfers heat from a warm object to a cooler object) and to a "stripper," where the carbon dioxide is driven out of solution by heat. A compressor forces the carbon dioxide overboard, and the renewed monoethanolamine is cooled by the heat exchanger and returned to the scrubber. The process is continuous or "regenerative" and requires only energy to keep it running.

In order for a closed environment to be suitable for habitation for many weeks, it is necessary to do more than remove carbon dioxide. Because crews are permitted to smoke, carbon monoxide enters the atmosphere in the submarine and is removed, along with other combustible hydrocarbons, by a catalytic burner, which operates by heating the air stream, passing it over a catalytic bed containing a mixture of copper and manganese oxides called hopcalite, cooling it again, and returning it to the cabin. Combustion products, mostly carbon dioxide and water, are handled by other components of the ship's system. Other organic contaminants, odours, etc. are removed by the carbon filter, a bed of activated charcoal through which the ship's air is circulated. Particulate matter and aerosols are removed by an electrostatic precipitator located in the galley, which is the largest source of such contamination after tobacco smoke.

Three other components complete the system. Temperature and humidity are controlled by air-conditioning equipment using conventional technology. As water is condensed on cooling coils, it takes with it soluble gaseous contaminants, mostly ammonia. Central to the entire life-support system is the atmosphere analyzer, which continuously provides readings of the levels of oxygen, carbon dioxide, hydrogen, carbon monoxide, etc. Finally, there is the matter of a source of oxygen. Some nuclear submarines are supplied by cylinders of compressed oxygen, which ultimately limits the length of time they can remain submerged. More modern vessels contain equipment for separating water into oxygen, which is used, and hydrogen, which is discharged overboard. An interesting difference between the submarine and the spacecraft is that in the submarine there is ample energy available to increase endurance by eliminating the use of a stored commodity, whereas in the spacecraft hydrogen and oxygen are used as a source of energy, via the fuel cell.

Research and working submersibles. Since about 1960 there has been an increasing interest in exploring the sea. One of the by-products of this interest has been the construction of dozens of small research submarines and industrial work boats. As a class these submersibles all have a fairly common approach to life support. Their missions are generally short enough so that little is required except carbon dioxide removal and the addition of oxygen. Almost universally they use oxygen supplied from cylinders and are equipped with Baralyme or lithium hydroxide carbon dioxide scrubbers powered by electric motors. Since virtually all of these small submersibles rely upon battery power, they are not ordinarily able to provide heating and humidity control. Pressure is kept close to that at sea level, and the normal earth atmosphere is generally maintained throughout the mission.

Bathyscaphes, which have been used to explore the deepest parts of the sea, may be compared in terms of life support to small submersibles. Dependent on battery power, they perform short missions, are not heated, and maintain pressure equal to that at sea level. A small cylinder of oxygen and a soda lime canister with a blower are all that is needed. More sophisticated systems involve automatic control and better atmosphere monitoring, but the principle is the same.

Jacques Piccard's "Ben Franklin"

One research submersible (the submarine "Ben Franklin," designed by the Swiss engineer Jacques Piccard) has a life-support system that can be considered unique. The "Ben Franklin" is designed to drift in an unpowered state, remaining submerged for several weeks at a time. The effectiveness of the system was tested on a 1,650-mile (2,640-kilometre) Gulf Stream drift mission from West Palm Beach, Florida, to the North Atlantic in 1969.

Features of the "Ben Franklin" system included: passive removal of carbon dioxide by lithium hydroxide absorption, low-temperature oxygen storage, preheated stored hot water, silica gel humidity control, and the use of the sea to absorb excess heat radiated from the vessel. The boat was designed to operate on minimal power.

The sea-level air environment in the "Ben Franklin" was maintained by using the original nitrogen in the atmosphere when the vessel was at surface and supplying oxygen from low-temperature storage. For a crew of six, oxygen consumption was 1.6 pounds (680 grams) per man per day, the equivalent of about 2,250 calories. This was consistent with the crew activities and food consumption. Carbon dioxide was removed by lithium hydroxide panels mounted in various places throughout the submarine; these depended only upon convective currents for their operation. In the 30-day mission, performance of the system was adequate, carbon dioxide levels never exceeded 1.5 percent, and absorber efficiencies were as great as 75 percent.

Temperature control was passive, in that no additional heat was supplied to the submarine other than 'that from the operation of lights and instruments and generated by the crew itself. This type of control is possible only in temperature between 55" and 75" F (13"–24" C), and operation of such a vessel in colder water will necessitate better insulation and the use of additional heat. In the Gulf Stream drift the temperature averaged 60° F (16" C), and the only thermal control exercised by the crew was change of clothing.

Although the humidity during this particular mission was never uncomfortable for the crew, humidity control is a major problem in a passive system; it was accomplished in this instance by absorption on silica gel. Some 2,400 pounds (1,089 kilograms) of silica gel were used over the 30-day period, and relative humidity was maintained between 65 and 75 percent. Condensation on the hull would accomplish water removal if the sea were 15° F (— 9"C) cooler than the cabin interior, but in the Gulf Stream this is not the case.

The mission was long enough to cause concern about trace contaminant buildup. A variety of contaminants were checked daily or weekly with gas detector tubes, small glass tubes that are filled with appropriate chemical agents through which a measured amount of gas is drawn and that change colour specifically according to the contaminants involved, with an intensity roughly proportional to the amount of contaminant present. The significant contaminants detected by this method were hydrogen, ammonia, carbon monoxide, methane, and ketones. Carbon monoxide had reached a maximum allowable level of 40 parts per million by the end of the 30-day experiment, indicating the need for a removal system on longer missions.

Since power was at a premium, the method for providing hot water was to store enough in vacuum-insulated tanks to last the entire mission. The vacuum insulation on half the water supply broke down, which created a burden in regard to the reconstitution of freeze-dried foods.

Since refrigeration was unavailable, specially packaged freeze-dried meals of about 3,000 kilocalories per day were supplied. The crew found these not entirely acceptable, primarily because the water available was not hot enough to restore the flavour properly. On future missions a greater effort will be directed toward providing palatable foods.

The waste system, like the others, was very conservative. Wash water from the sinks and showers was held in a storage tank and used for flushing the chemical toilet. Toilet waste was run through an electric grinder, mixed with a germicide, and stored. Toilet odours were forced by an electric blower through an odour-removal canister containing Purafil granules. Additional odour removal was provided by mixing activated charcoal with the lithium hydroxide used in the carbon dioxide removal panels. A battery-powered chemical adsorber and an oxidizer were provided for use when needed.

This 30-day mission indicated that the life-support system would be capable of sustaining a mission of up to 60 days with only minor modifications.

AMBIENT PRESSURE SYSTEMS

Diving. Ambient pressure life-support systems are those that involve the exposure of the users to the pressure of the surrounding water. Except for some unusual anthropomorphic pressure-resistant diving "suits," the only way an underwater worker can operate effectively in the underwater environment is to be himself at that pressure and be breathing gas supplied to him at a pressure very near to that of his lungs.

Simple diving devices

A great many reference works on the history of diving show drawings of devices whereby the diver obtains fresh air by means of a tube leading directly from his mouth to the surface. Such a system was described by Leonardo da Vinci about 1500, but neither his drawings nor his explanation of them is clear, and Leonardo may have realized that such a system would not work. The pressure differential across the chest wall, even at a depth of as little as two feet (60 centimetres) makes breathing for any length of time impossible; in addition, the circulatory system in the chest cannot tolerate this stress. This idea, however, has survived in one form in the snorkel used by skin divers. This is a simple, open tube about one foot (30 centimetres) long, curved to fit the mouth of a swimmer, and extending a few inches above the surface. This permits him to look toward the bottom without having to raise his head to breathe, and to swim easily when laden with other diving equipment.

Surface-supplied diving includes the traditional helmeted diver, trailing a stream of bubbles and peering out of a heavy brass helmet through a small round window. The life-support aspects of "hard-hat" diving are simple, though techniques for safe management of buoyancy, valves, lines, etc. are rather complicated. The hose from the surface supplies air to the helmet at a rate sufficient to flush out carbon dioxide exhaled into it by the diver, who controls this flow and by throttling the exhaust can adjust the amount of air in the suit and hence his buoyancy. The helmet and suit are fastened together as a single unit; cuffs permit the diver to use his bare hands (an important factor, since in muddy water most of the work has to be done by feel). He wears woollen underwear as insulation against the cold; if he is careful and wears a well-fitted suit, the inside remains dry.

A variation on the traditional helmet is one designed for "mixed-gas" diving, almost always with a mixture of helium and oxygen. Helium does not cause the narcosis (semiconsciousness) or "rapture of the deep" that becomes limiting when air or nitrogen–oxygen mixtures are breathed at depths beyond about 200 feet (60 metres). Gas supplied to a diver at great depths is necessarily under compression.

Other surface-supplied diving systems can be characterized as "demand" systems. The traditional helmet diver breathes from a space that is supplied with gas and either purged or scrubbed. In recent years a variety of other types of gear have been developed that have the common feature of a demand breathing regulator. This is similar to the regulators used by pilots; it supplies gas to the diver only when he inhales. The diver may wear a helmet that contains a mouthpiece or an oral–nasal mask, or he may wear a mask and a mouthpiece. For a suit he may wear a neoprene foam "wet suit," which provides about a quarter of an inch (six millimetres) of insulation but which is not sealed against water and allows a small amount of water to seep inside and be warmed to skin temperature. Alternatively he may wear a dry suit sealed to the helmet and insulated with woollen underwear, or the "constant-volume" suit, which is a dry suit with special provisions to maintain the same volume at different depths, thereby maintaining a constant buoyancy.

Scuba gear

Most of the demand-type systems are adaptable to self-contained breathing-gas-supply systems, or self-contained underwater breathing apparatus (scuba gear). These provide gas only when the suction of inhalation calls for it and can be used in any situation by merely providing a source of compressed air to the regulator.

The scuba gear used by sports divers and by some commercial divers consists of one or two tanks of compressed air, a regulator, a mouthpiece, and the necessary hoses and harness. Two basic types of scuba are available, identified as single-hose or two-hose systems. In the single-hose system the regulator is worn on the mouthpiece; this is connected to another regulator on the tank by a hose, and the exhaust gas is vented just beneath the chin of the diver. In the two-hose system the regulators (first and second stage) are mounted on the tanks on the back of the diver, who inhales and exhales through a set of two one-way valves integral with his mouthpiece. Two large corrugated hoses connect the valve system with the regulator. There are specific advantages to each of these two systems, but both are acceptable for most scuba-diving situations.

One characteristic of open-circuit scuba, the bubble trail given off by a diver, is disadvantageous in clandestine military applications; various methods have been used to solve this problem. The simplest and most expedient is a closed-circuit system containing pure oxygen and a carbon dioxide absorber. Although pure oxygen can be breathed safely only to a depth of about 25 feet (seven metres), such bubbleless systems were widely used in World War II and are still used in most of the world's navies. Pure oxygen systems are dangerous, however, and are not suitable for unqualified divers.

From another point of view the bubbles given off by a diver represent a loss of gas, limiting the duration of a dive. Attempts to extend the time a diver can work, using the amount of gas he can carry, have prompted the development of both closed and semiclosed mixed-gas systems. These maintain a breathing mixture containing an appropriate amount of oxygen and can be adjusted for use at any depth.

A semiclosed system recirculates the breathing-gas mixture through a carbon dioxide-absorbing canister and adds oxygen as needed. A mixture of inert gas and oxygen is bled at a fixed rate, sufficient to supply the oxygen requirements of the diver. A fraction of the gas expired by the diver is allowed to vent overboard. The system produces a few bubbles, but gets about ten times as much time out of a given amount of gas as would be available by open circuit alone.

The closed-circuit re-breathing apparatus

The most interesting development in scuba gear in recent years has been that of closed-circuit, mixed-gas rebreathing, which uses an electrochemical oxygen sensor to adjust the inflow of oxygen into the system. As in the other scuba, a canister of chemicals absorbs carbon dioxide. No gas at all is allowed to escape—the inert gas is reused for the entire dive time.

Still another innovation in extending the capabilities of scuba is the use of low-temperature, cryogenic gases. Although not quite as advantageous as in spacecraft, in which weight is at a premium, several hours' diving can be obtained from cryogenic open-circuit systems, and more sophisticated mixed-gas, cryogenic systems afford extremely efficient diving. These include modifications to provide carbon dioxide and humidity control by freezing these constituents out of the atmosphere breathed by the diver, and even a possible provision for a diver heat source using a catalytic magnesium burner.

Diving bells. Historically, the diving bell was used as a life-support system before the advent of adequate diving apparatus for diving to even rather shallow depths. In recent years, as the depth to which divers descend has been increased and consequently the decompression time as well, the diving bell once more has become a necessary adjunct to diving.

No longer operating like an inverted tumbler as did its ancestors, the modern diving bell, sealed at the surface, is lowered to the work site retaining surface pressure inside. The diver may take as long as he likes to inspect the job and obtain necessary tools and assistance from the surface; only when everything is ready does he compress the bell to ambient pressure. The diver then opens the hatch. exits into the water, does his job, climbs back into the bell, seals the hatch, and calls for a hoist to the surface. Gas is now vented from the bell at the controlled rate called for in the decompression tables. The bell must be stressed for both internal and external pressure, since it maintains atmospheric pressure inside initially, then holds the diver's bottom pressure or the appropriate table pressure inside during the ascent phase. The bell is lifted onto the deck of the support vessel or platform and attached to a deck decompression chamber. Pressure in the deck decompression chamber is brought up to that of the bell, the hatches between them are opened, and divers can then pass into the deck decompression chamber for a more comfortable decompression.

Submersible chambers are usually outfitted with a minimum of life-support devices, in the interest of simplicity. and to conserve weight and space. For short dives of an hour or two, no carbon dioxide scrubber is needed; usually a simple Baralyme canister with a blower is used. Oxygen for breathing is supplied directly or in mixture with an inert gas such as helium. Lights and communications are essential, and some systems are also equipped with heating and gas analysis equipment. There must of course be both pressurizing and decompression capabilities—the gas used may be attached to the bell in cylinders, or it may be supplied by hose.

Working from the bell, the diver may be supported in several ways. For short jobs, simple open-circuit scuba will work, but at continental-shelf depths these do not afford enough time. More common is the use of a lightweight open-circuit helmet supplied from the bell by hose. This method is named after the Turkish water pipe, the "hookah." Another form of hookah uses a recirculating pump to circulate gas from the bell to the diver and back, thereby permitting the bell's carbon dioxide scrubber and oxygen-makeup systems to serve the diver, and avoiding the wasteful process of open-circuit breathing.

The "hookah" system

Saturation diving. The longer a diver stays down, the longer he must decompress in order to return safely to surface pressure. For any given depth, however, there is a saturation point, at which body tissues are saturated with inert gas; after that, no matter how long a worker stays under pressure his decompression time does not increase. Reasonably safe and efficient decompression from saturation at depths up to 600 feet (180 metres) can be accomplished by a decompression at the rate of 15 minutes per foot (300 millimetres), or about 100 feet (30 metres) per day, so that no decompression tables are necessary.

In practice, saturation divers are compressed slowly to working pressure, generally in the deck chamber, and are then transferred as needed to and from the work site sealed in the diving bell. They may work several hours a day (this is usually limited by cold tolerance) and remain in the deck chamber for rest and recuperation.

Life support in the deck chamber is subject to many of the problems of closed systems, but the methods used are usually rather simple. A major constraint is fire safety; the end of every decompression will always involve an atmosphere in which flammability is greater than normal; sources of ignition and flammable materials are therefore kept to an absolute minimum. Carbon dioxide scrubbing is the major function that must be performed, along with the maintenance of a suitable oxygen level. Analysis for these two gases is an essential part of any saturation system. Humidity is controlled by condensation on cooling coils or with chemical absorbents, and temperature control is provided by conventional means, usually with a heat exchanger in the deck decompression chamber supplied with hot and cold pipes. Food and drink are passed in through a lock.

Warming the diver. Several layers of long woollen underwear under the heavy rubber suit have kept hard-hat divers warm for years. More recently, the advent of popular scuba diving has prompted the development of the wet suit, a layer of neoprene foam between the diver's skin and the water, though not sealed against water leaking inside. Insulation such as this suit affords is adequate for short, shallow dives of up to one hour in even the coldest water. But for two reasons deep helium–oxygen diving has introduced a new dimension to diver cold stress. First, at great depths the insulating value of conventional materials vanishes, since insulation utilizes dead gas space and these gas spaces are now compressed to a small fraction of their original dimensions. Second is the thermal conductivity of helium; helium removes heat several times faster than air does, both through the respiratory tract and through the skin. At 600 feet (180 metres) the heat removed by the respiratory system can be equivalent to the total metabolic heat generated by the man. So it is out of the question to try to support a deep diver by insulation alone. Some supplementary heat is essential.

Two general approaches are being pursued in the matter of insulation, a noncompressible insulation and restoring the gas space by addition of inert gas. Glass or plastic microspheres imbedded in a rubber matrix make a material with most of the insulating capability of neoprene foam but one that does not compress at depth. Such materials in 1970 were stiff, heavy, fragile, and expensive. Inert gas, such as Freon or carbon dioxide (gases that have a low thermal conductivity) can be injected between layers or in bladders of a suit to provide a dead-gas insulating layer. Though this is a feasible approach and is commercially available, it does involve added complexity and can create a buoyancy problem.

With or without adequate insulation, it is still necessary to supply heat to the diver. The most widely used method is an open-circuit hot-water system, whereby hot water is flushed through the inside of a suit and out at the extremities. Such a system requires considerable energy. The closed system has its appeal here, as in breathing. For one thing, an independent free-swimming diver cannot use the hot-water system because of the necessary hose. A closed system of tubes similar to the Apollo garment, but with warm water instead of cold, has been suggested, but the practical result is a suit made up of a series of interconected bladders rather than tubes, contained within a dry suit. One simple suit having space for about 40 pounds (18 kilograms) of warm water postpones for over an hour the cooling of the diver. Such suits require 1.2 to 2.5 kilowatts per hour to maintain a diver at 600 feet (180 metres) in water at freezing temperature. Several methods for supplying this heat are available.

Undersea habitats. There is considerable attraction in the idea of enabling underwater workers to live submerged for days or weeks at a time. Toward this end a number of undersea habitats have been designed, built, and used. The basic requirement of an undersea habitat is a ballasted container filled with a gas mixture appropriate for the depth, having an opening in the bottom for access to the sea, or a suitable system of locks. At the depths where air can be breathed it is feasible to use an open-circuit system to maintain a fresh atmosphere, merely by forcing air into the habitat and letting it bubble out. This method is not feasible where helium mixtures are used; here the minimal requirements are to provide scrubbing of carbon dioxide and maintenance of a suitable oxygen level—analysis and makeup. Temperature, humidity, and food and water are important factors in longer missions.

The first saturation dive

A new phase of man's utilization of the sea began in 1962, when a young Belgian diver remained at a depth of 200 feet (60 metres) beneath the Mediterranean Sea for 24 hours and was decompressed on deck in the same submersible chamber that took him to and from the sea floor. This experiment showed for the first time that the concept of saturation diving was valid. Though saturation diving had been in the thoughts of diving physiologists for years, the experiment demonstrated conclusively that decompression efficiency could be multiplied manyfold by extending bottom time beyond the point at which the body tissues were saturated with inert gas. Once a diver has become saturated, his decompression time is fixed at a maximum value, but he can extend his time at pressure virtually infinitely without increasing the decompression obligation.

Shortly after this experiment, Capt. Jacques-Yves Cousteau, a French biologist and co-inventor of the open-circuit demand-type scuba apparatus, operated his Conshelf I habitat about 33 feet (10 metres) below the Mediterranean Sea for seven days. In 1963 five oceanauts lived for a month in Conshelf II at a depth of about 36 feet (11 metres). A significant step occurred in 1964, when two men remained at 432 feet (132 metres) in the Caribbean for over 48 hours, in an inflated rubber "sausage," travelled to and from the bottom in the six-foot-long diving bell used earlier, and decompressed on deck in a deck decompression chamber built for the purpose. The atmosphere in their undersea house was scrubbed by an electric scrubber, odours were absorbed, and nothing was done about humidity. An electric heater was intended to provide some protection from the chilling effect of high-density helium. Gas composition was 3.6 percent oxygen, 5.6 percent nitrogen, and 90.8 percent helium. The divers were able to swim and work near their habitat by means of a "hookah" rebreather, which pumped chamber gas to a demand mask located on the diver's back and returned expired gas to the diving bell.

In the same year, the U.S. Navy started a series of experiments (Sealab I), keeping four men at 193 feet (58.8 metres) for nine days. They breathed a mixture of 79 percent helium, 17 percent nitrogen, and 4 percent oxygen. A temperature of 29" C (84" F) was found to be

The Sealab experiment

comfortable; humidity remained about 72 percent. No adverse physiological effects were noted in medical and biochemical monitoring conducted on the four divers.

Sealab II was set in 205 feet (62.5 metres) of water off the coast of California; it was a cylindrical tank 57 feet (17.4 metres) long and 12 feet (3.6 metres) in diameter. A total of 28 divers manned three 15-day tours, with one staying 30 consecutive days. The atmosphere was essentially the same as in Sealab I. The divers had free access to the sea, where temperatures were around 10° C (50" F); daily swimming in this water alone caused considerable stress. Though there were many minor changes in the blood and the biochemistry of some of the divers, none were serious or irreversible, and all could probably be associated with the general stress of the situation. For decompression, groups of divers were transferred in a sealed transfer capsule to the deck chamber. An air-conditioning and dehumidification system was equipped also with charcoal and lithium hydroxide scrubbers. Hopcalite was used in some scrubber canisters in an effort to reduce carbon monoxide concentrations that had begun to develop. Helium and oxygen were stored in compressed cylinders on board, and an umbilical cable provided access to these gases on deck of the support barge, as well as gas analysis to topside analyzers. Power and communications were also handled effectively through this cable.

Cousteau and the Conshelf III habitat

Coincidentally, at the same time the Sealab II aquanauts were in the Pacific, the Cousteau operation Conshelf III was 328 feet (100 metres) below the Mediterranean. A sphere 18 feet (5.4 metres) in diameter, the Conshelf habitat was to a large extent autonomous, though it had power and communication lines to a shore support station. Six oceanauts remained in this habitat not only for the 27 days at pressure but for a 3%-day decompression, which took place while the habitat was being towed back to its home port at Monaco. Sophisticated systems were tried in this experiment, including a liquefied gas generator designed to supply breathing oxygen and to remove humidity and carbon dioxide with the same machinery, and at the same time provide a deep freeze for food storage: helium infiltration of its insulation caused its eventual failure.

Finally, in contrast to the expense and complexity of some of the more spectacular operations, a number of simple and inexpensive shallow-water habitats have been run by students, schools, clubs, and individuals. These furnish a focal point for undersea activity and are used for study of marine biology, geology, and limnology. An example is Sublimnos, located in Lake Huron near Tobermory, Ontario, at a depth of 30 feet (9 metres). Supplied by air from a small compressor and electrical power from a generator, it is used by scientists from a number of institutions for continuous surveillance of the biology of the lake, even during the winter, when two feet of ice make routine underwater operations exceedingly difficult.

Underwater welding chambers. A recent engineering innovation has created a special set of life-support problems. Welding under dry conditions is being performed on undersea pipelines in place on the ocean floor. A chamber is placed over the pipe, the water inside is displaced with gas, and diver-welders are sent into the chamber. This situation involves at once many of the traditional problems of diving (such as decompression, inert gas narcosis, and oxygen toxicity), of welding (electric shock, eye protection, noxious fumes), of closed systems (toxic contaminants, carbon dioxide, oxygen makeup, gas logistics), of undersea habitats (humidity, helium speech, choice of gases, analysis), and of fire safety (oxygen levels, hypoxia, hydrocarbons, arcs). A major innovation helping to make safe underwater welding possible was the discovery of the "zone of no combustion." If oxygen is kept below about 6 percent, ordinary materials will not burn even if heated to their ordinary ignition temperature. At depths greater than about 60 feet (18 metres) this percentage of oxygen is enough to supply metabolic needs, and it is therefore possible to maintain an environment in which men can breathe but which will not support fire.

The "zone of no combustion"

IV. Other life-support systems

Firemen's suits. One of the most widespread of all life-support systems is the suit used by firemen. There are two principal hazards from which firemen must be protected, heat and smoke or atmospheres low in oxygen. Though heat is a very real problem, the killer in fatal fires is almost always lack of oxygen.

A suit made of closely woven asbestos alone is adequate for close approach to the periphery of a fire, and it is usually worn when entering a burning building; but by itself it does not afford protection against asphyxia, or more properly hypoxia, lack of oxygen. Neither does a conventional gas mask, since its function is only to remove (usually by means of activated charcoal) toxic agents from the air being breathed. Properly equipped firemen also wear breathing systems, by which oxygen is furnished to them in a system not easily contaminated by the noxious fumes and smoke. These may be of several types but usually are either a closed-circuit oxygen system containing a carbon dioxide absorber or an open-circuit method of breathing from a tank of compressed air. The mask usually covers the entire face because of the threat to the eyes, but it may contain a mouthpiece or mask inside the faceplate to reduce the dead space and prevent carbon dioxide buildup. The design usually includes a method of flowing the fresh gas over the faceplate to reduce fogging.

Anesthesia units. A type of closed-circuit breathing system in common use is the hospital anesthesia machine. This is a system of valves and tubes having a carbon dioxide absorber and a device for introducing a certain percentage of volatile vapour or gas, the anesthetic, into the circuit. It is usually attached to the patient by means of a tube inserted into the trachea and sealed by an inflatable cuff. The anesthetist monitors the condition of the patient, his degree of anesthesia, the amount of anesthetic, and the oxygen and carbon dioxide in the gas circuit.

Gas masks. Another system familiar to soldiers but rarely put to extensive use since World War I is the gas mask. The classical gas mask is little more than a faceplate and a set of tubes and valves that forces inspired air through a canister of activated charcoal and thereby removes aerosols and organic gases present in minute quantities. More modern versions may contain other substances specifically chosen for their ability to absorb the toxins that might be used. Such masks are of course useless in atmospheres lacking oxygen.

BIBLIOGRAPHY. Works treating the basic principles of life support in unfavourable environments include: K.E. SCHAEFER, *Man's Dependence on the Earthly Atmosphere* (1962); P. WEBB (ed.), *Bioastronautics Data Book*, NASA SP-3006 (1964); and J.H.U. BROWN (ed.), *Physiology of Man in Space* (1963).

Works treating aviation and space systems include: T.W. ALEXANDER, *Project Apollo: Man to the Moon* (1964); *The Closed Life-Support System*, NASA SP-134 (1967), *Conference on Nutrition in Space and Related Waste Problems*, NASA SP-70 (1965); V.V. ANTIPOV, "Biologic Investigations on Spaceships 'Vostok' and 'Voskhod'," in T.C. BEDWELL and H. STRUGHOLD (eds.), *Bioastronautics and the Exploratiotz of Space* (1965); H.G. ARMSTRONG, *Aerospace Medicine* (1961); R.C. ARMSTRONG, *Life Support Systems for Space Flights of Extended Time Periods*, NASA CR-614 (1966); I. ASIMOV, *Environments Out There* (1967); J.H.U. BROWN (ed.), *Physiology of Man in Space* (1963); A.C. CLARKE and the EDITORS OF LIFE, *Man and Space* (1964); B.E. FLAHERTY, *Psychophysiological Aspects of Space Flight* (1961); SIR EDMUND HILLARY, "Triumph on Everest," pt. 2, "The Conquest of the Summit," *Natn. Geogr. Mag.*, 106:45–63 (1954); SIR JOHN HUNT, "Triumph on Everest," pt. 1, "Siege and Assault," *ibid.*, 106:1–44 (1954); A.M. JOSEPHY (ed.), *The American Heritage History of Flight* (1962); K. KAMMERMEYER (ed.), *Atmosphere in Space Cabins and Closed Environments* (1966); S. LEVINE, *Appointment in the Sky: The Story of Project Gemini* (1964); U.T. SLAGER, *Space Medicine* (1962); M. SMITH and C.A. BERRY, "Dinner on the Moon," *Nutrition Today*, 4:37–41 (1969); D.J. WATANABE and G.E. LAUBACH, "Life Support," *Space/Aeronaut.*, 52:81–87 (1969); K.F. WEAVER, "First Explorers on the Moon," pt. 3, "The Flight of Apollo II: One Great Leap for Mankind," *Natn. Geogr. Mag.*, 136:752–787 (1969).

Works on underwater systems include: P.B. BENNETT and D.H. ELLIOTT, *The Physiology of Diving and Compressed Air*

Work (1969); J.Y. COUSTEAU, *World Without Sun*, ed. by J. DUGAN (1965); R.W. HAMILTON, JR. and H.R. SCHREINER, "Putting and Keeping Man in the Sea," *Chem. Engng.*, 75:263–270 (1968); T.A. HAMPTON, *The Master Diver and Underwater Sportsman* (1970); C.J. LAMBERTSEN (ed.), *Underwater Physiology* (1967); D.C. PAULI and G.P. CLAPPER, "Project Sealab Report," Dept. of the Navy, ONR Rept. ACR–124 (1967); K.E. SCHAEFER, *Man's Dependence on the Earthly Atmosphere* (1962); U.S. DEPT. OF THE NAVY, *U.S. Navy Diving Manual* (1970); G.A. WELLER, *The Story of Submarines* (1962).

(R.W.Ha.)

Light

Light, a basic aspect of the human environment, cannot be defined in terms of anything simpler or more directly appreciated by the senses than itself. Light, certainly, is responsible for the sensation of sight. Light is propagated with a speed that is high but not infinitely high. Physicists are acquainted with two methods of propagation from one place to another, as (1) particles and as (2) waves, and for a long time they have sought to define light in terms of either particles or waves. In the early 19th century a wave description was favoured, though it was difficult to understand what kind of wave could possibly be propagated across the near-vacuum of interstellar space and with the extremely high speed of 300,000 kilometres per second (186,000 miles per second). In the latter half of the 19th century a British physicist, James Clerk Maxwell, showed that certain electromagnetic effects could be propagated through a vacuum with a speed equal to the measured speed of light. Thus, in the second half of the 19th century, light was described as electromagnetic waves (see ELECTROMAGNETIC RADIATION). Such waves were visualized as analogous to those on the surface of water (transverse waves) but with an extremely short wavelength of about 500 nanometres (one nanometre is 10-metre). The analogy is valid up to a certain point but the experimental results obtained at the end of the 19th century and in the early years of the 20th century revealed properties of light that could not have been predicted from knowledge that was obtainable about other waves. These results led to the quantum theory of light, which in its primitive form asserted that, at least in regard to its emission and absorption by matter, light behaves like particles rather than waves. The results of certain important experiments on the spreading of light into shadows and other experiments (on the interaction of beams of light) that supported the wave theory found no place in a particle theory. For a time it was believed that light could not be adequately described by analogy with either waves or particles—that it could be defined only by a description of its properties. A reconciliation of wave and particle concepts did not emerge until after 1924.

Concept of light as both wave and particle reconciled

Two properties of light are, perhaps, more basic and fundamental than any others. The first of these is that light is a form of energy conveyed through empty space at high velocity (in contrast, many forms of energy, such as the chemical energy stored in coal or oil, can be transferred from one place to another only by transporting the matter in which the energy is stored). The unique property of light is, thus, that energy in the form of light is always moving, and its movement is only in an indirect way (usually to a small extent) affected by motion of the matter through which it is moving. (When light energy ceases to move, because it has been absorbed by matter, it is no longer light.)

The second fundamental property is that a beam of light can convey information from one place to another. This information concerns both the source of light and also any objects that have partly absorbed or reflected or refracted the light before it reaches the observer. More information reaches the human brain through the eyes than through any other sense organ. Even so, the visual system extracts only a minute fraction of the information that is imprinted on the light that enters the eye. Optical instruments extract much more information from the visual scene; spectroscopic instruments, for example, reveal far more about a source of light than the eye can discover by noting its colour, and telescopes and microscopes extract scientific information from the environment. Modern optical instruments produce, indeed, so much information that automatic methods of recording and analysis are needed to enable the brain to comprehend it.

From the standpoint of wave motion, blue light has a somewhat higher frequency and shorter wavelength than red. In the quantum theory, blue light consists of higher energy quanta than the red.

The subject of light is so wide and its associations are so numerous that it cannot be accommodated within one article of reasonable length. There are three main divisions of the subject of light: physical optics, physiological optics, and optical instrumentation. This article deals with physical optics; although electromagnetic theory is considered in some detail, further elucidation may be obtained in the article ELECTROMAGNETIC RADIATION. The articles entitled EYE AND VISION, HUMAN and PHOTORECEPTION include the physiological and psychological aspects of light. The article entitled OPTICS, PRINCIPLES OF is concerned with the basic theory and the technology of lenses, mirrors, and optical systems. Certain highly developed instruments constitute the subjects of specialized text articles (see TELESCOPE; MICROSCOPE; PHOTOMETRY, ASTRONOMICAL; LASER AND MASER). There is also a separate text article on SPECTROSCOPY, PRINCIPLES OF. The experimental evidence that led to the quantum theory of radiation is included in this article together with a brief statement of some of the basic ideas. The quantum theory of radiation, however, is so closely associated with the quantum theory of matter that the two must be considered together and this is done in the text article entitled MECHANICS, QUANTUM.

The following outline summarizes the contents that will be found below:

I. General considerations
    Historical survey
    Basic concepts of wave theory
    Light spectrum
    Velocity of light
II. Interference and diffraction phenomena
    Interference
    Diffraction
III. Polarization and electromagnetic theory
    Polarized light
    Electromagnetic-wave character of light
IV. The interaction of light with matter
    Reflection and refraction
    Dispersion and scattering
    Mechanical effects of light
V. Quantum theory of light
    Photons
    The wave–particle nature of light

I. General considerations

HISTORICAL SURVEY

Pythagorean hypothesis of light

From 500 BC to AD 1650. In this period, there were innumerable confusions and false starts toward an understanding of light. Sometimes an idea was stated, though not clearly, and then almost forgotten for centuries before it reappeared and was generally accepted. The uses of plane and curved mirrors and of convex and concave lenses were discovered independently in China and in Greece. References to burning mirrors go back almost to the start of history, and it is possible that Chinese and Greek knowledge were both derived from a common source in Mesopotamia, India, or Egypt. The formulation of general empirical laws and of speculation about the theory of light derives mainly from Mediterranean (Greek and Arab) sources. Pythagoras, Greek philosopher and mathematician (6th century BC), suggested that light consists of rays that, acting like feelers, travel in straight lines from the eye to the object and that the sensation of sight is obtained when these rays touch the object. In this way, the more mysterious sense of sight is explained in terms of the intuitively accepted sense of touch. It is only necessary to reverse the direction of these rays to obtain the basic scheme of modern geometrical optics. The Greek mathematician Euclid (300 BC), who accepted the Pythagorean idea, knew that the angle of reflected light rays from a mirror equals the angle of incident light rays from the object to the mirror. The idea that light is emitted by a source and reflected by an object and then enters the eye to produce the sensation of sight was known to Epicurus, another Greek philosopher of Samos (300 BC). The Pythagorean hypothesis was abandoned and the concept of rays going from the object to the eye was finally accepted about AD 1000 under the influence of an Arabian mathematician and physicist named Alhazen.

Angles of incidence and of refraction—*i.e.*, the change in direction of a light ray going from one transparent medium to another—were measured by an astronomer, Ptolemy, in the 1st century in Alexandria. He correctly deduced that the ray is bent toward the normal (*i.e.,* the direction perpendicular to a boundary plane, such as the plane separating air and water) on entering the denser medium. A Dutchman, Willebrord van Roijen Snell, discovered the so-called sine law that gives the index of refraction (a measure of the change in direction) for light in a transparent medium. The laws of reflection and refraction were brought together by a 17th-century French mathematician, Pierre de Fermat, who postulated that the rays of light take paths that require a minimum time. He assumed that the velocity of light in a more dense medium is less than that in a less dense one in the inverse ratio of the indices of refraction.

Fermat's principle

The idea of rectilinear propagation of light—that is, that it travels in a straight line—was applied to drawing and painting long ago. Euclid was familiar with the basic idea, but the main theory was developed by Leonardo da Vinci, and a complete description of shadows was given by the Danish astronomer Johannes Kepler in 1604. Kepler also was the first to apply the laws of rectilinear propagation to photometry (the measurement of light intensities).

From 1650 *to* 1895. At the beginning of this period, the result of the conflict between the corpuscular theory and the wave theory was in doubt. At the end of the period, the wave theory was generally accepted and seemed capable of explaining all known optical phenomena though, with hindsight, it can now be seen that there were some important difficulties.

Diffraction—*i.e.*, the spreading of light into shadows—was first observed in Italy in the 17th century. In England, a worker, who independently noticed diffraction, also observed the interference colours of thin films, which are commonly seen today in an oil film on a wet road surface or in the iridescent colours of a butterfly's wing. He believed that light consists of vibrations propagated at great speed. Christiaan Huygens, of Holland, greatly improved the wave theory. In England, Sir Isaac Newton did not attach much importance to the small amount of spreading of light, and he knew that strictly rectilinear propagation could not be reconciled with the wave .theory. Polarization phenomena (which can be accounted for by transverse wave motion in a single plane) discovered in the 17th century by a Danish physicist, Erasmus Bartholin, and by Huygens were not consistent with the theory of longitudinal waves (waves vibrating in the direction of propagation, like compression waves in a coiled spring), which was the only wave theory then considered. Newton therefore supported the corpuscular theory, although he did not reject the wave theory completely. He accepted a concept of a luminiferous ether, and he postulated that the particles had "fits of easy reflection" and "fits of easy transmission"; *i.e.,* he assumed that they changed regularly between (1) a state in which they were reflected at a glass surface and *(2)* a state in which they were transmitted. He thus introduced periodicity—one of the basic ideas of wave theory—in a form that anticipates the quantum mechanics. Newton, using a glass wedge, or prism, discovered that white light can be separated into light of different colours and took the first steps toward a theory of colour vision.

In the century following his death the great authority of Newton was quoted to uphold the corpuscular theory and to oppose the wave theory in a way that he probably would not have approved. It was not until the 19th century that the work of Thomas Young of England; Augustin-Jean Fresnel, François Arago, and Armand-Hippolyte-Louis Fizeau, all of France; Irish scientist Humphrey Lloyd; and German physicist Gustav Kirchhoff established the transverse-wave concept of light; *i.e.*, light is a wave vibration at right angles to the direction of travel. A universal medium pervading all space and called the ether was supposed to be some kind of elastic solid. This made it possible to accept the transmission of light through a vacuum, but there was no completely satisfactory theory of the ether or of the way in which light is modified by transparent materials like glass. The necessity for an elastic solid disappeared when Maxwell proposed an electromagnetic theory of light. He stated the laws of electromagnetism in a clear mathematical form and generalized the concept of an electric current. From his equations he predicted the existence of transverse electromagnetic waves having a constant speed c in *vacuo.* The constant c had a value of 300,000 kilometres per second and was derived from measurements on electrical circuits. It was known from the work of Ole Rømer, a Danish astronomer; Jean-Bernard-Leon Foucault of France; and others that the velocity of light was not much different from the velocity constant *c.* A.A. Michelson, a physicist in the United States, measured the velocity of light and showed that it is equal to *c* within a small margin of experimental error. This result, together with the work of a German physicist (Heinrich Rudolf Hertz) on electromagnetic waves of larger wavelength, confirmed Maxwell's predictions (see ELECTROMAGNETIC RADIATION). The existence of a connection between electromagnetism and light had, indeed, been demonstrated in England much earlier in the century by Michael Faraday, who observed the rotation of the plane of polarization of a beam of light by a magnetic field (Faraday effect).

Transverse-wave concept established

From 1900 to the present. Maxwell's theory is a theory of waves in a continuous (*i.e.*, infinitely divisible) medium. The energy of the waves is also infinitely divisible so that an indefinitely small amount can be emitted or absorbed by matter. Classical physical theories of the 19th century had predicted that in such a system the energy in equilibrium would be distributed so as to give an equal amount to each mode (frequency) of vibration. Because a continuous medium has an infinite number of modes of vibration, and the atoms (which constitute matter) have only a finite number, all the energy of the universe would be transformed into waves of high frequency. Maxwell understood this difficulty, which was later most clearly stated in the Rayleigh–Jeans law (after two English physicists, Lord Rayleigh and Sir James Hopwood Jeans) of the radiation of a blackbody (a body in which the intake and output of energy are in equilibrium). The German physicist Max Planck demonstrated that it is necessary to postulate that radiant-heat energy is emitted only in finite amounts, which are now called quanta. At first, it was hoped to retain, without modification, the theory of light as electromagnetic waves in free space and to use the quantum concept only in relation to the interaction between radiation and matter. In 1905, however, Einstein showed that, in the photoelectric effect, light behaves as if all the energy were concentrated in quanta—*i.e.*, particles of energy now called photons. In the same year, Einstein published the theory of relativity, which modified the whole of physics and gave a special role to the velocity constant *c.* Because light, in some situations, behaves like waves and, in others, like particles, it is necessary to have a theory that predicts when and to what extent each kind of behaviour is manifested. The main development of the quantum mechanics, which does precisely this, took place between 1925 and 1935.

The role of the velocity of light in relativity theory

Light from ordinary sources is emitted by atoms the phases of which are not correlated with one another, so that there is a random irregularity or incoherence between the waves emitted from different atoms. This places severe restrictions on the conditions under which the periodicity associated with wave theory can be observed. In England, Lord Rayleigh appreciated this effect and knew that, by the use of pinholes or slits and light of a narrow range of wavelength, effectively coherent light could be produced. For a long time, interest in this topic lapsed. About 1935 Frits Zernike, a Dutch physicist, and others extended the theory of coherence to include the concept of partial coherence. This appeared to be of practical importance only in a few rather special applications (*e.g.*, in the Michelson stellar interferometer; see below Interference). A theory of stimulated emission, attributable to the work of Einstein and an English physicist, Paul A.M. Dirac, postulated that under certain conditions atoms could be made to radiate in phase so that highly coherent radiation could be maintained indefinitely. The practical realization of these conditions, previously thought to be impossible, was achieved in 1960 (see article on LASER AND MASER).

Information transfer in optical systems

A second major development in the theory of light in this century is the application of so-called Fourier transform methods (a mathematical treatment of light waves) to a wide range of optical problems and, especially, to the transfer of information in optical systems (see article on OPTICS, PRINCIPLES OF).

Today, the theory of light has again reached a point at which all known terrestrial phenomena are included in one logical theory. The known unsolved problems concern the transmission of light over the vast distances of intergalactic space. Here the theory of light impinges on the science of cosmology.

BASIC CONCEPTS OF WAVE THEORY

In this section on the wave theory of light, those properties of light that are consistent with a wave theory are described using a minimum of mathematical formulation (see also WAVE MOTION). It is convenient to introduce the basic concepts of wave theory in relation to mechanical systems. Below, in the section on *Interference*, and beyond, it will be necessary to consider results obtained by more sophisticated mathematical methods, such as Fourier analysis.

General characteristics of waves. Periodicity *in* time and space. If one end of a stretched rope is vibrated, a wave will run along the rope. Figure 1 (top) represents a profile of the wave—*i.e.*, a "snapshot" of the displacement of the rope from its normal position. It gives the variation of this displacement (indicated by ξ) at different points (z) along the axis of propagation for one specific instant of time. Similarly. Figure 1 (bottom) shows the variation with time of the displacement at one arbitrary point on the axis. In Figure 1 (top) the distance between successive crests is constant and is called the wavelength (λ). Similarly, the constant time between crests in Figure 1 (bottom) is called the period (τ). The temporal frequency ($\nu_t = 1/\tau$) is the number of vibrations per unit time and the spatial frequency or wave number ($\nu_s = 1/\lambda$) is the number of waves per unit length. The wave shown in Figure 1 (top) may be represented by the cosine of an angle (ϕ) to give the displacement for a particular point on the axis at any instant of time:

$$\xi = A\cos\phi = A\cos 2\pi(\nu_t t - \nu_s z), \tag{1}$$

in which ξ is the displacement at any point z on the axis at a time *t*, A is the amplitude (the maximum displacement); the angle ϕ (phi) in this case is equal to $2\pi(\nu_t t - \nu_s z)$ and is called the phase angle, or simply, the phase.

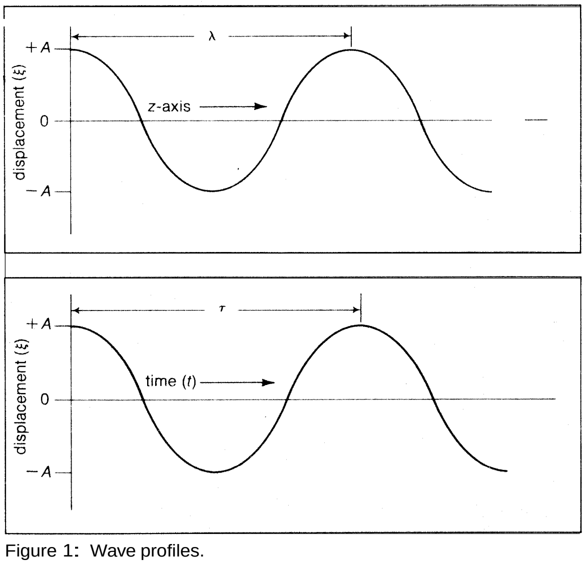


Figure 1: Wave profiles.
(Top) Variation with position at one time. (Bottom) Variation with time at one place (see text).

Energy. The energy per unit volume (W) stored in a wave motion is proportional to the square of the amplitude (A) so that, with a suitable choice of units, $W = A^2$.

Phase velocity. Any one crest moves forward a distance λ in a time τ; *i.e.*, with a velocity b of the wavelength divided by the period or the temporal frequency divided by the spatial frequency,

$$b = \frac{\lambda}{\tau} = \frac{\nu_t}{\nu_s} = \lambda\nu_t. \tag{2}$$

The velocity h is called the phase velocity because the phase angle ϕ will remain constant when the time t changes by an incremental amount t_0 and z changes by z,, $= bt_0$. (This may be seen by substituting $t = t_0$ and $z = z_0$ in the expression for this phase and using b $= \nu_t/\nu_s$.)

Relation of phase velocity to the medium

The velocity of light in vacuum (denoted by c) is the same for all frequencies; all colours travel through space with the same speed. The phase velocity (denoted by b) in a material medium, on the other hand, depends on the medium and on the temporal frequency and, hence from equation (2), on the wavelength.

Wave surfaces. Two-dimensional waves are formed by vibrating (dipping) the end of a rod up and down in the surface of a liquid. Waves spread from the point of origin (where the rod contacts the surface) and, at any moment, the phase at any point on a circle is the same; *i.e.*, if, at a given moment, the wave is at a maximum at one point on a circle then it is at a maximum everywhere on this circle, and the circle as a whole is a wave crest. Similarly, a trough is found at all points on another circle (the radius of which is $\lambda/2$ greater than that of the first circle). As the waves progress farther and farther from the origin, they become less strongly curved about the origin so that, at great distances, they are approximately plane waves.

Light waves are propagated in three dimensions and, for waves from a point source in an isotropic medium (*i.e.*, one in which the speed is the same along any radius), the phase is constant over spherical surfaces drawn about the point source as a centre. The surfaces of constant phase are called wave surfaces, and waves are called plane, spherical, ellipsoidal, and so on according to the shapes of the wave surfaces.

Reflection and refraction. The similarity between the behaviour of light waves and the surface waves of a liquid may be demonstrated with an apparatus called a ripple tank. For reflection of a train of surface waves incident on a flat object, it may be readily observed that the angle of reflection is equal to the angle of incidence. For waves that are refracted in passing from one medium of the ripple tank in which the phase velocity is b_1, to another in which the phase velocity is b_2, measurements of angles of incidence (θ_i) and refraction (θ_r) of the surface waves verify Snell's sine law of refraction for light; *i.e.*, that the ratio of the sines of the angle of incidence and refraction is a constant, or

$$\frac{\sin\theta_i}{\sin\theta_r} = \frac{b_1}{b_2} = n_{12}, \tag{3}$$

in which the constant n_{12} is called the index of refraction from medium 1 to medium 2. The index of refraction (n) from vacuum to a material medium is called the index of the medium and, for transparent mediums is always greater than unity (one). When n_{12} is less than unity, as happens when light is refracted as it passes from glass into air, the refracted ray grazes the surface if $\sin\theta_i =$

n_{12}, θ_i being the angle of incidence in the glass. At angles of incidence greater than this critical angle there is total reflection; *i.e.*, light, instead of penetrating into the air, is reflected back into the glass.

Dispersion. Newton found that, when a beam of white light is refracted by a glass prism, it is dispersed, or split, into beams of different colours. This phenomenon is now interpreted in the following way: the velocity of light in glass varies fairly rapidly with its wavelength, whereas its velocity in air varies little; rhus the index of refraction and hence the angle of refraction depend on wavelength. A beam of white light, containing as it does a wide range of wavelengths, is thus dispersed by a glass prism so that light of one wavelength emerges from it in a different direction from light of another wavelength. Because colour depends on wavelength, the emergent light forms a spectrum (see Plate). All material mediums are, to some extent, dispersive (*i.e.*, phase velocity varies with the temporal or spatial frequency).

Group velocity

Wave groups. When a stone is dropped into a quiescent pond, a few waves may be seen travelling out from the point of impact. This group of waves maintains its identity as it is propagated over a considerable distance, although it finally dies away. The velocity of tho group as a whole is called the group velocity. Careful observation shows that the group velocity is less than the phase velocity. Individual waves may be seen to appear at the back of the group, advance through it, and die out as they reach the front of the group. In a nondispersive medium the group velocity is equal to the phase velocity, while in a dispersive medium it may be greater than, less than, or equal. For light waves, the group velocity is almost always less than the phase velocity.

Interference. When two or more wave motions are present at the same place and time, the simplest assumption is that the resultant displacement (ξ_R) is the algebraic sum of the individual displacements (ξ_1, ξ_2, ξ_3, etc.), *i.e.*,

$$\xi_R = \xi_1 + \xi_2 + \xi_3 + \ldots + \xi_N. \qquad (4)$$

Nearly all observations on light are in accord with this equation, which is a statement of the principle of superposition. These phenomena constitute the subject of what is known as linear optics. The possibility that additional phenomena might be observed at high intensities of light has long been accepted, and the use of lasers in the attainment of the necessary high intensities has led to the discovery of frequency doubling and other effects that cannot be predicted from equation (4). These new observations constitute the material of nonlinear optics (see OPTICS, PRINCIPLES OF). Equation (4) is valid for all the phenomena of interference, diffraction, etc., which will be described in this article.

Two waves are said to be coherent if their phase difference remains constant during a period of observation. Figure 2 shows two equal coherent plane waves travelling across the same space, with the wave fronts inclined at a small angle α, ***AB*** representing a surface corresponding to a crest of one wave. (The surface must be assumed to be perpendicular to the page.) C_1 D_1, C, D_2, etc., represent surfaces that correspond to crests of the other wave. The intermediate dotted lines represent troughs. At points such as P_1 (and P_2, P,, . . .), a crest of one wave coincides with a crest of the other and according to the principle of superposition the displacement is twice that of either wave alone. At points Q_1, Q_2, etc., a crest of one wave meets a trough of another; so the displacements being equal and opposite, the resultant is zero. Thus, an observer looking at a plane that is perpendicular to the page and passes through A B sees a series of straight lines through P_1, P_2, P_3, etc., representing large displacement and a series of lines through &, *Q*,, Q„ etc., representing zero displacement.

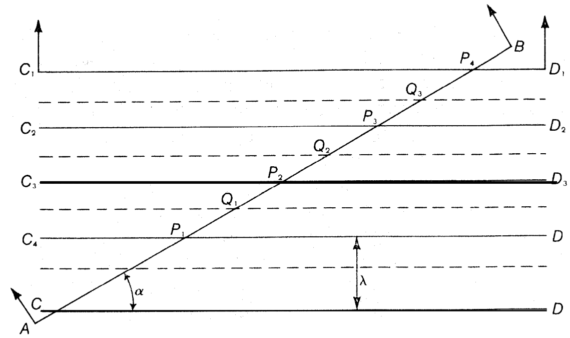


Figure 2: Interference of iwo plane waves *AB* and ***CD*** with directions inclined at an angle ***a***. The crests of ***CD*** are represented as ***C₁D₁, C₂D₂,*** etc., and the troughs are shown as broken lines (see text).

Interference fringes

There are many ways in which coherent beams of light can be made to cross at an angle of about one part in a thousand. The eye (or a low-power magnifier) can be focussed on a plane such as that through ***AB***. The resulting parallel light and dark lines are called interference fringes (Figure 3). From Figure 2 it may be seen that the separation (d) of two bright fringes is ***h/a*** or 1,000 λ if ***a*** = 0.001. When ***a*** has this value, d = 0.5 millimetre for blue-green light and this would imply that ***h*** is about 0.5 × 0.001 or 1/2,000 part of a millimetre (this is usually written 500 nanometres).

Milward T. Rodine

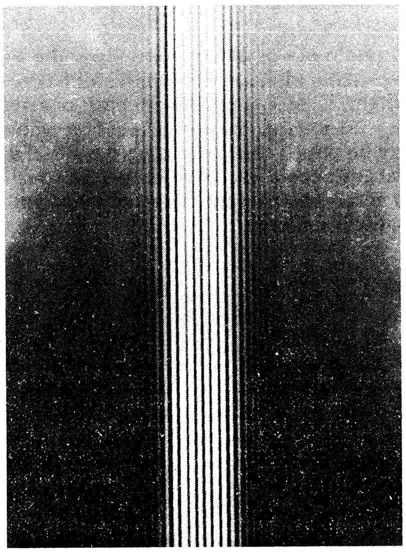


Figure 3: Two-beam interference fringes from Young's double slits or Fresnel's biprism (see text).

In this experiment the spatial periodicity of the light waves (about 2,000 waves per millimetre) has been made to produce fringes with periodicity of about two per millimetre. The spatial periodicity of a light wave is too high for the human eye, and it cannot be magnified directly. Interference methods effectively magnify it so that the resultant fringes can be seen by eye or with a convenient magnification. The following method of producing interference fringes, developed by Thomas Young, is now called Young's experiment.

In the arrangement shown in Figure 4, light of one wavelength passes through a slit S producing semicylindrical waves that are intercepted by two other slits P_1 and *P*,. The two slits P_1 and P_2 act as secondary sources of coherent, semicylindrical waves the combined effect of which is observed on the plane perpendicular to the page and designated ***AB***. In a typical case the separation (a) of P_1 and P_2 is a millimetre and the distances l_1 and l_2 are each about a metre. The slits are a centimetre or so long but are much less than a millimetre wide. They are accurately parallel to one another and, as represented in the drawing, are at right angles to the page. Because the waves from P_1 and P_2 are indirectly derived from the same small source, they are coherent. When they cross plane ***AB*** they are nearly plane because of the large radius, and they intersect at an angle ***a*** equal to 0.001. It may be shown that the intensity (I) for these fringes varies

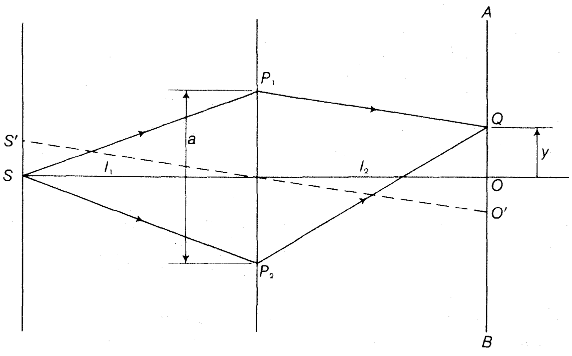


Figure 4: Young's experiment (see text).

from point to point along the line ***AB*** in the way shown in Figure 5 (curve A), which is in accord with the equation

$$I = 2A^2(1 + \cos\varepsilon) = 2I_0(1 + \cos\varepsilon), \qquad (5)$$

in which ***A*** is the amplitude of either wave, I_0 is the intensity of one wave acting alone and the phase difference $\varepsilon = 2\pi ya/\lambda l_2$. Bright fringes are seen in positions for which $\varepsilon = 2\pi p$ or $y = p\lambda/l_2a$ (in this case p is a whole number, which may be positive, zero or negative—0, ±1, ±2, ±3, etc.). Because cos ε varies from -1 to $+1$, I varies from $4I_0$ to zero. The average, in accordance with the law of conservation of energy, is 21.

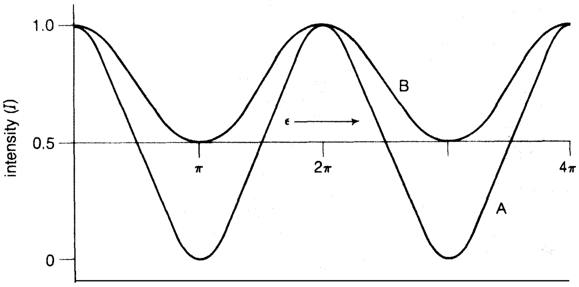


Figure 5: Interference fringes obtained in Young's experiment (see text).

Diffraction. Plane waves that pass through a restricted opening emerge as divergent waves. When the opening is less than one wavelength in diameter the emergent wave is nearly spherical. Whenever a beam of light is restricted by holes or slits or by opaque obstacles that block out part of the wave front, some spreading occurs at the edges of geometrical shadows. This effect, called diffraction, is also obtained with transparent obstacles that cause an irregularity in the wave front. Diffraction can be demonstrated by allowing a parallel beam of light to fall on a grating consisting of an array of equally spaced narrow slits. If the extent of physical separation of two adjacent slits is ***e***, then the path difference between any two adjacent rays emitted in a direction symbolized by e is $e \sin\theta$, and if this path difference is an integral number (p) of wavelengths,

$$e \sin e = p\lambda, \text{ or } \nu_s \sin\theta = pg, \qquad (6)$$

in which ν_s is the spatial frequency ($1/\lambda$) and g is the number of lines per unit width of the grating, then the waves from different slits have phases that differ by angles of $2p\pi$, and they reinforce one another. Thus, when lenses are employed with a grating, sharp lines are obtained for each wavelength at values of θ corresponding to integral values of ***p***. If white light is used, each line is drawn out into a spectrum of wavelengths because the direction of reinforcement depends on the wavelength.

Polarization. In the propagation of waves on a rope or across the surface of a liquid the displacement (as shown in Figure *1*) is in a direction perpendicular to the direction of propagation and the waves are said to be transverse. Sound waves in a gas consist of alternate dilation and compression and the displacement is in the direction of propagation. The waves are longitudinal. If a beam of longitudinal waves is propagated in a vertical direction, there is nothing to distinguish one azimuthal plane from another—everything that is true for an east–west plane is equally true for a north–south plane. With transverse waves the displacement may be in the east–west plane; in that case, there is no component in the north–south plane, and this should manifest itself in the form of a property that depends on the azimuth. Such an effect is called an azimuthal property. An ordinary beam of light from a thermal source does not exhibit any azimuthal property, but experiments show that light can have an azimuthal property and must be represented by transverse waves.

Azimuthal property

If an unsilvered glass plate has an index of refraction equal to *1.5* and the angle of incidence of a beam of light is *57°*, about *15* percent of the light will be reflected from the two glass surfaces of the plate (Figure *6*); this percentage will not be altered when the glass plate is rotated about an axis parallel to the beam of light so as to change the azimuth of the plane of reflection. If a second mirror $(G,)$, parallel to the first (G_1), is used to reflect the beam in the same plane as that of the original reflection, about 30 percent of the light incident on the second plate of glass will be reflected; but if the second plate is turned so as to reflect the light in a plane perpendicular to that of the first reflection—*i.e.*, out of the plane of the page—hardly any light will be reflected. Thus, after the first reflection, the beam of light will have acquired an azimuthal property—it will be reflected more strongly when the transverse displacement is in one azimuthal plane than when in another. Further tests will show that the transmitted light has a complementary azimuthal property; it is more strongly reflected in the perpendicular plane—though the difference is less marked

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

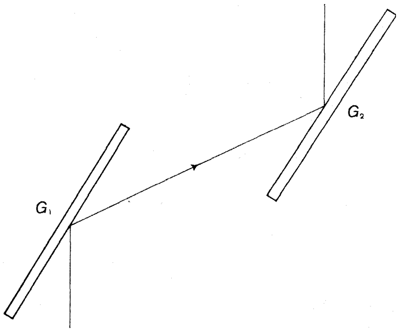


Figure 6: *Malus'* experiment.
Successive reflections at two unsilvered mirror surfaces, G_1 and G_2 (see text).

These results may be understood if ordinary light consists of a mixture of transverse waves with displacements in all azimuthal planes but only one component is reflected from a glass surface when the angle of incidence is 57°. The reflected light is said to be plane-polarized because all of the displacement of the wave is in one azimuthal plane. The transmitted light (about 85 percent of the whole) contains about *50* parts of *a* component that is polarized in a perpendicular plane and about 35 parts of light that is polarized in the same way as the reflected light. It is more strongly reflected in the plane of the page, but because it is only partially polarized, the azimuthal effect is less.

The above experiments do not show whether or not the reflected light has its displacement in the plane of reflection or perpendicular to it. It is a matter of choice whether the reflected light is said to be polarized in or perpendicular to the plane of reflection. Some controversy (and some difference of nomenclature) that formerly led to confusion was removed by the electromagnetic theory (see below). In this theory light is represented by two vectors (quantities that can be represented graphically by arrows that point in the field directions), a magnetic vector in the plane of reflection and an electric vector perpendicular to it. Confusion is avoided by specifying the plane of the electric vector instead of speaking of the plane of polarization.

The azimuthal property of reflected light at the surface of any medium—glass, plastic, a liquid—is most strongly manifested when the angle of incidence is so chosen that its tangent is equal to the index of refraction; that is, it satisfies Brewster's law (after Sir David Brewster, a British physicist), which states that, at the polarizing angle, the incident and refracted beams make an angle of 90° with one another: tan $\theta_i = n$, in which e, is the angle of incidence, called the polarizing angle, and n is the index of refraction of the medium. Nevertheless, there is some azimuthal difference after reflection at any angle except $\theta_i = 0$ or $\theta_i = 90°$. Other ways of producing polarized light are described in a later section.

Optical activity

It is found that the plane of polarization of a beam of polarized light is rotated when the beam is passed through certain mediums (especially sugar solutions). These mediums are said to be optically active. Most mediums do not normally rotate the plane of polarization, but do so when there is a magnetic field in the direction of propagation (the Faraday effect).

The wave equation. The expression for a plane wave, given in equation (1) and showing the relationship between displacement ((), the time span (t), and distance (z) along the wave, may be differentiated twice with respect to t and z; that is, to find out how the displacement changes with position and time. This operation yields the partial differential equation:

$$\frac{\partial^2\xi}{\partial z^2} = \frac{1}{b^2}\cdot\frac{\partial^2\xi}{\partial t^2}, \qquad (7a)$$

in which b is the phase velocity. For a three-dimensional wave the analogous expression is

$$\frac{\partial^2\xi}{\partial z^2} + \frac{\partial^2\xi}{\partial y^2} + \frac{\partial^2\xi}{\partial x^2} = \frac{1}{b^2}\cdot\frac{\partial^2\xi}{\partial t^2}. \qquad (7b)$$

There are many solutions of this basic equation. Some correspond to the sinusoidal plane waves, which have already been considered. Others correspond to groups of plane waves that differ slightly either in direction, or wavelength, or both. Yet another solution of the general wave equation is:

$$\xi = \frac{A}{r}\cos 2\pi(\nu_t t - \nu_s r), \qquad (8)$$

in which r is the magnitude of a radius vector drawn from the origin and A is a constant. This represents spherical waves.

Energy of a beam of light. The energy in a small volume (dV), through which plane waves are passing, is proportional to the product of the square of the amplitude (A), or its energy per unit volume (W), and the small volume; that is, $A^2dV = WdV$. The rate of transport of energy across a surface normal to the direction of propagation is proportional to the product of the energy per unit volume, the phase velocity, and a small area (dS) normal to the direction of propagation, or $WbdS$. For spherical waves, the rate of transport is inversely proportional to r^2, *i.e.*, $(A/r^2)dS$. Because the area of a sphere is $4\pi r^2$ in which r is its radius, this equation implies that the total energy crossing any sphere surrounding a point source is independent of the radius. Thus, inverse-square law for the intensity of radiation at a distance r from a point source is in accord with the law of conservation of energy—the total energy of a wave remains the same even though the wave is spread over a greater area.

Doppler–Fizeau effect. The length of a wave train emitted in one second by a stationary light source is equal to the velocity of light (c) times one second, which in itself is equal to the product of its frequency (ν_t) times its wavelength (λ)—*i.e.*, $c = \nu_t\lambda$. If the source moves away from the observer with a velocity (v) that is small compared with the velocity of light, then the length of the wave train increases so as to be numerically equal to the sum of the two velocities ($c + v$) and the number of waves remains the same. The wavelength h increases to h' by a factor $(c + v)/c$; that is h' $= (1 + v/c)\lambda$. This change was discovered by an Austrian physicist, Christian Doppler, in the 19th century in relation to sound waves and subsequently applied to light waves by Fizeau. It is called the Doppler–Fizeau effect. The Doppler–Fizeau effect is easily observed when part of the light from a gas laser is allowed to be scattered by a moving body and mixed with a little unscattered light. It is known from the study of sound waves that the beat frequency is equal to the difference between the frequencies of the two waves that are mixed. Although the frequency of light waves is extremely high (more than 10^{14} per second), the beat frequency may be a megahertz (10^6 cycles per second), which is easily detected by radio amplifiers, or even a few hundred cycles per second, which the human ear can detect. Thus, just as interference fringes provide a periodic phenomenon in which two light waves combine to produce fringes of low spatial frequency, so the Doppler–Fizeau effect produces beats the temporal frequency of which is a known, but very small, fraction of the temporal frequency of the light waves. In this way the periodicity of light in both space and time is exhibited and measured.

Beat frequency

LIGHT SPECTRUM

It was seen, in the preceding section, that white light can be dispersed into a spectrum by refraction, by diffraction, or by interference. Newton showed that if a suitably oriented slit is used to select a small region of the spectrum, the light that passes through the slit is much more homogeneous than the original white light, and he was unable to observe any further dispersion when passing this light through a second prism. Delicate methods of interferometry nevertheless show that this light is never entirely of one wavelength, however fine the slit, but covers a range (Δλ) of wavelengths. The ratio of the wavelength divided by this range, which measures the purity of the spectrum, may be a few thousand for a spectrum formed by a prism and up to a million for a spectrum formed by a large diffraction grating. It is never infinite, as it would be if Δλ were zero.

The spectrum of a hot body such as the solar photosphere is continuous (every wavelength is represented); but a German physicist, Joseph von Fraunhofer, early in the 19th century observed that the solar spectrum contains numerous dark lines appearing at certain wavelengths, which are attributed to wavelengths originally emitted by inner layers of the Sun but then absorbed by various elements (in gaseous form) in the cooler outer layers (see Plate). Emission spectra produced by electric sparks and arcs contain sharp bright lines which are characteristic of the elements in the electrodes (see SPECTROSCOPY, PRINCIPLES OF).

In monochromatic light, colour and wavelength are associated. Nevertheless, as Newton said, "the rays, to speak properly, are not coloured." Colour is a sensation in the human mind. Light of one wavelength can stimulate the visual system so that a certain colour sensation (*e.g.*, red) is produced. The way in which the visual system analyzes colour is entirely different from the way in which physical instruments form a spectrum (see EYE AND VISION, HUMAN).

There are a number of ways in which spectra are produced in nature. The rainbow is the most striking of these. The primary rainbow is formed by reflection and refraction of light in raindrops. The rays emerging from the drops are spread out, but for any given wavelength there is a minimum angle of deviation and there is a concentration of energy at this angle. For green light the minimum angle of deviation is about 138" and an observer with his back to the Sun sees the bow at an angle of 42°

Primary and secondary rainbows

to the direction of the Sun's rays. Because of the dispersion of water, the angles for different wavelengths are not exactly the same, and the red is seen on the outside and blue on the inside of the bow. A weaker rainbow is formed by rays that have been twice reflected. In this the colours are reversed. Still weaker supernumerary bows are caused by diffraction in droplets. A rainbow may be regarded as a spectrum of the Sun, but the purity is low.

VELOCITY OF LIGHT

The accepted value of the velocity of light *(c)* in vacuum is 299,792.5 ± 0.4 kilometres per second, or 186,282.4 ± 0.2 miles per second (see The Table). The velocity is the same for all wavelengths over the whole range of the electromagnetic spectrum from radio waves to gamma rays. Methods of measurement are of three types: (1) measurement of the time (T) in which a group of waves covers a known distance (*l*); (2) measurement of the frequency (ν_t) and wavelength (h) of monochromatic waves; and (3) indirect methods, such as measurement of the change of frequency or wavelength (Doppler–Fizeau effect) when a beam of light is reflected from a mirror moving with a known velocity.

The Constant c
(in kilometres per second)

	year	value
Derived from measurements of the velocity of light		
Michelson	**1927**	**299,796 ± 4**
Michelson, Pearson and Pease	**1935**	**299,774 ± 11**
Value accepted in 1941	**1941**	**299,773 ± 3**
Bergstrand	**1951**	**299,793.1 ± 0.2**
Bergstrand (mean value)	**1957**	**299,792.9 ± 0.2**
Derived from measurements on radio waves		
Essen (10' MHz)	**1950**	**299,792.5 ± 1**
Froome (2.4 and 7.5 × 10⁴ MHz)	**1951–8**	**299,792.5 ± 0.1**
Value adopted by 12th General Assembly of the Radio-Scientific Union	**1957**	**299,792.5 ± 0.4**
Derived from electrical measurements		
Rosa and Dorsey (ratio of units)	**1907**	**299,788 ± 30**
Mercier (Lecher wires)	**1923**	**299,795 ± 30**

Methods of type (3) have, so far, given an accuracy of only a few percent. Methods of type (2) cannot be used for light waves because the frequency is about 1.5 × 10^{14} hertz and is too high to be measured directly. The remainder of this section will review measurements of the velocity of light by methods of type (1) and compare the results of the best measurements with the results obtained for radio waves by methods (1) and *(2)*.

Astronomical measurements. In 1676 Rømer made careful measurements of the times at which satellites of Jupiter were eclipsed by the planet. The times observed did not agree with those calculated on the assumptions of a constant period of rotation and of instantaneous transmission of light. Starting at a time when the Earth was at its nearest to Jupiter, the apparent period increased and the eclipses became increasingly later than the calculated limes as the Earth receded from Jupiter. Similarly, the period shortened when the Earth was moving toward Jupiter. The observed times were consistent with a finite velocity of light such that the time for it to transverse the Earth's orbit is about 1,000 seconds. Taken with modern values of the size of the Earth's orbit, the derived value of the velocity is 298,000 kilometres per second. It is remarkable that this first measurement was even of the correct order; the most important conclusion was that the velocity of light is finite. An English astronomer, James Bradley (died 1762), obtained a similar value by the so-called aberration method, based on the apparent motion of stars as the Earth travels in its orbit about the Sun.

Early terrestrial experiments. In terrestrial experiments by method (1), the beam of light is periodically marked either by interrupting it at regular intervals or by modulating it (alternately increasing and decreasing its intensity). The marked beam is transmitted to a distant mirror and the return beam passes through the apparatus that interrupts or modulates the outgoing beam and then to a detector. If the time required for transmission to the distant mirror and return is ½, 3/2, 5/2, . . . , times the period of the interrupter (or modulator), then the amount that reaches the detector is small. It is usual to adjust either the path length or the period of the interrupter or modulator until the light registered by the detector is a minimum. In the earlier experiments a mechanical chopper was used as interrupter and the eye was the detector. The later experimenters used electronic modulators and photoelectric detectors.

The apparatus used by Fizeau in 1849 is shown in Figure 7, in which M_1 is a partially reflecting mirror and M_2 is a fully silvered mirror. As the speed of the wheel (which has 720 teeth) was increased from zero it was found that the light was first eclipsed by a tooth when the speed was about 12.6 revolutions per second—*i.e.*, when the time to make the round trip was 560 microseconds (0.00056 second), the length of the double path being 17.3 kilometres (about ten miles). The chief error in the measurement lay in the difficulty of determining the exact speeds at which the light received by the eye at E was at a minimum. Essentially the same method was used by others between 1874 and 1903. The accuracy gradually improved, and it was shown that the velocity is between 299,000 and 301,000 kilometres per second.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by Permission of John Wiley & Sons, Inc.

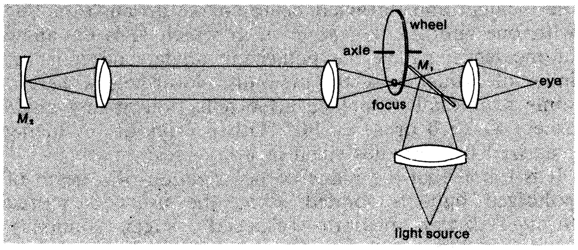


Figure 7: Fizeau's method for measuring the velocity of light.

In 1834 Sir Charles Wheatstone of England suggested a method incorporating a rotating mirror for interrupting the light that was later developed by Arago (1838) and Foucault (1850). It was considerably improved by Michelson, who made measurements over a span of over 50 years from 1879 to 1935.

Use of rotating mirror

Michelson's measurements. Figure 8 shows the arrangement used in 1927. The mirror M_3 **is** a little above the plane of the diagram, and M_3' is a little below. Light from the source *S* passes to one face of the octagonal mirror M_1 and then to M,, M,, and M,. From M_4 it goes to the mirror M_5 at a distance of about 35 kilometres (about 22 miles). It returns via M_6, M_4, M_3', and M' to the octagon. An image of *S* is seen in an eyepiece at E. The octagonal mirror rotated at 528 revolutions per second. It turned through approximately one-eighth of a revolution during the transit of the light. If the rotation were exactly one-eighth of a revolution the image would be undisplaced from the position it had when the mirrors were stationary. In some of Michelson's experiments the speed of rotation was slowly changed until this condition was obtained. In others the speed and distance were fixed, and a small displacement of the image was measured.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

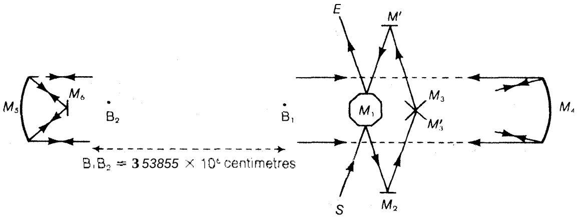


Figure 8: Michelson's Mount Wilson experiment, 1927.

It is difficult to estimate the accuracy of Michelson's

1927 and 1935 experiments, and it is no longer important to do so in view of the more accurate measurements made since 1945. His most important contribution to the measurement of the velocity was the proof that the velocity agreed with Maxwell's prediction to better than one part in a thousand. This gave confidence to those working on applications of the electromagnetic theory.

The electro-optical shutter. This device, based on the Kerr effect (see below), makes it possible to modulate a beam of light at frequencies more than 10,000 times the highest frequency of interruption used by Michelson and obtain values in reasonably good agreement with each other and with Michelson's later work. This method was greatly improved by E. Bergstrand in Sweden, who reduced the random errors by a factor of more than 30 and obtained a value for the velocity of light of 299,793.1 kilometres per second.

Radio-frequency measurements. The velocity of electromagnetic waves of radio frequency in vacuum has been measured by several methods. An English physicist, Louis Essen, measured (1950) the resonance frequency of a cavity resonator (an electromagnetic device) whose dimensions were also determined with high accuracy. Keith Davy Froome, a physicist in England, measured (1952 and 1958) the wavelength in air, corresponding to a known frequency, using a microwave interferometer. The results of these and other measurements are in agreement with those of Bergstrand to within a few parts per million. The velocity of radio waves in vacuum is thus equal, within this accuracy, to the velocity of light. The velocity of gamma rays is also the same, within the much lower accuracy of this last measurement. The Table summarizes the measurements of the velocity constant (*c*) and shows that there is now satisfactory agreement between results obtained over a wide range of conditions.

Recognition of the velocity of light as a fundamental constant

Since the publication of the special theory of relativity (1905), the constant c has been recognized as one of the fundamental constants of modern physics. For this reason, attempts will undoubtedly be made to measure it with even greater precision. The use of lasers may help, but a major improvement will require the establishment of better standards of length and time than those now available (see CONSTANTS, PHYSICAL).

Velocity in material mediums. All measurements of the velocity of light involve interruption or modulation of a beam of light so as to form groups of waves and the velocity measured is the group velocity. The difference in magnitude between the wave velocity and the group velocity of light in air is only about one part in 50,000, but in most glasses and in some liquids it is much larger. Michelson obtained 1.758 for the ratio of the velocity in air to the velocity in carbon disulfide. The inverse ratio of their indices of refraction is 1.64 and the value calculated from this for the ratio of group velocities is 1.745 for wavelength 580 nanometres, close to Michelson's observations. Bergstrand found that the ratio of the velocity in *vacuo* to the velocity in a certain glass was 1.550 ± 0.003. The refractive index of the glass was 1.519, but the ratio of c to the group velocity was 1.547. The experimental results thus agree with those calculated on the assumption that the measured velocity is the group velocity.

II. Interference and **diffraction** phenomena

INTERFERENCE

Quasi-monochromatic waves. A perfectly monochromatic wave, represented by equation (1), has constant amplitude and is not limited in space or in time. Sources of light (other than lasers) emit waves the amplitude of which varies with time. For example, a single undisturbed atom emits a damped wave (Figure 14A). Under favourable conditions the damping is so weak that 10^7 waves are emitted before the amplitude has fallen to half its initial value and the change of amplitude is not significant over a distance of several thousand wavelengths. Wave trains of this type are said to be quasi-monochromatic. Superposition of these waves gives interference when the path difference is not too large.

Photometric summation. Figure 5, curve A, shows the way in which the intensity of light varies from place to place when two monochromatic or quasi-monochromatic waves overlap. The intensity at a point in the region where the waves overlap may be expressed as the sum of two terms: (1) the sum of the intensities of each wave acting alone ($2I_0$ if each alone would give intensity I_0); (2) a term representing the interference of the waves. The second term varies from point to point along the direction of propagation between the values $-2I_0$ and $+2I_0$. Thus the total intensity varies from $4I_0$ (*i.e.*, twice the intensity sum) to zero. Now, when a large number of waves from different sources cross a certain space, the fringes caused by the interference of each pair of waves have their maxima in different places and the overall result is that, at any point, the interference terms are positive nearly as often as they are negative and their total sum is nearly zero. In this case the resultant intensity at any point caused by a number of sources is just equal to the sum of the intensities (at that point) of each source acting alone. This is the law of photometric summation and is used by illumination engineers in calculating the illumination on a surface that receives light from various sources. Interference fringes are obtained only when experimental conditions are such that the interference fringes caused by light emitted from different atoms all have their maxima in the same places (or near to the same places). The interference term then becomes a significant fraction of the summation term. This may be achieved either (1) by using two secondary sources (such as the two slits used in Young's interference experiment), which are both derived from the same primary source, or (2) by using a laser in which the source atoms are stimulated in such a way that the phase relations between them remain constant during the period of observation.

Visibility of **interference** fringes. The distribution of intensity in interference fringes, shown in Figure 5, curve A, represents an ideal that is closely approached in some experiments, but generally the distribution is such that the fluctuations that constitute the fringes are superposed upon a nearly uniform background. Michelson defined the visibility of fringes as the difference between the maximum and minimum intensity of a fringe divided by their sum, or

$$V = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \tag{9}$$

in which V is the visibility, $I_{\max}$ is the maximum intensity, and $I_{\min}$ is the minimum. The fringe visibility is thus always between zero and one. When the minimum intensity is zero, the visibility equals one. Obviously, fringes for which V is less than one are obtained when waves of unequal amplitude are superposed because the weaker cannot, at any point, annul the stronger. It is also found, however, that even when the intensities are equal, the visibility is usually less than one (as shown in Figure 5, curve B). Further consideration of Young's slit experiment leads to recognition of two conditions that must be fulfilled to obtain fringes of high visibility. These relate to their geometrical condition and spectral range.

Two conditions for high visibility

Geometrical conditions. In the arrangement shown in Figure 4, the centre of the fringe system is at a position *O* on the screen, on the straight line from the source slit *S* to a position midway between P_1 and P_2 (the slits are all assumed to be extremely narrow). If slit *S* is moved to *S′* then the centre moves to *O′*. If, instead of moving the slit *S* to this new position, it is gradually widened, the intensity at any point **Q** is found by adding the intensities of waves emitted by atoms behind different parts of the slit. Because the fringes on plane *O′OQ* produced by light from different parts of slit *S* are not in register, there cannot be zero intensity at any point in the pattern. As slit *S* is widened the fringes gradually become blurred—*i.e.*, the visibility falls from unity to zero. If $I_1 = I_2$, no fringes are seen when the width of slit *S* is about equal to the distance (d_f) between successive fringes.

Spectral range. In the case in which the slit *S* is extremely narrow and the light is not all of exactly one wavelength, the path difference and the phase difference will be zero at the centre of the fringe system for all wavelengths, so that for all wavelengths there is maximum intensity at the centre O of the fringe system. Because the separation of the fringes is proportional to the wavelength, the fringes produced by light of different wavelengths gradually go out of register as the path difference is increased. With white light, one clear fringe is seen in the centre. A few coloured fringes are seen on either side because the eye makes a certain degree of separation of the colours. If a filter is used to restrict the light to a band of say 50 nanometres wide, then about ten fringes may be seen on either side, and this number is increased if the wavelength range is further restricted.

These two causes of reduced visibility differ in that the geometrical condition affects all parts of the fringe system equally and the effect of the spectral range increases as the path difference increases. In discussing these phenomena it has been assumed, in accordance with the preceding discussion, that the intensity of light from different atoms obeys the law of photometric summation. It is also assumed that the photometric law applies when different wavelengths are superposed.

Partial coherence

Coherence. When two beams of light can interact so as to produce interference fringes the visibility of which is unity, they are said to be perfectly coherent. When their interaction produces no fringes (but only photometric summation) they are said to be noncoherent or incoherent. An elaborate mathematical theory of coherence that has been developed recognizes that coherence and noncoherence are extreme cases—between them lies "partial coherence." Zernike, who contributed a great deal to the development of the subject, defined the degree of coherence γ_{12} of two sources as equal to the visibility of the fringes obtained in the most favourable circumstances using light from these sources. It has been shown that the visibility of the fringes obtained in Young's experiment depends on the width of the slit S_1, and the following mathematical relation has been derived:

$$\gamma_{12} = \frac{\sin (2\pi ad/l_1)}{2\pi ad/l_1}, \qquad (10)$$

in which d is the width of the slit S_1 (Figure 4). If d is gradually increased from zero, γ_{12} falls from one (for d equal to zero) to zero for d $= l_1 a/\lambda$, a value equal to the separation of the fringes when $l_1 = l_2$. When the width d is further increased, fringes are again seen but they are of low visibility and are reversed (*i.e.*, there is now a dark fringe in the centre).

For the case of a slit source being inaccessible for measurement, its angular width (d/l_1) can be determined by measuring the visibility of the fringes while a, the separation between P_1 and P_2, is varied. Michelson used this method to obtain the angular diameter d of a star (serving as the slit source) from measurements of the visibility of interference fringes formed in the focal plane of a telescope that receives light from two small mirrors mounted in front of the telescope's objective, separated from each other by a distance a.

The concept of coherence that has been applied to light from two pinholes may be extended to a beam of light considered as a whole. A roughly parallel beam of light is incident on a thin sheet of metal normal to the direction of propagation. Then two pinholes may be made in the sheet at A and B (Figure 9) and the visibility of the resulting fringes measured so as to obtain the mutual coherence γ_{AB}. If A and B are initially coincident and are slowly separated then γ_{AB} falls gradually from one to zero. It is possible to define a region of coherence around any point A such that if point B lies within this region the coherence is good ($\gamma_{AB} > 0.7$). Similarly, by devices such as that described in the next section it is possible to measure the mutual coherence between A and a point A' that is, as it were, downstream from A and to define a "coherence length" I, such that coherence is good when $AA' < l_c$. When, (1) the region of coherence extends across the whole beam of light, and (2) the coherence length is large, the beam is said to be highly coherent because the mutual coherence between any two points such as B and A' is high. What qualifies as a "large coherence length" depends on the type of source and the conditions of the experiment; ten centimetres is a large coherence length for the kind of source considered in the next section, but a well-stabilized gas laser may give a beam with a coherence length of many metres.

Coherence length

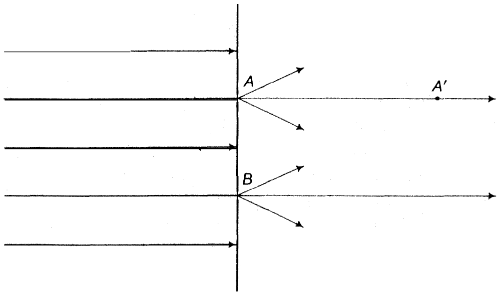


Figure 9: Two pinholes in an opaque sheet to illustrate mutual coherence between points A, A', and *B*.

In the wave equation for light already cited, the displacement and phase angle, represented by the variables ξ and ϕ, were used to specify a wave motion, but, for light, these quantities are not observable nor can they be inferred from any observations—because of the high frequency of the wave motion. The coherence γ_{12} and the phase difference, however, are observable quantities that characterize sources and beams of light. This makes them important both in theory and in practice.

Two-beam interference. In the Michelson interferometer, shown in Figure 10, the incident wave W is divided at the beam splitter BS so that part of the light is transmitted and part is reflected. After reflection at M_1 and M_2 the two parts form the wave fronts W_1 and W_2. These are copies of W, and, because corresponding points are superposed, coherence is obtained even with an extended source. The light from source S, selected by the filter at FF', is quasi-monochromatic. The plane R represents the image of M_2 that would be seen by reflection in BS. The phase differences between W_1 and W_2 are the same as if W, had been reflected from R, which is called the reference plane. M_1 may be traversed normal to itself and may also be tilted with respect to the reference plane. A compensating glass plate C having the same thickness as BS is used so that both wave fronts will pass through a total of three thicknesses of glass.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

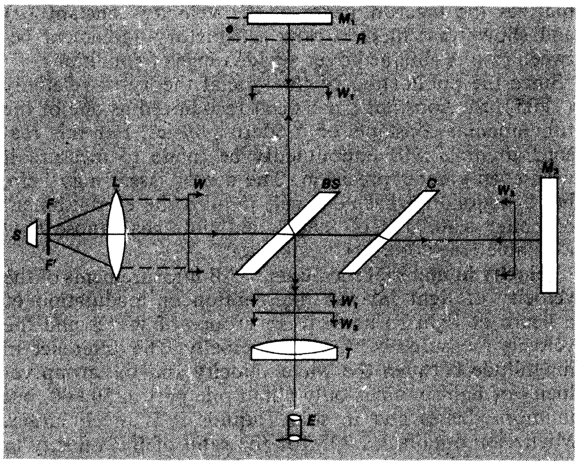


Figure 10: The Michelson interferometer.

Fringes will be formed when M_1 is adjusted to be exactly parallel to R and separated from R by a small distance e. For a hollow cone of rays, each ray will be incident on M_1 and on R at an angle 8. After passing through the instrument on their return trip, these rays will be focussed into a circular ring in focal plane FF' of the lens L. At each point on this ring two waves will be superposed and their path difference will be $2e \cos 8$. Bright rings are obtained for values of θ such that 2e $\cos \theta = p\lambda$, in which p is an integer. The appearance of these fringes is similar to that of Newton's rings (see Plate). These fringes are known as fringes of equal inclination because any one ring corresponds to a set of rays that all have the same inclination, θ, to the mirror M_1. They are conveniently observed by focussing an eyepiece E on the plane F. Because the lens T and the eyepiece E constitute a telescope focussed for an infinite distance, the fringes are said to be localized at infinity.

The apparatus may also be adjusted so that the mirror M_1 is inclined to R, the image plane of mirror M,, and nearly in coincidence with it. The incident light is rendered nearly parallel and normal to plane R. If the telescope is removed, straight line fringes can be seen by an observer who focusses his eye on the region between M_1 and R. A bright fringe is the locus of points for which $2l_p =$ ph. These fringes are called fringes of equal thickness.

Fringes of equal thickness may be formed by reflection at the two glass surfaces bounding an air film between two glass plates (Figure 11). Strictly speaking, this arrangement does not give two-beam interference because multiple reflections occur, as shown in the figure. Only the two beams A and B, however, need be considered for present purposes. Beams like B' and C', caused by multiple reflections, are weak, and unless the glass plates are fairly thin and of high optical quality, fringes formed by beams reflected from the outer surfaces of the glass plates are close together and are of poor visibility. If the arrangement is such that one of the plates is truly planar and that the other is spherical, as is the case for a convex lens lying on a glass plate, the resulting fringes of equal thickness are circles centred at the point of contact. They are known as Newton's rings (see Plate). In this situation, in which one surface is plane and the other is not, the fringes form a contour map of the nonplanar surface. They are then called contour fringes. This is a useful method for testing the flatness of a surface and determining the location of irregularities.

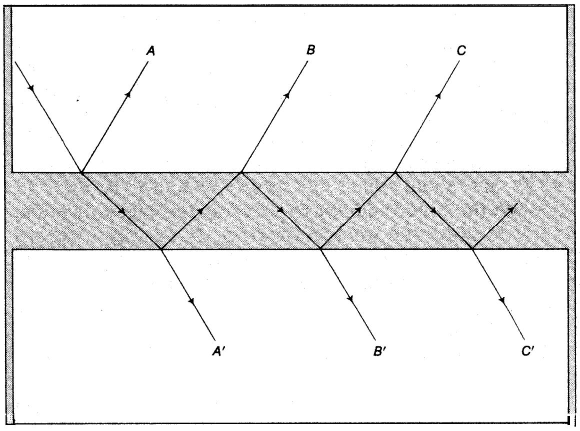


Figure 11: Interference in a thin film of air between two pieces of glass. The rays A, B, C, etc. interfere with each other, as do rays A', B', C', etc.

Newton's rings

Multiple-beam interference. If the two inner surfaces of the plates shown in Figure 11 are coated so as to make them reflect 80 percent or more of the incident light, then the resulting interference pattern will be caused by the superposition of many beams. Figure 12 shows an arrangement for producing the fringes of constant inclination by multiple beam interference. The amplitudes of successive beams are proportional to r, r^2, r^3, etc. (r is the ratio of the intensity of the reflected light to that of the incident light for one reflection). The phase differences are ϵ, 2ϵ, 3ϵ, etc., in which $\epsilon = (4\pi e \cos\theta)/\lambda$. These fringes are much sharper than those obtained with two-beam interference (see Plate).

With a large number of beams the intensity is extremely high when they are all in phase ($\epsilon = 0$), but, even when the phase difference between any two successive beams (*e.g.*, the first and the second) is quite small, the phase difference between the first and say the thirtieth beam is so large that the later beams in the series are in opposition to the earlier beams. Thus the intensity is relatively small except when the value of ϵ is close to one of the values $2p\pi$ (in which p is an integer). Multiple-beam fringes of constant inclination were used by Charles Fabry and Aifred Perot in France for resolution of spectral lines having only small differences of wavelength. Multiple-beam fringes of constant thickness have been used by an English physicist, Samuel Tolansky, to detect surface irregularities down to less than a nanometre.

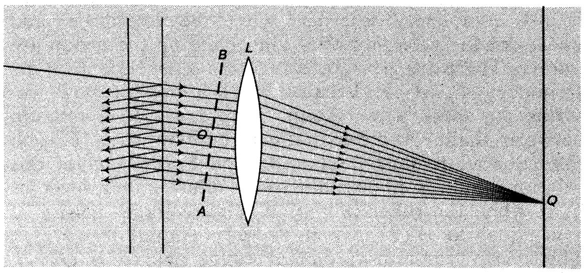


Figure 12: Multiple beam interference. Lens L concentrates all beams at focus Q with same phase differences they had while crossing a plane AB normal to *OQ*.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

Wave groups. If two pendulums that have frequencies ν_t per minute and $(\nu_t + 1)$ per minute are started together, they will gradually go out of step; after half a minute they will be moving in opposite directions and after a minute they will be together again. Over a long time they will move together once every minute. In a similar way, when two waves of slightly different frequency are moving in the same direction, they are sometimes in phase and sometimes out of phase so that the resultant is sometimes large and sometimes small, as shown in Figure 13. Two waves may be considered for which the spatial frequencies are ν_s and (v, + $\Delta\nu_s$) and temporal frequencies ν_t and $(\nu_t + \Delta\nu_t)$. The fluctuation represented by the envelope (dotted line in Figure 13) is called the beat wave. It has a temporal frequency equal to the difference $(\Delta\nu_t)$ of the temporal frequencies of the constituent waves and a spatial frequency $(\Delta\nu_s)$ equal to the difference of the spatial frequencies. It is therefore propagated with a velocity $U = \Delta\nu_t/\Delta\nu_s$. Many physical problems involve groups of waves that include a range of frequencies. It is found that, even in a dispersive medium, a group is propagated over a considerable distance as a recognizable unit. The velocity of this recognizable group is $U = d\nu_t/d\nu_s$.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

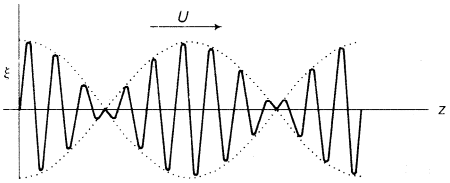


Figure 13: A simple beat wave of amplitude ξ moving with velocity U (see text).

Velocity of groups of waves

There is a certain kind of wave group for which the variation of the displacement with distance (z) along the path of propagation may be represented by the expression $h(z) \cos \nu_0 z$, in which $h(z)$ is a function that varies with z much more slowly than $\cos \nu_0 z$—*e.g.*, in Figure 13, $h(z)$ would be the function represented by the dotted line, and v, is the spatial frequency of the individual waves represented by the full line. These waves are called modulated waves. If $h(z)$ varies extremely slowly with z, the modulated wave is quasi-monochromatic in the sense described above; *i.e.*, it departs little (over distances long enough to contain many wavelengths) from a monochromatic wave. A modulated wave is completely described when $h(z)$ and ν_0 are known. It is also completely described when the amplitudes and phases of the various waves that make up the group are known. These are given by a function α dependent on the frequency ν_s, $\alpha(\nu_s)$. Because $h(z)$ and $\alpha(\nu)$ both describe the same wave group, there must be a relation between them. A mathematical theorem of a French mathematician, Jean-Baptiste-Joseph Fourier, gives this relation, making it possible to calculate either $h(z)$ or $\alpha(\nu)$ when the other is known. The average density at z is equal to $W(z)$, which is proportional to $(h[z])^2$. The energy per unit frequency range near ν_s is $G(\nu_s)$, which is proportional to $(\alpha[\nu_s])^2$.

When $h(z)$ varies very slowly with z, $\alpha(\nu)$ is large for a range of ν close to ν_0 and falls rapidly to near zero outside this range, as shown in Figure 14B. If this range in which $G(\nu_s)$ is large is ν_R, and if l_R represents a range of z over which $h(z)$ varies very little (as shown in the Figure), then it is found that ν_R and l_R are inversely proportional to one another and that their product is of the order of magnitude of unity. This represents the fact that the longer the wave train the more closely its properties agree with those of the ideal monochromatic wave, which is infinitely long and has a precisely defined frequency.

From (A) R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

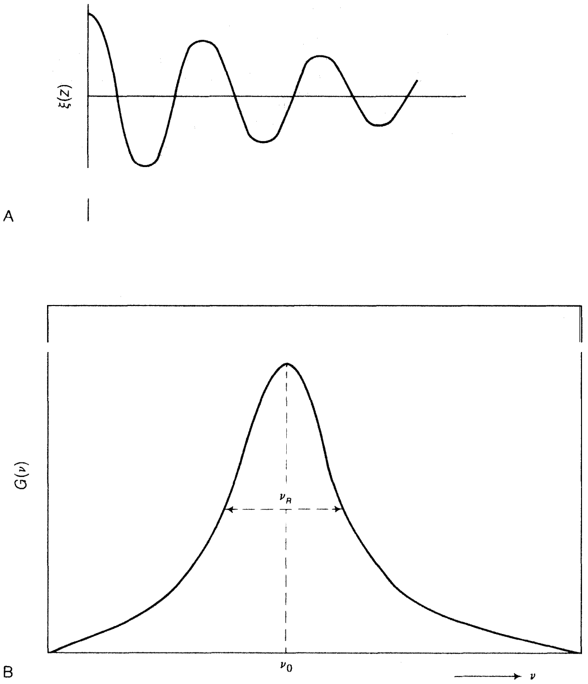


Figure 14: Damped waves.
(A) Amplitude, $\xi(z)$, as a function of distance, z. (B) Energy, $G(\nu)$, as a function of frequency, ν. The figure shows strong damping. For light emitted by free atoms, l_R would encompass 10' waves or more and ν_R would be correspondingly smaller (see text).

Undisturbed atoms emit exponentially damped waves the length of which is usually 10^7 waves or more so that ν_R is a small fraction of ν_0. Collisions increase the damping by a factor that is proportional to the pressure. The observed radiation is also modified by the Doppler–Fizeau effect, because the atoms that emit the light do not all have the same velocity. This increases the range $\Delta\nu_R$. Even when the effects of collision damping and Doppler–Fizeau effect are combined, the value of l_R for the wave trains emitted in low pressure electrical discharges is still about 10^5 wavelengths, and most of the energy is confined within a frequency range of order 10^{-5} ν_0 (corresponding to a wavelength range of less than 0.01 nanometre). These quasi-monochromatic waves are called wave groups. The light emitted by high-pressure lamps or by luminescent solids extends over a much wider range of frequency, and wave-group theory has little useful application to problems concerning non-monochromatic light from these sources.

Propagation of wave groups from stars

Wave groups in a dispersive medium. In vacuum, all components of the group have the same phase velocity, and therefore the phase relations between different members of the group are constant. A group advances as a unit without any change of the modulation function $h(z)$. In a dispersive medium, the phase relations change and $h(z)$ changes as the wave train advances, but this change is slower than might be expected. Over considerable distances the group is propagated as a recognizable whole with the group velocity U. The change of of $h(z)$ is small for passage through a gas and also for groups (for which ν_R is small) that represent sharp spectral lines. Thus these wave groups are propagated virtually unchanged through an optical instrument or, if they arrive from the Sun or stars, through the Earth's atmosphere.

In Young's experiment, Figure 4, the fringes have maximum visibility at the position O, corresponding to zero path difference and zero phase difference for all wavelengths. If a thin sheet of mica is inserted in front of slit P_1, the centre (or position of maximum visibility) is displaced upward. It was at one time thought that the new centre would be found at the point corresponding to zero phase difference for the mean wavelength in the wave group—*i.e.,* the point calculated for equal times from slits P_1 and P_2 allowing for the fact that the phase velocity in mica is less than that in air. This did not agree with experimental observation. It was found that the new centre is situated at the position where the times from P_1 and P_2 are equal when the group velocity is used to calculate the time required to traverse the piece of mica. At this position the wave train from P_1 exactly overlaps the wave train from P_2. At any other position, part of each wave train cannot take part in the interference because it does not coincide with any part of the other wave train. This light that cannot interfere forms a uniform background and so reduces the visibility of the interference fringes.

If white light is used and a fairly thick piece of mica is inserted, no fringes are obtained. This is because the wave train has changed shape so much in passing through the mica that it can no longer match the wave train that has travelled through air.

Michelson, using the apparatus shown in Figure 10, studied the decrease in visibility of interference fringes as the path difference between the two wave trains is increased. The reduction in visibility as the path difference Increases may be assigned either (1) to the fact that the parts of the wave trains that overlap are decreasing or (2) to the increasing difference between the positions of the bright fringes for different wavelengths in the group. As was seen above, the length of the wave trains and the range of the wavelength are inevitably linked, and so these alternatives (1) and (2) do not constitute two different theories. They are just two different ways of visualizing how wave trains (or wave groups) interfere.

DIFFRACTION

Theory of diffraction. Huygens assumed that every point on a wave front may be regarded as a source of spherical wavelets the envelope of which is the position of the wave front at a later time. Huygens was thus able to account for rectilinear propagation and for the laws of reflection and refraction. Fresnel added the hypothesis that the wavelets can interfere and this led to a theory of diffraction. Figure 15 shows how a coherent, monochro-

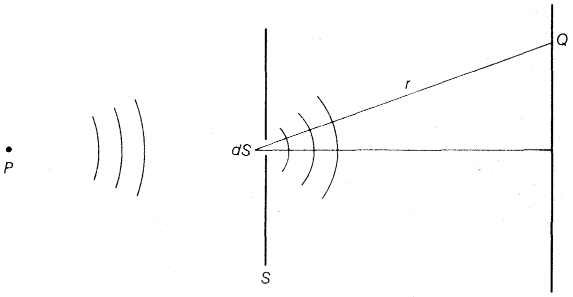
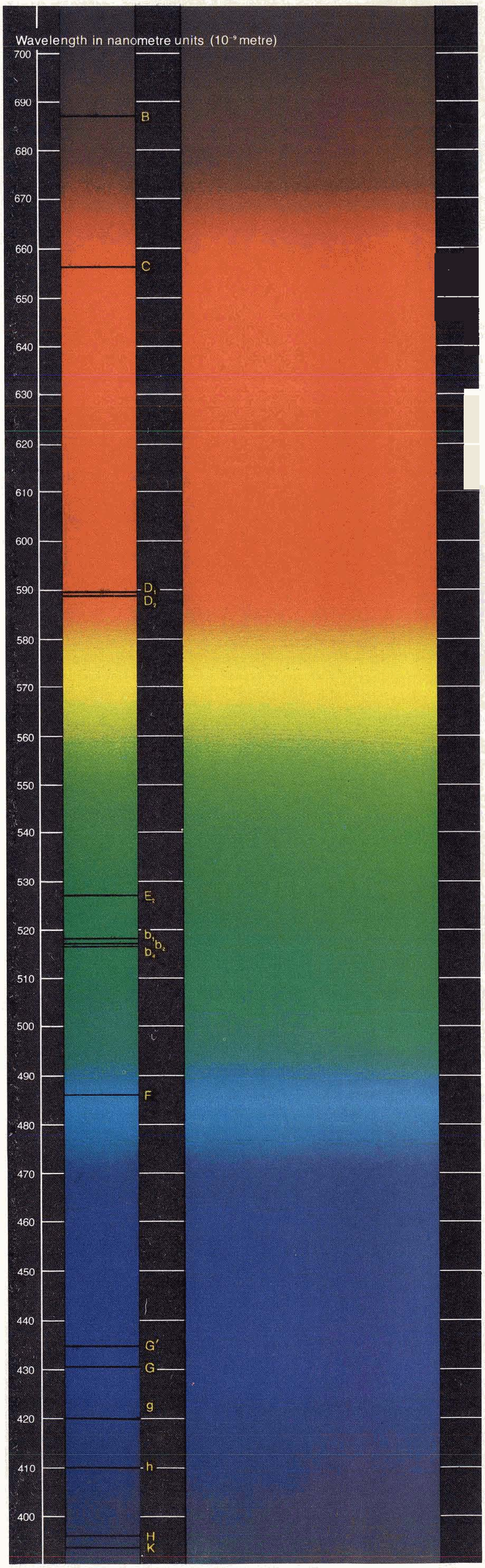
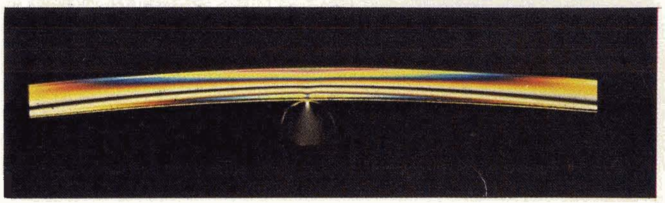


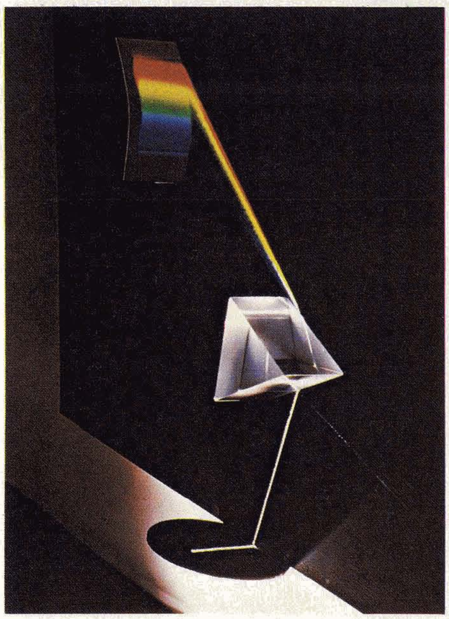
Figure 15: Principle of Fresnel's theory of diffraction (see text).



The visible solar spectrum (simulated), showing prominent Fraunhofer absorption lines.



Bending of transparent **plastic** bar between crossed **Polaroids** to show photoelastic strain patterns.



Spectrum of **white light** by (above) a **prism** and (below) a **diffraction grating.** **With a prism,** the red end of the spectrum **is** more compressed than the **violet end.**

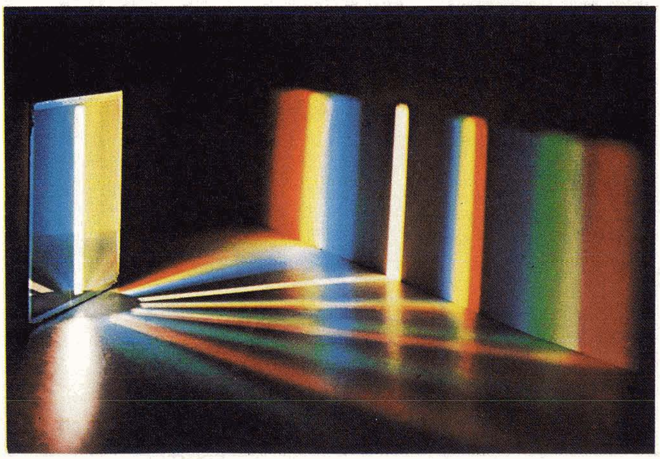
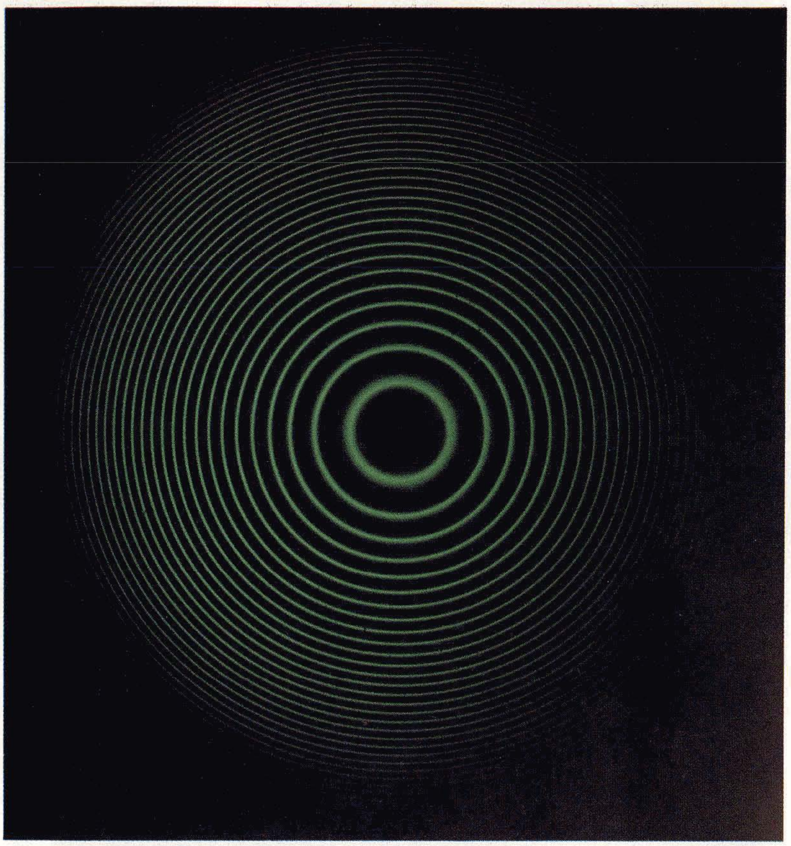


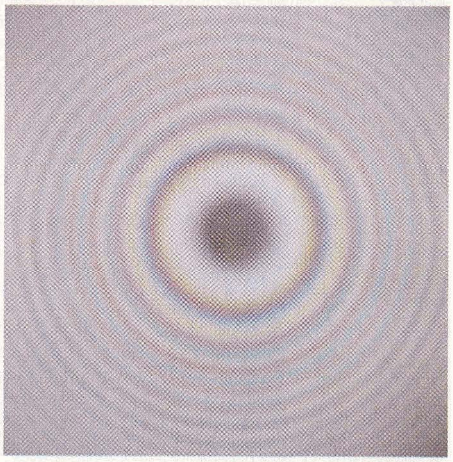
Plate 1: By courtesy of (bottom) Bausch & Lomb, Inc., Rochester, N.Y.; photographs, (left) EB Inc., (top) Charles D. Reilly—EB Inc., (centre) Fritz Goro—Life Magazine © 1944, Time Inc.

Plate 2 Light



Fabry-Pérot fringes.

Newton's rings.



Interference patterns obtained with crystals in convergent plane-polarized light (polarizer and analyzer crossed): (left) uniaxial crystal cut normal to the axis; (right) biaxial crystal cut perpendicular to bisector of the axes.

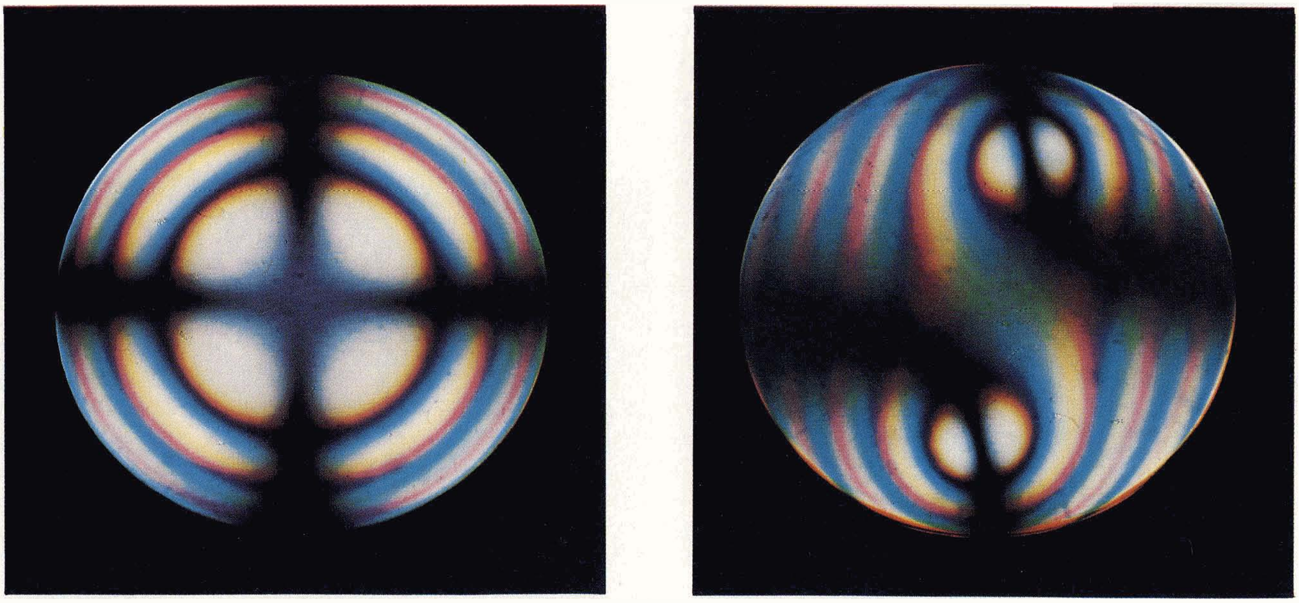


Plate 2: By courtesy of (top left, bottom left, bottom right) R.W. Ditchburn and Blackie 6 Son Limited, Glasgow and London; photograph, (top right) Charles D Reilly—EB Inc

matic wave from a point source P falls on the screen S, which is opaque except for an aperture dS. Fresnel assumed that the amplitude ($d\xi_Q$) of the wavelet at Q, originating from a small area dS, is:

$$d\xi_Q = \alpha \frac{A}{r} f(\chi) dS, \tag{11}$$

in which A is the amplitude of the incident wave, α is a constant, r is the distance from dS to Q, and $f(\chi)$ is a function of χ, the inclination factor; this factor was introduced by Fresnel because he believed that the effect of the element dS would be greater in the forward direction ($\chi = 0$) than in an inclined direction. The total effect at Q was obtained by superposing the wavelets from all parts of the aperture, allowing for phase differences caused by a variation of r and also for variation of the inclination factor, $f(\chi)$. Fresnel developed an ingenious method of dividing S into a series of zones of equal area and calculating the total effect as the sum of a simple series. This method applies only to circular apertures and obstacles and then only to points on the axis of symmetry, but Fresnel also developed integrals that are more generally applicable.

Fresnel zones

Fresnel predicted that there should be a bright spot at the centre of the shadow of a circular obstacle. The experimental verification of this unexpected result gave confidence in Fresnel's wave theory of diffraction.

Fraunhofer diffraction. When the source and pattern screen are sufficiently far from the slit, the phase differences corresponding to different parts dS of the slit opening vary linearly with x and y coordinates in the plane of the aperture (Figure 16). This situation is obtained when two spherical lenses L_1 and L_2 are introduced with source P at the focus of L_1 and Q in the focal plane of L_2. Spherical waves emanating at the focus of a lens are rendered plane wherever they encounter the lens. Plane waves are made spherical by a lens. They have the same radius of curvature as the focal length of the lens. The wave falling on S is a plane wave, and the total effect at Q may be regarded as caused by a plane wave leaving S. The same result is obtained if L_1 and L_2 are replaced by a single lens (situated near S) that forms an image of P at Q_0. This is known as far-field diffraction, or Fraunhofer diffraction, and is thus distinguished from near-field, or Fresnel, diffraction. It should be understood, however, that there is only one physical theory of diffraction that is derived from the ideas of Huygens and Fresnel. Fraunhofer diffraction is of great practical importance especially in regard to the performance of optical instruments.

From R.W. Dtchburn. *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, inc.

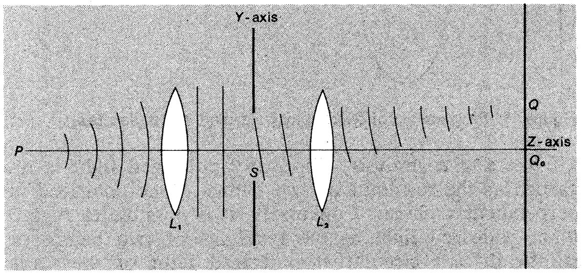


Figure 16: Arrangement for Fraunhofer (far-field) diffraction. The opening at S diffracts light from source P onto plane Q (see text).

Groups of waves with different directions. When a plane wave is incident upon a slit as shown in Figure 16 (so that its width is limited), the emergent light may be represented by a group of plane waves. All the waves of this group have the same spatial frequency but differ in regard to direction of propagation. It is possible to define a range of angles (in the plane of the page) within which most of the light is found. If this range is θ_R and the width of the slit is w, then it is found that sin θ_R is inversely proportional to w; *i.e.*, the narrower the slit, the greater is the angular spread—sin θ_R is roughly equal to λ/d. Using Fourier's theorem it is possible to derive an equation that gives the amplitude and phase of the light diffracted in any direction as a function of the width of the slit. Extension of the calculation to diffraction by apertures or obstacles of any shape involves more lengthy mathematics but no new physical principle.

Angular power spectrum. The energy diffracted in any direction is proportional to the square of the corresponding amplitude. This energy expressed as a function of the angles that define the direction is called the angular power spectrum. It may be measured and is found to agree with that calculated when the width of the slit (or, more generally, the shape and size of the apertures) is known. There are many problems in which it is desired to carry out an inverse calculation—*i.e.*, to calculate the shape and size of the apertures from measurements of the angular power spectrum. Unfortunately, this is not, in general, possible because measurement of the angular power spectrum does not give the phase of the diffracted light. It is found, however, that measurement of the angular power spectrum yields a function (known as the auto-correlation function) of the size and shape of the obstacles or apertures responsible for the diffraction. In X-ray analysis, this function gives important information about symmetry. A complete picture of the crystal may often be obtained by combining calculations from the angular power spectrum with information derived from other sources.

It is found that when diffraction is due to a number of apertures (or obstacles) that are similar in size, shape, and orientation, the angular power spectrum (G) is the product of two factors, F and f, in which F (called the form factor) depends only on the properties of the individual aperture and f (called the structure factor) depends only on the arrangement or spacing of the elements. When the apertures are irregularly arranged, f is just equal to N (the number of apertures). Thus the diffraction halos produced by an irregular distribution of small similar objects have the same intensity distribution as the pattern for a single particle. This principle is used in a device called an eriometer to determine the size of blood corpuscles and may also be used to calculate the average size of the small particles that cause a halo around the Moon.

Diffraction by similar apertures

When N similar elements are arranged in a regular pattern, the structure factor may vary from zero to N^2. A diffraction grating (a plate having parallel lines engraved across its surface) with N lines is such a pattern, and, for any of the directions θ_p defined by equation (6), the light from all elements (lines) is in phase; the amplitude is N times that given by a single element, and thus the energy and the structure factors are proportional to the total number of lines squared—*i.e.*, $f = N^2$.

Limits of resolution. Diffraction spreads the light in optical images; so that if two objects are too close to each other, the gap between them cannot be distinguished. The distribution of intensity with radius in the image of a point source is shown in Figure 17. Rayleigh showed, theoretically and experimentally, that the images of two point sources are just resolved when their separation is such that the centre of the pattern due to one image falls on the first minimum of the pattern due to the other (Figure 18). This implies that a telescope with a perfect objective of diameter D can just resolve two stars whose

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

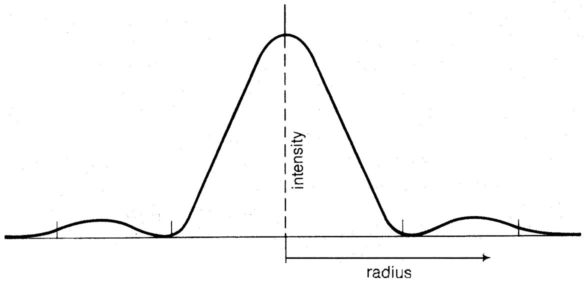


Figure 17: Illumination of a point source image modified by diffraction, shown as the variation of intensity with radius.

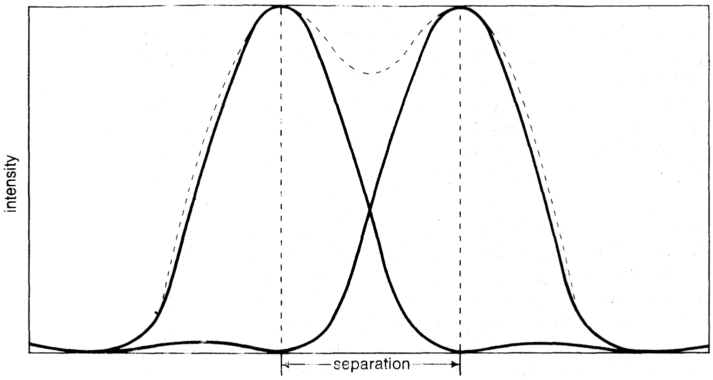


Figure 18: Overlapping images of two point sources. Full lines show how intensity varies with distance from a separate source, dashed line shows combined intensity.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

angular separation is 1.2 λ/D. Qualitatively, this agrees with a calculation that shows that most of the energy in the diffraction pattern of an aperture of width d lies within an angular range $\pm\lambda/d$.

The angular separation of the maxima resulting from light of two wavelengths λ and λ + AX in a spectrum formed by a diffraction grating is obtained by differentiating equation (6), resulting in $\Delta\theta = p\Delta\lambda/e$ cos θ. These maxima are just resolved if Ae = λ/M, in which M is the width of the beam diffracted by the grating. For a grating of N lines, this width is Ne cos θ, and the resolving power R = $\lambda/\Delta\lambda = pN$. A grating ten inches wide, for example, with 10^4 lines per inch and p = 10, has a resolving power R = 10^6.

The limit of resolution for a microscope depends on conditions of illumination and is at best about half a wavelength ($\lambda/2$), or about 250 nanometres for visible light.

III. Polarization and electromagnetic theory

POLARIZED LIGHT

Inter-
ference of
polarized
light

Interaction of plane-polarized beams. Fresnel and Arago, using an apparatus based on Young's experiment (Figure 4), investigated the conditions under which two beams of plane polarized light may produce interference fringes. They found that: (1) two beams polarized in mutually perpendicular planes never yield fringes; (2) two beams polarized in the same plane interfere and produce fringes, under the same conditions as two similar beams of unpolarized light, provided that they are derived from the same beam of polarized light or from the same component of a beam of unpolarized light; (3) two beams of polarized light, derived from perpendicular components of the same beam of unpolarized light and afterwards rotated into the same plane (*e.g.*, by using some device such as an optically active plate) do not interfere under any conditions.

Result (1) is to be expected because two displacements in perpendicular planes cannot annul one another, and result (2) is also easily understood. Result (3) shows that mutually perpendicular components of unpolarized light in a beam are non-coherent. Their phase difference varies in time in an irregular way. Unpolarized light has a randomness, or lack of order, as compared with polarized light (implying an entropy difference). This order (or lack of order), rather than the azimuthal property, is the most fundamental difference between polarized and unpolarized light. Perfectly monochromatic light is perfectly coherent and completely polarized.

Superposition of polarized beams. Two coherent beams of plane polarized light may be thought of as propagated in the *Oz* direction, one with its vector along Ox and the other with the electric vector along *Oy*; *i.e.*, the two vibrations are at right angles to each other as well as to the direction of propagation (Figure 19). If the beams have amplitudes *a*, and a_y and phases ε_x and ε_y, then, in general, the resultant vibration (R_1, R_2, and R,) may be represented in magnitude and polarization by a vector, or arrow, the tail of which touches the axis of propagation *Oz* while the point moves round the ellipse (Figure 19). It goes round once when the phase angle ϕ (see equation [1]) changes by 2π—*i.e.*, at any given place when *t* changes by y or for any one time when *z* changes by λ. The beam is said to be elliptically polarized. If the phase difference is $\pi/2$, then the axes of the ellipse are equal to *a*, and a, and are along O_x and O,.

Elliptically polarized light may be regarded as the most general type of polarized light. If the amplitudes of the two waves are equal, *a*, = *a*,, and the phase difference is still $\pi/2$, the ellipse becomes a circle and the light is said to be circularly polarized. If the phase difference ε_{xy} is not equal to $\pi/2$, the resultant is still elliptically polarized light, but the axes of the ellipse no longer coincide with the axes of coordinates. If the phase difference $\varepsilon_{xy} = 0$ or π, the ellipse shrinks to a straight line and the light is plane-polarized. If the representative vector, when viewed by an observer who receives the light, rotates in a clockwise direction, the light is said to be right-handed (or positive) elliptically polarized light. The opposite sense of rotation corresponds to left-handed (or negative) elliptically polarized light.

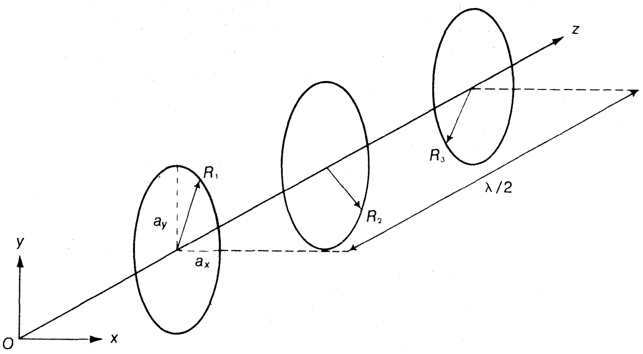


Figure 19: Progression of elliptically polarized wave (see text).

In the above analysis, elliptically polarized light is regarded as the resultant of two beams plane-polarized in perpendicular planes. Conversely, it is possible to regard plane-polarized light as the resultant of two beams of elliptically (or circularly) polarized light of the same wavelength, provided that the ellipses are similar in orientation and eccentricity, but one beam is right-handed and the other left-handed.

Double refraction. In the 17th century Bartholin showed that a ray of unpolarized light incident on a plate of calcite, unlike glass or water, is split into two rays, as shown in Figure 20. One ray, called the ordinary ray, is in the plane containing the incident ray and the normal to the surface. If the angle of incidence is varied, this ray is found to obey Snell's law of sines, equation (3). The other ray, called the extraordinary ray, is not in general coplanar with the incident ray and the normal; also, for it, the ratio of sines is not constant. The fact that Snell's law is not obeyed in certain directions implies that the velocity of light in such a medium, called anisotropic, depends on the direction of travel in it. The two rays are polarized in mutually perpendicular planes. This is known as double refraction, or birefringence.

Bire-
fringence

In order to apply Huygens' method of constructing wave fronts (see above Theory of diffraction), it is necessary to assume that, in an anisotropic medium, the wave surface from a point source consists of two sheets, or surfaces (Figure 21). The observation that one ray obeys both laws of refraction implies that one sheet must be a sphere—like the wave surface in an isotropic medium. Huygens assumed that the other sheet is an ellipsoid of revolution that touches the sphere either internally (Figure 21A) or externally (Figure 21B). There is one velocity of propagation for the direction defined by the line through the two points of contact (called the optic axis) and two velocities for any other direction corresponding to light polarized in two mutually perpendicular planes.

Crystals for which the wave surface has the form shown

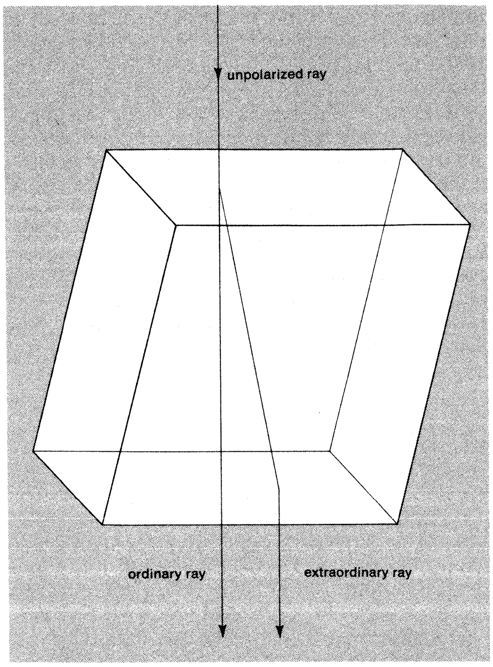


Figure 20: Double refraction showing two rays emerging when a single light ray strikes a calcite crystal at right angles to one face (see text).

in Figure 21A are said to be positive uniaxial crystals, and Figure 21B refers in a similar way to negative uniaxial crystals. Huygens thought all crystals were uniaxial, but later observations showed that the general form of the wave surface is more elaborate and is biaxial.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc , by permission of John Wiley & Sons, Inc.

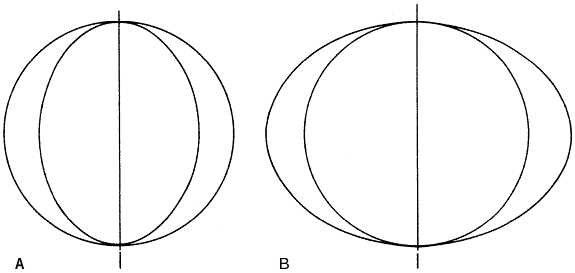


Figure 21: Wave surface for (A) positive and (B) negative uniaxial crystals.

When a parallel beam of plane-polarized light is incident normally upon a thin crystal plate and the crystal is rotated about an axis normal to the plate, two orientations will be found in which a single beam of plane-polarized light emerges; for one orientation, the light is called the ordinary ray, and for the other orientation, the extraordinary ray. Two lines may be drawn on the plate (or on its mount) to indicate the direction of the electric vector of the incident beam when the orientation is such that plane-polarized light emerges. The directions of these lines are called privileged directions for the given anisotropic plate; they are perpendicular to one another.

When light waves pass through an isotropic plate of thickness d, the phase change is $2\pi nd/\lambda$ (n is the index of refraction). For an anisotropic plate there are two indices, n_o and n_e, corresponding to the two privileged orientations, and there are corresponding phase changes, $2\pi n_o d/\lambda$ and $2\pi n_e d/\lambda$. If a beam of light, polarized in some plane other than one of the two privileged orientations, is incident on the plate, it may be resolved into two wave components (lying in two mutually perpendicular planes) that emerge with a phase difference ε_{oe} equal to $2\pi(n_e - n_o)d/\lambda$. The phase difference ε_{oe} is called the retardation of the plate. When the retardation ε_{oe} is equal to $\pm\,\pi/2$, the plate is called a quarter-wave plate, and when ε_{oe} is equal to $\pm\,\pi$, it is called a half-wave plate.

The maximum difference of $(n_e - n_o)$ for calcite, a crystal that is strongly birefringent, is about 0.17. Quarter-wave plates are often made by splitting sheets of mica. Inasmuch as the relevant difference of indices is about 0.004, a mica quarter-wave plate is about 60 wavelengths, or 0.03 millimetre thick.

When a parallel beam of plane-polarized light is incident normally on a quarter-wave plate, the emergent light is circularly polarized if the plane of the electric vector bisects the angle between the privileged directions. Otherwise the emergent light is elliptically polarized, the axes of the ellipse being parallel to the privileged directions of the quarter-wave plate. In a similar way, a suitably oriented quarter-wave plate may be used to convert elliptically polarized light to plane-polarized light.

Induced birefringence effects

Many substances that are normally isotropic become weakly birefringent when subjected to shear stress (photoelastic effect; see Plate). Birefringence is also induced by an applied electric field (Kerr effect). A restricted class of crystals that are normally birefringent shows a large change in birefringence when an electric field is applied. Weak birefringence is produced by an applied magnetic field.

Production of polarized light. Unpolarized light may be separated into two components, polarized in perpendicular planes, either (1) by reflection or (2) by refraction. A single reflection at the polarizing angle produces a well-polarized reflected beam of low intensity (*i.e.*, most of the waves are polarized in one plane) and a strong transmitted beam that is only partially polarized. A pile of glass plates provides many surfaces for reflection and thus gives a much stronger reflected beam that is still well polarized; it also considerably improves the degree of polarization of a transmitted beam.

A Nicol prism (Figure 22) is made by cutting a calcite crystal in a suitable plane relative to the crystal axes and cementing the two parts together with Canada balsam. The critical angle for total reflection is less for the ordinary ray than for the extraordinary ray. The crystal is cut at an angle (relative to the crystal axes) such that, for a cone of incident rays of divergence about 24°, the ordinary ray is totally reflected, whereas the extraordinary ray is transmitted. Both rays are almost completely polarized. A prism of this type with air instead of Canada balsam as the separation film was devised by Foucault for polarization of ultraviolet radiation.

From R.W. Ditchburn, *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

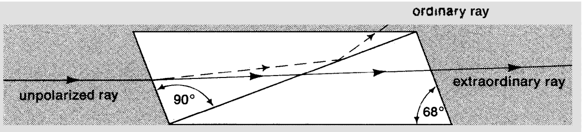


Figure 22: Nicol prism (see text).

Many crystalline substances have different absorption coefficients for the ordinary and extraordinary rays, but the difference is not large enough to provide a useful method of polarization. A series of artificial materials that polarize by absorption has been developed. Complex crystals are produced in a plastic matrix and oriented by stretching. One material of this type, called Polaroid H, transmits about 80 percent of plane-polarized light with the electric vector in one plane and less than 1 percent if the vector is in a perpendicular plane. This type of film offers the most convenient and inexpensive method of obtaining an intense beam of light that is nearly 99 percent plane-polarized.

Analysis of polarized light. Any polarizer has a preferred plane such that if an incident plane-polarized beam has the electric vector in this plane, a high intensity is passed, whereas the device rejects light polarized with the electric vector in a plane perpendicular to the preferred plane. This property implies that any polarizer may be used as an analyzer to test whether or not a given beam of light is plane-polarized. The analyzer is rotated in the beam and if the intensity I of the transmitted light is zero for some orientation, the beam is plane-polarized with the electric vector in a plane perpendicular to that for which the intensity is zero. If there is a position for which the analyzer gives a non-zero minimum, then either (1) the light is elliptically polarized or (2) it is partially polarized. Also if the intensity is not affected by rotating the analyzer, then the light is either (3) unpolarized or (4) circularly polarized. These ambiguities between (1) or (2), and (3) or (4), may be removed and the analysis of a mixture may be completed by using an analyzer and a quarter-wave plate together.

Stokes parameters and Poincairé sphere. Any mixture of different kinds of polarized light may be reduced to an equivalent amount of one kind of polarized light characterized by the orientation and length of axes of an ellipse. It may, of course, happen that the ellipse degenerates to a straight line or a circle. An alternative way of characterizing polarized light is by means of the Stokes parameters, named after a British physicist, George Gabriel Stokes, and defined as follows:

$$S_0 = \langle a_x^2\rangle + \langle a_y^2\rangle; \qquad S_1 = \langle a_x^2\rangle - \langle a_y^2\rangle;$$
$$S_2 = 2\langle a_x a_y \cos \varepsilon_{xy}\rangle; \qquad S_3 = 2\langle a_x a_y \sin \varepsilon_{xy}\rangle,$$

in which the polarized light is regarded as the resultant of two beams polarized with vectors *a*, and *a*, along the coordinates Ox and Oy, and ε_{xy} is the phase difference as in the section above. The brackets <> indicate time averages. If the light is completely polarized, then the square of S_0 is equal to the sum of the squares of the other three parameters. This relation was used by a French mathematician, Henri Poincairé, to produce an elegant representation of the properties of polarized light by means of a sphere, which is known as the Poincairé sphere. Each point on the sphere represents a different kind of polarized light; *e.g.*, the two poles represent right- and left-handed circularly polarized light, points on the equator represent plane polarized light; other points on the sphere represent different kinds of elliptically polarized light. The Stokes parameters constitute a matrix representation of polarized light. This kind of representation makes available certain powerful mathematical methods (which have been developed in other connections) for calculating how the state of a beam of polarized light changes on reflection or on passing through a crystal.

Colours of thin plates. When a parallel beam of white light (I_0) is passed through a thin slice of crystal placed between a polarizer and an analyzer, the intensity *I* of the emergent light may be expressed as the sum of two terms:

I = white term + colour term.

The white term represents the amount of light that would be transmitted if the crystal were not present. It is equal to I,, $\cos^2$ e, in which θ is the angle between the preferred directions of polarizer and analyzer. The colour term depends on the relation between the directions of the crystal axes and the preferred directions of polarizer and analyzer. It is also dependent on the retardation ($\varepsilon_{oe} = 2\pi d[n, - n_o]/\lambda$) of the plate. Because this retardation varies with the wavelength, the transmission depends on the colour. Some parts of the spectrum are transmitted more than others, and the emergent light is coloured. The first term, on the contrary, does not depend on the properties of the crystal, and it represents white light. It can be reduced to zero by setting the analyzer and polarizer so that their preferred directions are at right angles ($\theta = \pi/2$). The colour is then obtained most strongly, because any transmission of light results from anisotropy (*i.e.*, birefringence) in the slice. This provides a sensitive method of detecting strains in glass and may also be used to detect induced anisotropy when mechanical or electrical stress is applied. This method is sometimes used to detect regions of large strain in beams or girders. A model of the beam is made in plastic and placed between polarizer and analyzer. On applying a load, fringes appear in the regions of most strain. The Plate shows a picture obtained in this way. The example is chosen for simplicity. The useful application is to more complicated structures for which calculation is laborious.

A high-speed optical shutter called the Kerr shutter can be made by placing an isotropic material between crossed polaroids, one acting as a polarizer and the other as an analyzer, and applying an electric field. If a suitable material is used, the intensity of the transmitted light will reproduce the variation of electric field, even at frequencies over a megahertz. An improved shutter is produced by using the change of birefringence in some crystals.

The Kerr
shutter

Rings and brushes. The retardation of waves by a crystal plate depends on the relation between the perpendicular to the surface of the plate and the axes of the crystal from which it was cut. Observation of the colours of one plate with parallel light gives information in relation to one direction within the crystal. It is possible to obtain information about many directions within the crystal at one and the same time. This is done by taking the light that has passed through the polarizer and converging it strongly, using a microscope objective (reversed). It then passes through the crystal as a solid cone of light. It is rendered nearly parallel again by a second objective and passes through the analyzer. Finally it passes through an optical system that brings all rays that have passed through the crystal in one direction to a focus at a single point in a certain plane (a similar focussing is shown in Figure 12). Thus each point in this plane corresponds to a certain direction in the crystal and to a single value of retardation ε_{oe}. The differences in retardation ε_{oe} are caused partly by variation of (n_o — n,) with the direction in the crystal and, usually to a lesser extent, by a difference in the length of path within the crystal. This implies that, with monochromatic light, bright and dark fringes are observed. Each dark fringe is the locus of points for which the retardation is an odd number multiple of 180°—*i.e.*, ε_{oe} equals $p_0\pi$ (p_0 being an odd integer). If the slice is cut from a uniaxial crystal with its normal in the direction of the optic axis, the symmetry of the wave surface (Figure 21) requires that the fringes be circular rings (see Plate). The rings are interrupted at certain points for which the second term (in the expression of the intensity of the light transmitted by a crystal plate placed between polarizer and analyzer) is zero. This gives either a bright cross or a dark cross depending on the angle between the preferred directions of the polarizer and the analyzer; these crosses are known as brushes. The patterns obtained with a slice cut so that the normal is not in the direction of the optic axis (a crystal direction for $n_o = n_e$) or with biaxial crystals are complicated and sometimes beautiful (see Plate). Such patterns may be used to identify crystals present as fairly small inclusions in other materials.

ELECTROMAGNETIC-WAVE CHARACTER OF LIGHT

Maxwell's equations. Historically the theory of electricity and magnetism developed in the form of a number of empirical laws each of which was a generalization based on a series of experiments; *e.g.*, Coulomb's law dealt with the force between two stationary electric charges. Maxwell replaced all these laws by a single theory concisely stated in the form of a set of vector equations. It has been said that Maxwell's theory is Maxwell's equations, and indeed it is impossible to do justice to Maxwell's achievement without use of these equations. The content of these equations and their relevance to the theory of light will be described here in general terms. The reader may refer to standard texts on electromagnetic theory for the equations themselves.

Electric and magnetic fields are specified by means of the vectors E and H with which are associated the vectors D, B, and *J* (electric and magnetic induction and density of electric current). Maxwell's equations fall into two groups: (1) three constitutive equations and (2) four field equations.

Constitutive and
field
equations

All material bodies contain electrons. These are negative charges circulating around heavier nuclei that are positively charged. When an electric field is applied to a material body, the average positions of the negative charges relative to the positions of the positive charges are changed. This creates an internal electric field. Similarly the action of a magnetic field on a material changes the movement of the electrons and sets up an internal magnetic field. The constitutive equations state that effects within a material body are proportional to the applied fields so that the resultant fields within the body are proportional to the applied field. Certain constants are defined: E is the dielectric constant, μ is the magnetic permeability, and a is the electrical conductivity. There is no general agreement, however, concerning these constants, and therefore some authorities use a different nomenclature (for treatment of the constants in greater detail, see ELECTROMAGNETIC RADIATION and ELECTRICITY). But these are basic properties of the material: for free space the values are ε_0, μ_0, and zero respectively.

The first of the four field equations quantifies certain properties of the electric induction at the boundary of a volume that contains a net positive or negative charge. The second states that, since there are no free magnetic poles, a certain integral of magnetic quantities is zero. The third equation states that, when the magnetic flux through a surface changes, electrical voltages appear on the boundaries of the surface. The fourth states that electrical currents in conducting materials and changes of the electric induction in nonconducting materials produce magnetic effects.

The field equations (like the constitutive equations) are linear equations: they state that certain quantities are proportional to one another; *e.g.*, the third equation states that the electrical voltages are directly proportional to the rate of change of magnetic flux. No new constants (other than ε, μ, and σ), however, are introduced. In Maxwell's equations, electricity and magnetism are two aspects of one thing called electromagnetism. Maxwell's theory can indeed be stated in a way that does not mention electricity and magnetism separately.

Maxwell's new hypothesis

The three constitutive equations and the first three of the field equations are precise formulations of known empirical laws. In the fourth field equation Maxwell introduced a new hypothesis—that an electrical change in a nonconductor produces magnetic effects. This hypothesis—which can be verified by electrical experiments—leads to the theory of electromagnetic waves capable of being propagated through a vacuum.

Propagations of electromagnetic waves. When Maxwell's equations are combined, using standard mathematical methods, a new equation is obtained. This is similar in form to the wave equation (7) given above. It predicts the existence of electromagnetic waves, with well-defined properties which will now be described.

1. In free space (*vacuo*) electromagnetic waves are propagated with a phase velocity c $= (\varepsilon_0\mu_0)^{-1/2}$; the magnitude of c obtained from electrical measurements of $\varepsilon_0\mu_0$ is approximately equal to the measured velocity of light (see the Table). Plane waves are propagated without attenuation. Spherical waves have an amplitude inversely proportional to the distance from a small source.

2. For transparent nonconducting mediums such as water or glass, the phase velocity is b $= (\varepsilon\mu)^{-1/2}$, and the index of refraction is $n =$ c/b $= (\varepsilon_0\mu_0/\varepsilon\mu)^{1/2}$. This relation holds for radio waves and, with moderate accuracy, for infrared radiation. For visible light it holds moderately well for some mediums but not for others. The constants E and μ are associated with redistribution of charge in response to changes in the electromagnetic field. At low frequencies these movements follow the changes of the field in a simple way. At optical frequencies, when the field is reversing about 10^{14} times a second, the situation is more complicated. The effective value of μ at these frequencies is μ_0 and the effective value of E is less than the value measured with static fields, or fields that vary more slowly—apart from the possibility of resonance, which will be considered later.

3. In conducting mediums, the electromagnetic waves are absorbed as they are propagated. This absorption is exponential; *i.e.*, the fraction absorbed in a thin layer of thickness *d* is βd, in which β is a constant the value of which can be calculated when E and σ are known. For metals, β is very large so that a layer only a tenth of a micrometre in thickness absorbs about half of the incident light. This absorption is much stronger than that of nonconducting substances, which are normally regarded as opaque (such as ebony): these are usually quite transparent in thicknesses of a micrometre or so.

Although the simple theory is correct in predicting high absorption for metals in general, the values of β for different metals (calculated from E and σ) are not correct. This is because the effective value of n at high frequencies is not the same as that measured with direct current or low frequency alternating current.

4. The waves are transverse; the electric field vector (E) and the magnetic field vector (H) are perpendicular to one another and to the direction of propagation along the axis OZ as shown in Figure 23.

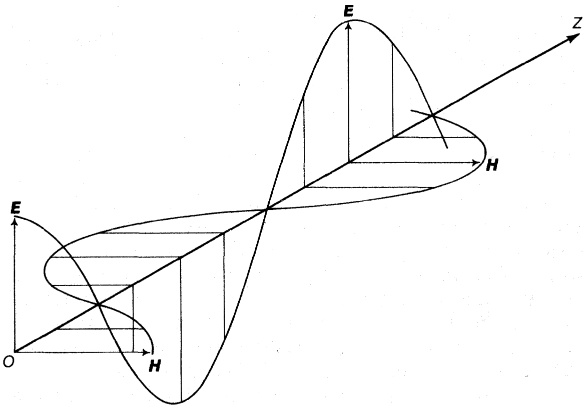


Figure 23: Electromagnetic wave, showing that electric field vector **E** and magnetic field vector **H** are in phase (see text).

5. The two quantities E and H, the magnitudes of vectors E and H, fluctuate in phase, so that when E is a maximum *H* also is a maximum (Figure **23**). This is implied in the equation

$$\varepsilon^{1/2}E = \mu^{1/2}H, \qquad (12)$$

which enables either E or H to be calculated when the other is known.

6. The electromagnetic field possesses energy, and the energy per unit volume (W) is equal to ½($\varepsilon E^2 + \mu H^2$). In view of equation (12) above, this implies that for electromagnetic waves,

$$W = \varepsilon E^2. \qquad (13)$$

The Poynting vector

7. The propagation of electromagnetic waves involves **a** transport of energy. A vector $S =$ E × H, called the Poynting vector (after an English physicist, John Henry Poynting), is used to describe this flow—*i.e.*, the energy flows in the direction of S (which is also the direction of wave propagation in an isotropic medium)—and the amount of energy crossing unit area (lying across the direction of flow) in unit time is equal to *S*—*i.e.*, to the magnitude of S. For electromagnetic waves S $= E \times$ H $= (\mu/\varepsilon)^{1/2} E^2$, and this relation, combined with equation (12), implies that $S = W/(\varepsilon\mu)^{1/2}$, or S $= bW$; thus, the mean value of *S* is b times the mean value of *W*—*i.e.*, the energy crossing unit area in unit time is equal to the velocity multiplied by the energy density.

The linearity of Maxwell's equations implies that the law of superposition applies to electromagnetic waves. The electromagnetic theory is therefore able to give a satisfactory account of all the simpler phenomena of interference and diffraction already described above. There are a few problems for which it is clearly superior to the "scalar-wave" theory that was described. These include the polarization of light transmitted through a slit the width of which is of the same order of magnitude as the wavelength and also the polarization of light scattered from microscopic particles. Calculations based on the electromagnetic theory are sometimes difficult mathematically, but the results are in agreement with experiment.

IV. The interaction of light with matter

REFLECTION AND REFRACTION

When light is incident in a perpendicular direction on the surface of a piece of glass, part is transmitted and part is reflected. This happens at any surface that forms the boundary between two transparent mediums of different refractive indices. For a particle theory of light, this phenomenon constitutes a serious difficulty. If a beam of light consists of a stream of particles all alike, because like causes produce like effects, there is no good reason why some should go through the surface and some be turned back. If the particles are not all the same (*e.g.*, if some have more energy and are therefore able to get through), then it would be expected that all those particles that do penetrate one surface would be able to pass through a second. This is never observed. In fact, the fraction of light transmitted at the second surface is the same as the fraction transmitted at the first (for light of one wavelength). Newton was aware of this fundamental difficulty and introduced the ad hoc hypothesis of particles with periodic oscillation between "fits of easy reflection" and "fits of easy transmission."

A wave theory of light, on the other hand, offers an explanation that follows in a natural and logical way from its premises. When a wave passes from one medium to another certain conditions must be fulfilled at the boundary. These conditions arise because the frequency of the transmitted wave must be the same as that of the incident wave. The velocities of light in the two mediums are different, and therefore, because velocity is the product of wavelength times frequency, the wavelength cannot be the same. By way of analogy, elastic waves can be imagined falling on a welded boundary between two solids—*e.g.*, steel and brass—with point P_1 just inside the first medium and P_2 extremely near P_1 and just inside the second medium. Then if P_1 and P_2 do not move exactly together the material at the boundary must tear. If the boundary is so strong that it does not tear, then the waves at the boundary must match each other in a most precise way. This match is possible only if there is a reflected wave as well as a transmitted wave: when the density and the elastic properties of the two mediums are known then the fraction reflected can be calculated.

For light, the fraction reflected can he calculated when the indices of refraction are known, and it is found that the fraction reflected when a beam of light is incident perpendicular to a boundary between air (refractive index nearly equal to 1) and a medium of refractive index n is $(n - 1)^2/(n + 1)^2$. Thus for glass (index 1.5) about 4 percent is reflected, but for diamond (index 2.42) 17 percent is reflected. This result is predicted by any reasonably self-consistent wave theory including the electromagnetic-wave theory. When light is incident on a surface at an angle other than perpendicularly, however, the calculation of the fraction of light reflected is more difficult. This fraction depends both on the direction of incidence and on the polarization of the light.

Fresnel relations

Fresnel derived certain expressions (known as Fresnel's relations) for the amount of light reflected under various conditions, using an elastic solid theory. To do this he assumed boundary conditions chosen ad hoc—*i.e.*, to produce the correct result. In the electromagnetic theory there is no choice in regard to the conditions that must be satisfied at the boundary of two dielectrics. The boundary conditions are fixed by the constitutive equations and by general principles such as the conservation of energy. They are derived from experiments on electricity, not from experiments on light.

These boundary conditions may be applied to electromagnetic waves with the following results.

1. The existence of a wave reflected so that the angle of reflection is equal to the angle of incidence and of a refracted wave the direction of which obeys Snell's law.
2. All the results that Fresnel had previously obtained for the amounts of light reflected and refracted under a variety of conditions, including Brewster's law.
3. The existence of total reflection and of certain observed results in regard to the change of polarization of light that has been totally reflected (when the incident light is polarized).
4. The amount of light reflected at the surfaces of metals (including the result that strong reflection is, in general, obtained with good conductors).
5. A detailed description of certain complicated results obtained when light is incident upon the surface of a transparent anisotropic crystal (*e.g.*, calcite).

DISPERSION AND SCATTERING

Oscillating dipoles. In electrostatics, a dipole consists of two equal and opposite charges situated a small distance apart. The line joining the charges is called the axis of the dipole, the dipole moment M is the name given to the product of one charge times separation. An oscillating dipole may be thought of as one in which charges move so that M fluctuates between $+M_0$ and $-M_0$ (*i.e.*, M = $M_0 \sin 2\pi\nu_t t$). Maxwell's theory shows that an oscillating dipole produces electromagnetic waves that are nearly spherical at distances large compared with the size of the dipole. The total energy emitted is proportional to the square of the ratio M_0/λ^2 (in which $A = c/\nu_t$).

Scattering by free electrons. When an electromagnetic wave (with an electric vector of amplitude E_0) operates on a free electron, it causes the electron to oscillate so as to produce an oscillating dipole with a moment proportional to E_0/h^2. This dipole emits waves in all directions. These waves are regarded as scattered light. The scattered light is strongest in directions perpendicular to the axis of the dipole, which is the same as the direction of the electric vector in the incident wave. It is polarized with its electric vector in the plane containing the dipole. The ratio of the energy scattered to the energy incident per unit area is: $k = (1/6\pi\epsilon_0^2)\ (e^4/m^2c^4)$, in which e is the charge and m is the mass of the electron. The energy scattered is equal to the amount of energy incident upon an area equal to k. For this reason k is called the scattering cross section or the scattering coefficient.

Dipole radiation

Scattering by bound electrons. A bound electron is subject to a restoring force when displaced from its equilibrium position, similar to a weight on the end of a spiral spring. An electron bound to an atom can oscillate with a natural frequency vo. Its motion produces an oscillating dipole and electromagnetic waves are emitted. Energy lost in this way constitutes damping, and the oscillation decays like the well-known damped harmonic oscillator of classical mechanics. The curve of decay is similar to the curve of a damped waye shown in Figure 14A. When a light wave is incident, a bound electron oscillates. The strength of the dipole produced and, hence, of the light scattered depends on the frequency (ν_t) of the light wave and also upon the frequency vo and the damping constant γ of the electron.

The scattering cross section (k') is found to be:

$$k' = \frac{\nu_t^4}{(\nu_0^2 - \nu_t^2)^2 + \gamma^2\nu_0^2}\,k; \qquad (14)$$

and it reaches a maximum value when v, is nearly equal to ν_0; *i.e.*, when the incident wave is in resonance with the natural frequency.

Molecular scattering. Because atoms and molecules contain bound electrons, they scatter light, but not just like single bound electrons. The scattered wave has an amplitude proportional to the amplitude (E_o) of the incident wave and inversely to the distance from the scattering centre; *i.e.*, the electric vector E_r of the scattered wave is $\eta E_o/r$, in which η is a constant. This value applies to the wave emitted at right angles to the electric vector of the incident light. The phase of the scattered wave is either the same as that of the incident wave or differs from it by an amount that is the same for each molecule. Figure 24 shows a beam of light incident upon a layer of gas at low pressure. The molecules are irregularly placed so that the light scattered from different molecules, such as M_1, M_2, M_r, M_i, etc., in any direction (except the forward direction) is incoherent. The law of photometric sammation (see above) applies, so that the total energy scattered by N molecules is just N times the energy scattered by one molecule.

A gas with N molecules per unit volume has an equivalent cross section (k_m),

$$k_m = \frac{8\pi}{3}\eta^2 N. \qquad (15)$$

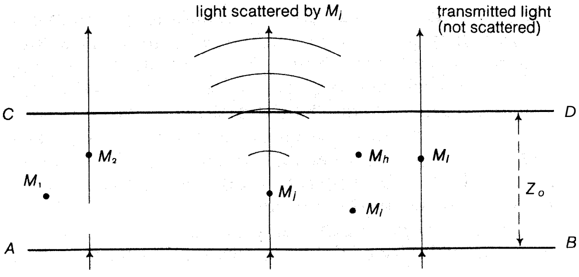


Figure 24: Scattering of light from molecules (see text).

In the forward direction the light scattered from different molecules is coherent because the total distance from the back edge of the gas layer (AB) to any molecule, such as M_j and M_l and thence to the front edge of the layer (*CD*) is equal to the thickness of the layer (z_0). Thus in the forward direction, the scattered waves can interfere with each other and with the wave representing the unscattered light. The resultant is a new wave in the forward direction but retarded—*i.e.*, with a phase lag $\eta\lambda N z_0$. In ordinary wave theory, however, the velocity of light in a medium of refractive index *n* is c/n and the phase lag due to passage through a layer of thickness z_0 is equal to $2\pi(n-1)z_0$. Equating these two estimates of the lag requires

$$2\pi(n-1) = \eta\lambda^2 N, \qquad (16)$$

and substituting for η from equation (16) into equation (15) implies that

$$k_m = \frac{32\pi^3}{3\lambda^4 N}(n-1)^2. \qquad (17)$$

This equation was derived by Rayleigh. He also showed that the brightness of the blue sky is about that to be expected if the light is produced only by molecular scattering, and it is not necessary to suppose that there is any appreciable scattering by dust particles—in the high atmosphere and on a clear day. The factor $1/\lambda^4$ implies that the scattering coefficient for blue light ($\lambda = 400$ nanometres) is about six times larger than that for red light ($\lambda = 640$ nanometres) and this accounts for the blue colour. Laboratory measurements of the amount of scattered light agree with that calculated from equation (17). For the rare gases (*e.g.*, argon) the scattered light is almost completely plane-polarized; but other gases show less complete polarization (*e.g.*, 90 percent for carbon dioxide) as would be expected if their molecules possess an electric dipole moment. The light of the blue sky is polarized but polarization is reduced because some of the light has been scattered more than once.

Dispersion. Equation (17) implies that (n − 1) may be calculated if η is known; *i.e.*, if the ratio of the amplitude of the scattered wave to that of the incident wave can be calculated. Dispersion formulas for gases have been obtained by assuming that the atoms or molecules contain bound electrons having a response that varies with frequency in the manner described above. The methods of calculation involve the properties of atoms and molecules rather than those of light and are not considered in this article. Certain general results are readily obtained—*e.g.*, that for most transparent mediums the refractive index increases regularly toward the blue end of the spectrum (this is called normal dispersion). It is also found that in a region in which there is a strong absorption, the index varies rapidly, and over part of this region the dispersion is anomalous (*i.e.*, the index increases in the direction of larger wavelengths). This derives from equations (14) and (17).

Normal and anomalous dispersion

Scattering of larger particles. **A** particle that contains N atoms but is small compared with the wavelength of light has a dipole moment approximately N times that for a single atom. The scattered energy is then proportional to N^2, or to V^2, or to R^6 (if V is the volume, and R is the radius). Thus a particle 20 nanometres in diameter scatters about as much light as 10^{12} separate atoms. For larger particles the scattering still increases though more slowly. When the particle is larger than a wavelength it must be considered as a diffraction object. The ratio of effective cross section for scattering to actual cross section fluctuates as the size increases. The theory of this variation, in relation to size, refractive index, and absorption has been worked out and verified experimentally.

Scattering by macroscopic solids and liquids. In crystalline solids the atoms are regularly arranged so that the wavelets scattered in a given direction have regularly varying phases. They interfere to give a small resultant except in the direction of the forward wave. Scattering of light is small in relation to the number of atoms per unit volume and results largely from crystal defects. For X-ray wavelengths smaller than the interatomic distance, the coherent waves scattered in certain directions reinforce one another.

For amorphous solids and liquids the number of atoms in any volume having a diameter of about one wavelength is large and the scattering per atom, although usually larger than in crystals, is still small in relation to the number of atoms involved.

The *Kramers–Kronig* relation. Any satisfactory theory of dispersion must comply with the condition that the scattered wave can never appear in advance of the incident wave that produces it. Physicists Hendrick Anthony Kramers of The Netherlands and Ralph de Laer Kronig of Germany showed that this basic causality condition implies that the dispersion (*i.e.*, the variation of refractive index with frequency) and the absorption are not independent. They derived equations enabling the absorption to be calculated when the dispersion is known (for all frequencies) and vice versa. It is not surprising that a relationship should exist, because dispersion and absorption are each related to the resonators described above in connection with scattering by bound electrons. The relationship has been found of great importance in many branches of pure and applied physics.

MECHANICAL EFFECTS OF LIGHT

When light is emitted or absorbed there are three mechanical effects: (1) an exchange of energy, (2) an exchange of linear momentum, and (3) an exchange of angular momentum. These are manifested as (1) heating (or cooling), (2) a pressure, and (3) a torque. If the conservation laws apply, then light must possess energy, momentum, and angular momentum. All these effects are predictable either from classical electromagnetic theory or from quantum theory. They can all be explored in relation to either exchanges between light and macroscopic pieces of matter or interaction between radiations and atoms. In this section the first will be treated, leaving the latter for the next section.

Light pressure

Equation (13) is an expression for light energy according to electromagnetic theory, and every detector of light is an energy transducer (converter). There is a general theorem, valid for light and sound, predicting that radiation must exert a pressure equal to W (the energy per unit volume) on a body that absorbs it. Twice this pressure is produced if the light is totally reflected at normal incidence. In electromagnetic theory, when light falls at normal incidence on a metal surface, a current (proportional to the field E) is produced in the metal; and, because this occurs in the magnetic field of the light, there is a reaction proportional to the vector product of the electric and magnetic field strengths (E × H) and normal to the surface—*i.e.*, a reaction proportional to the Poynting vector (*S*). For an insulator the induced current is the displacement current. Because the Poynting vector is proportional to W, the energy per unit volume, the pressure is proportional to W (further consideration shows that, in the usual units, it is equal to W).

The measurement of the linear momentum of light is difficult because, under normal experimental conditions, the pressure is less than 10^{-4} dyne per square centimetre. In Figure 25, light from powerful sources S_1 and S_2 falls on the mirror vanes A_1 and A_2 that are at the ends of a rod suspended by a quartz fibre. The pressure on the vanes produces a couple or twisting of the fibre and hence a rotation of the system that is measured by the usual

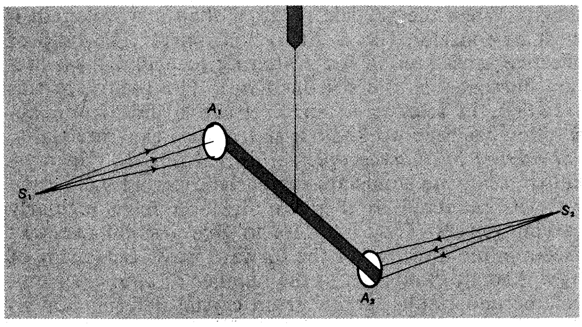


Figure 25: Apparatus for measurement of light pressure (see text).
From R.W. Ditchburn. *Light* (1963); Interscience Publishing, Inc., by permission of John Wiley & Sons, Inc.

mirror-and-scale method. The true radiation pressure can be measured only when a certain thermal action, called the radiometer effect, has been eliminated. In the radiometer effect, if radiation heats a surface, then molecules of gas that rebound from one side will move faster than those that rebound from the other, cooler side, and there will be a net pressure. In the period 1901 to 1905, when physicists were making the first accurate determination of light pressure, it was not possible to remove this effect by placing the system in a vacuum in which the number of molecules is at a minimum; instead, the effect was reduced by substituting the special cells for the vanes. Any radiometer effect within the cells produces no net force. The residual effects at outer surfaces are small and are the same whichever side of the cell is illuminated. The measured values (1) when radiation is absorbed and (2) when it is reflected agreed with theoretical values within 1 percent. By use of an alternative method of eliminating the radiometer effect, results in agreement with the theoretical values were also obtained.

It can be shown that a beam of circularly polarized light possesses angular momentum and thus should exert a torque upon a half-wave plate that reverses the rotation of polarization (*i.e.*, from left-handed to right-handed circularly polarized light or vice-versa). The experimental difficulties of measuring this torque (of about 10^{-11} dyne centimetre) are formidable and the overall error is about 10 percent. The result agrees with the theoretical value within this margin.

The density of radiation incident on the Earth's surface at its mean distance from the Sun is about 1.36 kilowatts per square metre (for a surface normal to the Sun's rays), and the corresponding light pressure is about 5×10^{-5} dyne per square centimetre. The total pressure on the Earth is about 6×10^{14} dynes, a minute fraction of the gravitational force exerted by the Sun.

Importance of the mechanical properties of light

The mechanical properties of light, difficult to measure in the laboratory, are important under certain astronomical conditions. Gravitational forces are proportional to the mass of a body and hence to the cube of its radius. For a body having large radius compared with the wavelength, the radiation pressure is proportional to the square of the radius. Whereas the attraction and repulsion may be equal and balanced for a medium size particle, for a smaller particle, the radiation pressure may be a large fraction of the whole force. For this reason, radiation pressure has a considerable effect in extending a comet's tail. In a star the outward flow of radiation from the centre produces a net force tending to expand the star, thus opposing the gravitational forces. Also, when radiation energy is transferred between two parts of a rotating star that are moving with different velocities, momentum transfers are involved. This produces an effect known as radiative viscosity because the resulting forces reduce the relative velocity of adjacent layers and so reduce the angular velocity of rotation.

V. Quantum theory of light

In the classical electromagnetic-wave theory, the absorption of light is a continuous process and there is no lower limit to the amount of energy that an atom can absorb from light of given frequency. Energy can be exchanged, in infinitesimal amounts between radiation and atoms, in an enclosure that is in thermal equilibrium. The total energy of the enclosure should, in equilibrium at temperature T, be distributed so that each degree of freedom has an equal amount, kT, in which k is Boltzmann's constant. The number of possible modes of vibration for transverse waves with frequencies between ν and $\nu + d\nu$ is $8\pi\nu^2 d\nu$ per unit volume. (In this section all frequencies are temporal frequencies and ν is used for pt). If $\rho(\nu)d\nu$ is the radiation energy per unit volume in this frequency range, and c is the velocity of light, then the radiation density is as shown in curve R–J of Figure 26. This law, derived by Rayleigh and Jeans, is the inevitable consequence of classical wave theory: It predicts that the radiation density increases without limit as the frequency increases, and this implies that, in equilibrium, all the energy in the universe is found in the high-frequency end of the electromagnetic spectrum and none is present in matter. This catastrophe may be avoided by assuming that there is some upper limit to the possible frequency of radiation, but the Rayleigh–Jeans law does'not agree with measurements of temperature radiation in the visible and ultraviolet regions of the spectrum (curve R–J, Figure 26). The experimental results are represented by curve P in Figure 26, and Planck suggested a formula, now called Planck's law, that fitted this curve. Originally this formula was probably an inspired guess, but Planck soon showed that it could be derived by assuming that a mode of frequency can change only by $h\nu$ (or by a multiple of hu) in which h is a new universal constant now called Planck's constant. The minimum "element of energy" (as Planck referred to it) is now called the quantum. For visible light the quantum is about 3.5×10^{-12} erg or 3.5×10^{-19} joule.

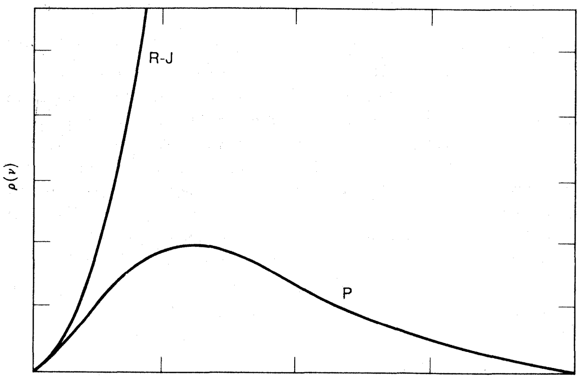


Figure 26: **Equilibrium** radiation density curve R–J (Rayleigh–Jeans); curve **P** (Planck; see text).

Properties of a photon

The theory in which Planck introduced the quantum was concerned with the interaction between radiation and matter. He expected that only minor modifications of the electromagnetic theory of light in free space and of the electron theory of matter would be needed. This expectation was not realized. Experiments, some of which will be described below, led Einstein to regard light as an assembly of entities, later called photons. A frequency ν, energy $h\nu$, momentum $h\nu/c = h/\lambda$, and angular momentum $h/2\pi$ are associated with a photon of circularly polarized light. In many situations the photons appear to be localized concentrations of energy and momentum—*i.e.*, to have the properties of particles. Yet the evidence that led to the rejection of a corpuscular theory in favour of a wave theory is still as valid and as compelling as it was in the time of Fresnel.

The theory, which includes logical descriptions both of those experiments that seem to require a corpuscular theory and those that support a wave theory, is known as quantum mechanics. This is a general theory of radiation and matter—not a special hypothesis about their interaction. The account given below of some aspects of quantum mechanics that are important in relation to experiments on light must be related to the general theory (see MECHANICS, QUANTUM).

PHOTONS

When light of suitable wavelength is incident upon the clean surface of a metal, electrons are emitted. If W_m is the maximum kinetic energy of one of those photoelectrons released by light of frequency ν, then it is found that the maximum kinetic energy is Planck's constant times the difference between two frequencies, or W, $=$ h(v $- \nu_0) = h\nu - W_0$, in which W_0 is written for $h\nu_0$. No electrons are emitted when ν is less than the critical frequency v,. It is known from other experiments that a minimum amount of energy called the work function is needed for an electron to escape from a surface. If W_0 is accepted as the work function for emission of a photoelectron, then it is easily understood that no electrons are emitted when frequency ν is less than the critical frequency and that the maximum energy of escape for an electron is $h\nu - W_0$. Some electrons may make collisions after absorption of energy $h\nu$ and either fail to emerge from the surface or else emerge with lower energy. Escape by gaining energy from two photons of frequency less than ν_0 is extremely improbable because the energy due to one absorption is almost always lost by collision before a second photon is absorbed.

The number of photons absorbed, and therefore the number of electrons released, are proportional to the energy density in the beam of light and to the time. The energy of individual electrons is independent of the energy density—*i.e.*, independent of the amplitude of the waves.

Bohr's postulate

Conservation of energy and **momentum.** Niels Bohr, a Danish physicist, postulated that an atom can exist only in one of a set of discrete stationary states of internal energy ($W_1, W_2, \ldots$). The frequency $h\nu_{nm}$ of light emitted in passing from state n to m is related to the energy difference by Bohr's frequency condition: that Planck's constant times the frequency is equal to the energy difference between two states; *i.e.*,

$$h\nu_{nm} = \text{Wn} - W_m.$$

For some pairs of states (W, $-$ W,) can be measured by experiments on collisions between electrons and atoms, and Bohr's condition is found to be satisfied within the limits of experimental error. Many other, less direct but more accurate, experiments are also in agreement with this equation.

In Bohr's theory, and subsequently in quantum mechanics, an angular momentum is associated with each atomic state. In the simplest situation, when polarized light is emitted in a magnetic field, the angular momentum about a known axis changes by $h/2\pi$. The changes are such that angular momentum is conserved if a photon of circularly polarized light has an angular momentum of $h/2\pi$ about the direction of propagation.

Compton scattering

It was shown by a United States physicist, Arthur Holly Compton, that when X-radiation is scattered by a nearly free electron, there is a wavelength change. A formula that gives the wavelength change as a function of the angle of scattering was derived by assuming that a "particle" of energy $h\nu$ and momentum $h\nu/c$ collides with a particle of mass m that is at rest, and that energy and linear momentum are conserved. The resting electron gains momentum and kinetic energy by changing its speed. The photon cannot change its speed, which is always c. It therefore must change its frequency when it changes energy and momentum. The measured change of frequency agrees with that calculated by Compton.

An atom that has absorbed a quantum of energy $h\nu_{mn}$ and changed from state m to state n may return to state m with emission of a photon of the same frequency (resonance radiation)—*i.e.*, $\nu_{mn} = \nu_{nm}$. If there is a state s intermediate in energy between states m and n, the return may be in two stages; n to s and s to m, with the emission of two photons. The sum of the two quanta $h\nu_{ns}$ and $h\nu_{sm}$ is then equal to a single quantum $h\nu_{nm}$ so that energy is conserved. In gases, these emissions usually follow the absorption in less than a microsecond, and the term fluorescence is used to include both resonance radiation and multistage processes. In solids and liquids there are sometimes intermediate states that are metastable so that emission occurs over much longer periods (phosphorescence; see LUMINESCENCE). In all these processes the frequency of the emitted radiation is not greater than that of the absorbed radiation in accord with an empirical law discovered in the 19th century. In quantum theory this law is a necessary consequence of conservation of energy.

Raman effect

In Rayleigh scattering, the frequency of the scattered light is equal to that of the incident light. The physicists Sir Chandrasekhara Venkata Raman of India and Leonid Isaakovich Mandelstamm of the Soviet Union independently discovered processes in which light is scattered by an atom or molecule that changes its state. There is a corresponding discrete change in the frequency of the scattered radiation. When the scattering atom or molecule is in its lowest state, the frequency of the unscattered radiation is always less than that of the incident light (Stokes lines). When light is scattered by a molecule that is not in its lowest state, photons of higher frequency may be found in the scattered radiation (anti-Stokes lines). The scattered photon then has more energy than the incident photon, but the molecule gives up energy by passing to a lower state. The Raman effect is observed when light is scattered by solids and liquids. There is also an additional kind of scattering in which photons exchange energy and momentum with the thermal vibrations of the solid or liquid (Brillouin scattering, named after a French physicist, Louis-Marcel Brillouin).

The experiments quoted are examples of a large number that are in accord with the view that, within the limits of measurement, energy and momentum are conserved in detail—*i.e.*, in each individual interaction between an atom and radiation.

Spontaneous and stimulated emission. Planck's theory gives the distribution of radiation energy with frequency in an assembly of atoms and radiation in thermodynamic equilibrium at a common temperature in an enclosure. An Austrian physicist, Ludwig Boltzmann, applying the same principles to the assembly of atoms, derived a formula that gives the distribution of atoms between different stationary states—*i.e.*, the ratio of the number N_n in any state n to the number N, in any other state m. In the enclosure now considered, Boltzmann's and Planck's laws must both be satisfied though each atom is changing its state frequently by emitting or absorbing a photon or by collision. Equilibrium can be maintained if, and only if, each individual reaction is balanced by its own reverse process. Einstein assumed that the equilibrium was maintained by the following process: (1) absorption of radiation at a rate $B_{mn}N_m\rho(\nu)$; (2) spontaneous emission at a rate $A_{nm}N_n$; and (3) stimulated emission at a rate $B_{nm}N_n\rho(\nu)$, in which A_{nm}, B_{mn}, and B_{nm} are constants, now known as Einstein coefficients. When the rate of absorption was equated to the rate of emission (*i.e.*, to the sum of spontaneous and stimulated emission), an equation of the same form as Planck's law was obtained. This equation became identical with Planck's law if

Einstein coefficients

$$B_{nm} = B_{mn} = \frac{c^3}{8\pi h\nu^3}A_{nm}. \tag{18}$$

If there were no stimulated emission (*i.e.*, if $B_{nm} = 0$), Planck's law and Boltzmann's law could not both be satisfied. Einstein postulated the existence of stimulated emission in 1905, but for more than 30 years it appeared to be only of theoretical interest. At the densities of radiation then available, the rate of stimulated emission was small compared with the rate of spontaneous emission. Equation (18) shows that in the microwave region of the electromagnetic spectrum (frequencies of a few thousand megahertz) the ratio $\mathbf{B}_{mn}/\mathbf{A}_{nm}$ is much larger than in the visible spectrum. Accordingly, it was much easier to develop microwave amplification by stimulated emission of radiation (maser) than the corresponding laser effect with radiation in or near the visible spectrum. The existence of stimulated emission can be fitted into a classical wave theory, like many other concepts that originated in relation to theories of photons.

Interference and diffraction of photons. It was at one time suggested that interference and diffraction phenomena are caused by collisions between photons considered as small particles, or at least to some kind of complex interaction between photons. A simple experiment performed by a British physicist, Geoffrey Ingram Taylor, in 1908 excludes this possibility. He photographed the diffraction pattern of a needle and reduced the illumination until long exposures were needed to obtain an image. When the chance of two or more energy quanta passing through the apparatus simultaneously was made extremely small, the diffraction pattern was exactly the same as that obtained with a strong source of light.

This experiment is supported by many later experiments, showing that interference and diffraction are to be associated with single photons. The discussion in the section on coherence above implies that when ordinary sources of light are used each photon can interfere only with itself and not with any other photon. An interference pattern can, of course, be photographed only by recording the effects of many photons because one photon can activate only one grain on the photographic plate, but it does not matter whether the photons all arrive over a time span of a microsecond or of several weeks.

The photograph of an interference pattern is both a wave phenomenon because it shows the characteristic spatial periodicity and a quantum phenomenon because the whole energy of a photon can be used to activate a single grain. It is not possible to trace the path of one photon (regarded as a particle) through the apparatus and, at the same time, obtain the interference pattern. In Young's experiment there is no way of finding out through which slit a given photon passes--except by covering one slit and thus losing the interference fringes.

THE WAVE–PARTICLE NATURE OF LIGHT

Uncertainty relations. Experiments with the photoelectric effect show that energy can be transferred from one atom to another in a way that suggests that photons are corpuscles—*i.e.*, localized concentrations of energy and momentum. Other experiments imply equally clearly that the light emitted from an atom, when it loses energy, must be represented by wave groups. Both of these sets of experiments are equally valid, and together they require that the photon must have both wave-group and particle properties at the same time.

It will now be assumed that the intensity of the waves gives the relative probability that a suitable detector will detect a photon at a given point. Experimentally, waves and particles are both abstractions, each describing the same physical system. Quantum mechanics does not seek a relationship such that particles are guided by waves according to the laws of classical physics—as in the theory of the English physicist J.J. Thomson, in which photons are pictured as closed tubes of electric force guided by Maxwellian waves. The waves and particles now to be considered do not have all the relevant properties of classical waves and particles. When a photon is absorbed, the whole of a wave group, possibly extending over a large volume, is annihilated. The position of a photon particle cannot be exactly specified. It is known only that it is within the wave group and most likely exists where the waves are most intense. Also the energy and momentum of the photon are determined by the frequency of the waves. Fourier analysis of wave groups shows that, when the wave group is short (so that the position of the photon is fairly well determined), the frequency of the waves and, hence, the energy and momentum of the photon are known only within a wide range. The product of the uncertainty in position and the uncertainty in momentum can never be less than a certain minimum value. It will now be shown how these relations appear in some experimental situations.

In one experiment a parallel beam of light is incident on a slit of width d, producing a diffraction pattern on a screen. The y-coordinate, at the moment when a photon passes through the slit, is known within an uncertainty $\Delta y = \pm d$. The direction of propagation when the photons emerge from the slit is uncertain because of diffraction, but most of the photons are observed within an angular range $\pm\theta$, in which $\sin\theta = \lambda/d$. There is therefore an uncertainty Δp_y in the y-component of the momentum (p) of about $\pm(h \sin\theta)/\lambda$, so that the product of the two uncertainties, Δp and Δy, is

$$\Delta p \Delta y \sim \frac{dh \sin\theta}{\lambda} \sim h,$$

in which the sign $\sim$ indicates is of the "order of magnitude of." In a second experiment, monochromatic light is passed through an optical shutter. The shutter is opened for a short time *At* so that a wave group of duration *At* and length $\Delta z = c\Delta t$ passes through the shutter. Then $c\Delta t$ is the length l of the wave train. This implies that the frequency ν varies within a range $\Delta\nu = 1/\Delta t$—*i.e.*, that the photon energy varies within a range $\Delta E \sim h/\Delta t$ so that $\Delta E\Delta t \sim h$. More detailed analysis shows that, if the uncertainties Δt and ΔE are the root mean square deviations of statistical theory, then their product is equal to or larger than $h/4\pi$; so that $\Delta E\Delta t \gtrsim h/4\pi$.

Uncertainty in energy of photons

Individual examples do not prove that the uncertainty relations are generally true any more than individual examples prove the general validity of the second law of thermodynamics. The fact that every example agrees, and that ingenious attempts to find exceptions all fail, constitutes a kind of proof by default, but the real proof is in the whole range of experimental work that demonstrates the equal status and inescapable association of wave and particle properties.

Maximum observations. The uncertainty relations show that precise measurement of both position and momentum are incompatible. It follows that a precise deterministic theory of particle dynamics is not appropriate because the precise initial condition of a particle is not observable. An extension of the uncertainty relations shows that the number of photons in a light beam and its phase cannot be simultaneously measured with precision, and thus a classical wave theory is also excluded.

In any experimental situation there are, in general, two or more observations that are compatible. Neither affects the other, and their results are independent of the order in which they are made. A complete assembly of such measurements is said to constitute a maximum observation. A state of a system is specified by giving the result of a maximum observation or in an equivalent way. The relative probabilities of future observations on a system can be calculated from the result of a maximum observation without reference to any previous results obtained on the system. For every system there is one certain prediction —that an immediate repetition of the maximum observation will give the same result.

Empirical questions. There is a certain type of question to which the quantum mechanics gives no direct answer because the question is not related to a possible observation. One may ask, for instance, what happens to a photon of right-handed circularly polarized light as it passes through a half-wave plate. It is known that the photon will emerge as a photon of left-handed circularly polarized light and that angular momentum is transferred to the plate. The wave theory gives an account of the progress of a wave through the plate. If, however, one tries to visualize the particle and asks for a mental picture of the way in which this particle alters as it goes through the plate, there is no ready answer. The theory has nothing to say unless an experiment is designed to determine what happens to the photon in passage through the plate. For example, a half-wave plate is split in two and the light is passed through the quarter-wave plate, which forms one half. It may subsequently be passed through a Nicol prism analyzer. There will be some orientation of the Nicol prism such that the photon will certainly pass through. If a Nicol prism is oriented at a to this direction, the probability for a single photon to pass through is $\cos^2$ a. This detailed theoretical prediction agrees with experiment. Thus the quantum mechanics correctly predicts the results of an experiment but leaves unanswered the original question about the way in which the photon, visualized as a particle, changes its angular momentum as it passes through the half-wave plate.

Assemblies of photons. The application of quantum mechanics to the electromagnetic field in an enclosure justifies a representation in terms of photons. These may be regarded as indistinguishable particles. Any number of these particles may exist simultaneously in the same energy state. The statistics of particles that obey these conditions is called Einstein–Bose statistics. The distribution of particles among different states may be calculated by the usual methods of statistical mechanics in which probability (or entropy) is maximized subject to certain limitations. In an assembly of material particles the total energy and also the number of particles are constants. With photons, the second condition is no longer valid because two photons of frequencies ν_{ns} and ν_{sm} may be replaced by one of frequency $\nu_{nm} = \nu_{ns} + \nu_{sm}$, thus decreasing the number of photons while conserving the total energy. The statistical calculation leads straightforwardly to Planck's law. It is found that for ordinary sources of light the probable number of photons in any given state is always much less than unity. For a laser source many photons are found in relatively few states, and, indeed, in an ideal single mode continuously operated laser, all photons would be in one state.

A considerable amount of experimental and theoretical work has recently been done on coincidences of photon detection in two or more detectors that are observing the same source. This kind of photon statistics measures the coherence of light from a star and hence the size of the star.

Assemblies of photons and atoms. In the Dirac theory of absorption and emission of radiation, quantum mechanics is applied to an enclosure that contains radiation and free atoms or molecules. Processes involving creation and annihilation of photons with corresponding atomic transitions must maintain equilibrium. The theory derives expression for the Einstein *A* and *B* coefficients of absorption and emission and for their ratios. It is also shown that, whereas spontaneous emission is random, stimulated emission is coherent with the radiation that excites it. The theory extends to include an account of Rayleigh scattering (and hence of dispersion) and of the Raman effect. All problems concerned with the absorption, emission, and scattering of light are, in principle, soluble, although the mathematical formulation, even of approximate formulas, is usually difficult.

Limitations of quantum mechanics

The quantum-mechanical theory is adequate at optical frequencies and at lower frequencies. It is also adequate at the higher frequencies associated with ultraviolet radiation and for X-ray scattering including the Compton effect. It is inadequate for photons of the still higher frequencies involved in the creation and annihilation of fundamental particles. It is to be expected that, in due course, a basic theory of electromagnetic forces and of their relation to gravitation and to the strong forces between nucleons will emerge. The present theory will then have the status of an approximation valid for a wide range of frequencies including the frequencies associated with visible light.

BIBLIOGRAPHY. A.C.S. VAN HEEL and C.H.F. VELZEL, *Wat is licht?* (1968; Eng. trans., *What Is Light?*, 1968), is a good book for the general reader; numerous illustrations are accompanied by a simple text. M. MINNAERT, *Licht en kleur in het landschap* (1937; Eng. trans., *The Nature of Light and Colour in the Open Air*, new ed., 1959), also recommended for the general reader, deals with optical effects such as rainbows, mirages, and many other effects of light and colour. MAX BORN and EMIL WOLF, *Principles of Optics*, 3rd rev. ed. (1965), is a comprehensive treatise on the wave theory of light, particularly strong on diffraction, scattering, polarization, and coherence—suitable for the reader with advanced mathematical knowledge. ROBERT W. DITCHBURN, *Light*, 2nd ed. (1963). is a text at the university level that includes both theory and experimental evidence in relation to wave phenomena and quantum optics. K.D. FROOME and LOUIS ESSEN, *The Velocity of Light and Radio Waves* (1969), is a critical review of modern methods of measuring the constant *c*. JOHN H. SANDERS, *The Velocity of Light* (1965), contains reprints of original papers by A.A. Michelson, F.G. Pease, F. Pearson, Louis Essen, A.C. Gordon-Smith, and E. Bergstrand, together with a review of other work on the subject. LEONARD I. SCHIFF, *Quantum Mechanics*, 2nd ed. (1955), is a systematic exposition of the physical concepts of quantum mechanics, including a description of the electromagnetic field and its interactions with matter. JOSEPH NEEDHAM, *Science and Civilisation in China*, vol. 4 (1962), includes an authoritative account of the study of light in ancient China and a detailed comparison with parallel developments due to Greek and Arab philosophers. EDMUND T. WHITTAKER, *A History of Theories of Aether and Electricity*, rev. ed., vol. 1 (1951), deals with the background to the development of the electromagnetic theory of light in the 19th century.

(R.W.Di.)

Light, Aberration of

Annual, diurnal, and planetary aberration

The term aberration in the context of this article refers to the various displacements in the apparent direction of a celestial body consequent on the finite speed of light in combination with transverse motions of the body and the observer. The displacement in apparent direction arises in much the same way as when a raindrop, falling straight down past the window of a moving automobile, seems to be falling in a direction slanted from front to back of the car. The fundamental aberration phenomenon with light results from the motion of the observer; the displacement, measured as an angle between the real and apparent directions of the light rays, is small and is in the direction of the observer's motion. The displacement due to the steady motion of the solar system, called secular aberration, is unobservable because it is constant in time, and it is usually disregarded. The displacement due to the orbital motion of the Earth around the Sun is the annual aberration, while the component due to the rotation of the Earth is the diurnal aberration. Motion of the body itself will produce an additional transverse displacement during the time that the light has taken to reach the Earth; this is called planetary aberration. The annual and diurnal aberrations have values of about 20 and 0.32 seconds of arc, respectively. The term stellar aberration was first used by James Bradley, later Astronomer Royal (1693–1762), who both discovered the phenomenon of annual aberration and explained it.

Discovery of aberration. The discovery of aberration resulted from a program of observations directed towards the solution of another problem, that of determining stellar parallax, the difference in direction of a stellar body observed from two diametrically opposite points in the Earth's orbit around the Sun. The possibility of computing the distance of a star by measuring the angles of the triangle formed by parallax (that is, by observing the slight and regular change in the star's direction due to the orbital motion of the Earth) had intrigued many astronomers from earliest times. It was not only a subject of great intrinsic interest, but it excited particular attention because it offered the possibility of a direct confirmation of the Copernican theory of the solar system. It is now realized that even the largest observed stellar parallax is extremely small, being less than one second of arc, and needs the greatest refinement for its observation. Nevertheless, some 17th-century astronomers did find comparatively large annual displacements that were sometimes ascribed to parallax. The method most frequently employed involved the measurement of the zenith distance of a star (that is, the angle between the zenith and the star) at intervals throughout one year. Robert Hooke observed the star Gamma Draconis in this way at his London observatory during 1669 and concluded from his observations that it has a parallax of about 30 seconds of arc. Jean Picard later observed Alpha Lyrae in a similar fashion but was unable to detect any parallax. John Flamsteed, from a subsequent study of the Pole Star over a period of eight years, supposed it to have a parallax of at least 40 seconds. That these large displacements are due to the phenomenon of aberration, and not at all to parallax, entirely escaped these earlier astronomers; and the way was open to Bradley to explain them correctly, and to make improved observations, during the following century.

Bradley's first program

In 1725 Bradley resolved to re-examine the problem of large anomalous annual displacements in star positions observed during the previous century, primarily in order to obtain the parallax of a star. His first program of work was done with his friend Samuel Molyneux, an amateur astronomer. They adopted the usual technique of observing the zenith distance of a star that culminates close to the zenith. A vertical telescope made for this purpose, 24 feet long with an aperture of 3% inches, was mounted against a chimney stack in Molyneux's house. The star Gamma Draconis was selected, and observations commenced on December 3, 1725. By December 21 Bradley and Molyneux had established that the star was apparently moving. Observations continued over a year showed that the star was apparently shifting position by about 20 seconds of arc, and that after a maximum of one year it returned to its original position. Bradley immediately realized that this effect could not be attributed to parallax because it bore the wrong phase relationship to the position of the Earth in its orbit. Bradley observed a maximum displacement when the Earth was in positions 1 and 3, as shown in Figure 1; but he realized that, had the displacements been due to parallax, the maximum displacements would have occurred at positions 2 and 4. As the telescope at Kew had only a small range of movement, it could observe only a small number of stars. Wishing to observe many more stars in a similar fashion, Bradley erected another shorter instrument in 1727. With this he was able to observe more than 200 stars, including Capella, and to establish the precise phase relation between the position of the Earth in its orbit and the displacement of the star. At this time he saw that his observations could be explained by supposing that the velocity of light could be combined in a vector diagram with velocity of the observer. The following demonstration of this combination is substantially that of Bradley.

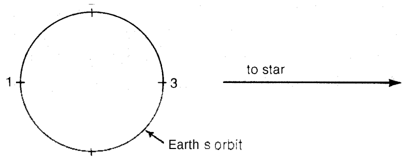


Figure 1: Orbit of the Earth. Maximum displacements of star's position are observed when the Earth is at positions 1 and 3, least at positions 2 and 4 (see text).

Development of theory. The nature of the aberration phenomenon is most easily understood if the celestial body is thought of as an emitter of photons of light, or packets of energy, rather than waves. In the diagram (Figure 2) the object is assumed to be emitting photons in all directions with velocity c. A stationary observer may obtain the direction of this object by sighting upon it through holes at A and B, distance *d* apart, which are so positioned that a photon entering at A may leave the sight at B. If the observer is also moving with velocity *V* transverse to the direction of the object, during the time that the photon has taken to move from hole A to hole B, hole B has moved a distance equal to d × V/c to one side; and the photon will miss it. But the photon could be caused to pass through both holes if the line joining them were to be tilted through an angle θ equal to V/c, because then hole B would have moved to the required position by the time that the photon reached it. This implies that the direction of the object is apparently displaced by an angle equal to V/c, measured in radians (a radian is that arc of a circle that is equal to the radius).

A small extension of this demonstration indicated that if the direction to the star is inclined at an angle θ to the line of sight the product of aberration ζ is equal to κ times sin θ, in which κ equals the fraction V/c radians (1 radian = 2.06264 X 1 0 seconds of arc). The value of κ found by Bradley was 20.25 seconds of arc (sec arc).

In spite of his careful work he never did achieve his original aim of measuring a parallax, but he remarked that if the parallax of one of his stars had been as much as one second of arc he could have detected it. His method was not a good one because the parallax appears as a small difference between two large quantities.

Bradley also noticed an apparently steady change in declination (an angular measurement of distance north or south of the celestial equator) that could not be attributed to aberration. He soon recognized that this effect is due to a rotation of the true pole of the Earth around the mean pole, caused by the attraction of the Moon upon the Earth's equatorial bulge. He continued his observations from 1727 to 1748 to confirm the explanation. This motion of the Earth's axis is called nutation (from the Latin "to nod" or "to rock"), and the circumstances of its discovery by Bradley illustrate the extreme care and the tenacity and patience that he applied to all his work.

Nutation

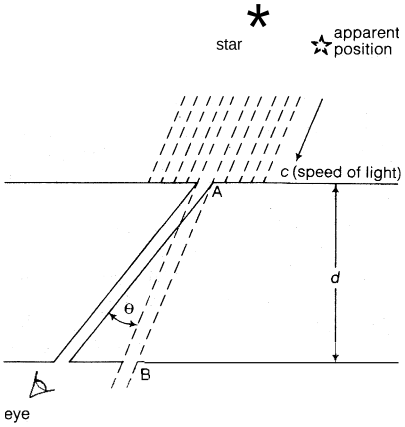


Figure 2: Idealized system of sighting on a star by alignment of two slits (A and B). The star's apparent direction is shifted because of the motion of the Earth (see text).

Although apparently simple in theory, the measurement of the constant of aberration is in practice one of the more difficult in astronomy because a comparison has to be made between star positions measured at a time interval of some six months. Since Bradley's measurements there have been many determinations. K.A. Kulikov, in his book *Fundamental Constants of Astronomy*, gives a tabulation of the results of 82 determinations up to 1951. Results of later measurements have been compiled. The value adopted by the International Astronomical Union in 1964, to be used to correct star places in ephemerides from 1968 onward, is 20.496 seconds of arc. The diurnal aberration, due to the Earth's rotation, has a maximum value of 0.319 seconds.

Constant of aberration

Astronomical effects. The annual aberration is one among several phenomena that affect the apparent position of a star in the sky. Other such phenomena are those of nutation, precession, and diurnal aberration. These effects require formal correction to star positions, and the full treatment is complex (see *Explanatory Supplement to the Ephemeris* [1961]); application of spherical trigonometry gives relations for the effect of aberration on the right ascension—angular measurement of east–west position in the sky—and declination of a star.

The planetary aberration is usually calculated on the assumption that both object and Earth move in straight lines during the time that the light takes to travel between the two. This assumption introduces an error of less than 0.001 second of arc for the outer planets. Although in the preparation of a star catalog corrections are made for nutation, precession, and aberration, no correction is made for the star's motion during the passage of light from it to the observer. An ephemeris for the Sun, however, includes the annual aberration.

The constant of aberration. The value of the constant of aberration, the direct measurement of which is discussed above, may also be calculated from the dimensions of the solar system. Taking into account the ellipticity of the Earth's orbit, the expected value of κ, which is related to the mean velocity of the Earth in its orbit, is given by the following equation:

$$\kappa = \frac{2\pi a}{cT\,(1 - e^2)^{\frac{1}{2}}} \text{ X } 2.06264 \times 10^5 \text{ sec arc,}$$

in which a is the semi-major axis of the Earth's orbit, e is its eccentricity, and T is the sidereal year expressed in mean solar seconds. If the equatorial radius of the Earth is denoted ρ, and the solar parallax is π_{Sun}, then these two quantities are related by the equation:

$$a = \frac{\rho}{\pi_{Sun}} \times 2.06264 \times 10^5.$$

Substituting for a in the equation for κ gives,

$$\pi_{Sun}\kappa = \frac{2\pi\rho}{cT\,(1 - e^2)^{\frac{1}{2}}} \times (2.06264 \times 10^5)^2.$$

Substitute the following values in this equation: $e = 0.01675$, $\rho = 6.378245 \times 10^3$ kilometres, $c = 2.997925 \times 10^8$ metres per second, then T = 3.155814954 X 10^7 seconds. The value for the velocity of light (above) is recommended by the XII General Assembly of the International Scientific Radio Union, which has assigned to it an uncertainty of ±0.0004 metre per second. This substitution leads to the following numerical relation between κ and π_{Sun}.

$$\pi_{Sun}\,\kappa = 180.245$$

Bradley was aware of such a relationship and used a similar one to verify his theory of aberration. Knowing the velocity of the Earth in its orbit, he was able to calculate: from his own observations, the time taken by light in its passages from the Sun to Earth. A comparison of this time with that deduced from the velocity of light, obtained from observations of the satellites of Jupiter, gave agreement to within 1 part in 493.

The relation may be used to deduce either the constant of aberration or the solar parallax. The solar parallax is now determined directly, knowing the velocity of light, by observing radar echoes from Venus. The accuracy of radar results greatly surpasses that obtainable by the best optical measurements. Probably the most accurate solar paraliax obtained by radar is that of D.O. Muhleman, who obtains 8.7941379 seconds of arc ± 0.0000015 second. Substituting this value into the previous relation gives $\kappa = 20.496$ seconds of arc. Note that this value for κ is independent of the velocity of light that is assumed.

Aberration, ether, and relativity. Although the phenomenon of aberration found a natural explanation in a particle theory of light, its explanation by a wave theory is not simple; and it was thought at the time that the amount of aberration might depend upon the thickness of glass or other transparent material in the telescope. Sir George Airy, then Astronomer Royal, remarked that this is "... a question which involves, theoretically, one of the most delicate points in the Undulatory Theory of Light." In order to test this point, Airy performed in 1871 a famous and far-seeing experiment. He filled a telescope with water and with it measured again the aberrational displacement, finding that use of the water-filled telescope gave the same result as before. At first sight, this is a rather surprising result, because in addition to refraction at the air–water interface, the light will pass more slowly between the slits if there is water between them, so that the tilt of the telescope should be increased. The magnitude of the expected effect can be readily calculated, if it is assumed that the water-filled telescope must be inclined at the new angle of aberration ξ. Because the light is refracted when it enters the telescope, it travels along the telescope inclined at an angle Ψ to its axis. As both ξ and Ψ are small angles, an approximate form of the refraction formula $\Psi \sim$ *3/12* (in which n is the refractive index of water) may be used to calculate * ▪. The velocity of light in the new telescope is equal to c/n, and the former condition for observation of the star must be modified to $\Psi = nv/c$. Hence, $\xi = n^2v/c$. Indeed, it should be possible to measure the absolute speed of the Earth, not just its orbital speed, by just such an experiment as this, for the change in angle of aberration on filling the telescope with water should be equal to

$$\xi - \zeta = \frac{n^2v}{c} - \frac{v}{c} = \frac{v}{c}(n^2 - 1).$$

The matter of why ξ should equal ζ can be resolved in a formal manner by following a suggestion first made by the French physicist Augustin-Jean Fresnel, that the water drags the light sideways with it at a velocity equal to fv, in which the factor f is called the Fresnel drag coefficient. A simple extension of the calculation given above shows that the null effect of filling the telescope with water can be accounted for if f is put equal to $1 - 1/n^2$. In 1871 the drag coefficient was already an established part of physics through the work of Armand-Hippolyte-Louis Fizeau, who had succeeded in demonstrating the effect quantitatively. Fizeau constructed an interferometer into the two beams of which were introduced tubes containing water. Interference fringes were obtained first with the water still, and then with the water in one tube following in the direction of the light, and in the other tube against the direction of the light. A small shift in the position of the fringes, increasing with the speed of the water, verified the expression for the drag coefficient.

The aberration experiment was not the only one to fail to give the absolute velocity of the Earth through the ether. Even more celebrated was that of Albert A. Michelson, and later Michelson and Edward E. Morley, who attempted to measure absolute velocity by using an interferometer to compare the speed of light in the direction of the Earth's motion with the speed at right angles to the motion; again the result was negative. Just as the aberration experiment apparently needed the drag coefficient for its understanding, so the Michelson–Morley experiment required the Lorentz–Fitzgerald contraction—a contraction of a moving body by a factor $1 - v^2/c^2$ in the direction of motion.

Theory of relativity

These facts found a natural expression in Einstein's special theory of relativity. For example, the Fresnel drag coefficient now emerges as the result of combining in a relativistic fashion the velocity of light in the medium, c/n, with the velocity of the medium. The aberrational effect can he calculated similarly by transforming the velocity components of a photon in a frame of reference, S, which might be the rest frame of the Sun, to the components in another frame of reference, S′, that of the Earth. If the velocity difference between the two frames of reference is V, then a velocity u_x in one system transforms into a velocity u_x' in the other, in which

$$u_{x'} = \frac{u_x + V}{1 + \frac{u_x \times V}{c^2}}.$$

Taking, as before, the angle between the line of sight to the star and the direction of the Earth's velocity as θ, so that $u_x = \cos\theta$, then the star appears at inclination θ' in frame S′ in which

$$\cos 8' = \frac{\cos\theta + V/c}{1 + V/c \cos 8'}$$

This is the relativistic form of the aberration formula, in which the change in direction of the star is equal to $\theta\text{-}\theta' = \xi$. The formula was derived by Einstein in his celebrated paper of 1905 on relativity. Putting, as is usual, $\beta = V/c$, and expanding this relation in terms of β, a second order is obtained in β, $\zeta = \beta \sin\theta - \beta^2 \sin 2\theta$.

As $\beta \sim 10^{-4}$, it is possible for practical purposes to neglect the term in β^2, when the formula reverts to the one obtained classically. If it is felt necessary, the calculation can be repeated for a telescope filled with water, with the same results. The Michelson–Morley experiment is technically an extremely difficult one to carry out, and it is satisfying that such a simple experiment as measuring the constant of aberration with a water-filled telescope can provide direct evidence for the theory of relativity.

BIBLIOGRAPHY. The discovery of aberration by Bradley is described in "A Letter from the Reverend Mr. James Bradley Savilian Professor Astronomy at Oxford, and F.R.S. to Dr. Edmond Halley Astronom. Reg. & c. Giving an Account of a New Discovered Motion of the Fix'd Stars," *Phil. Trans. R. Soc.*, 35:637–661 (1728). A general account of the circumstances of the discovery and of the life of Bradley is given by s . ~RIGAUD in his book, *Miscellaneous Works and Correspondence of James Bradley* (1832). The relativistic treatment of aberration was first given by EINSTEIN in his paper, "Zur Elektrodynamik bewegter Korper," *Annln. Phys.*, 17:891–921 (1905). The methods of correction of star places for aberration are given fully in the *Explanatory Supplement to the Astronomical Ephemeris and the American Ephemeris and Nautical Almanac* (1961); a full account of early work on aberration and a tabulation of early data is given by K.A. KULIKOV, *Fundamental Constants of Astronomy* (1964; orig.

pub. in Russian, 1956). An account of the 21st International Astronomical Union symposium, May 1963 is ***The System of Astronomical Constants,*** ed. by J. KOVALEVSKY (1965).

(D.E.Bl.)

Lighthouse

Lighthouses, the centuries-old function of which is to provide the mariner with an identifiable seamark by day and by night, giving him positional information and warning him of a hazard, have been influenced by the continuing advance of technology. More advanced and sophisticated aids, such as radio navigation systems, may be considered part of the subject treated in this article.

HISTORY

Lighthouses of antiquity. The forerunners of lighthouses proper were beacon fires, kindled on hilltops, the earliest references to which are contained in the *Iliad* and the Odyssey. The first authenticated man-made lighthouse was the renowned Pharos of Alexandria, which stood some 350 feet high. The Romans erected many lighthouse towers in the course of expanding their empire, and by AD 400 there were some 30 in service from the Black Sea to the Atlantic. These included a famous lighthouse at Ostia, the port of Rome, completed in AD 50, and lighthouses at Boulogne and Dover. A fragment of the original Roman lighthouse at Dover survived in the 1970s.

The Phoenicians, trading from the Mediterranean to Great Britain, marked their route with lighthouses, the present structure at La Coruña, Spain, being close to the site of an ancient Phoenician lighthouse (see Figure 1, right). These early lighthouses had wood fires or torches burning in the open, sometimes protected by a roof. After the 1st century AD, candles or oil lamps were used in lanterns with panes of glass or horn.

(Left) Ace Distributors Ltd., (Right) Douglas B. Hague

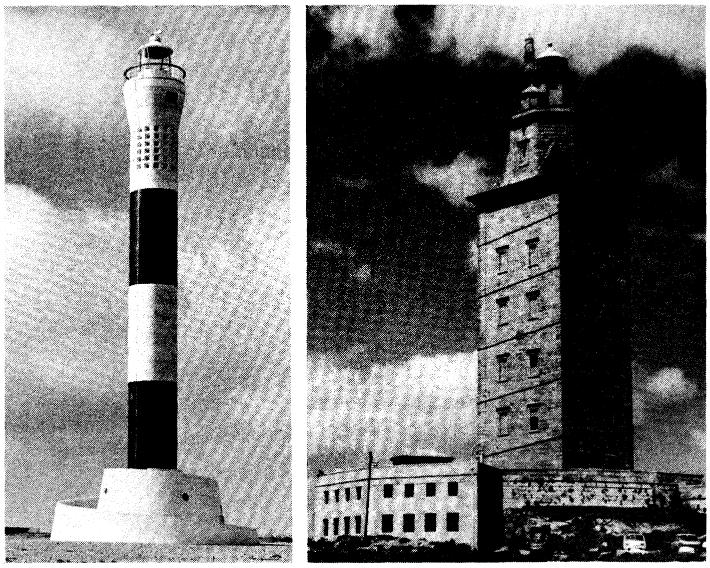


Figure 1: Shore stations.
(Left) Dungeness Lighthouse, Dunge Ness, England, a 135-foot tower constructed in 1981 from precast concrete rings. (Right) Tower of Hercules, at La Coruña, Spain, a 185-foot Roman lighthouse of the 2nd century AD, the exterior restored at the end of the 18th century.

Medieval lighthouses. The decline of commerce in the Dark Ages halted lighthouse construction until the revival of trade in Europe about AD 1100. The lead in establishing new lighthouses was taken by Italy and France. By 1500, references to lighthouses became a regular feature of books of travel and charts. By 1600, at least 30 major beacons were established.

Lanterna of Genoa

These early lights were similar to those of antiquity, burning mainly wood, coal, or torches in the open, although oil lamps and candles were also used. A famous lighthouse of this period was the Lanterna of Genoa, probably established about 1139. It was rebuilt completely in 1544 as the impressive tower that remains a conspicuous seamark today. The keeper of the light in 1449 was Antonio Columbo, uncle of the Columbus who crossed the Atlantic. Another early lighthouse was built at Meloria, Italy, in 1157, which was replaced in 1304 by a lighthouse on an isolated rock at Leghorn. In France the Roman tower at Boulogne was repaired by the Emperor Charlemagne in AD 800. It lasted until 1644, when it collapsed due to undermining of the cliff. The most famous French lighthouse of this period was one on the small island of Cordouan in the estuary of the river Gironde, near Bordeaux. The original was built by Edward, the Black Prince, in the 14th century. In 1584 Louis de Foix, an engineer-architect, undertook the construction of a new light, which was one of the most ambitious and magnificent achievements of its day. It was 135 feet in diameter at the base and 100 feet high, with an elaborate interior of vaulted rooms, richly decorated throughout with a profusion of gilt, carved statuary, and arched doorways. It took 27 years to build, due to subsidence of the apparently substantial island. By the time the tower was completed in 1611, the island was completely submerged at high water. Cordouan thus became the first lighthouse to be built in the open sea, and the true forerunner of such rock structures as the Eddystone Lighthouse.

The influence of the Hanseatic League helped increase the number of lighthouses along the Scandinavian and German coasts. At least 15 lights were established by 1600, making it one of the best lighted areas of that time.

During this period there was a fairly extensive use of lights exhibited from chapels and churches on the coast as a substitute for lighthouses proper, particularly in Great Britain.

Winstanley's wooden tower at Eddystone

The beginning of the modern era. The development of modern lighthouses can be said to have started from about 1700, when improvements in structures and lighting equipment began to come more rapidly. In particular, that century saw the first construction of towers fully exposed to the open sea. The first of these was Henry Winstanley's 120-foot-high wooden tower on the notorious Eddystone Reef, off Plymouth. Although anchored by 12 iron stanchions laboriously grouted into exceptionally hard red rock, it lasted only from 1699 to 1703, when it was swept away without a trace in a storm of exceptional severity; its designer and builder, on the lighthouse at the time, perished with it. It was followed in 1708 by a second wooden tower, constructed by John Rudyerd, which was destroyed by fire in 1755. Rudyerd's lighthouse was followed by J. Smeaton's famous masonry tower in 1759. Smeaton, a professional engineer, embodied an important new principle in its construction whereby masonry blocks were dovetailed together in an interlocking pattern. He also devised the curved hyperbolic profile, which became the classic design for the world's lighthouse builders. It was later modified to include a solid cylindrical base to break the main force of the sea and reduce the tendency of waves to sweep up the sides.

Due to the undermining of the foundation rock, Smeaton's tower had to be replaced in 1882 by the present lighthouse, constructed on an adjacent part of the reef by Sir James Douglass, engineer in chief of Trinity House. The upper portion of Smeaton's lighthouse was dismantled and rebuilt on Plymouth Hoe, where it still stands as a monument; the lower portion or "stump" can still be seen on the Eddystone Reef. Following the Eddystone, masonry towers were erected in similar open sea sites, which include the Smalls, off the Welsh coast; Bell Rock in Scotland; South Rock in Ireland; and Minots Ledge off Boston, Massachusetts. The first lighthouse of the American continent, built in 1716, was on the island of Little Brewster, also off Boston. After about 1775, lighthouse construction spread rapidly. By 1820 there were an estimated 250 major lighthouses in the world.

MODERN LIGHTHOUSES

Construction. While masonry and brick continue to be used, particularly for shore lighthouses, concrete and steel are the most favoured forms of construction, particularly offshore.

Structurally well-suited and reasonably cheap, concrete lends itself to aesthetically pleasing designs for shore-based lighthouses. The new lighthouse erected at Dunge Ness, England, in 1961 is a notable example of how concrete can provide strength without great mass. The slender 135-foot tower, rising from a spiral base, is admirably proportioned architecturally (see Figure 1, left).

Submerged caisson construction

Modern construction methods have considerably facilitated the building of lighthouses in the open sea. On soft ground, the submerged caisson method is used, a system first applied in 1885 to the building of the Roter Sand Lighthouse in the estuary of the Weser River in Germany, and then to the Fourteen Foot Bank light in the Delaware Bay, Delaware. With this method, a steel caisson or open-ended cylinder, perhaps 40 feet in diameter, is positioned on the seabed. By excavation of sand, it is sunk into the seabed to a depth of possibly 50 feet. At the same time, extra sections are added to the top as necessary so that it remains above high water level. The caisson is finally pumped dry and filled with concrete to form a solid base on which the lighthouse proper is built.

Two other main types of construction are used for offshore lighthouses where the seabed is reasonably firm and level. The first utilizes concrete to build a "float out" lighthouse: a cylindrical tower on a large hollow concrete base, which can be 50 feet in diameter. The tower is constructed in a shore berth, towed out to position, and then sunk to the seabed where the base is finally filled with sand. Weighing 5,000 tons or more, these towers rely on their weight for stability and require a levelled, prepared seabed. For greater stability during towing, the cylindrical tower itself often consists of two or more telescopic sections, raised to full height by hydraulic jacks after being founded on the seabed. This design has been pioneered largely in Sweden, where at least eight have been constructed.

Texas Towers

The other form of construction, using open steelwork, is largely based on the design of offshore oil and gas wells, including the so-called Texas Towers. A typical lighthouse of this type has an upper deck containing the plant, the equipment, and personnel accommodation. It is usually about 60 square feet, with a short tower on one corner to support the light. The deck also can provide a helicopter landing platform. The accommodation deck is supported on four to eight tubular steel piles, about two feet in diameter, driven into the seabed to a penetration of 150 feet or more. Before the piles are driven, a tubular steel braced framework, known as a "jacket" or "template," is placed in position on the seabed. The piles are driven down through the vertical tubular members of the jacket, which act as guides. Concrete grout is then forced into the annular space between pile and jacket tube. The deck is usually fabricated on shore, taken out on a barge, lifted into position on the substructure by floating crane, and welded on. The U.S. has built about 15 light towers of this type to replace lightships. The most recent is Ambrose Light off New York.

Illurninants. Wood fires were not discontinued until 1800, though after c. 1550 coal, a more compact and longer burning fuel, was increasingly favoured, particularly in northwestern Europe. A lighthouse in those days could consume 300 tons or more of coal a year. In full blaze, the coal fire was far superior to other forms of lighting, preferred by mariners to oil or candles. The disadvantage of both coal fires and early oil lamps and candles was the prodigious amount of smoke produced, which resulted in rapid blackening of the lantern panes, obscuring the light.

The Argand lamp

In 1782 a Swiss scientist, Aimé Argand, invented an oil lamp whose steady smokeless flame revolutionized lighthouse illumination. The basis of his invention was a circular wick, with a glass chimney that ensured an adequate current of air up the centre and the outside of the wick for even and proper combustion of the oil. Eventually, Argand lamps with as many as ten concentric wicks were designed. These lamps originally burnt fish oil, later vegetable oil, and by 1860 mineral oil was in general use. The Argand lamp became the principal lighthouse illuminant for over 100 years. A few are still in use.

In 1901, Arthur Kitson invented the vaporized oil burner, which was subsequently improved by David Hood and others. This burner utilized kerosene vaporized under pressure, mixed with air, and burnt to heat an incandescent mantle. The effect of the vaporized oil burner was to increase six times the power of former oil wick lights. The principle is widely used for cooking stoves, pressure lamps, etc.

Early proposals to use coal gas at lighthouses did not meet with great success. A gas-making plant at the site was usually impracticable and most of the lights were too remote for a piped supply. The use of acetylene gas, generated in situ from calcium carbide and water, increased following the discovery of the dissolved acetylene process, which by dissolving the acetylene in acetone made it safe to compress for storage.

Use of acetylene gas

Acetylene gas as a lighthouse illuminant had a profound influence on the advancement of lighthouse technology, mainly through the work of Gustaf Dalén of Sweden, who pioneered its application between 1900 and 1910. Acetylene produced a light equal to that of oil, burnt either as an open flame or mixed with air in an incandescent mantle. Its great advantage was that it could be readily controlled; thus for the first time automatic unattended lights were possible. Dalén devised many ingenious mechanisms and burners, operating from the pressure of the gas itself, to exploit the use of acetylene. Most of the equipment he designed is still in general use today. One device is an automatic mantle exchanger that brings a fresh mantle into use when the previous one burns out. Another, economizing on gas, was the "sun valve," an automatic day–night switch capable of extinguishing the light during the day. The switch utilized the difference in heat-absorbing properties between a dull black surface and a highly polished one, producing a differential expansion arranged by suitable mechanical linkage to control the main gas valve.

The acetylene system facilitated the establishment of many automatic unattended lighthouses in remote and inaccessible situations, normally only requiring an annual visit to replenish the storage cylinders and overhaul the mechanism. Acetylene equipment is still widely used today because of its simplicity, robustness, and ease of maintenance. Liquefied petroleum gas, such as propane, is also used.

Electric illumination in the form of carbon arc lamps was first employed at lighthouses at an early date, even while oil lamps were still in vogue. The first of these was at Dunge Ness, England, in 1862, followed by a number of others. The majority of these, however, were eventually converted to oil, since the early arc lamps were difficult to control and costly to operate. In 1913, the Helgoland Lighthouse was equipped with arc lamps and searchlight mirrors to give a light of 38,000,000 candlepower, the most powerful lighthouse in the world at that time.

Electric-filament lamps

The electric-filament lamp, which came into general use in the 1920s, is the standard electric illuminant for lighthouses today. The light output is some ten times that of the vaporized oil burner. Electric lamps in common use range from a maximum of about three kilowatts down to as little as six watts or so for small buoy and beacon lights. When a specially powerful light is required, modern arc lamp equipment is employed. This is usually the compact type, with the arc enclosed in a glass bulb to make the lamp replaceable. While mercury arc lamps are used, they have a distinct bluish colour. The xenon arc lamp, originally developed for colour film work, is preferred for the intense white light it produces. The most powerful lighthouse in the world today, at Creach, Ile D'Ouessant, France, is a carbon arc installation. An augmented light, giving a beam of 500,000,000 candlepower, is brought into service only during fog or misty weather.

Various lighthouse authorities are currently experimenting with the electronic xenon flash tube, a novel type of lamp used in photography. This lamp produces a very short flash of a few millionths of a second duration of extremely high intensity, imparting a distinctive character to the light and making it particularly conspicuous against other background lights.

The catoptric system

Optical equipment. With the advent of the Argand lamp, a reliable and steady illuminant, it became possible to develop effective optical apparatus for increasing the intensity of the light. In the first equipment of this type, known as the catoptric system, paraboloidal reflectors concentrated the light into a beam. William Hutchinson of Liverpool in 1777 produced the first practical mirrors for lighthouses, consisting of a large number of small facets of silvered glass, set in a plaster cast molded to a parabolic form. More generally, shaped metal reflectors were used, silvered or highly polished. These were prone, however, to rapid deterioration from heat and corrosion; the glass facet reflector, although not as efficient, lasted longer. The best metallic reflectors available in 1820 were constructed of heavily silvered copper in the proportion of six ounces of silver to one pound of copper (compare one-half ounce of silver to one pound of copper commonly used for plated tableware of the period). The cleaning cloths were kept for subsequent recovery of the silver. These mirrors could increase the intensity of an Argand lamp, nominally about five candlepower, almost 400 times.

Although the mirror could effectively concentrate the light into an intense beam, it was necessary to rotate it to make it visible from any direction. This produced the now-familiar revolving lighthouse beam, with the light appearing as a series of flashes.

Mariners were not favourably disposed to these early flashing lights, contending that a fixed steady light was essential for a satisfactory bearing. However, the greatly increased intensity and the advantage of using the pattern of flashes to identify the light gradually overcame their objections. The first revolving beam lighthouse was at Carlsten, near Marstrand, Sweden, in 1781.

The Fresnel dioptric system

In 1828 Augustin Fresnel of France produced the first apparatus using the refracting properties of glass, now known as the dioptric system. On a lens panel he surrounded a central bull's-eye lens with a series of concentric glass prismatic rings. The panel collected light emitted by the lamp over a wide horizontal angle and also the light that would otherwise escape to the sky or to the sea, concentrating it into a narrow, horizontal pencil beam. With a number of lens panels rotating around the lamp, he was then able to produce several revolving beams from a single light source, an improvement over the mirror that produces only a single beam. To collect more of the light wasted vertically, he added triangular prism sections above and below the main lens, which both refracted and reflected the light. By doing this he considerably steepened the angle of incidence at which rays shining up and down could be collected and made to emerge horizontally. Thus emerged the full Fresnel catadioptric system, the basis of all lighthouse lens systems today. To meet the requirement for a fixed all-round light, Fresnel modified his principle by producing a cylindrical drum lens, which concentrated the light into an all-round fan beam. Although not as efficient as the rectangular panel, it provides a steady, all-round light. Small drum lenses, robust and compact, are widely used today for buoy and beacon work, eliminating the complication of a rotating mechanism.

Prior to Fresnel's invention the best mirror systems could produce a light of about 20,000 candles, using an Argand lamp. The Fresnel lens system increased this to 80,000 candles, roughly equivalent to a modern automobile head lamp; with the pressure oil burner, intensities of up to 1,000,000 candlepower could be achieved.

For a light of this order the burner mantle will measure four inches in diameter. The rotating lens system would have four large Fresnel glass lens panels, 12 feet high, mounted about four feet from the burner on a revolving lens carriage. The lens carriage would probably weigh five tons, about half of it being the weight of glass alone. The rotating turntable floats in a circular cast iron trough containing mercury. With this virtually frictionless support bearing, the entire assembly can be smoothly rotated by weight-driven clockwork. If the illuminant is acetylene gas, the lens rotation can be driven by gas pressure. Installations of this type are still in common use, although many have been converted to use an electric lamp with an electric-motor drive. Modern lens equipment of the same type is much smaller, perhaps 30 inches high, mounted on ball bearings and electric-motor driven. With a 250-watt lamp, illumination of several hundred thousand candlepower can be readily obtained. Lens panels are now moulded in transparent plastic (Perspex), which is lighter and cheaper. Drum lenses are also moulded in plastic. In addition, with modern techniques high-quality mirrors can be produced easily and cheaply.

Intensity, visibility, and character of lights. The candlepower of illumination is expressed in terms of the international unit, the candle (also called candela). Intensities of lighthouse beams vary from thousands to millions of candles. The range at which a light can be seen depends upon atmospheric conditions and elevation. Since the geographical horizon is limited by the curvature of the earth, it can be readily calculated for any elevation by standard geometrical methods. In lighthouse work the observer is always assumed to be at a height of 15 feet, although on large ships he may be 40 feet above the sea. Assuming a light at a height of 100 feet, the range to an observer at 15 feet above the horizon will be about 18 miles. This is known as the geographical range of the light. In clear weather a light of 10,000 candles will be visible at 18 miles.

Known as the luminous range of the light, this distance is the limiting range at which the light is visible, in the prevailing atmospheric condition, disregarding limitations due to its height and the earth's curvature. A very powerful light, low in position, could thus have a clear-weather luminous range greater than that when first seen by the mariner on the horizon. Powerful lights can usually be seen over the horizon because the light is scattered upward by particles of water vapour in the atmosphere; the phenomenon is known as the loom of the light.

Atmospheric conditions have a marked effect on the luminous range of lights and are defined in terms of a transmission factor expressed as a fraction or a percentage up to a maximum of unity or 100 percent, assuming a perfectly clear atmosphere, never attained in practice. Clear weather in the British Isles corresponds to about 80 percent transmission, but in tropical regions it can rise to 90 percent, increasing the luminous range of a 10,000-candle light from 18 to 28 miles. Conversely, in mist or haze at about 60 percent transmission, a light of 1,000,000 candles would be necessary to maintain a luminous range of 18 miles. In dense fog, with the visibility down to 100 yards, a light of 10,000,000,000 candles could scarcely be seen at one-half mile. Because average clear weather conditions vary considerably from one region of the world to another, luminous ranges of all lighthouses by international agreement are quoted in an arbitrary standard clear-weather condition corresponding to a daytime meteorological visibility of 10 miles, or 74 percent transmission. This is known as the nominal range of a light. Mariners use nautical conversion tables to determine the actual luminous range in the prevailing visibility.

Sector lights

Fixed lights are still used in port, harbour, and estuarial areas where they give the mariner directional information by showing red or green over sharply defined sectors. Known as range or sector lights, they are often shown as fixed lights subsidiary to the main flashing light of the lighthouse. Another range system, sometimes called leading lights, consists of two lights at different elevations, about a half-mile apart. The mariner steers to keep the two lights in line one above the other.

Most lighthouses rhythmically flash or eclipse their lights to provide an identification signal. The particular pattern of flashes or eclipses is known as the character of the light, and the interval at which it repeats itself is called the period. International agreement restricts the number of different characters that can be used, through the International Association of Lighthouse Authorities in Paris, to which the majority of maritime nations belong. The regulations are too lengthy to quote in full, but essentially a lighthouse may display a single flash, regularly repeated at perhaps 5-, 10-, or 15-second intervals. This is known as a flashing light. Alternatively, it may exhibit groups of two, three, or four flashes, with a short eclipse between individual flashes and a long eclipse of several seconds between successive groups. The whole pattern is repeated at regular intervals of 10 or 20 seconds. These are known as group flashing lights. In another category, "occulting" lights are normally on and momentarily extinguished, with short eclipses interrupting longer periods of light. Analogous to the flashing mode are occulting and group-occulting characters. A special class of light is the isophase, which alternates eclipses and flashes of exactly equal duration.

The daymark requirement of a lighthouse is also important; lighthouse structures are painted to stand out against the prevailing background. Shore lighthouses are usually painted white for this purpose, but in the open sea or against a light background conspicuous bands of contrasting colours, usually red or black, are utilized.

Sound signals. Periods of bad visibility led fairly early to the idea of a supplementary audible warning. At first, sound signals were explosive, created by a cannon, or bells, Both were being used in the 1970s.

The explosive can be heard up to four miles. Charges are attached to a jib arm above the lighthouse lantern and detonated electrically at regular intervals of from two to five minutes. Sometimes the charges incorporate magnesium to give a bright flare. Often bells up to one ton in weight are arranged to sound at predetermined intervals. At times this action is automatic, with the use of a piston-operated striker powered by a compressed gas cylinder.

Fog signals

Fog signals that depend upon compressed air for their operation and emit a continuous note are most effective. The Reed operates like a reed musical wind instrument. It emits a rather high-pitched note at about 500 hertz. Although efficient, it is limited in power and not much used. The Siren has a slotted revolving rotor and fixed slotted stator at the throat of a suitable horn. The rotor is driven by an electric motor, or by the compressed-air supply, which is chopped by the rotating member. The Diaphone works on the same principle but uses a reciprocating slotted piston in a cylinder with matching ports. The Tyfon has a vibrating metal diaphragm operated by a system of valves and differential air pressure. Sirens, Diaphones, and Tyfons produce a low note of about 150 hertz.

The larger sizes of Diaphones are the most powerful fog signals in existence, with ranges up to eight miles. They consume as much as 40 cubic feet of air a second during blast, however, and require a large and powerful compressing plant of 50 horsepower or more. The Tyfon is an efficient, compact unit with low air consumption. It is common practice to mount a number of emitters in a vertical column, which concentrates the propagation of the sound in a horizontal direction and minimizes wasteful dispersion of sound vertically. Such an array can have an audible range of four to five miles.

With fog signals capable of emitting a continuous note, the station is distinctly identified, as with lights, by a predetermined number of short blasts regularly repeated. The Diaphone is much favoured by mariners because it has a distinctive sound, with a short, low-pitched characteristic grunt at the end of each blast. A compressed-air fog signal installation at a lighthouse can involve a formidable amount of machinery such as air compressors and storage reservoirs for the air, valves, piping, etc. Operating air pressures range from 30 to 60 pounds per square inch.

Electric fog signals, sometimes called Nautophones, are of the vibrating diaphragm type. A metal plate between the pole pieces of a magnet vibrates because alternating current passes through windings around the magnet. Nautophones usually produce a note of about 300 hertz. The most powerful emitters can handle one kilowatt of power, supplied from special alternators or solid-state, electronic-drive units. It is common practice to arrange these emitters in vertical columns, as for the Tyfon. Typical installations have a total power of four to six kilowatts and a range of three to four miles.

The propagation of sound in the open air is somewhat haphazard due to the vagaries of atmospheric conditions, which greatly affect its acoustic properties. Humidity, turbidity, and temperature all have an effect, and layers of air of different temperature can deflect the sound up or down. Ranges of fog signals therefore can vary greatly from day to day and quoted figures must be treated with caution, as it is impossible to guarantee performance.

A useful system around 1930 relied on the transmission of sound waves through the water, giving greater range and consistency. Underwater fog signals of this type were virtually abandoned but they were again being seriously reconsidered in the light of the modern technology of sonar (*q.v.*).

Radio aids. Sophisticated and complex radio-navigation systems such as Decca and Loran are not properly within the field of lighthouses, though in some countries they are operated as part of the lighthouse administration (see NAVIGATION).

Radio and radar beacons

Two aids that are strictly complementary to a lighthouse are the radio and radar beacons. Many lighthouses are fitted with medium-frequency radio-beacon equipment operating around 300 kilohertz. These transmit a characteristic signal lasting one or two minutes, with the station identification repeated continuously in Morse Code. A long dash of 20–25 seconds during the transmission enables a ship to take a bearing with its radio direction finder. Depending on the sector of the world, transmissions vary from five minutes every hour to continuously in busy waters. In Europe where there are a large number of radio-beacons, they are arranged in groups of three on one frequency to avoid interference. Each station in a group transmits for one minute of every three. This enables the ship to take three bearings in quick succession and to obtain a fix by triangulation.

Radar beacons are known as racons. Their purpose is to increase the strength of the reflected radar pulse from a seamark such as a lighthouse, lightship, or buoy and to enhance its presentation on the ship's radar screen. They also identify it positively from other echoes on the radar picture. In the passive method, polyhedral clusters of metal sheets are arranged to reflect as much as possible of the radar pulse emitted by the ship's radar. The racon is an active device, consisting of a small microwave transmitter-receiver. The radar pulse is picked up by the receiver and retransmitted in an amplified form. It is also processed so that it appears in an unambiguous form on the radar screen, usually as a short bright line immediately behind the target that can be broken up into an arrangement of dots and dashes to provide better identification.

Lightships and buoys. *Lightships.* Lightships originated in the early 17th century, arising from the need to establish seamarks in positions where lighthouses were at that time impracticable. The first lightship, established in 1732 at Nore Sand in the Thames Estuary, was rapidly followed by others. These early vessels were small converted merchant or fishing vessels showing lanterns suspended from crossarms at the mast head. Not until 1820 were vessels built specifically as lightships.

A modern lightship is constructed of steel, measures about 120 feet in length, and has a 25-foot beam (see Figure 2, left). A crew of seven is usual. The lightship carries a full range of lighthouse aids, including a powerful light, a compressed air or electric fog signal, radio beacon, and racon. The light is usually revolving-mirror equipment to save weight, and intensities up to 500,000 candles are common. Since a lightship rolls and pitches in the sea, lighting equipment is gimbal-mounted on a pendulum-stabilized platform to keep the light beams horizontal. Light vessels do not usually carry any means of propulsion; they are moored by a single chain and anchor from the bow. There have been a number of unmanned lightships, usually in the form of a boat-shaped float fitted with automatic acetylene light and gas-operated bell that are adequate in sheltered waters and where high-powered illumination is not required. A recent development is a large 40-foot diameter unmanned buoy, originally developed for oceanographic purposes, that can survive the worst open-sea conditions (see Figure 2, right). It carries an automatic diesel-engine power plant for its powerful light and fog signal.

Buoys. Buoys delineate channels in estuaries; approaches to ports and harbours; and mark isolated dangers, wrecks, and local areas of special significance. Con-

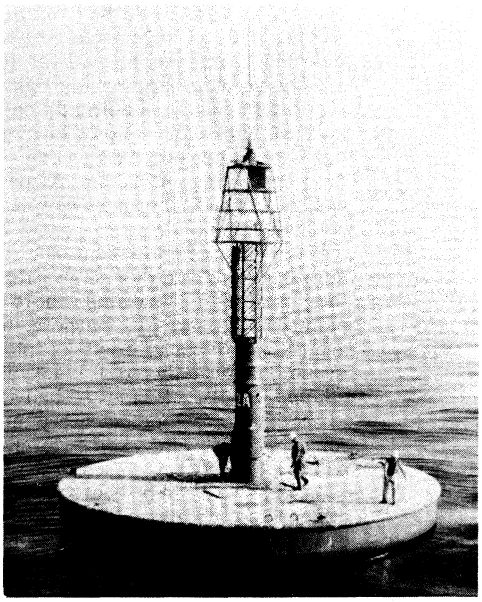


Figure 2: Modern *lightship* and *buoy*.
(Left) the "San **Francisco,"** a **manned** 128-foot-long lightship anchored off the entrance to the Golden Gate, San **Francisco,** with a mast top light visible for a maximum of 14 miles. (Right) The Scotland Sea Buoy, a 40-foot-diameter platform designed for unattended operation, off Sandy Hook. N.J.; its operating systems are monitored by radio.
By courtesy of the U.S. Coast Guard

structed usually of ½-inch steel plate, they vary in size from five to ten feet in diameter and from one to nine tons in weight. Buoys are moored to a two- or three-ton concrete or cast-iron sinker by a single length of chain, which is ordinarily about three times as long as the depth of water at the mooring location. Smaller sized buoys recently have been manufactured in glass-reinforced plastic (glass fibre). Buoys are inspected at regular intervals and removed periodically for cleaning, painting, renewal of moorings, etc.

Cardinal and Lateral systems

Colour, shape, appearance of a buoy (*i.e.*, its daymark), and the character of the light, if fitted, convey information to the mariner. Buoys must conform to one of two internationally agreed systems of buoyage, drawn up by the League of Nations in 1936. Known as the Cardinal and Lateral systems, they designate specific shapes and colours for the various buoy positions. With the Cardinal system the form of the buoy indicates the bearing of the significant feature it marks; with the Lateral system buoys are laid along routes or channels to indicate on which side of the buoy the ship should pass.

In the more widely used Lateral system there are three main daymark shapes: conical, cylindrical or can, and spherical, and two main colours, black or red, alternated with white or plain. Lights are white, green, or red, showing specified characters. For additional differentiation, topmarks in the form of can, cone, sphere, diamond, tee, or cross can be mounted above the buoy. With unlighted buoys the body profile incorporates the daymark shape. The top mark, if any, is carried on a staff fixed to the top of the buoy. Lighted buoys have a cylindrical body with the daymark shape carried on a superstructure. The light and sometimes a radar reflector are mounted above the superstructure. The buoy measures about 14 feet from keel to top of superstructure, carrying the light about nine feet above the sea. In high focal plane buoys an extra high superstructure increases the range of the light. A long, cylindrical tail tube attached to the bottom of the hull promotes stability and makes the buoy about *36* feet from top to bottom.

Lights can be acetylene, propane, or electric. A small drum lens six to eight inches in diameter provides a light of 100 to 200 candles with a clear-weather range of five to six miles. With electric lights a 12-volt, 6- or 12-watt lamp is typical.

Gas-storage cylinders or electrical batteries, enclosed within the cylindrical hull, power the buoy lights for up to two years without attention. The light can be extinguished during the day by a photoelectric day-night switch or a "sunvalve," in the case of acetylene equipment. Buoys are sometimes fitted with small electric fog signals with a range of about a half-mile. Bells can be actuated in a random fashion by the motion of the buoy or regularly by a striker operated from compressed carbon dioxide. Bouys also can be fitted with whistles, actuated by the motion of the sea, which draws air into a central tube and expels it.

Automatic lighthouses. A fair degree of automation was first achieved with the acetylene-gas system. Today, with the increasing use of electrical equipment, modern industrial automation techniques are being applied. In the automatic lampchanger, from two to six electric lamps are carried on a rotating carriage. The lamps are arranged to bring a fresh replacement into position and focus when the filament of the lamp in use burns out. Lighthouses are costly to operate and maintain. If lightkeepers can be eliminated and maintenance visits reduced, considerable money can be saved.

Much effort has been devoted to the development of fog-detecting systems, capable of automatically starting the fog signal when the visibility falls. A number of satisfactory instruments are in use.

Lighthouses, lightships, and buoys are by their very nature located in remote or inaccessible situations where it may be difficult to provide electric power from a public supply. For lighthouses and lightships, power is often generated in situ from diesel generators. For buoys and minor lights replaceable electric batteries are used. Alternative power supplies are being developed to provide long periods of unattended service. Wind generators have been successfully used. Fuel cells and radioisotope generators, the latter offering 10 years' operation without refuelling, are also being investigated. Solar cells have been successfully used in a number of cases for small lights in extremely inaccessible positions. Electric generators that utilize the energy of the oscillating motion of the buoy in the sea have been tried with some success.

National lighthouse systems. *Lists of* **lights.** All lighthouse administrations publish Lists of Lights that are comprehensive catalogues of all lighthouses, lightships, buoys, and beacons under their control. These lists give the mariner the needed information regarding location and characteristics of the various lights. **The** major maritime countries publish light lists in several volumes, covering all the lights in the world. In the U.S. this is undertaken by the United States Coast Guard, and in Great

Britain by the Admiralty Hydrographic Department. These bodies also publish charts on a worldwide basis with similar information. Light lists and charts are kept up-to-date with information supplied by the lighthouse authority concerned. The latter publicize changes to lighthouses, lightships, and buoys by issuing Notices to Mariners, available to all ships' masters. Where a change occurs at short notice, for instance in the case of an emergency or breakdown, radio broadcasts notify mariners of the change.

Lighthouse administration. In most countries lighthouse administration comes under the appropriate department of the central government, financed as part of the government budget from taxation revenue. In many countries a number of lights, buoys, etc., are operated and maintained by dock and harbour authorities, often financed from harbour dues.

In the U.S., lighthouses and other aids are administered by the Aids to Navigation Division of the United States Coast Guard under the Federal Department of Transportation. Altogether the Coast Guard operates and maintains 500 lighthouses and nearly 40,000 other aids to navigation.

In Great Britain lighthouses in England and Wales are administered by the Corporation of Trinity House, a public corporation independent of the central government. The system is almost unique in that the service is financed by "light dues" levied on all shipping at every port in the United Kingdom based on the registered tonnage of the vessel. It evolved from the early medieval practice where lighthouses were operated by private individuals as a business enterprise under a dispensation purchased from the Crown, the granting of patents to erect and operate a lighthouse being vested in the Monarch of the day. Trinity House evolved from an early guild or fraternity of Pilots. In addition to its lighthouse activities it is also the United Kingdom Pilotage Authority.

The arrangement in its present form originated with the Merchant Shipping Act of Parliament in 1894, which abolished all privately owned lighthouses. It converted Trinity House into a public corporation with responsibility for all seamarks. Annual revenue from light dues is about £6,000,000; the Corporation maintains about 90 lighthouses, 34 lightships, and 600 buoys. Lighthouses in Scotland and Ireland are separately administered by the Northern Lighthouse Board and the Commissioners of Irish Lights. These are semi-autonomous bodies, financed from the same fund of light dues as Trinity House, which remains the ultimate approving authority for all United Kingdom sea marks. In nearly all other countries the lighthouse administration is either a government body in its own right or operates as part of or in association with a main department or ministry, as follows: West Germany—Ministry of Marine; France—Lighthouse Service in association with the Ministry of Public Buildings and Works; Sweden—National Administration of Shipping and Navigation; Norway—Directorate of Lighthouse Services; Canada—Aids of Navigation Division, Department of Transport; Japan—Navigational Aids Department, Maritime Safety Agency; and Italy—Inspectorate of Lights in conjunction with the Public Works Department and Ministry for the Mercantile Marine and General Navigation.

BIBLIOGRAPHY. F.A. TALBOT, *Lightships and Lighthouses* (1913), a well-illustrated book describing lighthouses of the world from antiquity to the early part of the 20th century; J.P. BOWEN, *British Lighthouses* (1947), a short but usefully comprehensive summary of British lighthouse practice from the early 19th century to 1947; G.R. PUTNAM, *Lighthouses and Lightships of the United States,* rev. ed. (1933), an historical survey of most of the major lighthouses and lightships of the U.S.; D.A. STEVENSON, *World's Lighthouses before 1830* (1959), and an edited version of a detailed journal kept by Robert Stevenson during a tour of English lighthouses, *English Lighthouse Tours of Robert Stevenson* (1946), classic works by a member of the famous family of Scottish lighthouse engineers; F. MAJDALANY, *The Red Rocks of Eddystone* (1959), a readable description of the building of the four Eddystone lighthouses, emphasizing the human and dramatic aspects of the work rather than the technical; H.P. MEAD, *Trinity House* (1947), a discussion of the origins and history of Trinity House, the lighthouse authority for England and Wales; UNITED STATES COAST GUARD, *Historically Famous Lighthouses* (1950), and *Coast Guard History* (1949), two pamphlets that provide a brief insight into the work of the U.S. Coast Guard and some of its more famous lighthouses.

(I.C.C.)

Lighting and Lighting Devices

Lighting in modern times derives almost entirely from electric light sources; lamps with a wire filament in an evacuated glass bulb or long fluorescent tubes. A building is lighted throughout to levels comparable with daylight, and selectively placed light is focussed on work of special difficulty. Lighting devices of many kinds have been used through the long ages since man learned to control fire. Evidence of torches has been found at archaeological sites; the cave paintings at Altamira in Spain and the Dordogne in France are perhaps the oldest.

From its beginnings, artificial light has had two distinct purposes: to permit visibility in natural darkness and to create visual effects. The development of lighting technology has been characterized by invention of light sources with greater output and efficiency. With the introduction of each light source, technology for its control has advanced. The candle produced the candelabra that enhances light with prismatic glass adornments. The oil lamp brought in its train the development of various reflectors to concentrate the light. The gas mantle gave rise to the first efforts at precise optical control with prisms and reflectors, the principles of which were later applied to the electric light bulb, which, in turn, because of its safety and cleanliness, permitted the first designs of decorative and architectural lighting fittings and lampshades. The last major advance in light sources, the fluorescent tube, gave rise to the luminous ceiling, a characteristic feature of modern buildings.

Two important aspects of the development of lighting technology in recent years can be identified. First, lighting technologists and specialists in human vision have collaborated, and a science of lighting and seeing has grown up, based on the idea that artificial lighting must be not only decorative but efficient and comfortable in terms of human sight, as well. Second, the lighting technologist has learned from the architect, who has always concerned himself with natural lighting in buildings but rarely with its artificial counterpart. The lighting engineer has learned about the flow of light, modelling of people and objects, and the active creation of beauty by the interplay of light, form, and colour. Perhaps to these could be added a third development, the tardy recognition by building designers and engineers that large amounts of artificial light give off heat, which ignored can cause discomfort but controlled can heat a building. Understanding the relationship between light and heat has given rise to the concept of total environmental design. (For information on the physics of light, see LIGHT, and OPTICS, PRINCIPLES OF. The principles of human vision are covered in EYE AND VISION, HUMAN.)

Development of lighting technology

THE BEGINNINGS OF ARTIFICIAL LIGHTING

Until the development of electric lighting in the 19th century, all artificial lighting was based on some form of incandescence brought about by burning fossil fuels. Wood brands could give massive amounts of light but were dangerous and difficult to control. The introduction of vegetable and animal oils, waxes, and fats and the wick to control the rate of burning represented a major advance, gradually refined in a number of ways. There are numerous examples from every ancient civilization of lamps designed for utility, decoration, or both. Improvements were made in design and fabrication of light sources but little in their efficiency. In 18th-century England, for example, the labourers in the countryside still made artificial lighting from rushes and pork fat.

Rush light

Nonelectrical lighting devices. *Oil lamps.* During the 18th and 19th centuries, however, useful improvements were made in the design of oil lamps as a result of a better understanding of the physics of combustion. More reliable and powerful than its predecessors, the Argand oil lamp of 1784 used a cotton wick with a controlled current of air to obtain maximum oxygen for efficient combustion together with glass cylinders with or without simple optical control (bull's-eye lens) of the light output. The discovery of massive reserves of petroleum fuel, particularly in the United States, and of refining methods that produced an oil (kerosene or paraffin) that burned efficiently in an oil lamp, but without any danger of spontaneous combustion, made cheap, reliable, and safe lamps widely available and rendered the wax candle obsolete except for decorative purposes.

Gas light. Unquestionably the major development in artificial lighting during the 19th century was the introduction of methods for generating and distributing coal gas to all buildings in an urban community. The subsequent discovery of natural gas, again in the United States and Canada, reduced the cost to the fortunate citizens within reach of the natural-gas fields but did not materially alter the design of lighting equipment. The first development of coal-gas illumination is credited to William Murdock, a Scottish engineer working in Cornwall, who, in the tradition of many British inventors, tried his experiments on his own family first by lighting his own home by gas in full knowledge of the risk of explosion. No explosion occurred, and, a few years later, at the turn of the century, Murdock was commissioned to install gas lighting in London. The gas lighting in Pall Mall, installed in 1820, is reputed to be one of the oldest extant street-lighting installations in the world. For several decades, gas lighting made use of a simple burner in which the yellow light of the flame itself was the source of illumination. During the 1820s a new form of burner was introduced in which a controlled amount of air was admitted to the gas current, producing a high-temperature but nonluminous flame that heated a refractive, noncombustible material to a very high temperature. This became the source of light: the higher the temperature of the material, the whiter the colour and the greater the light output.

Welsbach gas mantle

The search for an optimum material that would be durable, infusible, and capable of being repeatedly heated to white heat, culminated, during the 1880s, in the development of the Welsbach mantle, which, in its final form, consisted of a network woven in cotton, of spherical, cylindrical, or linear form, impregnated with the salts of thorium and cerium. By suitable treatment these mantles could be made sufficiently strong to be transported and retailed. Once they were fitted to the lighting appliance, however, the material which gave them stability burnt off. and a slight jar could destroy the mantle. In spite of this disadvantage, the introduction of the Welsbach gas mantle resulted in gas lighting being employed everywhere within reach of the distributive network of coal- or natural-gas supply. The incandescent light from the thorium–cerium salts in the mantle was not truly white but was tinged a greenish colour with the luminous emissions of these salts, giving rise to the somewhat garish and ghostly appearance characteristic of gaslight. This was unfavourably contrasted with the pleasant, warm glow of the candle, and, as a result, the introduction of the gas mantle was also responsible for a revival in the use of candlelight as a luxury to be indulged in by the wealthy. Improvements in gas lighting itself, particularly by the use of pressure boosted above the safe level for the distribution mains, gave a considerable increase in lighting efficiency.

Coal gas was used only in urban areas in which the cost of distribution was low. This gave rise to the development of forms of local gas generation to confer the advantages of gas lighting in rural areas. Acetylene gas, generated by the dripping of water on calcium carbide, was a particularly valuable development, although its unpleasant smell gave rise to ventilation problems when used in domestic interiors. The natural acetylene flame, however, had a high luminous efficiency and a colour closely approximating natural daylight, which rendered it specially valuable in industrial processes requiring a high standard of fidelity to natural colours.

Electric lighting devices. *The* electric *lamp.* Despite the successful history of gaslight, modern lighting owes very little to anything that happened before the invention of the incandescent electric lamp about 1870. The incandescent lamp was not the first electric lamp; lighting devices employing an electric arc struck between electrodes of carbon had been developed and were in use earlier. Arc lamps were massive and complicated pieces of equipment that could be installed and maintained only by a skilled engineer. Used for street lighting, arc lamps had advantages, including reasonable reliability, high efficiency, and, above all, a pleasant colour, closely approximating natural sunlight. The light of the arc lamp was particularly kind to the colour of the human complexion and the stone of historic buildings. Partly for this reason, these devices were retained in dignified city settings—for example, in the City of London—long after more modern light sources had come into wide use.

Arc lamps

The importance attached to the invention of the incandescent filament lamp can be gauged by the degree of controversy which has arisen over attribution of its invention. Arguments and claims on both a national and individual basis have been made. While it is true that early lamps of a kind were made in France and Russia and that Sir Joseph Swan of England demonstrated and manufactured carbon-filament lamps of a sophisticated kind before anybody else, the true invention of electric lighting should be credited to Thomas Edison in the United States. To Swan, the lamp was an important technical and scientific development with a utilitarian purpose. To Edison, it was part of a grand scheme of electric lighting, which included the generation and distribution of electric power. The first incandescent lamps, both of Swan and Edison, consisted of a filament of carbon wire in an evacuated glass bulb, two ends of the wire being brought out through a sealed cap and thence to the electric supply. When the supply was connected, the filament glowed and by virtue of the vacuum, did not oxidize away quickly as it would have done in air. The commercial success of the lamps depended upon the production of a high degree of vacuum, and strength great enough for them to be transported without damage. Though these problems were solved by both Swan and Edison, it was Edison who made the major commercial development. Edison's firm arranged a historic demonstration of electric lighting in Menlo Park in 1879 and received contracts for generating plant, distribution networks, lamp outlets, and bulbs. It is important to realize that the whole of this concept, not merely the invention of the light bulb, was Edison's contribution. In particular, Edison employed the system of parallel rather than series installation, so that when a bulb failed, as often happened in early days, the rest of the installation did not go out as would have been the case with the series operation first used by Swan. In fact, the distribution system of electric lighting in use today is nothing more than a highly sophisticated version of Edison's original concept. Among early electric lighting installations was that at the Savoy Theatre in London, built for the Gilbert and Sullivan operas.

Edison's distribution system

Of the major improvements in electric lighting that followed Edison's first installations, the most important was development of metallic filaments, particularly the tungsten filament, now used almost exclusively in filament lamps. Tungsten was known to be particularly suitable for use in electric lamps, because, of all materials suitable for drawing into filament wires, it had the highest melting point. This meant that lamps could run at a higher temperature and therefore emit both whiter light and more light for the same electrical input than was possible with the less durable and less refractory carbon filament. Tungsten could not be used, however, until a process had been invented for drawing it into thin wire filaments; this development took place in the early 1900s.

From this time forward, light-bulb manufacture became a major industry; much money was spent on research in the physics of light production in the incandescent lamp in order to increase its efficiency, sturdiness, and life in use. Some of the leading technical physicists of the day were engaged by manufacturers, who profited considerably as steady improvements were introduced. The early tungsten lamps, like the carbon lamps, suffered from migration of filament molecules to the glass bulb, causing a blackening of the bulb, a loss in light output, and progressive thinning of the filament until it broke. The introduction of a small amount of inert gas (argon) reduced migration and enabled the filament to be run at a higher temperature, giving a whiter light, higher efficiency, and longer life. Further improvements, such as the development of the coiled filament and then the coiled-coil filament, improved tungsten-filament lamps to the point at which the manufacturer had to decide whether to market a lamp with almost infinite life, but low light output, running at a low colour temperature, or to sell the consumer a brighter and more efficient lamp with a relatively short life. The manufacturers settled the matter between themselves by agreeing to market a lamp that would last for an average of 1,000 hours, producing a light of colour considerably yellower than that of natural daylight but still whiter than that of an oil lamp or candle. This was judged to be the most acceptable compromise for the consumer, and in spite of occasional demands for lamps of longer life, little change has taken place over the past 40 years. The most recent development is the introduction of the tungsten–halogen-filament lamp, in which the presence of a small amount of iodine reduces the migration of tungsten, permitting a higher running temperature. Because of the chemical effect of the iodine on glass, the filament must be placed in a quartz rather than a glass bulb, increasing initial cost. Though these lamps have whiter colour and longer life, they have not yet been used for general domestic lighting. They are rapidly coming into universal use for automobile headlamps and other forms of light projectors calling for maximum light output from bulbs of minimum size.

Neon lights

Vapour-tubs **devices.** During the late 19th century, many physicists, particularly Sir William Crookes in England, experimented with methods of generafing radiation by striking an arc between electrodes in an evacuated tube to which small amounts of elemental vapour had been admitted. The physicist Georges Claude in Paris developed such a tube with neon gas as the filling; when a high voltage was applied to the two electrodes at either end of the tube, it emitted a deep red light whose value for decorative and advertising purposes was quickly recognized. Neon signs soon decorated the exteriors of commercial buildings in the cities of the world; experiments with other vapour fillings—particularly mercury vapour, which gave a blue light--enabled a variety of colours to be produced. The discharge tube, however, had little application in interiors until the development in the 1930s of the fluorescent tube, a long tube with a mercury-vapour filling, the inner wall of which was coated with a material which fluoresced white or near white when subjected to the radiation of the mercury discharge, and the quantity of light was multiplied a hundredfold. The lamp became a practical possibility for the illumination of interiors.

Somewhat before the fluorescent tube came into use, lamps with much shorter arcs, in which the vapour operated at a higher pressure, were introduced, including the high-pressure mercury-vapour-discharge lamp and the sodium-discharge lamp. Neither of these found much use for interior lighting, however, because of their colour rendering. They made people look like ghosts, and it was only for street lighting, where safety arguments could be advanced, that the mercury- or the sodium-discharge lamp was accepted. Application of these lamps for exterior lighting developed so rapidly that within 20 years almost all principal streets in cities of the world (excepting the United States) were lit by sodium-discharge lamps. The ghostly rendering of the human complexion was balanced by the pleasant appearance of the brick and stone facades of buildings. In the early 1950s, the pattern of lighting seemed clear: filament and fluorescent lamps for interiors, neon and similar strips for exterior advertising and signs, and sodium-discharge lamps for streets.

By the 1970s, however, developments suggested that this pattern might change radically. The mercury-discharge lamp, which, after a brilliant start, rapidly gave way to the sodium lamp, had undergone extensive development; new forms, becoming commercially available, suffered from fewer of the disadvantages of the parent. Such lamps may eventually find both internal and external use. Yet the day of the sodium lamp was not over, because equally new developments, particularly the high-pressure sodium lamp, had overcome to some extent the unpleasant colour rendering, and it was difficult to predict whether the mercury or the sodium lamp or both together would prevail. Lamp manufacturers continued to experiment with new devices, sometimes in the laboratory and sometimes with a commercial product. In this category might be included the xenon-discharge lamp, a high-powered source with colour almost identical to natural sunlight. The electroluminescent lamp, by virtue of which walls and ceilings could be made self-luminous, was once believed to be the lighting of the future, but has made little recent progress.

Optic fibres

Not all lamp development is concerned with general lighting on a massive scale. So much is now known about the physics of incandescent lighting, of discharge lighting, and of other forms of generating light, that any particular requirement can almost certainly be met by a lamp tailored to the purpose. Great lamps of massive light output have been made, using the arc principle, the vapour-discharge principle, and the incandescent-filament principle for lighthouses, searchlights, and the floodlighting of large areas, such as airport landing runways. Microscopic filament lamps have been created for insertion in the body to enable the surgeon, aided by suitable optical equipment, to make visual inspection to supplement other methods of diagnosis. The development of flexible optic fibres, thin wires of transparent plastic that "conduct" light by internal reflection from a suitable lamp at one end to the target at the other, have also been used for medical and other purposes. Other lamps for special uses include highly actinic lamps (*i.e.*, radiating energy in the visible and ultraviolet spectral regions and producing chemical changes thereby) for photography, colour television, and cinematography or lamps with emission in selected parts of the spectrum for industrial uses, infrared lamps for paint drying, and both infrared and ultraviolet lamps for therapeutic use. There are probably few branches of applied physics in which progress has been so rapid during the past 50 years as in the development of artificial light sources.

MODERN ADVANCES IN LIGHTING TECHNOLOGY

Lighting as an applied science was founded during the 1890s and 1900s when—with development of the incandescent-filament lamp, the carbon arc, and the maintenance for some years of the supremacy of the Welsbach mantle—the extension of the working day into the hours of darkness became feasible. A massive expansion in artificial-lighting technology resulted.

Lighting technology was advanced partly by unskilled and semiskilled inventors and mechanics, and partly by trained physicists and engineers who explored basic principles. The first group produced rapid results, while the second developed the theoretical knowledge upon which subsequent technology could be based.

Edison screw

Fixtures and voltage. The first requirement was to accommodate the light bulb where it was required on the ceiling, desk, or wall. The Edison-screw lamp terminal (male) and power outlet (female) probably had as important an effect upon the development of lighting for domestic and industrial use as any other single invention. All that was required to install electric lighting was for an electrician to affix the Edison-screw power outlet to the appropriate wall or ceiling surface with the necessary twin wires back to the supply terminal and then for the householder to screw into it a light bulb with an Edison-screw cap. Installation was safe, simple, and reliable. The various electric-power companies, Edison's being the pioneer, steadily refined and introduced new methods of power distribution that simplified electric lighting and made its use more safe and reliable. Individual light switches for each room and each lamp soon became sufficiently inexpensive to be universal. Controversy arose, not yet settled today, over the best method of distribution to avoid accident by electrocution. The same amount of light can be generated by a lamp carrying a small current (amperes) but subjected to a high electrical pressure (voltage), as would be given by a different lamp carrying a much heavier current but subjected to a lower voltage. It is much cheaper to distribute electricity at a high voltage, because the current can be less and hence the amount of expensive copper in the conductors can be reduced. On the other hand, the risk of death by electrocution is greater at the higher voltage and experience unfortunately showed that, in spite of instructions, people would occasionally touch the two terminals of the electrical supply together and subject themselves to the full electrical pressure. In the United States a decision was taken to supply households at no more than 110 volts. This meant that the distribution network had to be heavier and therefore more expensive. The United Kingdom, on the other hand, decided to use about 200 volts, thereby approximately halving the cost of domestic distribution but considerably increasing the risk. Over 70 years, neither country has seen fit to make a change. After a period of uncertainty, the whole of Europe followed the British rather than the American lead.

Safety devices. All forms of electric lamps are liable to failure, sometimes gradual, sometimes immediate, sometimes catastrophic. Gradual failure only results in loss of light which may often be imperceptible over a period. Modern discharge lamps, for example, last indefinitely but slowly lose their light output. Immediate failure, such as the breaking of the filament of an incandescent lamp, causes no more than temporary inconvenience. Sometimes, however, a filament lamp or a discharge lamp may fail, probably due to a manufacturing flaw, in such a way as to give rise to a surge of current against which the distribution system must be protected. Every circuit must have some form of fuse or contact breaker to cut off the current if it exceeds safe levels. With a fuse, one need only detect the cause of the fault, remedy it, and replace the fuse. The contact breaker is a switch which is tripped by an excess of current; it can be reset at once without providing another fuse. The past pattern was to use fuses in domestic circuits and contact breakers in larger installations where a professional electrician was available to detect the cause of failure before resetting the contact breaker. Modern usage relies on the circuit breaker. (For power generation and transmission, see ELECTRIC POWER.)

Directing light. The next stage in the development of lighting devices was to provide means by which the light from the electric lamp could be directed where it was wanted. At this point, the technology of lighting for seeing and lighting for effect parted company. In the first type of lighting, it is not desirable for the light source to be visible. Instead, maximum light is directed onto the work and none allowed to escape elsewhere. In lighting for effect, however, it is sometimes desirable to see the source for its brilliance; in particular, the light is directed to create effects, modelling, shadows, reflections, and high spots without considering whether people can see to read or work. The lighting in a church, for example, may be designed to enhance the appearance of the architecture.

Lighting for seeing and for effect

To achieve lighting for seeing, two principles of optics—reflection and refraction—were used, either separately or in combination. Light, when received at a smooth, polished surface, is reflected in a predictable and calculable way. Likewise, light received at the (incident) surface of a transparent medium, such as glass, is changed in direction (refracted), and, if the other (emergent) surface of the glass is not parallel to the first surface, the emergent ray of light is changed in direction. Thus, by suitable choice of incident and emergent surface, light can be refracted into any desired direction. The principle of lighting devices employing reflection or refraction is therefore to wrap such a reflecting device (mirror) or refracting device (lens or prism) around as much of the light source as possible, collecting all the light emitted in unwanted directions and redirecting it where desired.

Both forms of device can be used for a multiplicity of purposes. The searchlight reflector and the lighthouse refractor are both designed to concentrate the maximum intensity in the desired angle but for different purposes. The searchlight concentrates the light in order to illuminate and make visible a distant object. The lighthouse, on the other hand, concentrates the beam to permit the maximum visibility of the source itself from the greatest possible distance.

In the home, factory, or office, on the other hand, a narrow, concentrated beam is rarely required. The purpose is to illuminate the whole of a writing desk or an entire working area. Such lighting devices, therefore, must produce a wide but uniform beam, since it is irritating and fatiguing to work where light is bright at one point and dull at another. The lighting engineer aims at uniformity of illumination and uses reflector contours that calculation and experience indicate will give the most satisfactory result.

High skills were soon brought to bear on the design of optical controls for light, particularly in such areas as hospital operating theatres, where the maximum amount of entirely shadowless lighting is directed onto the surgeon's task; or in street lighting, in which light emitted from a small source is directed along a narrow band of street and sidewalk with perhaps 200 feet or 60 metres between adjacent lampposts. Street-lighting equipment in particular has made use of refraction as the controlling principle, in part because of its greater accuracy but chiefly because the refractor units could be made in the form of glass bowls or dishes that could be tightly sealed to the lamp-holding device to keep out weather and insects.

For interior lighting, however, the need for precise optical control progressively diminished as light became cheaper and more plentiful because of the more efficient generation and distribution of electric power, and because lamps became more efficient in producing light rather than heat. Interior lighting is now a matter of lighting the whole environment instead of only the working task. Fixtures have employed new methods of controlling light, such as the use of optical diffusion. Special techniques were developed, first in central Europe and subsequently elsewhere, that permitted a white opal glass to be blown into attractive shapes, such as spheres and cylinders, in which the lamp could be placed as part of the fixture and its light diffused to give a uniform white-light source of large size and pleasant appearance. The Bauhaus School of Architecture and Design in Germany under Walter Gropius combined with the skill of the central European glassmakers revolutionized interior lighting design during the 1930s. Other diffusing materials have since supplanted glass, and hence there exists an immense range of plastic diffusing light fixtures of all possible shapes and sizes. The reflecting light fixture still has its uses in general room lighting. A contemporary fashion among architects is to inset lighting into the ceiling and direct it through small holes, almost invisible during the day; this has led to development of ceiling-inset spotlights, which use all the old skills of the light-reflector designer.

Diffusing light

The lighting designer. The most essential tasks of the lighting designer are to spread adequate light on the work for efficient and comfortable seeing, and to light the environment so that people can see well and enjoy what they see. The determination of the amount of light for efficient and comfortable seeing is a matter for the specialist in physiological optics and in the relation between lighting and seeing. Experimental work in England, Germany, and the United States showed that visual performance could be analyzed into three basic factors: sharpness of vision (visual acuity), detection of contrast, and detection of movement. International agreement was eventually reached on a method by which necessary levels of lighting for specific tasks could be prescribed. By breaking down a visual task—for example, that of threading a needle—into its component parts, the amount of light necessary for efficient and comfortable seeing could be prescribed. This prescription forms the basis of the codes of recommended lighting practice that have been adopted in most countries.

The factors governing discomfort from bright light sources are also known and a glare prescription can be written, called a visual comfort index in the United States and a glare index in most other countries. Codes of lighting practice can therefore lay down the glare index based on the levels of tolerable brightness of lighting devices in their particular environment in order to avoid discomfort from glare.

Although this article is intended to be concerned primarily with the physical aspects of lighting design, in other words lighting technology or illuminating engineering, it is notable that lighting is concerned with human sensations; that is, human vision. It must be kept in mind that the eye can adapt to the kind of lighting in which it finds itself. The level of lighting by moonlight, for example, is one-1,000,000th of that by sunlight, and yet the human eye can adapt sufficiently to moonlight to perform many tasks.

Apparent brightness

This adaptation to light not only affects the range of brightness but also has a profound effect upon the sensation given by any particular physical amount of light. A surface of a certain brightness (luminance) seen in a dimly lit street at night, for example, will appear very bright, whereas a surface of the same brightness (luminance) seen in daylight looks like a very dark shadow. Lighting design must take this into account. Thus it is not physical brightness, but apparent brightness that determines the subjective effect. Though apparent brightness is the ultimate criterion for the designer, experience and existing technology let him undertake almost the whole of his lighting design in terms of physical measures.

Current lighting practice

The brief summary of current lighting practice below attempts to strike a balance among practices in different countries and by different schools of design. This is difficult; there is the "illuminating engineering" school of lighting, and there is the "architects or interior decorators" school, quite different in approach and results and almost irreconcilable. They accord most closely in Britain and are farthest apart in the United States.

INTERIOR LIGHTING

Interior lighting in most countries tends to follow similar patterns, dictated by the availability of light sources, by architectural fashion, and by the special lighting requirements of the task or the environment or both. The amounts of light now considered necessary for efficient and comfortable vision are very much higher than they were even ten years ago. Lighting levels have in fact increased up to 50-fold in the last 50 years. This is not because human sight has deteriorated. Ample evidence indicates that in spite of the fact that more people now wear spectacles than ever before, human sight remains much what it was in the past, and there is certainly no deterioration due to the use of artificial light. It is simply that people demand perfect vision and thus wear correcting glasses for minor visual defects that were previously neglected. In the same way, people want to be able to see excellently everywhere, whereas previously they were satisfied to move closer to the candle or other light source to read or perform other tasks.

Dwellings. Light installed by a skilled lighting designer and interior decorator can do much to create the character of an interior environment. In modest homes a central pendant fitting supplemented by one or two wall fittings or desk lamps, or lamp standards, remained in the 1970s the best way of achieving good lighting on sound visual principles with the greatest economy.

The eye functions most efficiently (with greatest visual acuity and comfort) when the task is lit to a slightly higher brightness than the immediate surroundings, which in turn should be slightly brighter than the general environment. This standard is usually achieved in a home when a moderate level of general lighting is provided from a central fixture and a high level of working light from a desk or floor lamp. Less satisfactory is the use of a desk lamp alone in a room, because this leaves the room too dark and gives rise to visual fatigue.

Recognizing this, the Illuminating Engineering Society of the United States, as long ago as the 1930s, designed a study lamp which provided both working light on the desk and general light in the room; this unit is probably still the most efficient lighting fixture for the economic lighting of working rooms in dwellings.

Fluorescent lighting has not yet found its way into homes to a great extent, apart from utility areas, such as kitchens and bathrooms. This is partly due to the large size of the lamp, which is out of scale with normal rooms in small houses, and partly to the fact that the light which it emits is less compatible with the social character of dwellings than the warmer, yellower filament lamp. People apparently do not want to prolong the colour of daylight into their social lives after dark.

Lighting for television viewing

Lighting for television viewing has been a source of some controversy, but it has been demonstrated that people with normal sight can light television-viewing rooms as they please without harming their eyes, though eye muscles may become fatigued more quickly in a darkened room. The colours of a colour-television screen are adulterated by too much white light falling upon the screen, and so, for watching colour television, the level of light should preferably be somewhat less than that for watching black-and-white. Lighting for television is more a matter of common sense and setting the light as the viewers want it, avoiding annoying reflections. Persons, with subnormal sight, however, may need care in the prescription of lighting for television viewing.

Schools. With changes in education that have taken place over the past 20 years, schools have been made more attractive and welcoming than ever before. This development has been most marked in England where school building and school lighting corresponds much more closely to that in homes than it does in other countries. In continental Europe and the United States, emphasis is still placed primarily on lighting the task. American schools are almost always lit by fluorescent lighting, inset in the ceiling or in the form of an overall luminous ceiling, providing shadowless lighting to a high level over the whole working space. In England, on the other hand, the emphasis is on the lighting of the environment. Continental Europe, with the exception of Scandinavia, tends to follow the American pattern.

Offices. Ceiling-mounted fluorescent lighting is used in offices to create a uniform level of light over the whole of the working area, allowing maximum flexibility in office arrangement. The levels provided range from about one fifth to one tenth of the light from a natural sky. This appears to be about the optimum required for work on white paper. Improvements in the colour of fluorescent lighting are constantly being made, and the distortions in the appearance of the human complexion have been largely overcome.

The concept of office landscaping, created in West Germany, has emphasized the need for a uniform level of lighting over the whole working space. The essence of the office-landscaping principle is to permit every task to be placed anywhere in the office space and planned in relation to the flow of work demanded, and, for this, uniform lighting is essential.

Lighting practice in offices nowadays sometimes makes use of the principle of permanent supplementary artificial lighting of interiors (PSALI) originating at the Building Research Station in England, in which the light from the windows and the light from the artificial-lighting system are planned together as part of an integrated whole. Before this, the artificial lighting was planned primarily for operation after dark, while the windows provided the working light during the daytime and had to be very large if light was to penetrate into a large deep office. PSALI means that windows no longer must provide all the working light during the daytime but are important simply because they provide a view and an awareness of the world outside. PSALI has been developed primarily in Britain and the Scandinavian countries and, to a lesser extent, in continental Europe. In the United States the contrary tradition has developed, of working by artificial lighting entirely, by day as by night, and even drawing blinds over the windows by day to eliminate glare.

Hospitals. Here, again, different approaches have been taken. American hospitals tend to be much more "institutional" in character, making little use of colour on the walls and hangings and providing high levels of uniform fluorescent lighting from ceiling-mounted fixtures and from fixtures placed immediately above the bedheads. British practice, on the other hand, employs much lower levels of light, on the theory that sick people often want to rest; colour is used extensively both on walls and in draperies. Until quite recently there was a strong prejudice in England against the use of fluorescent lighting in hospitals, and many modern hospitals are still lighted by filament bulbs. Hospital day spaces are designed to resemble a home both in the use of light and of colour, as are visitors' waiting rooms. The essential requirement in an operating theatre is that high-intensity light should be concentrated upon the operating area so that, no matter where the surgeon places his hands, no shadows are cast. For shadow-free lighting, one very large source is needed, which may either be a suspended lighting unit (the least costly solution) or a luminous ceiling as developed in France. A luminous ceiling may be an enormous elliptical reflector, with the light source at one focus and the operating area at the other, or it may consist of a multiplicity of small projectors concentrated on the operating area, possibly with an array of push buttons at the surgeon's control to change the direction and distribution of the light. It is important to light the whole theatre as well as the operating area, both to enable the assistant staff to work efficiently and to minimize visual fatigue on the part of the surgeon.

Lighting the operating theatre

Factories. In industrial processes, the prime requirement is to light the task efficiently, since people are concentrating on their work and are less concerned with the environment. Factory lighting is provided almost universally by ceiling-mounted fluorescent equipment to a high level of illumination, ranging from one tenth to one fourth that given by the natural sky on a bright day. It is important to bear in mind that the factory-lighting problem is different from that of offices. Most office work is concerned with white paper, whereas most factory processes are concerned with objects which reflect much less light than white paper; therefore, the same degree of visual comfort results from a very much higher level of illumination. The higher level of illumination assists acuity and contrast and colour discrimination, thereby improving visual performance. Experience over the years has shown that higher levels of lighting give rise to higher productivity and that good lighting pays for itself.

Some critical visual tasks demand additional lighting, both to reduce visual fatigue and to give added visual efficiency. Assembly of fine instruments, watches, and cameras demands light on the work approximating that given by the bright sky, which is better achieved by selective lighting. Similarly, in drawing offices, many draftsmen like to have a high level of general illumination in the office, plus an individual desk lamp for selective lighting on the work.

Factory light fixtures

The factory-lighting designer often must choose between using totally enclosed and open fixtures. Though the totally enclosed fixture may be readily cleaned and maintained, open fittings can be so designed that the heat of the lamp creates an upward draft through the fitting, which carries dirt with it and so reduces the amount which settles on the reflecting and transmitting surface of the fixture. Thus, open fixtures are said to require less maintenance than enclosed ones. With the increasing tendency to build factories as totally enclosed artificial environments—with heat supplied from high levels of lighting, from machinery and industrial processes, and from the workers' bodies—the use of the open-type fixture is indicated, coupled with efficient filtering of the recirculated air. A further disadvantage of the enclosed fixture is that in a factory it may overheat, and, since fluorescent lamps rapidly drop in light output with increase of temperature, the light output from the fixture will fall.

The building lighting in a factory is usually arranged either as high-bay lighting, in which the fixtures are symmetrically positioned at maximum height from the floor, or as continuous-run lighting, in which fixtures are placed end to end across the width of the factory. The high-bay lighting system is better suited to spaces like aircraft hangars because it gives a uniform distribution of illumination with minimum interference with the volume of the space. The continuous-run system can be related directly to an assembly line or row of benches. Furthermore, it can be closely integrated with the daylighting, the continuous run of fluorescent lighting fittings being set along the same lines as the continuous run of glazing in the roof, either to replace it after dark, or to act as a permanent supplement during the daytime. Luminous ceilings are also employed in factories, particularly in those that demand a very close control over the interior climate. The luminous ceiling may consist of an uninterrupted plastic surface behind which the fluorescent lamps are arranged at regular intervals to give a uniform distribution of brightness and to help maintain a uniform temperature in the space below. In certain specialized manufacturing operations, such as the assembly of delicate watch mechanisms or submarine repeaters, in which avoiding dust is important, the overall sealed ceiling lighting system has obvious advantages.

Few factories are lighted today by uncorrected mercury-vapour or sodium-discharge lamps. During the 1930s and 1940s many such installations were made and some still remain, but they are being rapidly replaced by fluorescent lighting. The two main complaints against the use of mercury vapour and sodium were the unpleasant colour and the fact that flicker and stroboscopic effects were marked. Such effects could cause dangerous misjudgments of speed with rotating machinery. Modern fluorescent lighting gives less trouble, but, from the nature of the method of light production in the lamp, on alternating current some residual flicker, greater than that from filament lighting, is present. In poorly engineered lighting installations, lamp flicker may be acutely annoying. Fluorescent lamps near the end of their useful life also tend to flicker. Flicker can be reduced by use of special circuits or special lamps with a long "afterglow."

Flicker in fluorescent lighting

EXTERIOR LIGHTING

While much interior lighting, particularly in nonindustrial buildings, can be undertaken, with some technical assistance, by the architect or interior designer, most exterior lighting demands the services of the specialized lighting engineer. Lighting an airfield, for example, is recognized as a highly specialized and skilled job that would be dangerous to leave to any but an experienced professional. Not only must the airstrip itself be illuminated so the incoming pilot can judge his position and distance accurately, but the approach lighting must be designed so that the pilot can orient himself on to the right flight path with great accuracy from a great distance. The principles of central and peripheral vision have been studied in order to understand how people orient themselves in space and time with no other visual clues than points of light arranged in patterns in space. Despite great advances in instrumented landing equipment, airfield lighting promises to continue indefinitely to be an important field for the lighting engineer.

The lighting engineer also has a special role in such public places as streets, sports arenas, and parks and in the decorative floodlighting of buildings. All these forms of exterior lighting call upon the use of specialized equipment more or less tailored to the job.

Streets. Artificial lighting was perceived at an early date as a potential crime deterrent. The first installations of gas street lighting were made in London and Paris in the early 19th century with this purpose in mind. Toward the end of the century large cities began to recognize the prestige value of street lighting in the centre of their cities and developed more sophisticated street lighting, with ornate cast-iron lampposts housing fixtures mounted 25 to 30 feet (eight to nine metres) above the roadway with powerful filament lamps or arc lamps lighting not only the roadway but the prestigious buildings around. This phase of street lighting lasted approximately up to World War I.

With the development of the internal-combustion engine and the increase of traffic on the roads at night, street lighting acquired new importance. Levels of illumination were considerably increased with the aid of mercury- and sodium-vapour-discharge lamps. At this stage practice in the United States, in Great Britain, and in continental Europe diverged. The United States were most backward in the development of fixed street lighting, relying more upon the light from automobile headlamps and from shop windows, so that by the end of the 1930s the United States was at least ten years behind Europe. In Great Britain silhouette vision formed the basis for street-lighting design, the principle being that light from the street lamps was reflected off the "sheen" of the road surface to provide a background brightness against which objects were seen black in silhouette. Objections that this was an unnatural way to see things were largely overruled by the fact that far better revealing power could be achieved this way for the same expenditure of electrical power than by any other method. Continental Europe, on the other hand, adhered to the principle of natural vision, lighting the object preferentially and trying to keep the background dark. Revealing power by the continental system, however, was far less. The disadvantage of the British system was that it required by its very nature a high level of light to reach the observer's eye because the optical system of the lantern could not be controlled with sufficient precision to confine the light to the road surface alone. The continental system, on the other hand, because it did not depend upon reflection, was far more comfortable and less glaring.

Silhouette vision

After World War II, work undertaken in a number of laboratories showed that one of the major causes of street accidents was skidding and that much could be done to minimize this by the better design of road surfaces combined with tire-tread improvements. The road surfaces with best skid-resistance characteristics had exactly the wrong reflection characteristics for the silhouette principle of street lighting. Attention was therefore directed everywhere to the further development of the continental street-lighting system of illuminating the objects directly. With greater affluence, the much higher levels of lighting necessary for the proper operation of the continental system could be afforded. The silhouette principle, of immense value at its time, has now been abandoned, although in practice, much vision on the streets at night is still by silhouette.

Street lighting today employs sodium-discharge equipment almost universally for traffic routes because of its high luminous efficiency and relative ease of optical control, and fluorescent-tube lighting or colour-corrected mercury-vapour-discharge lamps in the centres of cities. Filament lighting, like gas-lighting, has been almost entirely abandoned. The use of fluorescent lighting necessarily gives rise to very large lighting fixtures, which may be mounted on standards or hung over the roadway, of a size and proportion that would not have been tolerated in the early part of the century.

Sports fields and arenas. Lighting of these installations has developed rapidly since World War II. As soon as lighting engineers could demonstrate that the cost of lighting a large arena to the levels necessary for games or athletic performances was economically feasible, sports-ground lighting developed rapidly. The basic principle is almost always that of employing massive banks of light projectors around the periphery of the ground, floodlighting the arena, rather than suspending lighting units above the arena itself. This has led to the development of powerful projectors based on the searchlight principle, and, for this, the development of the tungsten–halogen projector lamp, with its high light output and relatively long life, has been crucial to success. Arc lamps, the basis of earlier searchlight design, have been almost entirely supplanted, though the development of new forms of colour-corrected mercury-vapour-discharge lamps or the further development of the xenon-discharge lamp may prove eventually to be of value in sports-arena lighting.

Floodlighting of buildings. This field has also been transformed by the availability of filament lamps of high light output and long life. As with the lighting of sports arenas, the principle is to project light onto the facade of the building from powerful projectors located sufficiently far from the building not to impede circulation around it. The aim in floodlight design is to distribute the light, in amount and direction, so that all parts of the facade are adequately illuminated and the sculpture and form of the building are modelled either as they are in daylight or in such a way that other aspects of the design make themselves apparent. These principles have also been applied to the lighting of the interior of monumental buildings, such as cathedrals.

Measurement and calculation of lighting

FUNDAMENTALS OF LIGHT MEASUREMENT AND CONTROL

Lighting engineering is based on a few simple principles of the physics of radiation. A light source is said to radiate light because part of its emitted radiation causes a particular sensation when it strikes the retina of the human eye. Said to be outside the visible spectrum, radiation, which does not give rise to a sensation of light, may have short wavelengths (ultraviolet radiation) or longer wavelengths (infrared radiation).

The name light flux is given to the rate of emission from the source; the unit of light flux is the lumen. The light output from any lighting fixture is expressed in lumens. A wax candle emits about 13 lumens, a 100-watt filament lamp about 1,200 lumens, and a five-foot (1.5-metre) fluorescent tube about 5,000 lumens.

The lumen

Illuminance, reflectance, and utilance. The lighting level—that is, the amount of light spread over the area receiving the light—is called the illuminance (sometimes the illumination) and is expressed in lumens per unit area. One lumen per square foot is the amount of illuminance of a surface one foot (30 centimetres) from the light of one ordinary wax candle. (Hence, in the United States the unit of illumination is the footcandle, identical with the lumen per square foot.) Unobstructed daylight from a bright, overcast sky gives about 1,000 lumens per square foot. The unit of illuminance in metric units is the lux—one lumen per square metre. One lumen per square foot equals 10.76 lux. Unobstructed daylight from a bright, overcast sky gives about 10,000 lux. Overall fluorescent lighting to the standard required for factory work is about 1,000 lux.

Light falling on a surface is reflected off it according to the ability of the surface to reflect light. This reflection factor, or reflectance, relates the light falling on the surface as seen by the observer. A white surface has a reflectance of nearly 100 percent, while a black surface has a reflectance of about 2 percent and a medium-gray surface about 40 percent. If the surface diffuses light equally in all directions (*e.g.*, a matte surface paint or a dull velvet), the brightness of the surface will be related directly to the illumination falling upon it. The physicist and engineer use the word luminance to express the physical measure of brightness. The unit of luminance used in the United States is the footlambert. Luminance, illuminance, and reflectance are related by the equation: Luminance (footlamberts) equals illuminance (footcandles) times reflectance. Thus a diffusing off-white surface of 50 percent reflectance receiving ten footcandles will have a luminance of five footlamberts. Illuminance to a level of 100 footcandles falling on a desk top of reflectance of 30 percent on which is a white paper of reflectance of 80 percent would thus give a luminance of 80 footlamberts to the paper and 30 footlamberts to the desk top.

In order to calculate the amount of light necessary to produce a certain level of illuminance in a room, the concept of utilance, or coefficient of utilization, must be introduced. This is the proportion of light that is actually received on the surface under consideration. If in the room the total light from all the lighting fixtures in the installation is directed onto the working reference surface and nowhere else, the illuminance would obviously be given by the density of light flux in lumens over the area of the reference surface; that is, illuminance equals total flux (lumens) divided by area of reference plane (square feet or metres).

Calculating the utilance

Much of this light flux from the fixtures is, however, lost in various ways. Some light is absorbed in the fixtures themselves, while more light falls on the ceilings, walls, furniture, and elsewhere and is absorbed there and never reaches the reference surface. The coefficient of utilization, or utilance, is the measure of the ratio of the light flux actually reaching the reference plane to the light flux initially emitted by the lamps of the installation. It is determined by the loss of light by absorption at the surfaces of the room, the absorption of light in the fixtures, and the effect of the proportions of the room. Calculation of the utilance is a complicated process, but it is made easy by the publication of tables of utilance. Computer programs for its calculation are also available. In practice, the utilance varies from about 0.15 for totally indirect lighting (which hits the ceiling first and is reflected downward onto the working surface) through about 0.3 or 0.4 for general lighting (emitted in all directions) to about 0.6 for direct lighting (directed by means of reflectors or refractors directly on to the working surface).

If, for example, the required average level of illuminance in a factory 100 × 30 feet (30 × 9 metres) is to be 100 footcandles, and the coefficient of utilization has been found from tables for the particular space, the reflection characteristics of the walls, floor, and ceiling to be 0.3, then the total amount of light flux in lumens required from the lighting installation will be: required light flux (lumens) equals illuminance times area utilance, equals $\frac{100 \times 3000}{0.3}$ equals 1,000,000 lumens.

If the installation is to use five-foot (1.5-metre) fluorescent tubes, each giving 5,000 lumens, 200 tubes will be needed. These might be arranged in continuous trunking along the length of the factory, 20 tubes per line, requiring ten lines across the width of 30 feet (nine metres), or, more economically, the tubes might be twinned giving 40 tubes per line and hence five lines spaced, say, five feet (1.5 metres) apart, across the width of the factory.

This calculation does not allow for the fact that the light output will drop during the life of the installation, due to the inevitable accumulation of dirt and to the fall-off in light output as the lamps get older. It is usual in lighting calculations to step up the light output by about 25 percent to allow for depreciation losses.

The above calculation is very simple and does not require the skills of the lighting engineer. It is no more than a first attempt to indicate how much lighting is required. More often the lighting designer must direct light specifically where it is wanted for efficient and comfortable seeing. The light from bare lamps must be controlled according to his plans. In order to design a fixture that will, when correctly mounted, direct the light to where it is wanted, he must first know the light output of the source (the bare lamp) in all directions. This directional light output must be expressed in terms of the amount of light flux (lumens) being emitted by the source within a given cone (solid angle). Light flux per unit solid angle is expressed in candelas (formerly candlepower) and is called the intensity of the light source. An ordinary wax candle has an intensity of approximately one candela.

The intensity distribution of the bare lamp must be measured in order to determine where the light from the lamp is going and, thence, to redirect it where it is wanted. The intensity distribution is measured by placing a light receptor at a fixed distance from the lamp and in all necessary directions. The amount of light received on the receptor surface will then enable the intensity of the source in the particular direction to be calculated by means of the inverse-square law.

Inverse-square law

The inverse-square law of radiation states that the illuminance, E, at a point varies directly with the intensity, I, of the source and inversely as the square of the distance, D, between the source and the point. If the surface at the point is normal to the direction of the incident light, the inverse-square law gives: E = 1/02.

This is true if the source is a point and is approximately true if the distance, D, is great compared with the dimensions of the source. If, therefore, in measuring the intensity distribution of a point source of light, the illuminance at a given point at a given distance from the source is obtained, the light receptor being perpendicular to the direction of the source, the intensity I is given by: $I = E \times D^2$.

If the light receptor is moved around the light source on, for example, a rotating arm so that it is always at the same distance from the light source and always normal to it, the intensity in any given direction will therefore be proportional to the illuminance of the receptor, which can then be calibrated directly in units of intensity (candelas). Such an instrument employed in this way is called a light-distribution photometer.

Light distribution. The light distribution from a source can be expressed in different ways. If it is known that the light source is symmetrical about an axis (an ordinary general service pearl bulb, for example, is to all intents and purposes symmetrical about the vertical axis through the cap of the lamp, and a fluorescent tube is to all intents and purposes symmetrical about an axis along the centre line of the tube), the light distribution can be expressed in two dimensions in the form of a polar curve, a graph on circular coordinates. The light output can equally be expressed on a graph in rectangular coordinates, but the polar distribution curve gives a better visualization of the light-output characteristics of the lamp.

Light output of a complete fixture

The same procedure is used for measuring the light output of the complete fixture. After the fixture has been designed to direct light where required, the light output from the complete fixture must be measured in order to proceed with the next stage of lighting design. Here, again, if the light output from the fixture is symmetrical, a simple polar curve in two dimensions will give all the necessary information. If, however, the light output is not symmetrical—as would be the case, for example, in a street-lighting fixture where the purpose is to direct light in two directions only—it is then necessary to express the light output of the fixture in three dimensions. This is often done by means of a special graph based on principles of trigonometry and called an isocandle diagram. The information in the polar-curve intensity diagram, for a symmetrical fixture, or in the isocandle intensity diagram, for a nonsymmetrical fixture, gives all that is necessary to proceed on lighting design.

If, for example, the requirement is to produce an illuminance of 10 footcandles on a surface normal to the fixture and distant 100 feet (30 metres) from it, the inverse-square law will give the necessary intensity as being: 10 × (100 × 100) = 100,000 candelas in the direction of the surface. The intensities in other directions can be found in the same way. Hence, the required light distribution of the fixture can be determined from the illuminance requirements. Equally, in reverse, if a given fixture is available for the job, the illuminance values in significant directions and at given distances from the fixture can be found and compared with requirements.

Light distribution categories

This type of point-by-point lighting design is necessary when it is required to find or to provide specific illuminance at a specific point. A great deal of lighting design, however, particularly of interiors and also of large areas out-of-doors—such as a building facade—is more concerned with average illumination over a whole area. For this determination, it is better to know the light distribution from the fixture in broad zones. Thus the average intensity may be tabulated for a symmetrical fixture in zones. Even more simply, the light-output distribution of the fixture may be classified into three distribution categories—direct, general, and indirect distributions. In a direct distribution, all the light is concentrated downward onto the working surface. In indirect distribution, the light is directed upward onto the ceiling and reflected back to the working surface. A general distribution is one with equal intensities in all directions, so that some light reaches the working surface directly and the rest reaches it only after reflection.

These different distributions not only determine the calculation of the lighting design but also in large measure the character of the lighting. Totally direct lighting on the working surface, while concentrating the attention on the job, gives a dark ceiling often considered unpleasant and believed to lead to visual fatigue. It is a type of lighting formerly characteristic of industrial areas when light was expensive and had to be directed exclusively onto the work.

Indirect lighting

Indirect lighting is pleasant and restful and has often been advocated as the ideal form of lighting since it essentially imitates the diffuse light from the sky. It is not entirely satisfactory for work lighting, however, because the eyes tend to be drawn upward to the bright ceiling. It also lacks life and tends to be soporific. Indirect lighting, provided with a little sparkle from, for example, small reflecting surfaces and supplemented by local lighting on the work, can provide a very pleasant illumination. The best lighting, however, is usually provided by some form of general lighting combined with some direct lighting on special areas.

Glare, or visual comfort, index. Glare control, a recent development in lighting technology, takes two forms, one very simple, one very complex. The simplest means of glare control limits the perceived luminance of the fixture in the directions of view most likely to cause trouble. A typical example of such control is in the regulations for school buildings in England and Wales, where it is required that the luminance over an angle up to 45° above the line of sight should not on the average exceed 1,000 footlamberts. Otherwise it is necessary to determine the glare index for the whole environment.

Precalculated tables of glare index are published in Great Britain, similar to the tables of utilance, which enable the lighting engineer to find the glare index for his installation, provided he knows the light output and light distribution of his fixtures, as well as their number, position and mounting height, and the dimensions and reflectance characteristics of the room surfaces. Computer programs also exist.

Light-measuring instruments. The commonest photometric devices are based on the changes that occur in the electrical properties of certain substances when light strikes them. Simple measurements of illuminance made by the lighting engineer usually employ a photoelectric photometer with a light receptor such as a selenium photocell, connected to a suitable instrument that measures the current generated by the cell when light falls upon it.

Another light-sensitive material, cadmium sulfide, has been used in photometers: the change in its electrical conductivity upon illumination causes a change in the current produced by a battery. The instrument is calibrated at a standardizing laboratory in terms of the international standard of light.

Recommended Lighting Values*

class	examples	recommended illuminance		limiting glare index
		lux	footcandles	
Visual task				
Little difficulty	stores	200	20	—
Ordinary tasks	general offices	400	40	—
Some difficulty	business machines	600	60	—
Fine detail	tailoring	900	90	—
Very fine detail	gem cutting	local, up to 3,000	300	—
Environment				
Glare of little consequence	boiler houses	—	—	28
Secondary consequence	machine shops	—	—	25
Limited glare	shops	—	—	22
Close control	general offices	—	—	19
Closer control	school rooms	—	—	16
Minimum	hospital wards	—	—	13
No glare tolerable	art galleries	—	—	10

*Values taken from current lighting codes in Great Britain.

Very simple light meters, which can be carried in the pocket, are of use to the lighting engineer as a rough guide. They cannot, however, match in accuracy a precision instrument.

STANDARDS OF LIGHTING

Standards of lighting as laid down in regulations are in terms of illuminance (footcandles) and glare index, and these standards vary from one visual task to another and from one type of environment to another.

The Table gives a very brief selection of recommended values taken from current lighting codes in Great Britain.

American recommendations are at present higher than those for other countries. Differences in actual practice in the different countries are narrowing, however. America sets a high target, while European countries tend to set a lower target, which is often surpassed.

So far as the control of glare is concerned, there has not yet been international agreement on the method of specifying glare, and it is only in the United Kingdom that a universal and comprehensive generally used system of glare control is available. Typical values of recommended maximum, limiting glare index on the British system only, are therefore given.

BIBLIOGRAPHY. There are two quite different kinds of books written about lighting, those written by architects and interior decorators who use light as a tool to create atmosphere and character, and those written by illuminating engineers. Occasionally, as in DEREK PHILLIPS, *Lighting in Architectural Design* (1964), and R.G. HOPKINSON and J.D. KAY, *The Lighting of Buildings* (1969), the attempt has been made to marry the two different aspects of lighting, the former by an architect who practices as a lighting consultant and the latter a collaborative effort by an illuminating engineer and an architect working together. Generally, however, the architectural lighting books say relatively little about the hardware, and the illuminating engineering books say very little about the effects of light. Probably the two most useful reference books are the ILLUMINATING ENGINEERING SOCIETY, *Lighting Handbook* (issued every five years); and the ILLUMINATING ENGINEERING SOCIETY OF GREAT BRITAIN, *Recommendations for Good Interior Lighting.* The former is a compendium of every aspect of lighting and also includes a mass of useful trade material, whereas the latter is a brief summary and probably the best introduction to the subject for both the specialist and nonspecialist.

The best books on illuminating engineering were written some time ago, when the subject was a major discipline in schools of electrical engineering. The *Theory and Design of Illuminating Engineering Equipment,* by L.W. JOLLEY, J.M. WALDRAM, and G.H. WILSON (1930), is a "lighting engineer's bible" so far as principles and design are concerned. So are H.H. HIGBIE, *Lighting Calculations* (1934); and D.H. JACOBS, *Fundamentals of Optical Engineering* (1943). More recent books on the hardware of lighting are S.T. HENDERSON and A.M. MARSDEN (eds.), *Lamps and Lighting,* 2nd ed. (1972), and J.M. WALDRAM, *Street Lighting* (1952); J.W.T. WALSH, *Photometry* (1958), gives not only the fundamentals of photometry but also the basis of lighting engineering insofar as this relates to the measurement of light.

Later books on lighting and seeing are H.C. WESTON, *Sight, Light and Work* (1962); and a short book written primarily for the informed layman, R.G. HOPKINSON, *Lighting and Seeing* (1969). There are a great many specialist books about particular aspects of lighting—the NUFFIELD PROVINCIAL HOSPITALS TRUST, *Studies in the Function and Design of Hospitals* (1965), is of particular value to architects in showing the relation between natural and artificial lighting and building design. Much of the most recent work on lighting, however, has not been gathered between hard covers yet, and so it is necessary to refer to the original papers. These appear particularly in *Illuminating Engineering,* published by the Illuminating Engineering Society of New York; and *Lighting Research and Technology,* published by the Illuminating Engineering Society of Great Britain.

(R.G.H.)

Lightning

A natural electrical discharge within the atmosphere is called a lightning flash. Such flashes are usually associated with cumulonimbus clouds (thunderclouds) but also occur in nimbostratus clouds, snow and dust storms, and sometimes in the erupting gas of an active volcano. The flash occurs because separate regions of net positive and net negative electric charge are formed. This separation produces an electrical potential, and when the potential exceeds the dielectric strength (insulator strength) of air, a lightning flash occurs. Within a thunderstorm, the flash can occur between clouds (cloud-to-cloud discharges), within a cloud (intracloud discharges), between a cloud and air (air discharges), or from a cloud to the ground (ground discharges).

Cloud-to-ground discharges are composed of one or more strikes that frequently can be perceived by the eye as a flicker. Although these discharges represent only about 20 percent of the lightning flashes, such occurrences cause approximately 150 deaths, $20,000,000 of property damage, and set 10,000 forest fires destroying $30,000,000 worth of marketable timber each year in the United States alone.

Dangers to man

A direct or nearby lightning strike frequently causes death to man, but medical evidence indicates that prompt cardiopulmonary resuscitation and prolonged artificial respiration sometimes can revive a person who apparently has been "killed" by lightning. The effect of direct strikes to man-made objects can range from no damage to extensive destruction, depending upon the physical characteristics of the lightning flash and the degree to which the object struck provides a safe electrical conducting path to the ground. The same range of damage is experienced by trees, although some species show characteristic damage. Conifers, for example, usually have their barks split with a shallow furrow a few inches wide that spirals along the trunk, exposing the outermost layers of sapwood.

Intracloud discharges occur more frequently than ground discharges and almost always expend their energy in a harmless way. Strikes that pass through airplanes are not unusual and cause little or no damage. An exception, however, was the loss of a Boeing 707 in December 1963 that was attributed to lightning from an unseasonable thunderstorm. The most famous lightning flash in modern times was the strike to Apollo 12 on November 14, 1969, as the Saturn/Apollo space vehicle passed through clouds above Cape Kennedy. Fortunately, there was no serious damage.

Lightning is becoming less mysterious as it becomes better understood. The cloud-to-ground lightning flash is the simplest type of lightning to study because it is unobscured by its source, the cloud. Light emitted by the components of the flash can be subjected to photographic and spectroscopic examination to determine physical properties of the flash, such as velocity, temperature, and pressure. High currents flowing within the flash produce time-varying electric and magnetic fields whose characteristic signatures (patterns) reveal the movement of electric charge. Direct strikes to instrumented towers permit measurement of the magnitude of lightning currents; and the subsequent thunder has an acoustical energy spectrum that can be related to the energy per unit length in the flash. All of these measurements, combined with laboratory data on analogous electrical sparks, yield information that reveals the physical characteristics of lightning. Such knowledge, in turn, helps to reduce the deaths and damage due to lightning.

This article treats the physical properties of the several types of lightning, the ways in which these properties are determined, and the occurrence and distribution of lightning strikes. For additional information on electrical phenomena in the atmosphere, see CLOUDS; THUNDERSTORMS; ATMOSPHERE; and AURORAS. See also ELECTRICITY for basic information on the nature and propagation of electric charges and on electrical potential and allied topics.

KINDS OF LIGHTNING

A lightning flash occurs because regions of net charge are produced by charge separation processes within clouds to give an electrical dipole structure (like the magnetic dipole of the earth, with one end positive relative to the other) to the cloud. A South African authority, D.J. Malan, has proposed a cloud charge distribution that can schematically represent the distribution in a typical mid-latitude cumulonimbus cloud (see Figure 1). **A** net positive charge is located at an altitude of ten kilometres (where the temperature is about −45° C) and a net negative charge at five kilometres (−15°C); both charges are of the same magnitude, usually 40 coulombs (an absolute coulomb = 0.1000 electromagnetic unit). A smaller positive charge is located at an altitude of two kilometres (+5° C). The electrical potential between the cloud and ground is on the order of 10^8 volts. At this time the charge resides on ice particles or water drops or both. Discharges can occur between any of the oppositely charged regions or between the cloud and the surrounding air.

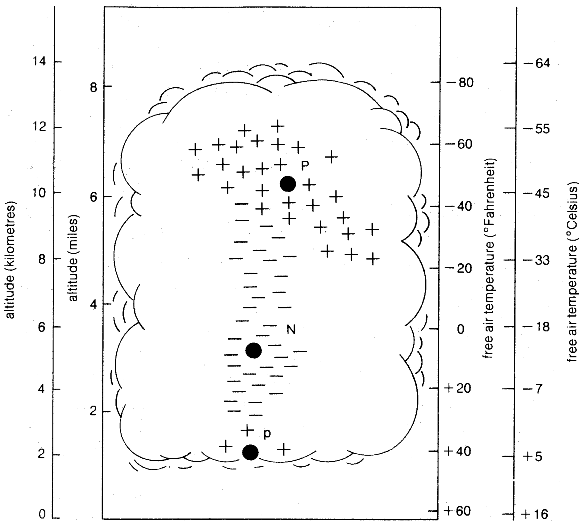


Figure 1: Probable distribution of the thundercloud charges P, N, and p (according to D.J. Malan). Point charges are indicated by solid black circles.

From **M.A.** Uman, *Lightning* (1969); McGraw-Hill Book Co.

Cloud-to-ground lightning. The flash of cloud-to-ground lightning is initiated by electrical breakdown between the small positive charge region and the negative charge region. On a time scale measured in fractions of a second, high-speed cameras can record luminous events in the flash (see Figure 2). A faint luminous process in regular distinct steps, typically of 50-metre length, at time intervals of 50 microseconds, descends in a downward branching pattern toward the ground. Carrying currents on the order of hundreds of amperes, this stepped leader or initial stroke propagates at a typical velocity of 1.5×10^5 metres per second, or about one two-thousandth the speed of light. Diameter estimates for the stepped leader range from a few centimetres to a few metres. The current carrying core is on the order of one or two centimetres, and photographic measurements indicate that a sheath of

From M.A. Uman, *Lightning* (1969); McGraw-Hill Book Co.

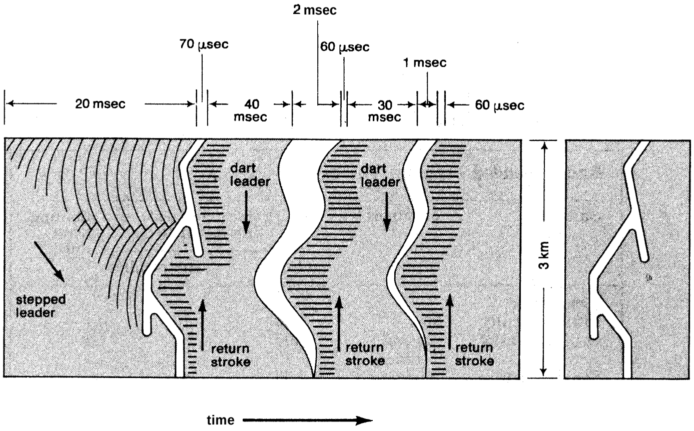


Figure 2: Time-resolved *lightning* flash between cloud *and* ground.
(Left) Flash of lightning as it would be recorded by moving film.
(Right) The same flash recorded by stationary film.

electric charge (corona sheath) with a diameter of one to ten metres exists around the core. As the branching process nears the ground, approximately five coulombs of charge have been deposited on the channel, inducing an opposite charge on the ground and increasing the electric field between the leader and the point to be struck. An upward discharge occurs from the ground, church steeple, house, or other object to meet the stepped leader about 50 metres above the surface. At this moment of junction the cloud is short-circuited to the ground and a highly luminous return stroke of high current occurs. Stepped leaders that have not reached the ground become the branches of the return stroke, and charge on the branches flows into the main channel. The five coulombs of charge typically deposited on the stepped leader flow to ground in a few hundred microseconds to produce peak currents that are usually on the order of 20,000 amperes but may range from a few thousand to 200,000 amperes. Temperatures in the channel are on the order of 30,000" K (50,000" F). Due to the junction process occurring near the ground, the time to peak current measured at the ground is typically ten microseconds. As the charge avalanches toward the ground, the luminosity (return stroke) propagates toward the cloud base at 5×10^7 metres per second, or approximately one-tenth the speed of light, and the high current carrying core expands to a diameter of a few centimetres. Laboratory experiments suggest that when pressure equilibrium is attained between the channel and the surrounding air, the channel approximates a high-current arc characterized by a current density of 1,000 amperes per square centimetre. In the rapid passage from ground to cloud the luminous return stroke is observed to pause at points where the branches join the main channel, and the channel is observed to brighten as charge from the branch flows into the channel. The stroke then continues its upward propagation, reaching the cloud base in about 70 microseconds; the downward propagating stepped leader traverses the same distance in 20 milliseconds.

By courtesy of (top) the University of Arizona, Tucson, photograph, (bottom) Icelandair

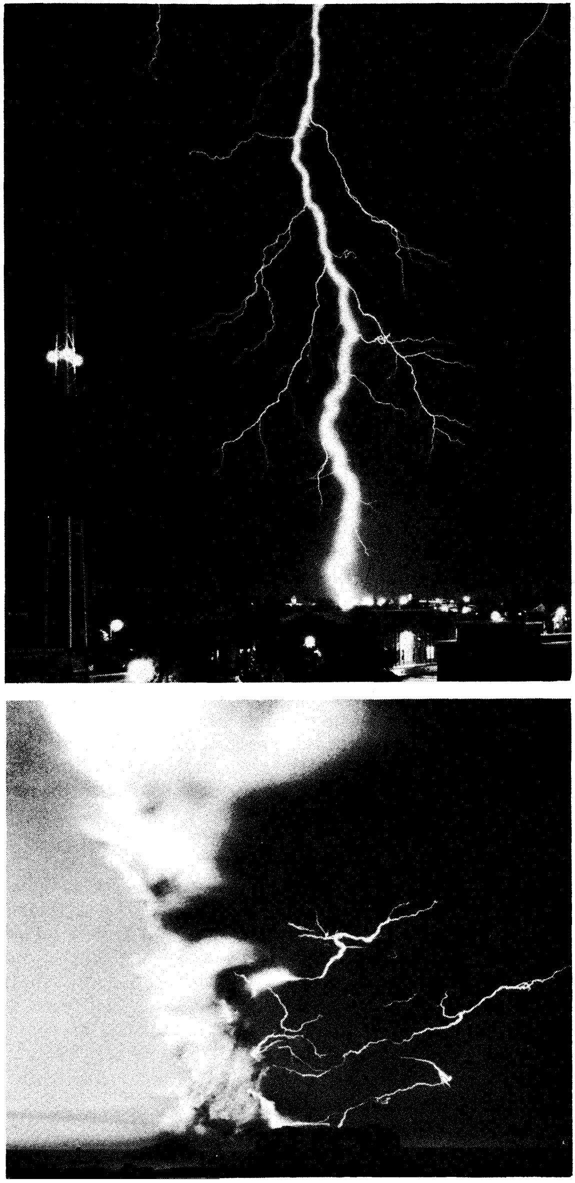


Figure 3: Naturallightning occurrences.
(Top) Cloud-to-ground lightning flash. (Bottom) Lightning in a volcano cloud on Surtsey, an island off the Icelandic coast.

Strokes and energy dissipation

At this moment there is a pause for tens of milliseconds and the channel cools to a few thousand degrees Kelvin. If a second stroke occurs it begins with the appearance of a dart of light, perhaps 50 metres in length, propagating down the channel of the return stroke. At a speed of 2×10^6 metres per second (about one one-hundredth the speed of light) the dart leader carries a current of the order of 1,000 amperes toward the ground. Once again a charge centre in the cloud is short-circuited to the ground and the return stroke occurs. This sequence of dart leader–return stroke typically occurs three to four times, although a flash to the ground that had 26 strokes and lasted two seconds has been reported. When a flash does have more than one stroke the subsequent return strokes draw charge from successively higher charge centres in the cumulonimbus clouds.

During the return stroke stage approximately 10^5 joules (1 joule $= 9.482 \times 10^{-4}$ BTU) of energy per metre are dissipated within the lightning channel. This energy is divided among the dissociation, ionization, excitation, and kinetic energy of the particles, the energy of expansion of the channel, and radiation. Spectroscopic measurements reveal that the air particles, principally nitrogen, oxygen, and water molecules, are split into their respective atoms and that, on the average, one electron is removed from each atom. The conversion from air molecules to a singly ionized plasma occurs in a few microseconds. At this time the temperature of the plasma is at least 30,000" K and the pressure is greater than ten at-

Velocities of Lightning Components and Related Data

	minimum*	representative	maximum*
Stepped leader			
Length of step (m)	3	50	200
Time interval between steps (ysec)	30	50	125
Average velocity of propagation of stepped leader (m/sec)†	1.0×10^5	1.5×10^5	2.6 X 10^6
Charge deposited on stepped-leader channel (coul)	3	5	20
Dart leader			
Velocity of propagation (m/sec)†	1.0×10^6	2.0×10^6	2.1 X 10^7
Charge deposited on dart-leader channel (coul)	0.2	1	6
Return strokef			
Velocity of propagation (m/sec)†	2.0×10^7	5.0×10^7	1.4 X 10^8
Current rate of increase (ka/μsec)§	<1	10	>80
Time to peak current (μsec)§	<1	2	30
Peak current (ka)§	—	10–20	110
Time to half of peak current (μsec)	10	40	250
Charge transferred excluding continuing current (coul)	0.2	2.5	20
Channel length (km)	2	5	14
Lightning flash			
Number of strokes per flash	1	3–4	26
Time interval between strokes in absence of continuing current (msec)	3	40	100
Time duration of flash (sec)	10^{-2}	0.2	2
Charge transferred including continuing current (coul)	3	25	90

*The words maximum and minimum are used in the sense that most measured values fall between these limits. †Velocities of propagation are generally determined from photographic data and thus represent "two-dimensional" velocities. Since many lightning flashes are not vertical, values are probably slight underestimates of actual values. ‡First return strokes have slower average velocities of propagation, slower current rates of increase, longer times to current peak, and generally larger charge transfer than subsequent return strokes in a flash. §Current measurements are made at the ground.
Source: Uman, *Lightning,* 1969.

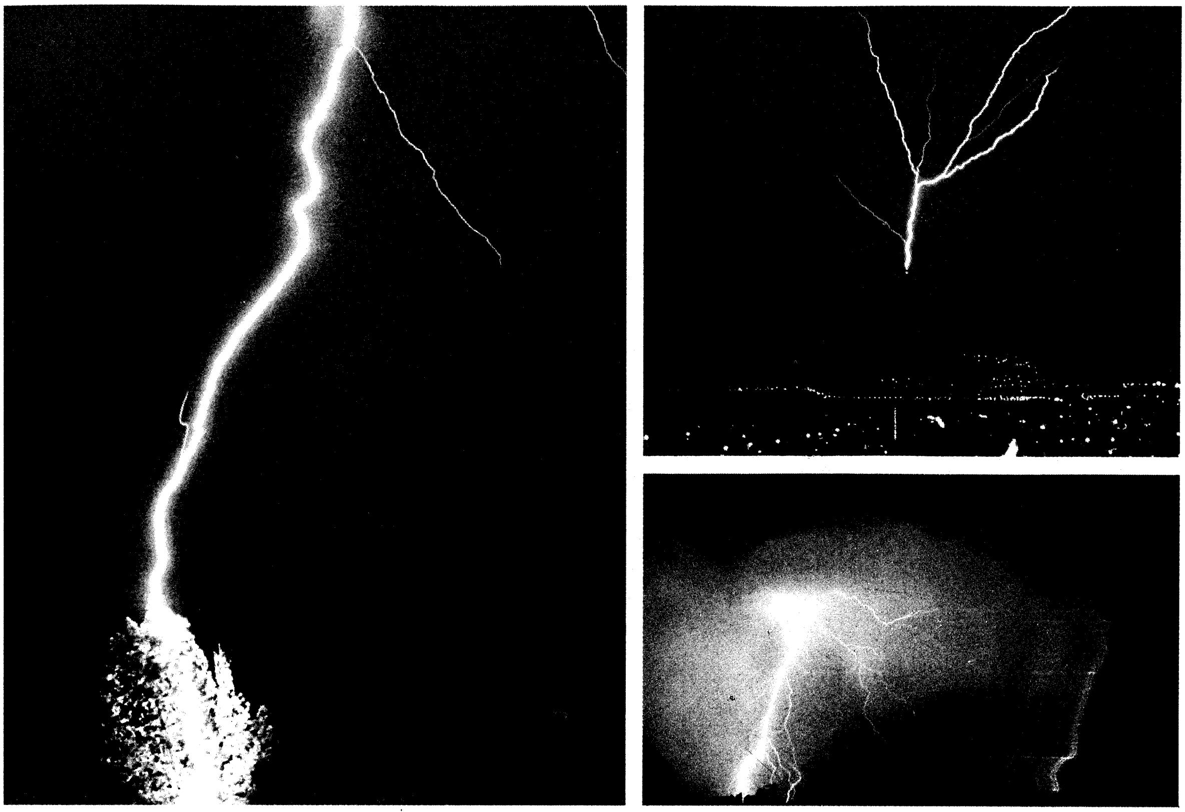


Figure 4: Occurrences of lightning involving ground *objects*.
(Left) A close lightning flash striking a tree at a distance of 60 metres. (Top right) "Triggered lightning"; the discharge is triggered by the presence of the tall tower atop Mt. San Salvatore, near Lugano, Switzerland. (Bottom right) Spectrum of a lightning flash.
By courtesy of (left and top right) Richard E. Orville. (bottom right) William S. Bickel, Universityof Arizona. Tucson

mospheres. The high pressure greatly exceeds the ambient (surrounding) pressure. The return stroke channel expands at a supersonic rate, and the shock wave decays to a sound wave that is eventually heard as thunder. Because it is estimated that 1 percent of the input energy is stored in the particles and less than 1 percent emitted as radiation in the visible and infrared region (4,000 to 11,000 A, where $A = 10^{-8}$ centimetres), it is probable that most of the energy dissipated goes into the energy of channel expansion, a process requiring no more than 10 to 20 microseconds.

A small percentage of discharges between the cloud and ground are actually initiated at the ground and propagate upward to a charged region in the cloud. These discharges often are initiated (or triggered) by tall structures or by towers on hilltops. These discharges have branching in the upward direction, giving the impression of a cloud-to-ground lightning flash that is upside down.

Some physical properties of cloud-to-ground lightning are summarized in the Table. The words maximum and minimum are not absolute but refer to the fact that most measured values occur within these limits.

Occurrence, duration, and current flow of intracloud lightning

Cloud-to-cloud and intracloud lightning. True cloud-to-cloud lightning is rare because most lightning flashes occur within the cloud. Of the total number of flashes occurring, South African studies report that the majority were intracloud, English records indicate 40 percent were intracloud, and studies in Sweden reported 20 percent were intracloud. The lower percentage of intracloud discharges may be associated with lower cloud bases. The first lightning flash in a thundercloud is typically an intracloud discharge. When an intracloud discharge occurs the cloud becomes luminous for approximately 0.2 second. The discharge is initiated by a leader that propagates between charge centres. The 0.2-second illuminosity is continuous and has several pulses of higher luminosity of one-millisecond duration superimposed upon it. This situation suggests minor return strokes as the leader contacts pockets of charge, but the similarity ends there. During the 0.2 second, the amount of the charge transfer is probably similar to the amount involved in a ground discharge: 20 coulombs, with a range from 0.3 to 100 coulombs. The mean velocity of propagation of the intracloud flash is $1 \times 10'$ to 2×10^4 metres per second. Electrical currents associated with the luminous brightening are probably in the range of 1,000 to 4,000 amperes. Strikes to aircraft indicate peak currents of only a few thousand amperes, an order of magnitude less than currents in ground flashes. Rise times to peak currents are measured in milliseconds, three orders slower than rise times in return strokes. The energy of intracloud flashes is unknown.

AURAL AND VISUAL PHENOMENA

Thunder. When air is crossed by a spark or a lightning flash, the air is heated rapidly and the cylindrical column expands at supersonic speed. Within a metre or two the shock wave decays to a sound wave. The sound heard as thunder comes from the entire channel length and is modified by the intervening medium. The result is a series of sounds that are variously described as peals, claps, rolls, and rumbles. These can conveniently be condensed into claps, which are sudden loud sounds, and rumbles, which are all other sounds of thunder. At short distances of a few hundred metres the thunder begins with a sudden clap followed by a long rumble. At extended distances the thunder begins with a rumble. Because light travels at 186,000 miles per second and sound travels at about 1,100 feet per second, observation of the time of arrival of thunder permits calculation of the distance to the flash; the sound will travel one mile in five seconds. For close lightning the elapsed time until the beginning of the thunderclap gives the minimum distance to the flash and the time duration to the rumble is a minimum estimate of the channel length. A thunderclap heard two seconds after the flash, followed by a rumble of 20 seconds, for example, indicates a distance to the flash of approximately 2,000 feet and a channel length at least four miles long. For reasons that include the atmospheric temperature lapse rate, wind shear, and terrain features, thunder is rarely heard at distances greater than 15 miles. The energy spectrum of thunder has been recorded and shows that most of the energy is in the low-frequency audible range with a maximum at 50 hertz (cycles per second).

Types of lightning flashes. Lightning types have been discussed thus far according to a scientific classification. Before the components of the lightning flash were known or studied. however, man devised his own method of classifying the bright flashes. Names such as forked, streak, sheet, heat, hot and cold, ribbon, and bead lightning all have their explanations in terms of ground or intracloud flashes.

Forked lightning refers to a many-branched cloud-to-ground lightning flash; the forked part is just the extensive branching created by numerous stepped leaders that propagate toward the ground until one fortuitously contacts the earth. The stepped leaders that have not reached the ground become the branches of the first return stroke. If there are very few branches, the flash is referred to as streak lightning.

An intracloud discharge, the most frequent type of lightning, usually illuminates a cloud without the channel being visible. The cloud has the brief appearance of a white sheet, hence the origin of the term sheet lightning.

Heat lightning is a type of distant cloud-to-ground discharge that occurs on warm summer evenings and is characteristically red or orange. All colours of the visible spectrum are emitted by the lightning channel, from the short-wave blue radiation to the long-wave red radiation. Blue light is scattered more efficiently than red so that relatively more red light propagates to the distant observer to give the impression of red lightning. (For this same reason the setting sun appears red.) Thunder cannot be heard because this type of lightning occurs at a great distance from the observer.

Hot and cold lightning refer to the observation that some ground discharges cause forest fires and others do not. Although the nature of the material struck is important, the return current characteristics of the return stroke are more important. Flashes containing strokes with a continuing current of a few hundred amperes and lasting for milliseconds after the high current phase have been shown to set forest fires and are consequently termed hot lightning. Flashes to trees without a continuing current usually cause only explosive damage and are called cold lightning.

Several rare forms of lightning include ribbon and bead lightning. These are also a form of ground discharge. If a flash consists of several strokes and a strong wind is blowing perpendicular to the line of sight, the channel is blown sideways and subsequent strokes are displaced. Due to the residual image on the retina of the eye, the flash appears to be a ribbon of several strokes. The same effect can be photographically recorded by slowly moving a camera from side to side, producing a slow time resolution of the component strokes in a flash. It was in this manner, in fact, that the stroke components of the flash were first identified around 1900. Bead lightning has been photographed on several occasions; the main lightning flash breaks into luminous sections, or beads, as the light intensity of the channel decays. The time-extended luminosity of the section can be caused by sections of the channel being viewed end-on or by sections that intrinsically have a larger diameter and, therefore, take longer to cool.

Ball lightning

The most rare and mysterious form of lightning is ball lightning, a ball of luminosity that usually occurs near the impact point of a flash, moves horizontally at a few metres per second, can penetrate closed windows, usually is accompanied by a hissing sound, has a diameter of about 20 centimetres, and has a lifetime of several seconds. The colour is quite variable, and the ball often ends with an explosion, leaving an odour of smoke. Ball lightning theories can be classified into those that propose an external energy source and those that propose an internal energy source. Unfortunately, none of the theories has succeeded in explaining the mobility of the ball or why the ball does not rise, as a hot plasma gas in pressure equilibrium should. An authentic photograph of ball lightning has never been published, but considerable research has been devoted to this phenomenon.

THE STUDY OF LIGHTNING

Problems and techniques of measurement. The transient, unpredictable nature of the lightning flash makes it one of the most difficult atmospheric phenomena to study. Occurring in a time scale measured in millionths of a second, the discharge strikes without warning, carries currents from a few tens to a few hundred thousand amperes, and has components that emit light with intensities varying over several orders of magnitude. Despite the harsh experimental conditions existing during the thunderstorms that produce lightning, it is possible to obtain data using photographic and spectroscopic techniques, electric and magnetic field measurements, current measurements, and acoustical measuring techniques.

Photography is the easiest and perhaps the most used process for recording the lightning flash. A time exposure at night during a thunderstorm invariably yields a few photographs. If the camera is moved slowly from side to side, it is possible to time resolve a ground flash into its component return strokes. Investigators in South Africa in the 1930s used high-speed photographic techniques, principally a rotating film drum and stationary optics, to obtain microsecond resolution. This procedure was sufficient to record successfully the faint stepped leader process, dart leaders, and return strokes. These data, when analyzed, yielded the two-dimensional velocities of the leaders and return strokes and the relative intensities of the same components.

Magnetic and electric measurements

The amount of electric charge involved in the entire flash and its polarity are determined by monitoring the electric field produced by the net charges in the cloud and the charges of the opposite sign induced on the ground, The motion of the charges within the cloud and between the cloud and ground constitutes an electric current; this, in turn, produces a magnetic field that can be measured at the ground. From this measurement the amount of current in the lightning channel can be deduced. A direct strike to an instrumented tower, on the other hand, will allow the current to flow through a known resistance. If the voltage versus time is measured and Ohm's law (voltage = current × resistance) holds, the current in the lightning channel can be directly calculated at the ground. This value, however, is not necessarily the lightning current in the channel above the tower. The simplest current measurement to perform, and probably the most difficult to interpret, is the experiment using magnetic links. Because the residual magnetism induced by the unidirectional magnetic field in the magnetic links depends on the maximum current, it is possible to measure peak flash current by placing the links around transmission towers.

The rapid rise of current in the flash produces a large variation in charge and current in a short time interval; the result is a time-varying electromagnetic field. Measurements of these fields are correlated with the visible part of the flash. Leaders emit primarily higher frequencies in the ten-megahertz range, and radiation in the one–ten-kilohertz range is predominantly from the return stroke. Studies of the microwave radiation from lightning show that the dart leader is the main source of this radiation.

At the other end of the electromagnetic frequency scale, the emission in the ultraviolet, visible, and infrared spectral regions provides a source for detailed recording and analysis. The simplest procedure for recording these emissions is to use a transmission diffraction grating (a ruled optical grating that transmits light according to wavelength) in front of a camera lens. If colour film is used, the spectrum will be recorded in colour in addition to the white image of the flash. The most intense lines are from singly ionized nitrogen atoms (atoms that have lost one electron), which are ions in an excited state from which light is emitted. A strong red line, the characteristic H-alpha line of the hydrogen atom, is derived from the dissociation of water vapour in the path of the lightning flash. Just as time-resolving cameras reveal the luminous components of a flash, time-resolving spectrographs reveal the spectral features of the flash components. Analysis of the recorded relative line intensity produced by emissions from singly ionized nitrogen atoms indicates that the return stroke temperature is at least 30,000" K, or five times hotter than the surface of the sun. A measurement of the H-alpha line width reveals that the electron density in the channel is 10^{24} m^{-3}, while the density of the channel is about one-hundredth of the normal atmospheric value. Ionization within the channel is approximately 100 percent, implying that every atom has lost one electron. In tens of microseconds the channel expands and cools. The rapidly expanding channel produces the sound wave with an acoustical spectrum that can be recorded. Analysis of the spectrum reveals the dominant peak in the acoustical spectrum of thunder: 50 cycles per second. Furthermore, theory relates this value to the energy deposition per unit length within the channel.

Occurrence and distribution of lightning. Lightning occurs throughout the world, and it is estimated that 1,800 storms producing 100 flashes per second are in progress at any given moment. On a per day basis the earth experiences 44,000 thunderstorms, which produce over 8,000,000 lightning flashes. Most of the flashes occur in the equatorial regions between 30° N and 30° S. The relative frequency of thunderstorms over land and ocean is not well-known, but initial satellite observations in the equatorial region at night suggest that ten times as many thunderstorms occur over land areas as over the sea.

Most of the estimated 8,000,000 flashes occur harmlessly, but others damage airplanes, kill and injure people, set fires, and cause accidents. Although lightning does not usually strike objects more than once, once is usually sufficient. Associated with a ground strike are voltage, heating, and pressure phenomena. A strike to a transmission line with a current rise of 60 kiloamperes per microsecond will produce a voltage difference in the transmission line of 100 kilovolts per metre.

Strikes to sand, on the other hand, will sometimes cause the sand particles to fuse along the path of the channel, forming a fulgurite, a permanent record of the lightning's path in sand. Large pressures are generated when the lightning strikes objects, instantly evaporating water in cracks to split trees or cause explosions. At 100,000 amperes, blocks of stone weighing five tons may be torn loose and rocks weighing 50 pounds may be thrown 20 yards or more:

Direct lightning strikes to humans usually cause death, and in fact do cause more deaths than any other weather phenomenon. Seventy percent of the deaths are single events and hence receive little publicity. Outdoor recreation activities, however, are increasing so that the number of deaths each year can be expected to increase. Common safety precautions during thunderstorms include staying off the water in small boats, staying inside cars, not standing under trees. A number of golfers have been killed or injured because their shoes were equipped with metal cleats, making their bodies more effective lightning rods. The telephone is considered by many people to be safe, yet lightning strikes during calls produced telephone casualties of four killed and 36 injured between 1959 and 1965.

Evidence indicates that lightning strikes to people need not be fatal if prompt cardiopulmonary resuscitation and prolonged artificial respiration are applied. A person struck by lightning is rendered unconscious, and heart action and breathing cease. Recovery can be complete, however, except for possible impairment or loss of sight or hearing.

Protection of structures

A satisfactory building protection system involves lightning rods to intercept the lightning flash, down conductors to transfer the charge safely to the earth electrode, and a grounding system to provide a safe conducting path into the ground. The lightning rod, invented by Benjamin Franklin in 1752, has undergone remarkably little change. Its purpose is to intercept a lightning discharge before it can contact a building and pass the lightning current safely to ground. A commercial proposal to increase the effective height of a lightning rod by using radioactive material to ionize the air is scientifically unsound. The additional ionization created by the radioactivity is several orders of magnitude less than the natural ionization produced by a natural lightning rod in the presence of a thunderstorm electric field. No artificial means are known to increase the range of protection afforded by a lightning conductor. Furthermore, the acceptance of a fixed value for the area protected by a lightning rod is not justified. The area protected is approximated by a circle whose radius is the height of the rod. This is only a statistical quantity as there are reports of strikes within the presumed area of protection.

BIBLIOGRAPHY. B.F.J. SCHONLAND, *The Flight of Thunderbolts,* 2nd ed. (1964), a general, popular account of lightning discharge and associated research; D.J. MALAN, *Physics of Lightning* (1963), a highly technical discussion; J.A. CHALMERS, *Atmospheric Electricity,* 2nd ed. (1967), and M.A. UMAN, *Lightning* (1969), advanced, up-to-date treatments.

(R.E.Or.)

Li Hung-chang

Li Hung-chang (in Pin-yin romanization Li Hongjang) was a Chinese statesman under the Manchu (Ch'ing) dynasty (1644–1912) who exemplified many Confucian ideals and in some ways symbolized Confucian China's inability to adjust to the modern world. He rose through the Confucian examinations to the highest offices and seemed to be omnicompetent, excelling in both civil and military affairs. But he added new elements. For long periods he was virtually in charge of relations with the Westerners, who had forced a new era of treaties on China; equally unprecedented for a Confucian official were his enterprises in commerce and industry.

Charles Phelps Cushing

Li Hung-chang.

Early life and career

Li was born on February 15, 1823, at Hofei, Anhwei Province. Both his father and Tseng Kuo-fan, who became his mentor, took terminal degrees in the examinations. Li started on his official career in 1844 under Tseng's guidance in Peking, the capital; in 1847 he earned his terminal degree.

In 1850 the Taiping Rebellion, a great national religious–political upheaval, broke out and threatened to topple the dynasty. When their homeplace was threatened, Li and his father organized a local militia. Li became so involved that he stayed (unofficially) at his post even when his father died in 1855, in defiance of the traditional Confucian mourning retirement. He earned a judgeship in 1856.

Tseng Kuo-fan, in 1860 governor general of the Liang-kiang provinces (central China), was organizing irregular anti-Taiping forces, and Li later joined his staff. In 1862 he was made acting governor of Kiangsu Province and travelled to Shanghai with his own troops in rented steamers. Hitherto, steamers had given the West a great advantage in two China wars (1839–42 and 1856–60), from which came the so-called unequal treaties, whereby China unilaterally surrendered such things as tariff autonomy and extraterritorial jurisdiction.

At just under 40 years of age, Li enjoyed high civil provincial rank and independent military power, a combination that had been forced on the central government by the exigencies of the rebellion.

For the next few years, Li worked partly with foreigners and their weapons in the anti-Taiping effort around Shanghai. Best known of these Westerners was Charles ("Chinese") Gordon, then a 30-year-old English army officer who led the "Ever-Victorious Army," a force, later put at Li's disposal, made up of foreign mercenaries. Although Westerners tended to credit this alien force with putting down the rebellion, it was really Tseng Kuo-fan and subordinates such as Li. The immediate director of the 1864 campaign against Nanking, the Taiping capital, was a brother of Tseng Kuo-fan, but Li disregarded his orders to assist in that terminal action because he felt that jealousies might arise—a delicacy that hinted at Li's own eminence.

Between 1865 and 1870, Li was heavily involved in various high official assignments in central, north, and west China, mostly to suppress other rebellions. He kept his interest in the Western-style arsenals he had established in Nanking and Shanghai, because he wanted to strengthen China against the West—or encroachment by Japan, whose modernization in the late 1860s was increasingly alarming to Li.

Appointment as governor general of Chihli

In 1870 came appointment as governor general of the capital province, Chihli. Li was also a grand secretary (a central government post), and superintendent of trade for the north, responsible for supervising Western trade out of the so-called treaty ports north of the Yangtze River. Although there was a new Peking agency for China's Western diplomatic relations, the Tsungli Yamen, Li became the Chinese negotiator most familiar to foreigners. Thus, he had positions in both the central and the provincial structures, military forces at his disposal, growing prestige abroad, and, it developed, an unprecedented 25-year term of office in Chihli.

During this long tenure, Li interested himself in several major modernizing projects: another arsenal at Tientsin and improved fortifications there, the sending of young Chinese to the United States to learn new skills, a commercial steamship line, Western-built warships, a coal mine, a railroad, a telegraph line, a cotton mill, a military academy, a modern mint, and two modern naval bases. He even talked about change in the procrustean Confucian examinations. There were few other Chinese officials interested in such projects.

During these years, Li also engaged heavily in negotiations with the Japanese, the British, the French, and other treaty powers. If his efforts were not in sum successful, this was largely a reflection of China's continuing relative military weakness, of which Li, throughout his "self-strengthening" efforts, was acutely aware. China sent a mission of apology to Great Britain in 1876 after the murder of a British official; in the same year, Japan made a treaty with Korea that ignored China's traditional suzerainty over the peninsula, and Li was not able, in a later treaty of commerce between the United States and Korea that he tried to manipulate, to get American recognition of the old relationship; in 1879 China lost to Japan her suzerainty over the Liuchiu (Ryukyu) Islands; Li sought French acceptance of Chinese suzerainty over Annam, but the result of the Sino-French War (1884–85) was that French suzerainty was substituted for it; Li's efforts to recover China's prestige in Korea were undercut by Japan, and, in 1885, Li and the leading Japanese statesman, It6 Hirobumi, in effect agreed to a joint protectorate over the contested peninsula. Indeed, in 1894 Japan went to war with China over Korea.

In this war, Li's northern fleet bore the brunt; virtually no assistance came from China's two other modem fleets. Again China lost a modem naval war and had to cede Formosa and the Liaotung Peninsula to Japan, to recognize Korean independence, to open new treaty ports, to pay a large indemnity to Japan, and to grant to the Japanese all of the advantages hitherto pre-empted by Westerners under the unequal treaties. Li had tried to avoid this war, but his influence, nonetheless, suffered because of it. He personally opened peace negotiations in Japan in March 1895 but was wounded by a Japanese fanatic—and. ironically, it was this attack, which excited Japanese sympathy, that somewhat ameliorated the harsh peace terms. (Certain Western powers, including Russia, forced the retrocession of the Liaotung Peninsula—which Russia in effect appropriated in 1898 anyway.)

State visits abroad

Still, to many Westerners, Li was the leading Chinese. In 1896, he attended the Tsar's coronation, and, while in Russia, negotiated a secret alliance, China's first equal treaty. Although he journeyed in state through western Europe and was received in Washington, D.C., by President Cleveland, his homecoming was chilly; probably the Empress Dowager had to use her influence to protect him (it was said that he bribed her for the favour). He kept his trade superintendency and in 1899 was made acting governor general of the Liangkwang provinces. His prestige was still such that he was selected to negotiate with the aggrieved Western powers after the 1900 Boxer Rebellion fiasco. Again Li had to preside at a national humiliation. He died on November 7, 1901.

Li Hung-chang did not exemplify all of the Confucian virtues. He did not have the reputation for financial disinterest enjoyed by Tseng Kuo-fan, and much of his innovating enterprise was made with an eye to personal profit. But the terminal crisis of the dynastic Confucian system in China cannot be explained thus. Li and a few contemporaries modernized parts of China's forces to protect the old system, but within that system, with its peculiar values and organization, the modern devices could not give full service. Li saw something of this contradiction; he hoped that examination reform would give prestige to the scientific modes, but his proposals were truncated, and he was nearly alone in making them. He did support the education mission noted above, but, when conservatives at court decried it as subversive to Confucian norms, Li did not stake his reputation on it, and in 1881 it was terminated. Li was adroit in manipulating the system and was unquestionably loyal to it.

BIBLIOGRAPHY. WILLIAM J. HAIL, "Li Hung-chang," in ARTHUR W. HUMMEL (ed.), *Eminent Chinese of the Ch'ing Period (1644–1912),* vol. 1, pp. 464–471 (1943); JOHN L. RAWLINSON, *China's Struggle for Naval Development, 1839–1895* (1967); STANLEY SPECTOR, *Li Hung-chang and the Huai Army* (1964); MARY C. WRIGHT, *The Last Stand of Chinese Conservatism (1957).*

(J.L.Ra.)

Liliales

The order Liliales is a vast assemblage of plants whose flowers usually have three petals and three sepals—the sepals usually resembling the petals in shape and colour—and whose leaves are generally linear or strap-shaped with the veins running parallel. These characteristics—along with one leaf in the embryo—are common to most monocotyledonous plants. Within this large order, exemplified by the lily, are more than 400 genera and 8,000 species, including erect perennial herbs, climbers, shrubs, and a few trees. The group contains many common garden plants and vegetables, such as lily, hyacinth, narcissus, amaryllis, asparagus, onion, and yam. Other economically important members include the agaves, fibre crop plants; *Colchicum,* from which the drug colchicine is extracted; and Smilax, from which the flavouring sarsaparilla is derived.

Contemporary authorities recognize from five to 20 families in the Liliales, an indication of the lack of consensus among botanists as to what constitutes the order. In this article 20 families are recognized, according to a recent classification scheme; the validity of this outlook is critically appraised in the last section.

GENERAL FEATURES

Diversity. Species of the Liliales are usually perennial herbs, with fleshy stems arising from various kinds of underground stems. Herbaceous climbers are found in the yam family (Dioscoreaceae) and in the families Alstroemeriaceae and Alliaceae. Climbing shrubs are known from the families Smilacaceae and Petermanniaceae. The order also contains succulents such as the Aloe and Agave species. Agaves include some of the largest members of the Liliales: the flower stalk of the so-called century plants (several Agave species) may attain six metres (about 20 feet) in height. Yucca and Dracaena species are often woody and treelike, such as the Joshua tree (Yucca brrvifolia) and the dragon tree (Dracaena draco).

Aquatic Liliales are rare. They include erect or floating herbs in the family Pontederiaceae, particularly in *Pontederia* species and the water hyacinth (*Eichhornia crassipes*), and rarely in the Amaryllidaceae, notably *Crinum thaianum*, native to Thailand.

Distribution. The lily order is cosmopolitan in distribution, although most of its members are found in temperate and subtropical regions. The extreme range is as far north as northern Greenland (*Tofieldia* coccinea) and as far south as the southern tip of South America and New Zealand (Luzuriaga *marginata*). In numbers of species and individuals the order is best represented in southern Africa, North America, western Asia, and Australia.

Among the widest in distribution are the families Liliaceae. Amaryllidaceae, and Pontederiaceae. The lily family is worldwide in distribution, but is most abundant in temperate and subtropical regions. The amaryllis family ranges from southern Africa to Asia and the Western Hemisphere; and the family Pontederiaceae, common in tropical freshwaters, ranges as far north as southern Canada.

Importance. The lily order is notable for its important food and industrial plants, on the one hand, and its many showy cultivated ornamental plants, on the other.

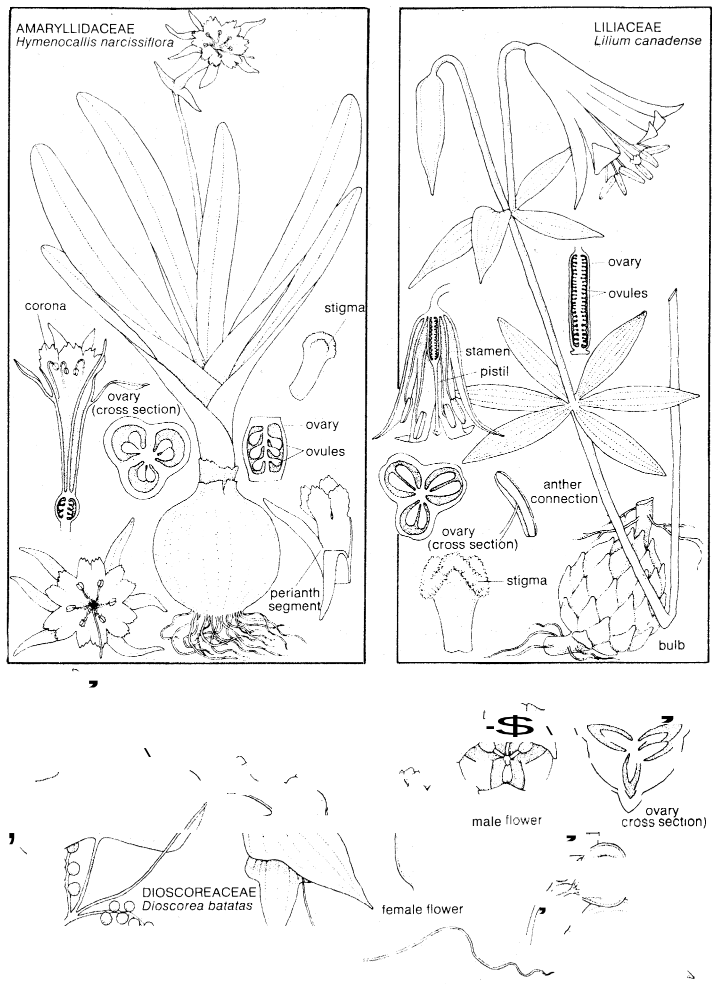
Edible products. Among food crops, the outstanding ones are onion (*Allium* cepa), garlic (*A.* sativum), leek (A. *porrum*), chive (A. *schoenoprasum*), rakkyo, or *ch'iao* (A. chinense), and Chinese chive (*A. tuberosum*). The most important of these is the onion, the others being grown mostly for local consumption or in limited quantities.

Other edible crops of the lily order are asparagus (*Asparagus officinalis*), of the family Liliaceae, and various species of Dioscorea, of the family Dioscoreaceae, of which white yam (D. *alata*), negro yam (*D. cayennen*sis) and cush-cush, or yampi (D. *trifolia*), are considered among the best. Used locally as important starchy food crops throughout the tropics, they are twining plants that produce tubers rich in starch and containing poisonous alkaloids; the latter are easily destroyed by boiling.

Elephant's-foot, or Hottentot bread (*Testudinaria ele*plzantipes), of the family Dioscoreaceae, produces a very large edible rhizome (rootlike structure) used as a food locally in Africa. Black bryony, or black bindweed (*Tamus* conzmunis), a perennial vine of the family Dioscoreaceae, is occasionally cultivated and consumed as a vegetable from southern and western Europe to Kurdistan in Asia. The roots are also used medicinally. Another source of starch is arrowroot, from Tacca *pinnatifida*, of the family Taccaceae. It is used locally in the South Sea islands as food. Starch from the tuberous roots of *Alstroemeria* species, of the family Alstroemeriaceae, is used locally as a food by the Andean Indians of South America.

Sarsaparilla, the flavouring extract, is obtained from the dried roots of several species of *Smilax* native to South America. The elongated berries of species of *Gethyllis* and *Apodolirion*, of the Amaryllidaceae, known as *kukumakrankas*, are collected and consumed locally in South Africa. In China and elsewhere, the buds of daylilies (*Hemerocallis* species) are harvested and used as a vegetable.

Industrial products. The 300 species of Agave, ranging from the United States to tropical South America, and the related 30 species of Furcraea provide important commercial fibres, obtained by processing the leaves. Sisal hemp is obtained from Agave *sisalana*, native to Yucatán; henequen from Agave fourcroydes, Mexico; mauritius hemp or flax from Furcraea *gigantea*, the tropics; and Ceylon bowstring hemp, from *Sansevieria* zeylanica, Ceylon.



Representative plants from three major families of the order **Liliales.**

Drawing by M. Moran

Many other Agave species are cultivated or harvested from native stands mostly for local fibre needs or for fermented beverages. Several species are used in making pulque, the national drink of Mexico; the alcoholic liquor mescal is distilled from pulque.

The water hyacinth, with its bladderlike leaf bases, has been recommended as a source of cellulose for papermaking and other uses.

Dracaena cinnibari, a tree native to the island of Socotra, in the Indian Ocean, is the source of a red resin called dragon's blood, or Socotra resin, which is used in varnishes and also as an astringent and to stop hemorrhage. *Dracaena draco*, a tree native to the Canary Islands, is the source of a red resin, also known as dragon's blood, which is used in the varnish industry, for pigment in papermaking, and in medicine.

Drug products. Drug products include many crude drugs used locally. Aloe juice, obtained from various Aloe species, is used as a purgative. The flowers of lily of the valley (*Convallaria majalis*) are used as a cardiac tonic in place of digitalis; squill, from the bulbs of *Urgirzea scilla*, is, among other things, a cardiac stimulant. White hellebore (*Veratrum album*), *min-mun-tung*, from the tubers of *Ophiopogon japonicus*, and China root, from the dried rhizomes of *Similax* china, have been used medicinally in China since ancient times. The seeds and corms of the meadow saffron (Colchicum *autumnale*) are used to treat rheumatism and also as a diuretic. The active chemical is the alkaloid colchicine, which is one of the most efficient chemicals yet discovered for causing the doubling of chromosomes in cells; it has played an important role in plant and animal science since 1937.

Ornamental plants. The lily order is outstanding for its many ornamental plants, cultivated from most of the 20 families; the more important ones, however, are found in the following families: Liliaceae (*e.g.*, tulip, hyacinth, crocus, lily, daylily, aloe); Amaryllidaceae (*e.g.*, *Amaryllis*, Narcissus, Lycoris); and Agavaceae (*e.g.*, Agave, Yucca, *Hosta,* Dracaena, Cordyline, Sansevieria).

In The Netherlands there is a considerable export business based on certain bulbous and cormous plants of the Liliales. In the United States, particularly in Oregon, lily (Lilium) breeding for the domestic and export trade has developed into a considerable industry. Locally, the growing of plants of the lily order is active in season, particularly the forcing and marketing of lilies, tulips, hyacinths, narcissus, amaryllis, and others for holidays.

NATURAL HISTORY

The life cycle of members of the Liliales is similar to that of other monocotyledonous plants (plants with one seed leaf). Pollination is usually by insects or by other small animals, but cleistogamy, a condition in which self-fertilization takes place within unopened flowers, occurs in the genera *Heteranthera* and Hydrothrix, of the Pontederiaceae.

Food and water storage. One of the outstanding features of a large number of the Liliales is the manner in which they have adapted to the seasonal peculiarities in different parts of the world. The most primitive Liliales are usually herbaceous perennials having short rhizomes with fibrous roots; they are adapted to humid climates. Organs for moisture and food storage have been evolved to carry the plant over unfavourable periods when growth is inhibited by dry or cold weather. These include creeping rhizomes, specialized thickened stems on or just below the ground surface, as in daylilies (Hemerocallis), and lily of the valley (Convallaria); bulbs, as in Lilium, Allium, and *Amaryllis;* corms, swollen bases of the stem, as in Biodiaea, *Milla*, and Tecophilaea; tubers, short underground stems, as in Gloriosa; and tuberous roots with buds at the stem end, as in Alstroemeria and Dioscorea.

In the agaves and aloes, the problem of surviving in a dry desert climate has been successfully solved by the evolution of tough, thick succulent leaves and a tough, thick stem for food and moisture storage. The American Agave species, family Agavaceae, have assumed the outward shape, within limits, of the thick succulent-leaved African Aloe and allied genera of the family Liliaceae. It is interesting to note that the related genera, *Hosta*, native to humid parts of China and Japan, and Polianthes (tuberose), native to Trinidad and humid Mexico, have not developed the succulent habit.

Extreme drought resistance is exhibited by such amaryllids as Leucocoiyne, native to the extremely dry Atacama Desert of Chile, where scant rainfalls occur years apart. When limited rains do occur after prolonged droughts, the bulbs, dormant for years, quickly sprout, come into flower, and produce seed. Another interesting adaptation to an arid climate is found in certain species of Vellozia, in which the upper parts of the stems are covered with fibrous sheaths of the leaves, and the lower parts with a covering of aerial roots inches deep. When water falls on the aerial roots, it disappears as if into a sponge. The plant is thus able to supply itself with moisture in the form of dew and occasional light rainfall as well as from moisture condensed on the aerial roots.

Reproductive adaptations. In many species of the Amaryllidaceae, the seeds soon lose their viability, but some, such as Chlidanthus boliviensis and C. *fragrans*, which grow in dry mountain habitats, retain their germinating power over a period of years. Amaryllis petiolata, native to the forests on the river banks in southern South America, produces no viable seeds but has evolved a novel means of dispersal. The parent bulb produces many small, nearly round bulblets on the bulb scales. They remain dormant almost a year. During the spring floods they are carried long distances.

Apogamous reproduction

Apogamy, a form of parthenogenesis (reproduction in which seeds are developed without fertilization), occurs in such genera of the Amaryllidaceae as *Habranthus* and Zephyranthes. Copious seed production by such means has allowed certain species of these genera to spread over very large areas of Texas and northern Mexico.

Vivipary, in the sense of bulblets sprouting from buds in a flower cluster (inflorescence) almost to the total exclusion of the regular florets in the inflorescence, is found in some Allium species. European wild garlic (Allium *vineale*) by such means has been naturalized as a noxious weed over wide areas in North America and elsewhere. Protandry, the maturing of the stamens before the pistil, is found to a slight degree in many Alliaceae and Amaryllidaceae and to an extreme degree in *Alstroemeria* species, of the family Alstroemeriaceae, in which the stamens mature and shed their pollen before the pistil begins to enlarge and mature. These processes favour cross-pollination.

Miscellaneous adaptations. The Pontederiaceae are adapted for life in an aquatic habitat. In *Eichhornia crassipes*, for instance, the swollen petioles enable the plant to float; and floating has led to very wide distribution of this species. In species of the family Smilacaceae, the prickly shrubs and climbers are protected to some extent against animals. Seeds of many species in the order Liliales are winged, enabling them to be carried by the wind.

FORM AND FUNCTION

The rootstock varies from creeping rhizomes, bulbs, corms, and tubers, all of which are modifications of the stem, to thickened storage roots with buds at the stem end. The plants are usually perennials, very rarely annuals, ranging from herbs to shrubs and climbers or very rarely to types with a woody stem that is branched and treelike. The leaves are usually linear or strap-shaped, with longitudinal veins; rarely, the leaves are broader, and with stalk petiole, with longitudinal veins separated by netted venation. Latex-bearing vessels (laticifers) are present in the Alliaceae.

Floral variations

The flower varies from small and in loose clusters to large, solitary, and showy, with the showy perianth (petals and sepals collectively) usually six-parted, rarely three-parted or less; the perianth segments are similar, and are sometimes united below into a tube. Occasionally there is a floral structure called the paraperigone, consisting of inward extensions of the lower part of the perianth tube; or scales or bristles may be present at the base of the filaments. The stamens usually number six, rarely are there three or less; each stamen consists of an anther—in which pollen is produced—and a filament, or stalk. The filaments may be variously modified, united into a tube below or into a conspicuous staminal cup. The pistil—usually one to a flower—consists of a three-parted ovary formed of three greatly modified leaves (carpels) and a stalk (style) capped by a pollen-receptive tip (stigma). Pollination is accomplished by small animals, including birds, insects, and arachnids. Fertilization—the union of a sperm cell and the egg cell—follows pollination. Most of the Liliales depend on cross-pollination, and thus fertilization follows flower opening; rarely are the flowers cleistogamous, a condition in which there is fertilization without opening of the flowers. The fruit is usually a dry, three-parted (rarely a one-celled) capsule, sometimes a fleshy berry, very rarely a nut; the seeds are sometimes winged.

EVOLUTION

In the absence of a fossil record, any estimate of the evolutionary relationships must be based on the characteristics of the living species. The evolutionary course within the Liliales, as well as in the superorder Lilianae, including several orders as well, is only partly understood; enough is known, however, so that the general evolutionary course may be described to a limited extent on the basis of gross morphology, chromosomal complement, anatomy, and biochemistry.

Morphological evidence. The general evolutionary course of some prominent morphological features in the Liliales, is listed in the accompanying table.

Evolutionary Course in the Liliales

organ	original form	evolved form
Stem	stem with taproot and fibrous secondary roots	rhizome, bulb, corm, tuber, or storage root with buds at the stem end
	herbaceous	shrubby or climbing, rarely woody, rarely treelike and branched
Leaves	linear, with longitudinal veins (originally evolved from petiolate leaves)	secondary petiolate leaves, with longitudinal veins, and reticulate veins between
Flower	perianth segments six, free	perianth segments three or fewer; perianth segments united below into a tube, with six or three segments above
	perianth tubular	perianth tubular with accessory structure (paraperigone)
	carpels free	carpels partially or fully united
	ovary three-parted	ovary with single compartment or two
	ovary superior	ovary inferior
	stigma three-parted	stigma three-lobed, capitate, or pointed
	stamens six, free	functional stamens four or fewer, filaments modified in various ways

Chromosomal evidence. Evidence from the chromosomal makeup of the cells has helped to solve difficult problems in phylogeny. The basic chromosome set, with a few large and many small chromosomes, in the *Yucca-Agave* group is very stable. The chromosome complement in *Hosta* is similar, on which basis it has been suggested that *Hosta* be recognized as the type of a separate tribe, Hosteae, to be placed next to the tribe Yucceae (*Yucca* and its allies) in the family Agavaceae. The evolutionary course is irreversible, and the chances that such a unique chromosome complement originated a second time are remote. The suggestion is therefore valid, indicating that these two tribes originated from the same ancestral stock. The difference in gross morphology can be attributed to differences in habitat: *Hosta* evolved in humid parts of China and Japan; *Yucca* and its allies in the arid parts of southwestern North America. Later, apparently a few species of *Yucca* spread to the more humid southern United States and the West Indies, but only after leathery leaves and other characters had evolved in the arid climate.

The presence of lacticifers in the Alliaceae indicate the distinctness of this family and its presumed early evolutionary divergence from the rest of the Liliales.

The significance of saponins

Biochemical evidence. Biochemistry and other scientific disciplines have yielded evidence that suggests relationships among the flowering plants. On the basis of the presence of one substance—saponin—the Liliales may be divided into two groups:

1. Families in which the plants usually contain saponin compounds, including the Liliaceae and its allies, Xanthorrhoeaceae, Alphyllanthaceae, Philesiaceae, Smilacaceae, and Asparagaceae; the Dioscoreaceae and its allies, Stemonaceae, and Taccaceae; the Agavaceae; and the Alliaceae.
2. Families in which the plants lack saponins, including the Pontederiaceae and its ally Philydraceae; and the Amaryllidaceae and its allies, Alstroemeriaceae, Haemodoraceae, Hypoxidaceae, and Velloziaceae.

Chelidonic acid is present in many families of the Liliales. Alkaloids are present in a number of families, including the Amaryllidaceae. Thus, this group is singled out as a very distinct entity. Alkaloids have also been reported for *Dioscorea* species, *Tacca chantrieri*, and *Stemona* species. Colchicine has been extracted from *Colchicum, Merendera,* and *Gloriosa* species of the Liliaceae. The occurrence of these and other distinct biochemical compounds in Liliales suggest the potential value of this approach in solving outstanding phylogenic problems.

Most authorities on the monocotyledons place the order Liliales in a central position from which other orders have arisen. A hypothetical, now extinct, group with nutritive tissue (endosperm) in the seeds, two-celled pollen grains, and separate or partially united carpels and numerous stamens presumably gave rise to three main lines.

The Liliales, Iridales, Zingiberales, and Orchidales are derived directly from the Liliales ancestral stock. The order Iridales is probably derived from the subfamily Melanthioideae of the Liliaceae. The order Zingiberales was probably derived directly from the Liliales and is claimed to have much in common with the Amaryllidaceae and the subfamily Asphodeloideae of the Liliaceae. The order Orchidales shows closest connections with the family Hypoxidaceae of the Liliales, especially the genera *Hypoxis* and *Curculigo*. The connecting link between these and the Orchidaceae is the most primitive subfamily of the Orchidaceae, the Apostasioideae. The Liliales stock may have given rise to two additional superorders, the Juncanae and Commelinanae.

CLASSIFICATION

The shift in diagnostic features

Distinguishing taxonomic features. In the past, the position of the ovary with respect to the perianth was considered very important in distinguishing groups within the Liliales but is no longer considered as reliable in delimiting larger groups. Currently, the number of chromosomes, the presence of such internal features as laticifers (vessels bearing latex), and the presence of chemical compounds such as saponins and alkaloids are considered of greater significance in grouping of families within the Liliales.

Annotated classification. A currently considered classification lists 20 families in the order Liliales.

ORDER LILIALES

Monocotyledonous plants usually with storage rootstocks. Herbs to trees. Leaves often clustered at base. Flowers usually showy and bisexual (having male and female parts); components in 6 (3 petals, 3 sepals) or 3 parts. The order includes 20 families, about 430 genera, and 8,000 species; distribution is worldwide, but members are concentrated in temperate and subtropical regions.

Family Liliaceae

Erect or climbing, herbs, rarely soft-wooded shrubs, with rootstock a rhizome, bulb, corm, or tuber. Flowers usually bisexual and regular, flower segments 6, rarely 4 or more, often tubular below, with petals and sepals similar. Stamens usually 6, sometimes fewer, rarely more; anthers with 2 compartments (locules). Ovary superior (rarely semi-inferior), usually with 3 locules, style (stalk) entire or divided (rarely free); ovules usually numerous per locule, rarely solitary; fruit a capsule or fleshy berry; seeds with copious nutritive tissue (endosperm). One of the largest families of flowering plants, with 250 genera and 3,700 species, most abundant in temperate and subtropical regions.

Family Xanthorrhoeceae

Perennials with thick rhizomes. Stem when present woody and sometimes tall, single, or somewhat branched. Leaves often sheathing, usually linear. Flowers usually small and in clusters (rarely solitary and larger); flower segments 6, free or shortly united below. Stamens 6, anthers with 2 locules. Ovary superior, with 1 to 3 locules; fruit a capsule or rarely a 1-seeded nut; seeds with hard endosperm. Includes 8 genera and 66 species, in Australia, New Caledonia, and New Zealand.

Family Aphyllanthaceae

Perennial herbs, with short rhizomes. Leaves reduced to sheaths. Flowers bisexual, 1 to 3 in short terminal spike; flower segments 6. Anthers united at bases. Ovary with 3 locules; ovules solitary; fruit a capsule. Consists of a single species, mostly in the western Mediterranean region.

Family Alliaceae

Rootstock usually a bulb or corm. Leaves usually basal, long and thin. Stem leafless (scapose). Flowers in clusters (umbel), subtended by 1 or more spathe-valves; flower bisexual, flower segments 6, usually similar, free or united below into a tube; paraperigone (corona), inner petallike appendages, may be present. Stamens 6 (rarely 3 or 2); filaments sometimes united into a staminal cup. Ovary superior, with 3 (rarely 2) locules; style entire or 3-lobed; fruit a capsule; seeds with fleshy endosperm. About 30 genera and 670 species, worldwide except for tropical regions, Australia, and New Zealand.

Family Agavaceae

Rootstock usually a rhizome. Stem short, often woody and arborescent. Leaves basal, narrow, lancelike, often thick or fleshy (succulent), sometimes with prickly teeth on the margin. Flowers regular or somewhat irregular, in clusters, with the floral branches suhtended by bracts; flowers segments 6, united into a tube below. Stamens 6, inserted at the base of the segments or on the tube; anthers with 2 locules. Ovary with 3 locules; ovules numerous or solitary in each locule; fruit a capsule or berry; seeds with fleshy endosperm. Encompasses about 20 genera and 720 species; especially prevalent in the tropics and subtropics and ranging into the temperate zone.

Family Amaryllidaceae

Perennial herbs, with bulbs or rhizomes. Leaves few to several, sometimes solitary, from the base of the stem, or apex of the bulb; linear, usually strap-shaped, lanceolate. Flowers usually showy, bisexual, regular or somewhat irregular, in clusters (sometimes 2 or 1 by reduction); usually each floret subtended below by a bracteole, and the entire inflorescence subtended below by *2* or more, rarely 1, spathe-valves; fiower segments 6, in 2 series, usually similar, often united below into a tube; paraperigone sometimes present. Stamens 6 (rarely more and rarely reduced to 3), opposite the segments, filaments free, or variously modified, or joined to form a staminal cup; anthers with 2 locules. Ovary inferior, with 1 to 3 locules, style slender, stigma capitate, 3-lobed or 3-parted; ovules usually numerous per locule; fruit a capsule or, rarely, a berry; seeds with copious endosperm. About *65* genera, including about 835 species, worldwide in distribution, mostly in the tropics and subtropics.

Family Alstroemeriaceae

Rootstock a rhizome. Stems usually erect, often climbing, leafy. Leaves crowded or scattered, the petiole usually twisted and thus reversing the surfaces. Flowers bisexual, irregular, in terminal clusters, or in irregular racemes, rarely solitary; segments 6, often narrowed to the base or spoon-shaped, in 2 series. Stamens 6, inserted on top of the ovary, filaments free. Ovary infer·ior, with 1 to 3 locules, style threadlike (filiform), shortly 3-lobed, usually maturing after the anthers; ovules numerous per locule; fruit a capsule; seeds small with copious endosperm. Four genera, including 200 species, ranging from Central America to southern South America, mostly in the Andes mountains.

Family Haemodoraceae

Herbs with short rhizomes, or rounded tubers and fibrous roots. Leaves clustered at the base, linear, or sword-shaped, sheathing at the base; stem leaves reduced or absent. Flowers bisexual, segments 6, regular or somewhat irregular, in clusters, often hairy. Stamens 6 to 3, filaments free, anthers with 2 locules. Ovary with 3 locules; style usually filiform; ovules 1 to many in each locule; fruit a capsule; embryos small with copious endosperm. About 15 genera, including 75 species, ranging from South Africa and Australia to tropical America.

Family Hypoxidaceae

Herbs with tuberous rhizomes or corms. Leaves usually basal, mostly with prominent veins, often with long hairs. Flowers bisexual, segments 6, regular or somewhat irregular, in clusis short or in the form of a long beak on top of the ovary; segments 6, similar in shape and colour. Stamens 6, rarely 3; anthers with 2 locules, entire or 2-lobed at the base. Ovary inferior, with 3 locules; single short style or 3 separated styles; ovules numerous per locule; fruit a capsule; seeds small; with copious endosperm. Includes 7 genera and 120 species, common except in Europe and northern Asia.

Family Velloziaceae

Arborescent and shrubby plants. Leaves tufted at the end of branches, narrow, often sharply pointed. Flowers bisexual, solitary, regular; flower tube absent or very short, segments 6, equal. Stamens 6, or numerous in 6 groups of 2 to 6; anthers linear and joined at their bases. Ovary inferior, with 3 locules; style slender; stigmas capitate or shortly trifid; ovules numerous per cell; fruit a capsule; seeds with abundant endosperm. Two genera, including 170 species, ranging from the Arabian Peninsula, tropical and southern Africa, and Madagascar to tropical South America.

Family Philesiaceae (including **Petermanniaceae)**

Rootstock a slender branched rhizome, supporting shrubs or tall woody climbers, rarely semi-epiphytic. Leaves without stipules, oblong or ovate. Flowers bisexual, regular; flower segments 6. Stamens 6. Ovary superior, with 1 to 3 locules; style slender; stigma 3-lobed or capitate; ovules numerous or a few per locule; fruit a berry. Consists of 8 genera, including 10 species, ranging in the Southern Hemisphere from New Guinea, the Pacific Islands, New Caledonia, Australia, and New Zealand to temperate South America and southeast Africa.

Family Tecophilaeaceae

Upright herbs, with tunicated corms or thick, flattened tubers. Leaves basal. or near the base of the flowering stems. Flowers bisexual, regular or slightly irregular. in clusters; perianth segments 6, free or sometimes united below. Stamens 6 or 4 (or 2 functional); anthers with 2 locules. Ovary semi-inferior, with 3 locules; style awl-shaped to filiform; stigma more or less 3-lobed; ovules numerous per locule; fruit a capsule; seeds with fleshy endosperm. Seven genera, including 25 species, ranging in a scattered pattern from Africa to southern California.

Family Cyanastraceae

Rootstock a tuber or a tuberous rhizome. Leaves basal, linear. Flowers bisexual, regular; solitary or in clusters, flower segments 6. Stamens 6, filaments free. Ovary partly inferior, with 3 locules; 2 ovules in each locule; fruit I-seeded. One genus, 6 species, confined to tropical Africa.

Family Asparagaceae (including Ruscaceae)

Rootstock a rhizome. Stems herbaceous or woody, erect or climbing. Leaves reduced to very small scales, bearing in their axils green, needlelike modified branchlets that function as leaves. Flowers solitary or in clusters; flower segments free or united into a tube below. Anthers 2-lobed. Ovary with 1 to 3 locules; ovules 2 or more in each locule; fruit a globose berry; seeds solitary, or few in each locule. Four genera, including about 310 species, ranging over Asia, Europe, and Africa to the Azores, Madeira Islands, and Canary Islands.

Family Smilacaceae

Rootstock a rhizome, supporting climbing shrubs. Stems and branches prickly, often with tendril-like petioles. Flowers usually unisexual (rarely bisexual), small, in cluster; flower segments 6, free (rarely united below into a tube). Stamens 6; anthers with 1 locule. Ovary superior, with 3 locules; ovules 1 or 2 in each locule; fruit a 1- to 3-seeded berry; endosperm hard. Comprises 4 genera, including 375 species, ranging from the tropics to temperate regions.

Family Stemonaceae (**including** Croomiaceae)

Erect or climbing perennial herbs, with rhizomes or tubers. Leaves often smilax-like, with parallel main nerves and closely parallel cross veins. Flowers in clusters, bisexual, regular, sometimes unpleasantly scented; flower segments 4. Stamens 4, resembling petals or sepals; filaments free or nearly so; anthers with 2 locules. Ovary superior to semi-inferior, with 1 locule; style simple, ovules numerous to 2 per locule; fruit a capsule; seeds with copious endosperm. Three genera, including 30 species, ranging from northern Australia, India, and Japan to the southeastern United States.

Family Dioseoreaceae

Rootstock thick and woody or a tuberious rhizome, usually supporting climbing stems, or sometimes shrubs. Leaves net-veined, often arrow-shaped or heart-shaped, usually alternate. Flowers regular, 6-lobed, inconspicuous, and in clusters. Stamens 6 (sometimes only 3 functional); filaments free or shortly united below; anthers with *2* locules. Ovary inferior, with 3 locules; styles 3, sometimes 3-winged; ovules 2, rarely numerous per locule; fruit a capsule or a berry; seeds with endosperm, often winged. Consists of 6 genera, including 750 species, confined to the tropics and warm temperate regions.

Family Taccaceae

Perennial herbs, with tubers or rhizomes. Leaves basal, broad- or much-lobed, often long-stalked. Flowers bisexual, regular, in clusters; flower in the form of a tube, with 6 lobes above. Stamens 6, inserted on the fiower tube; filaments short; anthers with 2 locules. Ovary inferior, with 1 locule; style short; the 3 stigmas often petallike and reflexed over the style; ovules numerous; fruit a berry, opening by 3 valves; seeds with abundant endosperm. Two genera, 31 species, ranging from China to tropical Asia.

Family Pontederiaceae

Freshwater plants, floating or rooted and erect. Leaves with floating or emmersed blades. Flowers bisexual, mostly regular, in clusters subtended by a spathelike leaf sheath, bracts minute or absent; flower 6-lobed. Stamens 6 or 3 (rarely 1), inserted on the blossom, filaments free; anthers with 2 locules. Ovary superior, with 3 locules and numerous ovules, or with 1 locule and a single ovule; style entire or 3-parted. Flowers sometimes cleistogamous. Fruit a capsule, seeds with abundant endosperm. Six genera and 30 species, in the Old World and in the American tropics and subtropics.

Family Philydraceae

Erect herbs with short rhizomes. Leaves narrow, basal, 2-ranked. Flowers bisexual, irregular, solitary in the axil of bracts; flower segments 4, corolla-like, free. Stamen 1; filament flattened; anther with 2 locules. Ovary superior, with 1 to 3 locules; style slender; ovules numerous; fruit a capsule. Four genera, 5 species, ranging from China and New Guinea to eastern and southwestern Australia.

Critical appraisal. The classification presented above is the product of several contemporary botanists, prominent among them being A. Cronquist, an American, and A. Takhtajan, a Soviet citizen. It is the nature of a taxonomic system that it can be improved, and, as new and more critical data are amassed, it is expected that changes will be made in the disposition of the families.

Dissenting opinions

The inter-relationships of the families listed above are determined to a considerable extent on the basis of gross morphology, as offered by earlier workers (see in the *Bibliography* the works of Engler and Prantl, Hutchinson, and others). This approach has yielded important results but is now being enlarged by data from other disciplines, including (1) chromosome number and morphology, such as the presence of a unique and stable basic chromosome complement in the Agavaceae; (2) evidence from anatomy, such as the presence of laticifers in the Alliaceae, and (3) evidence from biochemistry, such as the presence or absence of saponins, alkaloids, and other compounds.

In view of these data, the twenty families listed above may be accommodated as follows (see also in the *Bibliography* the works of Traub, Sterling and Huang, and others): the families Tecophilaeaceae and Cyanastraceae have been placed under the new order Tecophilaeales, which is transferred to the subclass Iridae, leaving 18 families—plus two new families under the new order Alliales—again making twenty families to be accommodated under subclasses Liliidae (saponin containing) and Amaryllidae (saponin free).

Subclass Liliidae: Order Liliales, laticifers absent, chromosomes not of *Yucca-Agave* type; families Liliaceae, Xanthorrhoeaceae, Aphyllanthaceae, Philesiaceae, Smilacaceae, and Asparagaceae. Order Dioscoreales, laticifers absent, unique alkaloids present; families Stemonaceae, Discoreaceae, and Taccaceae. Order Agavales, laticifers absent, unique and stable basic chromosome complement of *Yucca-Agave* type present; family Agavaceae. Order Alliales, laticifers present, and unique alliaceous scent usually present; families Hesperocallaceae, Milulaceae, and Alliaceae.

Subclass Amaryllidae: Order Pontederiales; families Pontederiaceae and Philydraceae Order Amaryllidales; families Amaryllidaceae (unique alkaloids present), Alstroemeriaceae, Haemodoraceae, Hypoxidaceae, and Velloziaceae.

BIBLIOGRAPHY. I.H. BURKILL, "The Organography and Evolution of Dioscoreaceae, the Family of the Yams," J. *Linn. Soc. (Bot.)*, 56:319–412 (1960); A. CRONQUIST, *The Evolution and Classification of the Flowering Plants* (1968), an important modern outlook on the classification of flowering plants, similar to that of Takhtajan (below), but developed independently; G.L. DAVIS, *Systematic Embryology of the Angiosperms* (1966), the only available summary of the embryology of the angiosperms, with all families of the Liliales included; A. ENGLER, *Syllabus der Pflanzenfamilien*, 12th ed., vol. 2, *Angiospermen*, ed. by HANS MELCHIOR (1964), this revised edition of Engler and Prantl's *Die Natuerlichen Pflanzenfamilien* (1887, 1930), includes a classification of the order Liliales that is in accord with other modern arrangements; E.B. GRANICK, "A Karyosysyematic Study of the Genus *Agave*," *Am. J. Bot.*, 31:283–342 (1944), an important paper on the relationship of *Agave* and *Yucca* on the basis of chromosome complement; S.M. HUANG and C. STERLING, "Laticifers in the Bulb Scales of *Allium*," *Ant. J. Bot.*, 57:1000–1003 (1970), a report on important work concerning laticifers in *Allium* species; and "Notes on the Laticifers of *Allium, Caloscordum, Nothoscordum, Tristagma*, and *Tulhaghia*," *Pl. Life*, 28:43–46 (1972); R. HEGNAUER, *Chernotaxonomie der Pflanzen*, vol. *2, Monocotyledoneae* (1963), an important summary of the research in chemistry of the monocots, correlating it with the gross morphology of the plants; WILLIAM HERBERT, *Amaryllidaceae* (1837), a facsimile reprint of Herbert's *Amaryllidaceae*, 1837, preceded by an introduction by H.P. TRAUB (1970), a still valid classic of the Amaryllidaceae in the wider sense; J. HUTCHINSON, *The Families of Flowering Plants*, vol. *2, Monocotyledons*, 2nd ed. (1959), a definitive grouping of the Monocotyledons, including the Liliales, on the basis of gross morphology; A TAKHTAJAN, *Flowering Plunts: Origin and Dispersal* (1969; Eng. trans. from the 2nd Russian ed., 1961), an important modern book on the grouping of flowering plants, including Liliales, similar to that of Cronyuist (above), but developed independently; H.P. TRAUB, "The Tribes and Genera of the Agavaceae," *Pl. Life*, 9:134–137 (1953) and 28:137–138 (1972), the grouping of *Hosta* with the Agavaceae; *The Genera of Arnaryllidaceae* (1963), a late revision of the family Amaryllidaceae with information also on the Alliaceae; "The Subgenera, Sections and Subsections of *Allium* L.," *Pl. Life*, 24:147–163 (1968) and 28:132–137 (1972), grouping of the species of *Allium* into subgenera and sections; "The Order *Alliales*," *ibid.*, 28:129–132 (1972); and "Class Liliida, of Superclass Monocotyidra," *ibid.*, vol. 30 (1974), disposition of 20 families formerly grouped under order Liliales; T.W. WHITAKER, "Chromosome Constitution in Certain Monocotyledons," J. *Arnold Arbor.*, 15:135–143 (1934), an important paper on the chromosome complement of nine species of the Agavaceae and *Hosta*.

(H.P.Tr.)

Lima

Lima, the national capital of the Republic of Peru, is also the country's commercial and industrial centre. It is located on the south bank of the Rio Rimac, just inland from the Pacific Ocean port of Callao. Its name is a corruption of the Quechua Indian name Rimac, meaning Talker. A vast concentration of individual towns, greater Lima houses a population of more than 3,000,000 (1971). The city forms a modern oasis, surrounded by the Peruvian coastal desert and overshadowed by the neighbouring Andes mountains.

Lima, a focus for the wealth of contemporary Peru, is a centre of attraction for the country's rural population. Provided with a reserve of inexpensive labour, a nearby ocean port, and good transportation links to the interior, the city's industrial strength is assured. It is also, however, beset with the problems of overcrowding, air pollution, economic and social stratification, and inadequate public utilities. The gravest difficulty attends the provision of adequate food and water supplies for an ever growing population. Such supplies must be obtained from distant sources in the Andes and from the coastal valleys, and it may be difficult to meet future demands.

Spanish origins

History. Lima was founded in 1535 by the Spanish explorer Francisco Pizarro. It replaced Cuzco as the capital of the new Spanish kingdom of Peru because its coastal location facilitated communications with Spain. The city, walled in during the 1680s, was defended from incursions by pirates by the fortress-port of Callao.

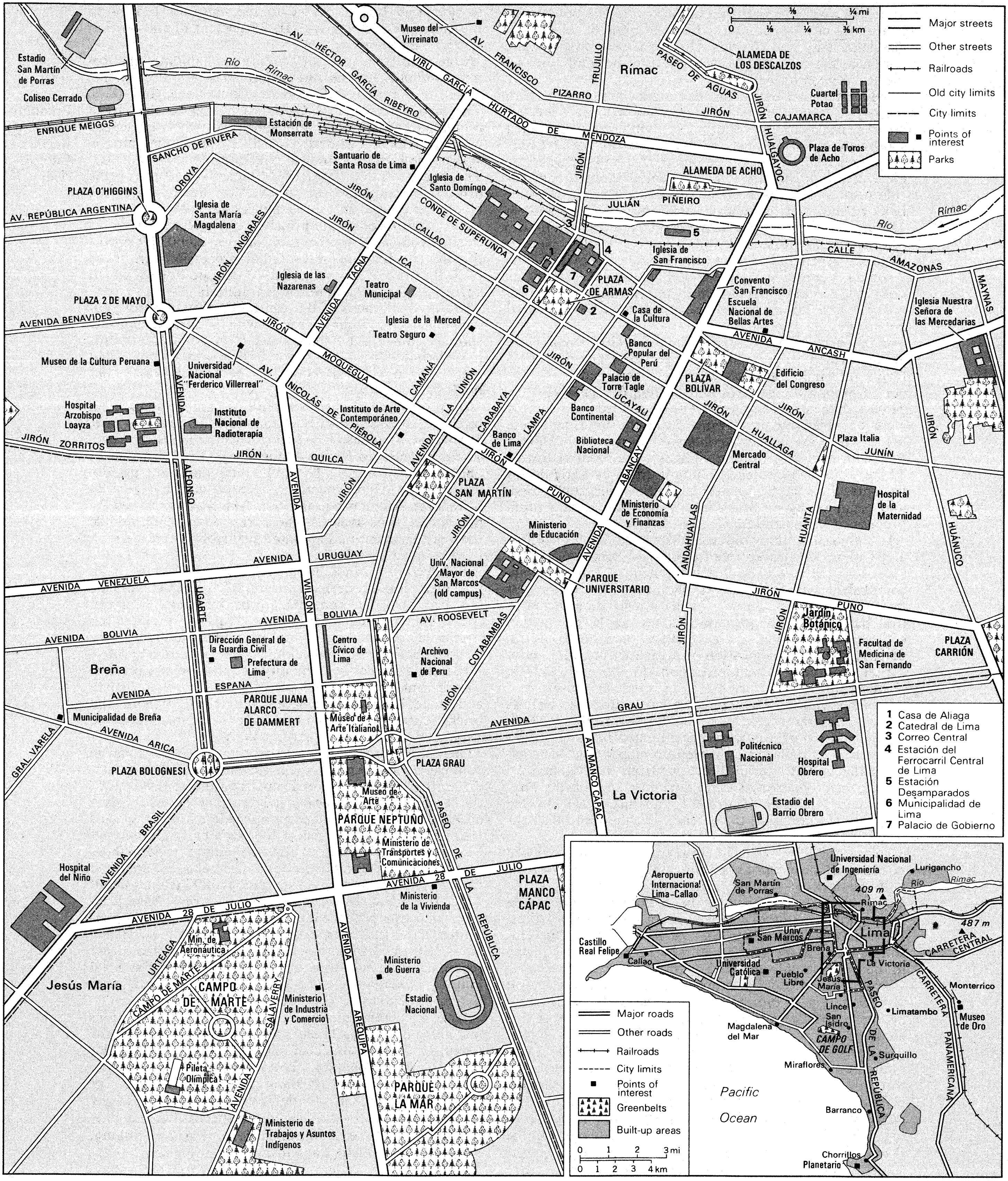
In planning the city, squares or blocks of 4,500 square yards were drawn, resulting in a checkerboard pattern. The founders set aside about 530 acres for the governor's house, for churches, and for the mansions of the colonists. In 1614 the city covered 776 acres and housed more than 25,000 inhabitants. The population grew steadily, reaching 54,000 in 1775; 64,000 in 1812; and 100,000 in 1876. By 1940 Lima had over 528,000 inhabitants and covered 5,560 acres.

During the 19th century, railways were built linking the city to Callao and to La Oroya in the Andes. The modern city then began to emerge as the major administrative, commercial, and industrial centre of Peru.

The contemporary city. Lima is located on the central Peruvian coast about eight miles inland from its Pacific port at Callao, at an elevation of 466 feet (154 metres) above sea level. Suburbs extend in all directions except to the east, where city growth is limited by the Andes. The city proper and the old suburbs cover an area of 27 square miles (70 sq km), while the province of Lima extends over 1,005 square miles (2,604 sq km).

The environment. Although Lima is in the tropical zone, the cooling effect of the offshore Peru (Humboldt) Current creates a temperate climate. The average temperature is between about 60" to 64" F (16" to 18° C) in the winter months from May to November and between about 70° to 80° F (21" to 27° C) in the summer months of December to April. Precipitation does not occur as rain but as a result of a condition known as Garúa, or vaporization of the atmosphere. The air is laden with constant mists that obliterate the sun during most of the winter. Annual precipitation averages between one and two inches. There are no strong winds, but light breezes blow from the sea during the day and from the Andes at night.

The cloud cover, the excessive humidity, the lack of winds, and the lack of direct sunlight combine to produce a high degree of air pollution. This condition is further complicated by exhaust fumes from vehicles.



Central Lima and (inset) its metropolitan area.

There is a paucity of plant and animal life in the vicinity. The plants are characteristic of deserts, while there is scanty vegetation on the river banks. There is, however, an extraordinary amount of animal and vegetable wealth found in the Pacific Ocean. The Peru Current has given the Peruvian sea quantities of plant and animal plankton that provide the ecological basis for plentiful marine life, which includes anchovies, flounder, shrimp, lobster, seals, and whales.

The city plan. The old city retains its original checkerboard pattern, and the streets have retained their original names; many have also acquired unofficial names relating to events that have occurred along them. The boundaries of the old city—once marked by the ruins of its walls—are defined by the avenues Alfonso Ugarte to the west, Unión to the east, and Miguel Grau to the south, and by the Rio Rimac to the north.

The first urban extension was made in the early 19th century to Magdalena Vieja (now Pueblo Libre), located on the coast to the southwest. Settlement continued southward to Miraflores and Barranco, ending at Chorrillos, where the river valley, the desert, and the seashore meet. These towns are both residential suburbs and seaside resorts. The northern suburb of Rimac, on the right bank of the river, is opposite old Lima. During the 20th century, the city expanded greatly to the west along the old seven-mile horse and wagon route to Callao. Named Colonial Avenue, it is now the Avenida Benavides. The Avenida Venezuela runs parallel to it to the south, as does the Avenida Argentina to the north. The first industrial and large commercial operations were located between Lima and Callao.

In the open country surrounding the city, small rural towns formed around chapels or parishes. The numerous farms or small tracts of cultivated land that were dispersed among the suburbs and barren, dry land also became urbanized. They were either divided into lots or were occupied by immigrants from the interior. These areas, which are called *barriadas* (slum neighbourhoods) or *pueblos jóvenes* (young villages), are squatter towns. There were also clusters of huts lived in by agricultural workers. Sixty large ranches were, however, subdivided, and occupied by *pueblos jóvenes*, thus initiating a rapid process of urbanization.

Although the industrial establishments between Lima and Callao are concentrated along Argentina and Benavides avenues, the overall development of Lima was accomplished without integral planning. The pattern of land use has given rise to a haphazard location of industrial buildings in inappropriate zones, to the occupation of green and productive areas without plan, and to a lack of planning in the allocation of areas for recreation, education, and central administration. In the early 1970s, however, the city was being zoned for industrial use.

Transportation. The Lima–Callao railroad as well as the electric tramway connecting Lima, Miraflores, Barranco, and Chorrillos were discontinued in 1963. These systems were replaced by a well-developed highway network that radiates in all directions from the old city. To the north the highway system is connected with the Pan-American Highway (Carretera Panamericana), to the south with Chorrillos, to the east with the Central Highway, and to the west with Callao.

During the 1960s, the internal traffic of Lima reached unexpected proportions, complicated by narrow streets and the lack of parking space. The congestion in the city's streets was relieved by the opening of the Paseo de la República (Republic Drive) from Lima to Chorrillos.

The international airport is located at Callao. The Central Railway of Peru, the highest standard-gauge railway in the world, climbs the Andes from Lima northeast to La Oroya. The railway line from the city to Callao was built in 1850 and is the oldest in South America. There is also a northern coastal railway to Paramonga and a short line south to Lurín.

The people. In 1971 the population of greater Lima numbered 3,130,000. The inhabitants were distributed throughout 32 contiguous districts, as well as in satellite towns separated from the principal urban nucleus. The population density varies from district to district. It is high along the main roads in such areas as Comas in the northwest, Agustino in the east, and Atocongo in the south. When migrants first arrive in Lima, however, they move to the centre of the city and accumulate in the districts of Surquillo, La Victoria, and Rimac, which now have up to 129,500 inhabitants per square mile. The average density in the city as a whole is 31,000 inhabitants per square mile.

Almost 75 percent of the population is classified as either white or mestizo. (In Peru the term mestizo applies to those of mixed white and Indian parentage as well as to urbanized Indians.) Due to the influx of Indians to the cities since World War II, 25 percent of the population is of Incan origin.

Predominance of Roman Catholicism

More than 97 percent of the city's inhabitants are Roman Catholic. Of the remainder, most are Protestants; there are a very few Buddhists, Confucians, and Jews. The high percentage of Catholics has given the city a traditional, conservative atmosphere, which is evidenced by large religious gatherings such as the annual procession of the Lord of Miracles (El Señor de los Milagros). The image of the "Lord of Miracles"—a copy of the original fresco painted on the wall of the church of the Nazarenes more than three centuries ago by a black slave—is carried through the principal city streets. The enormous accompanying procession of over 2,000,000 people fills the city for three days. Other processions are those of Santa Rosa of Lima and San Martin de Porras, which perpetuate the colonial religious tradition.

Housing. Because of the constant influx of rural people to the city, Lima is faced with a chronic housing shortage. Since 1960 there have been attempts to promote building in, and to supply services to, overcrowded areas. The unplanned nature of the city, however, has made the task difficult.

In the *barriadas,* houses are built of bamboo mats, mud brick, or collapsed oil tins. Sanitation facilities are rare, and water often has to be brought to the areas by truck. There are also *callejones*—long alleys lined with small one-story apartments. The individual units are usually overcrowded. Upper income residential suburbs—such as Miraflores, Monterrico, and San Isidro—are marked by fine homes, wide avenues, and parks. Since 1965, middle income apartment buildings and two-story dwellings with gardens have been built.

Architecture. Lima has blended colonial styles with the advances of modern architecture. The rainless climate has permitted the construction of buildings with flat roofs, while the frequency of earthquakes in the area has precluded the building of multistoried structures. Large enclosed balconies in the colonial style often offer protection from the excessive humidity and the cool sea breezes. The ancient building materials of adobe, bamboo thatch, and wood have been replaced with steel, cement, and bricks; structural innovations have been introduced to avoid disasters caused by earthquakes.

There are many buildings remaining from the Spanish colonial period. Among them are the Cathedral (Catedral), the temples of San Francisco, and the Palace of the Tagle Tower (Palacio de Torre Tagle). The Paseo de Aguas and the tree-lined promenade of the Descalzos (barefoot priests) also reflect the city's early elegance.

Commerce and industry. At the end of 1965 there were almost 2,800 industrial establishments in Lima and Callao. The labour force of about 500,000 includes about 125,000 workers engaged in manufacturing. The principal industries produce textiles, plastics, timber, pharmaceutical products, chemicals, and synthetic fibres. Heavy industry is also important. Food processing and drink manufacturing industries are numerous. The fish-meal industry in Callao is of considerable national and international importance.

Banks are represented by both state credit institutions and private capital. There are 22 banks, including the Continental Bank (Banco Continental) and the Popular Bank (Banco Popular). The Lima Bank (Banco de Lima), the National City Bank, and the Credit Bank of Peru (Banco de Crédito del Perú) are supported by French, U.S., and Italian capital, respectively. The state banks encourage agriculture, cattle raising, and mining. There are many finance and insurance companies.

The cooperative system is well developed in Lima, having 12 corporations engaged in construction; there are also numerous mutual benefit and savings associations.

The main commercial districts are Limatambo, Miraflores, and San Isidro. There are over three dozen food markets and a wholesale produce market. The itinerant food markets, or *paraditas,* which sell at retail prices, are characteristic of Lima. Meat and milk are often of short supply and must be rationed.

Water supply

Public utilities. Lima consumes 850,000 quarts of water daily. Local supplies are, however, inadequate because the Rio Rimac is low nine months of the year. To augment the water supply a six-mile tunnel was built under the cordillera of the Andes to obtain water from the lakes located in the high sierra east of Lima. It is estimated that, by 1980, Lima will lack water if other sources, such as the Rio Montaro to the east, are not utilized.

Water is also needed for the generation of electricity. In 1966 the installed capacity for the city was about 200,000 kilowatts. By 1971 the capacity installed for public service had a potential of 605,000 kilowatts.

Health, safety, and education. In 1969, health facilities included 22 hospital centres, 18 state and social security hospitals (caring for those covered by social security), and 22 private clinics.

Service in the fire brigades that man the city's 14 fire stations is honorary and voluntary. The state subsidizes the acquisition of fire-fighting equipment.

Public order is maintained by the police or civil guard and to a smaller degree by the municipal police. Schools, which include some religious establishments, include about 2,800 primary schools, a number of state technical schools, and more than 700 private schools. There are about 1,240,000 children of school age.

University City

There are five national and seven private universities attended by 120,000 students. The most important is the National University of San Marcos (Universidad Nacional Mayor de San Marcos de Lima); founded in 1551, it has a new campus outside Lima in the University City. The agrarian and engineering universities also have extensive campuses awaiting development.

Cultural life. The Academies of Language and of History are the principal cultural centres. The government's House of Culture (Casa de la Cultura del Perú) encourages the intellectual, literary, and artistic elements in national life. Other institutions are the National School of Fine Arts (Escuela Nacional Superior de Bellas Artes), the National Symphonic Orchestra (Orquesta Sinfónica Nacional), the National Conservatory of Music (Conservatorio Nacional de Música), and the National Theatre.

There are about 40 museums in the Lima metropolitan area. These include the Museum of Gold (Museo de Oro); the Museum of Art; the Museum of Italian Art; the Archaeological Museum (Museo Arqueologico "Rafael Larco Herrera"); the National Historical Museum; and the Museum of the Viceroys. The restored castle of Real Felipe in Callao is also a museum. There is also an astronomical observatory and planetarium, a botanical garden, and the "Park of Legends," which contains examples of Peru's varied animal and plant life. The Convent of San Francisco, with its catacombs and its gardens with tiled patios and fountains, is now open to the public after four centuries. It has a valuable library of incunabula, as well as works of religious art dating from the time of the viceroys. Numerous restored houses containing valuable furnishings and artwork are open to the public. The House of Aliaga contains the sword of the conquistador Francisco Pizarro and historical objects dating back to the time of the city's founding.

The National Library, founded by José de San Martin, the liberator of Peru, contains copies of the first printed works in Peru. The National Archives of Peru (Archivo Nacional de Perú) include judicial and historical collections. There are libraries attached to the universities, government ministries, and private associations, as well as 40 public libraries.

There are a dozen daily newspapers. *El Comercio,* founded in 1839, is Peru's oldest paper. *Ojo,* a morning paper, has the largest circulation. *El Peruano* is the official state gazette, while *Expreso* is the leading opposition daily. The headquarters of the foreign press agencies in Peru are located in Lima. There are also numerous periodicals and reviews covering cultural affairs, economics, and current events.

Lima is served by more than 30 radio stations. Most of them are commercial stations, two are government stations, and eight are cultural. Television programs are broadcast by five commercial stations.

Recreation. Recreational facilities include the racetrack, the bullfight plaza, cockpits, and the football (soccer) stadiums. There are three principal theatres, as well as numerous experimental university theatres, drama centres, and dramatic schools. Among these, the Theater of the University of San Marcos (Teatro Universitario de San Marcos) has a high reputation.

Favourite sports include football, golf, tennis, polo, swimming, and surfing. The popular spectator sports are bullfighting, cockfighting, and horse racing. There are more than 100 cinemas.

In addition to some of the most luxurious dining rooms in Latin America, there are also many popular Oriental restaurants, as well as those that serve typically Peruvian meals to the accompaniment of music and folk dances. The taste for folk music is evidenced by the existence of large coliseums in which national music and dance programs are attended by huge numbers of spectators.

BIBLIOGRAPHY. JUAN BROMLEY, *La fundación de la ciudad de los reyes* (1935); and with JOSE BARBAGELATA, *Evolución urbana de Lima* (1945), are the most important works on the foundation and growth of the city of Lima—they summarize Spanish-language historical works of the mid-20th century. Other features of the city are discussed in RAUL PORRAS BARRENECHEA (comp.), *Pequeña antologia de Lima (1535–1935)...* (1935); JOSE GALVEZ, *Calles de Lima y meses del año* (1943); JOSE MATOS MAR, *Estudio de las barriadas limeñas* (1955); and ALLAN G. AUSTIN and SHERMAN LEWIS, *Urban Government for Metropolitan Lima* (1970). JOSE CARLOS MARIATEGUI, *7 ensayos de interpretación de la realidad peruana* (1928), contains a chapter on the "centralism" of Lima that argues that the centralization of Peruvian politics benefits the city to the detriment of the rest of the nation; EMILIO ROMERO, *Geografía económica del Perú,* new ed. (1939), argues that the prosperity of Lima is founded in its geographic and economic background rather than in its privileged political situation.

(Em.R.)

Limestones and Dolomites

Limestones and dolomites are sedimentary rocks (those formed by the accumulation of sediment in air or water) that consist of more than 50 percent carbonate minerals, mainly calcite (calcium carbonate, $CaCO_3$) and dolomite (calcium magnesium carbonate, $CaMg(CO_3)_2$). For this reason, they are commonly referred to as carbonate rocks.

Although their distinction is based on their rather uniform mineralogical composition, limestones and dolomites are formed by a wide variety of processes. A limestone may be produced by the cementation of carbonate sand grains derived by erosion of a former carbonate landform (*e.g.*, island, hill, plateau); in this case it is a clastic rock, a sandstone, whose grains have been transported and deposited in accord with the laws of hydraulics. On the other hand, a limestone might result from direct precipitation in seawater, in which case it is a chemical rock, like salt, whose formation has been governed by such factors as water temperature and pressure, and the concentration of the solution. In still another category are carbonate bodies, such as reefs, that have been constructed principally by sedentary organisms; in this case the rock-forming process is skeletal secretion by physiological means. Since the three types of carbonate sediments may be mixed while they are still in the depositional environment or may be modified by various postdepositional changes, the final product may be a rock with a very complicated history.

Significance and use of carbonate rocks

Limestones and dolomites have long fascinated geologists and paleontologists because of their rich fossil content. Much existing knowledge of the Earth's chronology and evolution has been derived from study of fossils included in carbonate rocks. They represent about 20% of all sedimentary rocks and occur in all continents, in strata of every age—from Precambrian to Quaternary.

In the last few decades enormous petroleum reservoirs have been discovered in carbonate rocks in the Middle East, the Canadian Rockies, west Texas, and many other parts of the world. Some of these oil fields are localized in fossil reefs, such as Norman Wells (northern Canada), Leduc (Alberta), the Permian Basin of Texas, New Mexico, and the Tampico area (Mexico).

Carbonates have a large variety of economic uses. As building stone, limestones and dolomites are used for monuments, exterior and interior facings, and flooring of several kinds. Crushed rock is used for railroad ballast, riprap fill (broken rock used to protect structures from natural processes of erosion) around the bases of dams and piers, filter beds in sewage treatment, and surfacing for airports. The heating of limestones and magnesian limestones to temperatures of 900°–1,000° C (1,650"–1,800" F) will dissociate calcium carbonate and yield carbon dioxide and lime for commercial use. Lime has major applications in the construction industries, for the manufacture of glass, and for agricultural purposes. Carbonate rocks are also common host rocks for many ore deposits, because they are easily attacked by high-temperature solutions that precipitate ore minerals.

Treated in this article are the composition of limestones and dolomites and their textures, structures, origin, and geological significance. For further information on the mineral constituents see CARBONATE MINERALS; and for a discussion of the relation of limestones and dolomites to sedimentary rocks in general see SEDIMENTARY ROCKS. See also CAVES AND CAVE SYSTEMS; CORAL ISLANDS, CORAL REEFS, AND ATOLLS; DURICRUSTS; and EVAPORITES for the occurrence of carbonate rocks in natural environments, and OCEANS AND SEAS and ELEMENTS, GEOCHEMICAL DISTRIBUTION OF for treatment of carbonate geochemistry.

PHYSICAL AND CHEMICAL CHARACTERISTICS

Mineral and chemical composition. The most important rock-forming carbonate minerals are calcite, aragonite, and dolomite. Other carbonate minerals of the calcite family are magnesite ($MgCO_3$), rhodochrosite ($MnCO_3$), and siderite ($FeCO_3$); these occur in restricted environments and are quantitatively of limited importance.

Calcareous sediments forming today in tropical and subtropical shallow seas consist predominantly of aragonite and magnesium-rich calcite. Both minerals are metastable, however, and change quickly with geological time. All limestones of Precambrian to Tertiary age (*i.e.*, older than 2,500,000 years) consist of normal calcite; that is, calcite with a low magnesium content. This is because the rocks have undergone diagenetic (postdepositional) processes that changed their original mineralogical character. Only limestones formed in deep or cold marine environments have undergone no transformation; they were made of normal, low-magnesium calcite from the outset.

Minor constituents

Minor constituents also are commonly present in limestones and dolomites. These may be finely disseminated throughout the rock (*e.g.*, clay minerals, iron oxides, bituminous matter), present as individual grains (*e.g.*, glauconite, pyrite, quartz, and feldspar), or segregated into large nodules (*e.g.*, chert or flint). Compositional gradations are common, and many classification schemes have been used. In the chemical sense, however, carbonate rocks are rather simple because they reflect their mineralogical composition. Calcium and magnesium oxides and carbon dioxide are the most important constituents, forming more than 95 percent of the rock in many cases. When impure carbonates are involved, other constituents, such as silica, alumina, iron oxides, phosphorus, and sulfides, may contribute notably to the bulk composition.

Depositional textures. The two main classes of carbonate rocks are mechanically deposited carbonates, those with particulate texture, and carbonates grown in place, either by biological activity (reefs) or chemically accreted (tufa, travertine, caliche, and speleothems, or cave deposits). The mechanically deposited carbonates constitute the most important group; the other types are of more limited extent.

The particulate material consists of a mixture of various particles (grains) immersed in a finer matrix (lime mud). The rigid grain framework creates a pore system that permits the transmission and storage of fluids. Subsequently, this pore system may become partially or completely filled by a chemical precipitate (cement). The particles and their matrix produce varying rock textures, depending upon their arrangement, and the textures of limestones and dolomites exert a controlling influence on porosity and permeability and, in turn, on rock suitability as a reservoir for oil or water (see further PETROLEUM; GROUNDWATER). Five major types of grains are recognized in carbonates; namely, skeletal, detrital, pellets, lumps, and coated grains.

Skeletal *grains.* Grains consisting of whole and fragmented skeletons of marine plants and animals are called skeletal grains. Unlike other kinds of grains, they usually have a precise internal structure resulting from a regular crystal arrangement. A wide variety of marine organisms (mollusks, algae, corals, echinoderms, sponges, and foraminifera) of quite different character contribute their skeletal remains. Accordingly, different kinds of skeletal sediments may form. The skeletal grains can be studied and classified according to their environmental occurrence or to their mineralogy, biology, or sedimentological character. In this last case they are treated as grains and not as organisms or fossils; their shape, hydraulic behaviour, and relative resistance to breakdown of the whole skeleton are important factors. Six varieties of skeletal grains can be distinguished: (1) sheets and spicules, formed by disintegration of green algae (*Penicillus* species), sponges, and alcyonarians; (2) segments, formed by disintegration of segmented organisms such as green and red algae (*Halimeda* species), crinoids, and asteroids after their death; (3) branches, formed by mechanical or biogenic breakage of branched animals and plants, such as some species of red algae, corals, and bryozoans; (4) chambers, formed from hollow or partly hollow skeletons of gastropods, brachiopods, pelecypods, and foraminifera; (5) crusts formed on surfaces by such plants and animals as red algae, bryozoans, foraminifera, and annelid worms; and *(6)* massive grains—that is, those formed by colonial corals and coralline algae.

Varieties of skeletal grains

Detrital grains. Grains made of debris derived from pre-existing rocks are called detrital grains. They may originate during deposition as pieces broken from poorly consolidated sediments (intraclasts) or as pieces broken from rigid rocks (lithoclasts). Detrital grains usually have sharp angular corners, may show an internal texture derived from the source rock, and a single specimen may be essentially of one or of varied rock types.

Intraclasts form after the sediment becomes sufficiently coherent if there is an agent able to break it; they are abundant where erosion, dynamic energy—generally in the form of waves and currents—and burrowers occur. Intraclasts can be formed from carbonate mud when it is compacted or when it dries out, but in carbonate sands, rapid cementation (filling of the pores by a chemical precipitate) must occur. It is easier for sediments to become compacted or desiccated than cemented; thus, intraclasts usually have a fine-grained (mud) texture.

Lithoclasts form when carbonate rocks are exposed and eroded. They usually are related to land areas, such as a carbonate tropical island. They may represent minor unconformities, or interruptions in the depositional sequence.

Pellets. Grains of mud lacking significant internal structure are called pellets; they usually have an ovoidal or subspherical shape. They are a polygenetic group of grains, because some of them are of fecal origin (fecal pellets); whereas others (grains of matrix) may represent processes of physical accretion, diagenetic (postdepositional) alterations of skeletal grains, and recrystallization; and still others may represent bits of lime mud torn from the sea floor and rolled around before coming to rest. Genetically, the latter should belong to the intraclast category, but because of their small size it is impossible to recognize their detrital origin.

Fecal pellets, which are the most important type, range in size from tens of microns (one micron equals 0.001 millimetre) to a few millimetres. Their shape may be ovoidal, spherical, cylindrical, or discoidal, and some exhibit external sculpture. The smallest are spherical and the largest rodlike. Fecal pellets are produced by animals that eat bottom sediments (worms, holothurians, sea stars) and by filter feeders (oysters, tunicates, sponges,

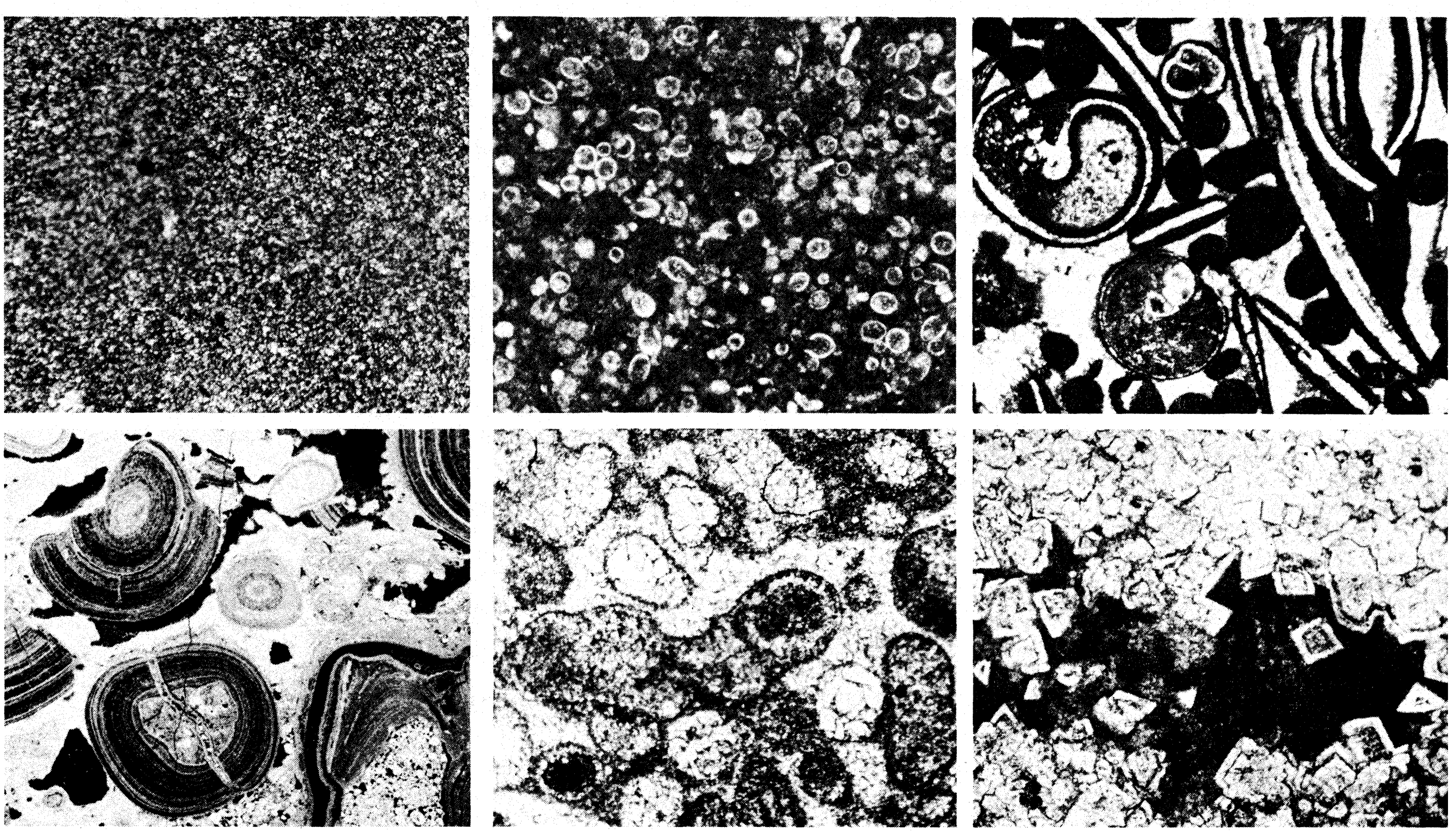


Figure 1: Photomicrographs of carbonate rocks.
(Top left) Micritic limestone, Triassic (magnified 18 x). (Top centre) Micritic skeletal limestone. Upper Jurassic (magnified 25 ×). (Top right) Skeletal oolitic limestone, with clean calcite cement, Lower Triassic (magnified 18 x). (Bottom left) Pisolitic dolomite, Upper Triassic (magnified 5 x). (Bottom centre) Lump limestone, recrystallized, Permian (magnified 18 x). (Bottom right) Diagenetic dolomite growing at the expense of existing micritic material, Middle Triassic (magnified 15 ×).
BY courtesy of A. Bosellini

some gastropods, and some worms). Their consistency covers a wide spectrum, from gelatinous to compact to very hard, almost lithified.

Significance of pellets

These grains are of considerable geologic importance, because they are very abundant in sequences of carbonate strata. In making them, the animals change the size of the original sediment: sand is produced from mud through biological elaboration.

Lumps. These are composite grains (*i.e.*, aggregates of particles, generally pellets), in which the single element protrudes from the principal body and gives it a characteristic lobate outline. In Recent sediments (those formed during the last 10,000 years) various forms can be distinguished (grapestones — forms resembling bunches of grapes — botryoidal lumps, friable aggregates); in ancient carbonates they are regarded as a single group. Lumps form in environments in which grains lying on the sea bottom tend to cement together; they occur in carbonate sand tongues controlled by tidal currents that flow through inter-key channels and are frequently associated with oolitic sands described below.

Coated ***grains.*** Grains having concentric or enclosing layers of calcium carbonate around a central nucleus are called coated; the most common types are oolites, pisolites, and algae-encrusted grains. Oolites (from the Greek *don,* "egg," and *lithos,* "stone") are subspherical grains less than two millimetres in size; if bigger, they are called pisolites. The nucleus is generally a skeletal grain, a pellet, or, more rarely, a quartz grain; some oolites may display a radial structure.

Origin of oolites

There are various theories of the origin of oolites. The first, introduced a century ago, is called the "snowball" theory. According to this theory, seawater is presumed to be crowded with suspended aiagonite crystallites in a particular place; as they roll on the bottom or float in the water, particles of whatever origin, possibly coated with adhesive mucilaginous organic material, capture the crystallites, which adhere to their surfaces. As a result of the continuous movement caused by waves and tidal currents, the small grains grow just like a snowball rolling downhill. **A** second theory explains oolite growth by chemical precipitation producing crystals perpendicular to the grain surface; as the grain rolls, these crystals break and lie on the surface, trapped by the organic mucilage. A third theory attributes oolite origin to the presence of proteinic matter that supposedly controls precipitation and orientation of crystals.

Matrix material. This is generally a very fine lime mud, now almost universally called micrite (originally from the contraction of the words microcrystalline calcite). Micrite, which applies to both consolidated and unconsolidated carbonate mud, is the carbonate analogue to the mud that occurs between the framework grains of many land-derived sands. There is considerable controversy over the size limit that is to distinguish matrix (micrite) from grains; four, 20, 30, 60, and 125 microns (one micron equals 0.001 millimetre) have been proposed. The origin of micrite also is controversial; two different interpretations exist. According to the first, lime mud is derived from the production of aragonite needles by precipitation from waters of abnormally high salinity and carbonate saturation. According to the second, micrite is essentially derived from the skeletal parts of marine organisms (corals, mollusks, calcareous algae, foraminifera, nannoplankton), mainly as a result of: (1) production of aragonite needles by postmortem disintegration of calcified green algae (*Penicillus* species); (2) production of mud-sized skeletal debris by predominantly physical processes of particle-size reduction in agitated environments; (3) production of mud-sized skeletal debris other than aragonite needles by predominantly biological processes of particle-size reduction in quiet-water environments (activity of boring micro-organisms, mastication and ingestion by the vagrant benthos); (4) production of lime ooze by accumulation of microscopic organisms.

Although direct chemical precipitation from saturated waters cannot be excluded, much research indicates that lime mud is largely of skeletal origin.

Stromatolites, bioherms, and reefs

Structures in carbonate rocks. Mechanically deposited carbonate rocks show the same structures that are present in terrigenous (land-derived) rocks: cross-bedding (inclined and intersecting layers), graded bedding (the grain size changes from coarse to fine, vertically), ripple marks, mud cracks, laminations (very thin sedimentary layers), and cut-and-fill structures (pockets formed by erosion and subsequently filled with sediment) may be well displayed in many limestones. There are some structures that are peculiar to carbonate rocks, however; stromatolites (fossil algae), stylolites (columns or veins that are inclined to the beds), nodular bedding, and various solution features are a few examples. Stromatolites are organic-sedimentary structures composed of flat, undulated, or cabbage-like laminations; these are built by dense mats consisting primarily of blue-green algae, which selectively trap and bind sediment particles among their mucilaginous filaments. Stromatolites range in age from the early Precambrian (perhaps 2,500,000,000 years old) to the Recent, although they are most abundant in the Precambrian and Lower Paleozoic (about 400,000,000 to 570,000,000 years ago). Recent stromatolites are forming in littoral (near shore) environments in south Florida and in the Bahamas; spectacular examples are those of Shark Bay (Western Australia), where stromatolitic columns more than one metre tall occur. These have an elliptical cross section, the long axis being oriented with the prevailing current. Similar current-oriented stromatolites have been discovered in the Precambrian of the Northwest Territories, in Canada.

Biohermal and reef structures are other typical structures of carbonate rocks. Bioherms are domelike, moundlike, or otherwise circumscribed masses built exclusively or mainly by sedentary organisms and enclosed in a normal rock of a different lithologic character. They vary in size and shape and may be as small as a few centimetres or impressive structures several kilometres across and hundreds of metres thick. Algae, stromatoporoids, coral, crinoids, and mollusks are among the most common constituents.

ORIGIN OF LIMESTONES AND DOLOMITES

Source material and its transportation. Limestones originate mainly through lithification (transformation to rock) of loose carbonate sediments. Exceptions are deposits like travertine or reefs, which are lithified in their original state. The primary origin of most limestones is attributable to biochemical or physicochemical processes.

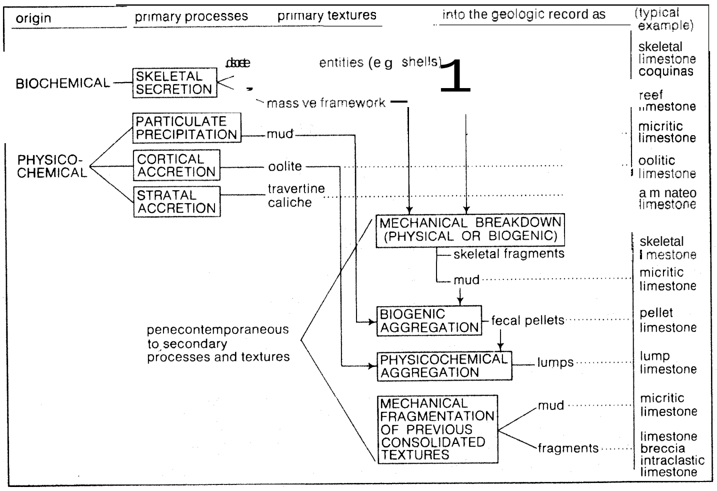


Figure 2: The origins of limestones.

Skeletons of marine plants and animals are produced biochemically, through skeletal secretions. If cemented, these skeletal sediments may be transferred directly into the geologic record as skeletal limestones, reefs, or bioherms. Otherwise, they may undergo early diagenetic processes like biogenic breakdown or biogenic and physicochemical aggregation. In this way, muds, fecal pellets, and lumps can be produced from carbonate matter of biochemical origin, and such rocks as micritic limestones, pellet limestones, and lump limestones will be formed.

Lime muds, oolites, and travertine are produced physicochemically, through particulate precipitation and layered accretion, although travertine may also be controlled by algal activity. These carbonate deposits may become micritic limestones, oolitic limestones, and laminated limestones, respectively.

Both biochemically derived and physicochemically derived carbonate sediments may be mechanically fragmented after early consolidation, and clastic sediments and rocks such as limestone breccias, intraclastic limestones, and carbonate mudstones may be produced.

The dolomite problem

Poor knowledge of the origin of ancient dolomites, which are widespread, together with the apparent lack of dolomite in Recent sediments, created the "dolomite problem," which has been debated for many years. The greater frequency of occurrence of dolomites in Lower Paleozoic and Precambrian strata had led at one time to the hypothesis of ancient "dolomitic seas" that were presumably quite different from modern ones. All of this was abruptly changed by the development of X-ray diffraction techniques. In the last few years dolomite has been discovered in Recent sediments, and an understanding of its origin and occurrence in the geologic record has been gained.

All the Recent dolomite occurs in a particular environment; namely, hypersaline lakes or lagoons, and supratidal areas (Persian Gulf, Bahamas, Florida, The Netherlands Antilles). Hypersalinity is the hallmark.

Hypersalinity in marginal marine environments and in intermontane basins may result because of excess of evaporation over rainfall; interstitial waters in the sediments transpire upward and evaporate at the sediment–air interface. This process is common in intertidal and supratidal flats in the tropics. Hypersalinity may also result when evaporation exceeds precipitation plus runoff, thus increasing the concentration and density of the waters in lagoons and shallow coastal seas. The heavy brine that is formed sinks and flows seaward down the sloping shelf. If, however, this reflux to the sea is prevented by natural barriers such as reefs or sills, the brine migrates to the lowest possible topographic depressions and seeps slowly through the underlying sediments, which are progressively dolomitized.

Hypersaline brines are also responsible for a subsurface process by which older calcareous rocks can be transformed into dolomites. Volcanic activity, faults, and special subsurface traps or other conditions control the migration of the brines and the dolomitization process.

Two major categories of dolomites, recognizable by their petrographic characters, are present in the geological record. Primary (syngenetic) dolomites are formed by chemical precipitation or substitution of originally calcareous sediments at the time of deposition. They retain their original texture (micrite, pellets, skeletal grains, and intraclasts) and structures (laminations, desiccation features), are poorly fossiliferous (ostracods, foraminifera, algae) and usually fine grained. Secondary (diagenetic) dolomites have a crystalline texture and are caused by replacement of former calcareous rocks or by recrystallization of primary dolomites. They are generally coarsely crystalline, porous, structureless, and massive and are often related to faults and volcanic activity. Relict textures (those originally present in the rocks) are frequently recognizable.

A third kind of dolomite that is petrographically recognizable is a chemical precipitate that partially or totally fills cavities in primary dolomites. It has crystalline texture and forms quite early in the genetic history of the rock. Cavities in primary dolomites may be lined by rounded masses of dolomite (colloform texture).

Carbonate sediments, unlike terrigenous sediments (*i.e.*, silicate muds, silts, and sands), usually form entirely within a local sedimentary basin—they are deposited and accumulated in the same area in which they are produced. In contrast with the terrigenous sediments, there is relatively little net lateral transportation, because of irregular physiography (shoals, islands, reefs) and, especially, the low gradient of the shallow-water carbonate platforms. Sometimes in the presence of steep submarine slopes or scarps, which occur along Recent and ancient carbonate platforms, shallow-water carbonate sands may be carried into other environmental regimes (oceanic bottoms) by turbidity currents (density currents that flow along the bottom by reason of their high sediment concentration). These displaced deposits are known from Recent deposits (Tongue of the Ocean, Bahamas) and from the Jurassic (136,000,000 to 190,000,000 years ago) troughs of the Mediterranean area.

High-energy and low-energy environments

Depositional environments. The hydrodynamic significance of the texture of carbonate sediments is that lime mud accumulation depends upon local conditions of water agitation. In a particular rock, the relative proportion of mud matrix to cement is an index of water agitation or mechanical energy. Depositional areas of "high energy" are inferred to produce clean, well-sorted, coarse-grained carbonate sediments, whereas areas of "low energy" are held to be responsible for muddy deposits. High-energy sites are beaches, surf zones, dunes, and tidal channels. Carbonate muds generally accumulate in protected areas such as lakes, lagoons, deep-sea basins, and areas on the lee side of major islands situated on oceanic banks.

Textural studies of Recent carbonate sediments indicate that the size, shape, and sorting of skeletal sands are greatly influenced by the nature of the skeletons. Even in quiet-water environments such as the lagoons of many Pacific atolls, for example, the green calcareous alga *Halimeda* can produce very coarse sand because it is segmented; when the alga dies the various segments are released. If the same coarse skeletal sand were made of mollusk fragments, however, it would be considered to represent a high-energy deposit—a beach, for example.

Carbonate rocks form in various environments; namely, continental, marine, or transitional. Most of them are formed in shallow marine and transitional zones of the tropical seas. Criteria that help to distinguish different environments of deposition may be found in grain types, sedimentary structures, geometry of sedimentary bodies, fossil content, and stratigraphic relationships with the rocks that are adjacent, both laterally and vertically. Diagnosis usually is achieved by the association of various distinctive elements.

Continental environments. Carbonate rocks such as crusts or soils may form in a semi-arid climate in which the predominant direction of movement of soil moisture is upward, because of the excess of evaporation over rainfall. Caliche, well known in the southwestern United States, is a limestone that precipitated from evaporating soil moisture; similar deposits are known in the Mediterranean region, where they are called nari and calcrete. Caliche-like deposits, containing pisolites and concretions formerly thought to be algal in origin, have been recognized in Mesozoic and Paleozoic carbonate sequences.

Spring, cave, and dune sand deposits

Another carbonate rock of continental origin is travertine, the genetic term for all organic or inorganic, nonmarine limestone accumulations that formed in lakes, rivers, springs, and caves; it is colloform (in rounded masses) or accretionary in individual layers, and highly porous.

Dunes composed of carbonate sands are abundant along the tropical coasts, where beaches may consist of skeletal fragments, oolites, and other carbonate grains. If cemented, these eolian deposits form eolianites, which display spectacular cross-bedding. Many famous islands, such as Bermuda and the greater part of the Bahamas, consist entirely of calcareous dune deposits.

Transitional and shallow marine environments. Knowledge of these particular and very important environments has been gained from studies made in tropical regions such as the Indo-Pacific islands, the Australian coasts, the Persian Gulf, the Bahama Banks, southern Florida, and other areas of the Caribbean and Gulf of Mexico. There is a vast range of subenvironments that all fit into the general pattern of an open sea with relatively deep water on one side; a marginal zone of high energy with turbulent and shallow water, where waves break on a reef, barrier beach, or sandy shoals; and a sheltered lagoon of variable width and depth that passes to the mainland through tidal marshes, mangrove swamps, or rocky shores. Sometimes there is no mainland but only a wide shallow bank or lagoon that grades into a turbulent marginal zone and beyond that into deep water on all sides.

In the high-energy marginal belt, oolitic and skeletal sands are the typical sediments, the latter being frequently associated with barrier reefs. The sands, kept continuously in motion by waves and tidal currents, accumulate in bars and submerged dunes that form elongated shoals. These are cut perpendicularly to the bank edge by tidal channels, and the bars and dunes surface in places during low tide. The sands are very clean, with no micritic matrix; abundant current structures such as ripple marks and cross-bedding occur. The infauna—animals living in the bottom sediments—is very scarce or absent because the sand is continuously in motion and migrating.

Oolites form on the edge of the platform where the open ocean water is slightly warmed as it flows up on the shallow banks; the increase in temperature produces the supersaturation of carbonate necessary for precipitation, and the movement of the bottom sediments produces the specific oolitic form. All Recent oolitic sediments are found in shallow water, less than 15 metres (50 feet) deep, and within range of exchange with deeper water areas nearby. Elevated salinity, however, is not required for the formation of oolites.

Deposition in the lagoon environment

Toward the internal lagoon the dynamic energy decreases considerably, and the lime mud, whatever its origin, collects in the most sheltered places. Bottom sediments are quiet and are stabilized by organic films, algae, or grass; organic activity is therefore abundant, and animals burrowing and feeding on the bottom continuously rework the sediments. In these environments micritic, pellet–micritic, and micritic–skeletal deposits will form; current structures will be absent, and traces of burrowing activity are quite common. In areas that are intermediate between oolitic shoals and lagoons, physicochemical conditions (supersaturation and quietness) may favour the development of lump sediments.

Farther toward land the lagoon grades into the littoral belt. When the shore is rocky, skeletal and intraclastic sands occur on the surf zone and on the beach. These sediments are beautifully laminated and may present very low-angle cross-bedding. On the shore, cross-bedded calcareous dunes may develop. When the shore is flat, tidal flats, tidal marsh, and mangrove swamps may occur. Tidal flats reveal a great variety of textures and structures because of the peculiar situation of intermittent exposure and submergence. They receive sediments mainly during storms, and these sediments are largely pelletiferous and muddy, though coarse skeletal sediments occur in the main channels as lag deposits or bars; stromatolitic laminations, mud cracks, and intraclastic sands and breccias (rocks consisting of coarse angular fragments) are quite common.

Open-sea environments. Open-sea environments are also the sites of distinctive associations of carbonate sediments. Depth and distance from land are the main controlling factors; life is precluded for many organisms, and the physical effects of current and waves are extremely limited.

Open-sea carbonate deposits may result from the exclusive vertical sedimentation of planktonic organisms or from the combined operation of such vertical sedimentation with the lateral effects of gravity displacement.

Pelagic carbonate deposits consist of the tests (hard coverings or internal supporting structures) of planktonic invertebrates or plants and the skeletal remains of swimming organisms that have sunk to the bottom. Recent deep-sea calcareous oozes consist of tests of foraminifera, pteropods, and nannoplankton. Ancient analogues of these pelagic carbonates occur in the Alpine geosynclines (depositional basins that serve as loci for subsequent mountain building).

Material of shallow-water origin may be introduced into deep-water depositional sites by submarine slumps, slides. and turbidity currents. The deposits, composed of skeletal sands, oolites, and intraclasts, become graded and are interbedded with the pelagic sediments. They occur near submarine scarps and slopes of carbonate banks, platforms, and islands.

Solution of carbonate and the compensation depth

The mineralogical composition of deep-sea sediment is controlled partly by the composition of the planktonic skeletal remains that are the chief component and partly by dissolution of magnesium-rich calcite and aragonite. ***Globigerina*** tests and coccoliths, for example, which are among the most important planktonic sources of deep-sea sediments, consist of magnesium-poor calcite. Under deep-sea conditions, as a result of the lower temperature and higher pressure, all the carbonates are removed by dissolution. The depth beyond which no carbonates are present is called the "compensation depth," and it occurs at 4,000–5,000 metres (13,000–16,000 feet). Aragonite and magnesium-rich calcite are less stable than magnesium-poor calcite and go into solution at shallower depths; magnesium-poor sediments prevail to the "compensation depth." This phenomenon explains why many deep oceanic basins are barren of carbonate oozes, even though the calcareous plankton sinks continuously to the bottom.

Diagenetic changes. The carbonate rocks observed in the geologic record have undergone diagenetic processes over long periods of time, altering them partially or even completely. The late diagenetic changes occurring in a sediment are a response to pressure, temperature, the nature of the fluids moving through the pores, and the physical and chemical properties of the original sediment.

The first change occurs in grain mineralogy when magnesium-rich sediments pass to magnesium-poor calcareous rocks. The result is a complete change of mineralogy of the grains, but with retention of their original depositional texture. The exact mechanism by which removal of magnesium takes place is not yet fully understood.

Cement is the most important textural diagenetic element and the most important single element in the lithification (rock-forming) process. It consists of a chemical precipitate that fills voids within the sediment. The pore space may have existed originally between the grains or may have been created by desiccation or selective dissolution of the lime mud matrix. The most common mineral cement is calcite, but dolomite, anhydrite, and siliceous cements also occur. Cement appears as a clear, coarse mosaic and commonly begins as euhedral (well-formed) crystals that grow radially outward from the grains or from the cavity walls. Sometimes cementation is incomplete, and the available space may be filled with finer diagenetic sediment, called "crystal silt."

Recrystallization is another important diagenetic process that is responsible for the obliteration of the original depositional textures; a mass of large interlocking crystals, which may strongly resemble cement, is the final product. The process, which consists of grain or crystal enlargement, may operate in either fine-grained or coarse-grained sediments, and it may be selective, involving only parts of the material, or complete. Apparently, lime mud matrix is the textural element most sensitive to recrystallization.

SIGNIFICANCE OF CARBONATE ROCKS IN THE GEOLOGIC RECORD

Geochemical evolution. Carbonate rocks represent a natural response to the geochemical evolution of the Earth and to major steps in the course of biological evolution. A first milestone in the carbonate history is the appearance of the first organisms that began to use carbon dioxide and to liberate oxygen as a by-product of sugar synthesis. This removal of carbon dioxide from solution in seawater must have increased the alkalinity and led to the precipitation of calcium carbonate. The first carbonate fossils, stromatolitic algal structures, are found all over the world in Precambrian rock formations, the oldest of which are more than 2,700,000,000 years old.

Paleozoic and Mesozoic occurrences

During Paleozoic and Early Mesozoic times (from 570,000,000 to about 200,000,000 years ago), limestones and dolomites were deposited in a belt extending from North America through western Europe and the Soviet Union to Australia. These carbonates were largely the product of sedimentation in shallow seas. In the Late Mesozoic (65,000,000 to 150,000,000 years ago), pelagic carbonate organisms burst out in an immense biological explosion all over the world. This great biological event changed the Earth's carbonate economy and shifted a large part of the carbonate deposition from the continental shelves to the oceanic depths. The earlier, shallow-water limestones were likely to be recycled—destroyed by erosion and solution during regressions of the sea and then redeposited during transgressions—whereas the pelagic carbonates are oceanic and almost permanently withdrawn from surface circulation. The range of marine carbonates today reflects conditions that have been valid only since Cretaceous times, and the earlier distribution requires a quite different standard for appraisal.

Paleoclimates. Current opinion holds that carbonate rocks record warm climatic conditions and, consequently, the ancient distribution of land and sea with respect to the Equator. This paleoclimatological significance of carbonates assumes, however, that the climatic conditions of the geologic past were similar to present conditions. This is not exactly true: the "normal" geological paleoclimate was warmer than the present climate by about 3°–5° C (5°–9° F) in equatorial latitudes and more than 20" C (36" F) in polar latitudes.

Some types of carbonate rocks can be considered climate indicators, however, because they are precisely restricted to a climatic range of occurrence. These rocks include (1) marine oolites, with a bimodal (doubly peaked) distribution at latitudes 25° north and south, water temperature of 20"–30" C (68°–86° F) and normal salinity; (2) skeletal limestones whose organisms have living representatives with distinctive climatic ranges; (3) calcareous soils and crusts that indicate temperate climates with moderate winter precipitation and strong seasonality; (4) calcareous eolian dunes, whose origin requires a warm (15"–30" C, or 59°–86° F), shallow sea with vigorous productivity of carbonate-shelled organisms, suitable beach conditions, and a relatively long dry season to aid dune building; and (5) beachrock, a skeletal sand rapidly cemented on beaches in the intertidal belt by inorganic aragonite; the formation of beachrock occurs between 30" and 40" north and south, with water temperatures of 15°–30° C and normal salinity.

Paleogeography. The changing water levels that coincided with the formation and subsequent retreat of extensive shallow seas during the geologic past provided favourable marine environments for the accumulation of carbonates. High sea levels favoured the carbonates, whereas low sea levels did not. The causes of sea level change include (1) major episodes of mountain building and uplift of the Earth's surface; (2) slowly pulsating movements that produce broader, regional uplift of the Earth's surface, and (3) eustatic changes, those changes of sea level that result from glaciations and deglaciations.

Influence of tectonism

Tectonism (movements of the Earth's crust) exerts a strong effect on carbonate sedimentation. With increasing tectonic activity the relief becomes more pronounced, continental shelves tend to be narrower, and the shelves are smothered by continental sediments transported to the sea by rivers. Moreover, the turbidity of the waters may eliminate organisms capable of rapid carbonate secretion. Carbonate sedimentation is generally restricted to areas of tectonic stability, such as stable interior platforms and shelves. Carbonates occur in these environments together with some continental sediments. The most common groups include stromatolitic limestones and dolomites, biostromes and bioherms, and skeletal, pellet, and oolitic limestones; pebble conglomerates and mud-cracked micrites are also quite common. Carbonate sediments generally are restricted to the early stages of the geosynclinal (mountain-building) cycle, when the relief is not yet pronounced and the main influx of continental sediments has not yet started. Cherty micritic limestones with planktonic fossils are typical in the geosynclinal troughs, and carbonate sediments affected by slumping and turbidity currents may also occur,

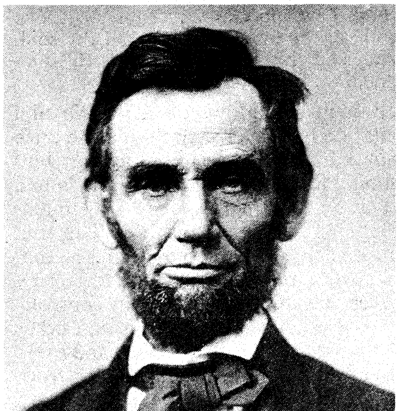
BIBLIOGRAPHY. L. CAYEUX, *Les Roches sédimentaires de France: roches carbonatées (calcaires et dolomies)* (1935), a classic work on carbonate rocks; G.V. CHILINGAR, H.J. BISSELL, and K.H. WOLF, "Diagenesis of Carbonate Rocks," in G. LARSEN and G.V. CHILINGAR (eds.), *Diagenesis in Sediments, pp. 179–322* (1967), a complete, well-illustrated discussion of many diagenetic processes, with many references; G.M. FRIEDMAN (ed.), *Depositional Environments in Carbonate Rocks* (1969), contains some very significant papers on the deep-water, shallow water, and transitional carbonate environments; W.E. HAM (ed.), *Classification of Carbonate Rocks: A Symposium* (1962), presents some of the most significant classifications of carbonate rocks; A.S. HOROWITZ and P.E. POTTER, *Introductory Petrography of Fossils (1971)*, well-illustrated introduction to the study of fossiliferous carbonate rocks; L.C. PRAY and R.C. MURRAY (eds.), *Dolomitization and Limestone Diagenesis: A Symposium* (1965), documents the most important occurrences of Recent dolomite.

(A.Bo.)

Lincoln, Abraham

The 16th president of the United States, Abraham Lincoln preserved the Union during the Civil War and brought about the emancipation of the slaves. Among American heroes, Lincoln continues to have a unique appeal for his fellow countrymen and also for people of other lands. This charm derives from his remarkable life story—the rise from humble origins, the dramatic death—and from his distinctively human and humane personality as well as from his historical role as saviour of the Union and emancipator of the slaves. His relevance endures and grows especially because of his eloquence as a spokesman for democracy. In his view, the Union was worth saving not only for its own sake but also because it embodied an ideal, the ideal of self-government, which was of interest to the people of the entire world. Hence the universality of his continuing appeal.

By courtesy of the Library of Congress, Washington, D.C.



Lincoln, 1863.

LIFE

Born on Feb. 12, 1809, in a backwoods cabin three miles south of Hodgenville, Kentucky, Lincoln was two years old when he was taken to a farm in the neighbouring valley of Knob Creek. His earliest memories were of this home and, in particular, of a flash flood that once washed away the corn and pumpkin seeds he had helped his father plant. The father, Thomas Lincoln, was descended from a weaver's apprentice who had migrated from England to Massachusetts in 1637. Though much less prosperous than some of his Lincoln forebears, Thomas was a sturdy pioneer. On June 12, 1806, he married Nancy Hanks. The Hanks genealogy is difficult to trace, but Nancy appears to have been of illegitimate birth. She has been described as "stoop-shouldered, thin-breasted, sad," and fervently religious. Thomas and Nancy Lincoln had three children: Sarah, Abraham, and Thomas (died in infancy).

Childhood and youth. In December 1816, faced with a lawsuit challenging the title to his Kentucky farm, Thomas Lincoln moved with his family to southwestern Indiana. There, as a "squatter" on public land, he hastily put up a "half-faced camp"—a crude structure of logs and boughs with one side open to the weather—in which the family took shelter behind a blazing fire. Soon he built a permanent cabin, and later he bought the land on which it stood. Abraham helped to clear the fields and to take care of the crops but early acquired a dislike for hunting and fishing. In afteryears he recalled the "panther's scream," the bears that "preyed on the swine," and the poverty of Indiana frontier life, which was "pretty pinching at times." The unhappiest period of his boyhood followed the death of his mother in the autumn of 1818. As a ragged nine year old, he saw her buried in the forest, then faced a winter without the warmth of a mother's love. Fortunately, before the onset of a second winter, Thomas Lincoln brought home from Kentucky a new wife for himself, a new mother for the children. Sarah Bush Johnston Lincoln, a widow with two girls and a boy of her own, had energy and affection to spare. She ran the household with an even hand, treating both sets of children as if she had borne them all; but she became especially fond of Abraham, and he of her. He afterward referred to her as his "angel mother."

Mother's death

This stepmother doubtless encouraged Lincoln's taste for reading, yet the original source of his desire to learn remains something of a mystery. Both of his parents were almost completely illiterate, and he himself received little formal education. He once said that, as a boy, he had gone to school "by littles"—a little now and a little then—and his entire schooling amounted to no more than one year's attendance. His neighbours later recalled how he used to trudge for miles to borrow a book. According to his own statement, however, his early surroundings provided "absolutely nothing to excite ambition for education. Of course, when I came of age I did not know much. Still, somehow, I could read, write, and cipher to the rule of three; but that was all." Apparently the young Lincoln did not read a large number of books but thoroughly absorbed the few that he did read. These included Parson Weems's *Life and Memorable Actions of George Washington* (with its story of the little hatchet and the cherry tree), *Robinson Crusoe*, *Pilgrim's Progress*, and Aesop's *Fables*. From his earliest days he must have had some familiarity with the Bible, for it doubtless was the only book his family owned.

Migration to Illinois

In March 1830 the Lincoln family undertook a second migration, this one to Illinois, with Lincoln himself driving the team of oxen. Having just reached the age of 21, he was about to begin life on his own. Six feet four inches tall, he was rawboned and lanky but muscular and physically powerful. He was especially noted for the skill and strength with which he could wield an ax. He spoke with a backwoods twang and walked in the long-striding, flat-footed, cautious manner of a plowman. Good-natured though somewhat moody, talented as a mimic and storyteller, he readily attracted friends. He was yet to demonstrate whatever other abilities he possessed.

After his arrival in Illinois, having no desire to be a farmer, Lincoln tried his hand at a variety of occupations. As a "rail splitter" he helped to clear and fence his father's new farm. As a flatboatman, he made a voyage down the Mississippi River to New Orleans. (This was his second visit to that city, his first having been made in 1828, while he still lived in Indiana.) On his return he settled in New Salem, a village of about 25 families on the Sangamon River. There he worked from time to time as storekeeper, postmaster, and surveyor. With the coming of the Black Hawk War (1832), he enlisted as a volunteer and was elected captain of his company. Afterward he joked that he had seen no "live, fighting Indians" during the war but had had "a good many bloody struggles with the mosquitoes." Meanwhile, aspiring to be a legislator, he was defeated in his first try and then repeatedly re-elected to the state assembly. He considered blacksmithing as a trade but finally decided in favour of the law. Already he had taught himself grammar and mathematics, and now he began to study lawbooks. In 1836, having passed the bar examination, he began to practice law.

Prairie lawyer. The next year he moved to Springfield, Illinois, the new state capital, which offered many more opportunities for a lawyer than New Salem did. At first he was a partner of John T. Stuart; then of Stephen T. Logan; and finally, from 1844 on, of William H. Herndon. Nearly ten years younger than Lincoln, Herndon was more widely read, more emotional at the bar, and generally more extreme in his views. Yet this partnership seems to have been as nearly perfect as such human arrangements ever are. Lincoln and Herndon kept few records of their law business, and they split the cash between them whenever either of them was paid. It seems they had no money quarrels.

Early legal career

Within a few years after his removal to Springfield, Lincoln was earning from $1,200 to $1,500 annually, at a time when the governor of the state received a salary of $1,200 and circuit judges only $750. He had to work hard. To keep himself busy he found it necessary not only to practice in the capital but also to follow the court as it made the rounds of its circuit. Each spring and fall he would set out by horseback or buggy to travel hundreds of miles over the thinly settled prairie, from one little county seat to another. Most of the cases were petty and the fees small.

The coming of the railroads, especially after 1850, made travel easier and practice more remunerative. Lincoln served as a lobbyist for the Illinois Central Railroad to assist it in getting a charter from the state, and thereafter he was retained as a regular attorney for that railroad. After successfully defending the company against the efforts of McLean County to tax its property, he received the largest single fee of his legal career—$5,000. (He had to sue the Illinois Central in order to collect the fee.) He also handled cases for other railroads and for banks, insurance companies, mercantile and manufacturing firms. In one of his finest performances before the bar, he saved the Rock Island Bridge, the first to span the Mississippi River, from the threat of the river transportation interests that demanded the bridge's removal. His business included a number of patent suits and criminal trials. One of his most effective and famous pleas had to do with a murder case. A witness claimed that, by the light of the moon, he had seen Duff Armstrong, an acquaintance of Lincoln's, take part in a killing. Referring to an almanac for proof, Lincoln argued that the night had been too dark for the witness to have seen anything clearly, and with a sincere and moving appeal he won an acquittal.

By the time he began to be prominent in national politics, about 20 years after launching upon his legal career, Lincoln had made himself one of the most distinguished and successful lawyers in Illinois. He was noted not only for his shrewdness and practical common sense, which enabled him always to see to the heart of any legal case, but also for his invariable fairness and utter honesty.

Private life. While residing in New Salem, Lincoln was acquainted with Ann Rutledge. Apparently he was fond of her, and certainly he grieved with the entire community at her untimely death, in 1835, at the age of 19. Afterward stories were told of a grand romance between Abraham and Ann, but these stories lack the support of sound historical evidence. A year after the death of Miss Rutledge, Lincoln was carrying on a halfhearted courtship with Mary Owens. Miss Owens concluded that Lincoln was "deficient in those little links which make up the chain of woman's happiness." She turned down his proposal.

Marriage to Mary Todd

So far as can be known, the first and only real love of Lincoln's life was Mary Todd. High-spirited, quick-witted and well-educated, Miss Todd came from a rather distinguished Kentucky family, and her Springfield relatives belonged to the social aristocracy of the town. Some of them frowned upon her association with Lincoln, and from time to time he too had doubts whether he ever could make her happy. Nevertheless, they became engaged. Then, on a day in 1841 that Lincoln recalled as the "fatal first of January," they broke the engagement, apparently on his initiative. For some time after that, he was overwhelmed by a mood of terrible depression and despondency. Finally the two were reconciled and on Nov. 4, 1842, were married.

Four children, all boys, were born to the Lincolns. Robert Todd, the eldest and only one to survive to adulthood, was never very close to his father. Edward Baker was nearly 4 when he died, and William Wallace was 11. Thomas, affectionately known as "Tad," outlived his father; Tad, who had a cleft palate and a lisp, was Lincoln's favourite. Lincoln left the upbringing of his sons largely to their mother, who was alternately strict and lenient in her treatment of them.

The Lincolns had a mutual affectionate interest in the doings and welfare of their boys, were fond of one another's company, and missed each other when apart, as existing letters show. Like most married couples, the Lincolns also had their domestic quarrels, which sometimes were hectic but which undoubtedly were exaggerated by contemporary gossips. Mrs. Lincoln suffered from recurring headaches, fits of temper, and a sense of insecurity and loneliness that was intensified by her husband's long absences on the lawyer's circuit. After his election to the presidency, she was afflicted in spirit by the death of her son Willie, by the ironies of war that made enemies of Kentucky relatives and friends, and by the unfair public criticisms of her as mistress of the White House. She lost all money sense and ran up embarrassing bills. She also put on some painful scenes of wifely jealousy. At last, in 1875, she was officially declared insane, but that was after she had undergone the further shock of seeing her husband murdered at her side. During their earlier married life, Mrs. Lincoln unquestionably encouraged her husband and served as a prod to his own ambition. During their later years together she probably strengthened and tested his innate qualities of tolerance and patience.

Religious views

With his wife, Lincoln attended Presbyterian services in Springfield and in Washington but never joined any church. He once explained:

> When any church will inscribe over its altar, as its sole qualification for membership, the Saviour's condensed statement of the substance of both Law and Gospel, "Thou shalt love the Lord thy God with all thy heart, and with all thy soul, and with all thy mind, and thy neighbor, as thyself," that church will I join with all my heart and all my soul.

Early in life he had been something of a skeptic and freethinker. His reputation had been such that, as he once complained, the "church influence" was used against him in politics. When running for Congress in 1846, he issued a handbill to deny that he ever had "spoken with intentional disrespect of religion." He went on to explain that he had believed in the doctrine of necessity—"that is, that the human mind is impelled to action, or held in rest by some power over which the mind itself has no control." Throughout his life he also believed in dreams and other enigmatic signs and portents. As he grew older, and especially after he became president and faced the soul-troubling responsibilities of the Civil War, he developed a profound religious sense, and he increasingly personified necessity as God. He came to look upon himself quite humbly as an "instrument of Providence" and to view all history as God's enterprise. "In the present civil war," he wrote in 1862, "it is quite possible that God's purpose is something different from the purpose of either party—and yet the human instrumentalities, working just as they do, are of the best adaptation to effect His purpose."

Lincoln was fond of the Bible and knew it well. He also was fond of Shakespeare. In private conversation he used many Shakespearean allusions, discussed problems of dramatic interpretation with considerable insight, and from memory recited long passages with rare feeling and understanding. He liked the essays of John Stuart Mill, particularly the famous one on liberty, but disliked heavy or metaphysical works.

Though he enjoyed the poems of Lord Byron and Robert Burns, his favourite piece of verse was the work of an obscure Scottish poet, William Knox. Lincoln often quoted Knox's lines beginning: "Oh! why should the spirit of mortal be proud?" He liked to relax with the comic writings of Petroleum V. Nasby, Orpheus C. Kerr, and Artemus Ward, or with a visit to the popular theatre.

Early politics. When Lincoln first entered politics, Andrew Jackson was president. Lincoln shared the sympathies that the Jacksonians professed for the common man, but he disagreed with the Jacksonian view that the government should be divorced from economic enterprise. "The legitimate object of government," he was later to say, "is to do for a community of people whatever they need to have done, but cannot do at all, or cannot do so well, for themselves, in their separate and individual capacities." He most admired Henry Clay and Daniel Webster among the prominent politicians of the time. Clay and Webster advocated using the powers of the federal government to encourage business and develop the country's resources by means of a national bank, a protective tariff, and a program of internal improvements for facilitating transportation. In Lincoln's view, Illinois and the West as a whole desperately needed such aid to economic development. From the outset, he associated himself with the Clay and Webster party, the Whigs.

Member of Illinois State Legislature

As a Whig member of the Illinois State Legislature, to which he was elected four times from 1834 to 1840, he devoted himself to a grandiose project for constructing with state funds a network of railroads, highways, and canals. Whigs and Democrats joined in passing an omnibus bill for these undertakings, but the Panic of 1837 and the ensuing business depression brought about the abandonment of most of them. While in the legislature he demonstrated that, though opposed to slavery, he was no abolitionist. Resolutions were introduced, in 1837, in response to the mob murder of Elijah Lovejoy, an antislavery newspaperman of Alton. Instead of denouncing lynch law, these resolutions condemned abolitionist societies and upheld slavery within the Southern states as "sacred" by virtue of the federal Constitution. Lincoln refused to vote for the resolutions. Together with a fellow member he drew up a protest against them. This maintained, on the one hand, that slavery was "founded on both injustice and bad policy" and, on the other, that "the promulgation of abolition doctrines tends rather to increase than to abate its evils."

During his single term in Congress (1847–49), Lincoln, as the lone Whig from Illinois, gave little attention to legislative matters as such. He proposed a bill for the gradual and compensated emancipation of slaves in the District of Columbia, but the bill was to take effect only with the approval of the "free white citizens" of the district. It displeased abolitionists as well as slaveholders and never was seriously considered.

Involvement in presidential politics

Much of his time Lincoln devoted to presidential politics, to unmaking one president, a Democrat, and making another, a Whig. He found an issue and a candidate in the Mexican War. With his "spot resolutions" he challenged the statement of Pres. James K. Polk that Mexico had started the war by shedding American blood upon American soil. Along with other members of his party, Lincoln voted to condemn Polk and the war while voting supplies for carrying it on. At the same time he laboured for the nomination and election of the war hero Zachary Taylor. After Taylor's success at the polls, Lincoln expected to be named commissioner of the general land office as a reward for his campaign services, and he was bitterly disappointed when he failed to get the job. His criticisms of the war, meanwhile, had not been popular among the voters in his own congressional district. At the age of 40, frustrated in politics, he seemed to be at the end of his public career.

For about five years he took little part in politics, and then a new sectional crisis gave him a chance to re-emerge and rise to statesmanship. In 1854 his political rival Stephen A. Douglas manoeuvred through Congress a bill for reopening the entire Louisiana Purchase to slavery and allowing the settlers of Kansas and Nebraska (with "popular sovereignty") to decide for themselves whether to permit slaveholding in those territories. The Kansas–Nebraska Act provoked violent opposition in Illinois and the other states of the old Northwest. It gave rise to the Republican Party while speeding the Whig Party on the way to disintegration. Along with many thousands of other homeless Whigs, Lincoln soon became a Republican (1856). Before long, some prominent Republicans in the East talked of attracting Douglas to the Republican fold, and with him his Democratic following in the West. Lincoln would have none of it. He was determined that he, not Douglas, should be the Republican leader of his state and section. In their basic views, he and Douglas were not quite so far apart as they seemed in the heat of political argument. Neither was an abolitionist, neither a proslavery man. But Lincoln, unlike Douglas, insisted that Congress must exclude slavery from the territories. He disagreed with Douglas' belief that the territories were by nature unsuited to the slave economy and that no congressional legislation was needed to prevent the spread of slavery into them. He declared (1858): *"A house divided against itself cannot stand. I* believe the government cannot endure permanently half slave and half free." He predicted that the country eventually would become "all one thing, or all the other." Again and again he insisted that the civil liberties of every U.S. citizen, white as well as black, were at stake. The territories must be kept free, he further said, because "new free states" were "places for poor people to go and better their condition." He agreed with Thomas Jefferson and other founding fathers, however, that slavery should be merely contained, not directly attacked. In the Lincoln–Douglas debates of 1858, while contesting for Douglas' seat in the United States Senate, he drove home the inconsistency between Douglas' "popular sovereignty" principle and the Dred Scott decision (1857), in which the U.S. Supreme Court held that Congress could not constitutionally exclude slavery from the territories. Though he failed to obtain the Senate seat, Lincoln gained national recognition and soon began to be mentioned as a presidential prospect for 1860.

Lincoln–Douglas debates

On May 18, 1860, after Lincoln and his friends had made skillful preparations, he was nominated on the third ballot at the Republican Convention in Chicago. He then put aside his law practice and, though making no stump speeches, gave his full time to the direction of his campaign. His "main object," he had written, was to "hedge against divisions in the Republican ranks," and he counselled party workers to "say nothing on points where it is probable we shall disagree." With the Republicans united, the Democrats divided, and a total of four candidates in the field, he carried the election on November 6. No one in the Deep South voted for him and no more than 40 out of 100 in the country as a whole. Still, the popular votes were so distributed that he won a clear and decisive majority in the electoral college.

President Lincoln. After Lincoln's election and before his inauguration, the state of South Carolina proclaimed its withdrawal from the Union. To forestall similar action by other Southern states, various compromises were proposed in Congress. The most important, the Crittenden Compromise, included constitutional amendments (1) guaranteeing slavery forever in the states where it already existed and (2) dividing the territories between slavery and freedom. Though Lincoln had no objection to the first of these amendments, he was unalterably opposed to the second and indeed to any scheme infringing in the slightest upon the free-soil plank of his party's platform. "I am inflexible," he privately wrote. He feared that a territorial division, by sanctioning the principle of slavery extension, would only encourage planter imperialists to seek new slave territory south of the American border and thus would "put us again on the highroad to a slave empire." From his home in Springfield he advised Republicans in Congress to vote against a division of the territories. The proposal was killed in committee. Six additional states then seceded and, with South Carolina, combined to form the Confederate States of America.

Secession of the Confederate states

So, before Lincoln took office, a disunion crisis was upon the country. Attention, North and South, focussed in particular upon Ft. Sumter, in Charleston Harbor. This fort, still under construction, was garrisoned by U.S. troops under Maj. Robert Anderson. The Confederacy claimed it and, from other harbour fortifications, threatened it. Foreseeing trouble, Lincoln, while still in Springfield, confidentially requested Winfield Scott, general in chief of the U.S. Army, to be prepared "to either *hold*, or ***retake***, the forts, as the case may require, at, and after the inauguration." In his inaugural address (March 4, 1861), besides upholding the Union's indestructibility and appealing for sectional harmony, Lincoln restated his Sumter policy as follows:

> The power confided to me, will be used to hold, occupy, and possess the property, and places belonging to the government, and to collect the duties and imposts; but beyond what may be necessary for these objects, there will be no invasion—no using of force against, or among the people anywhere.

Then, near the end, addressing the absent Southerners: "You can have no conflict, without being yourselves the aggressors."

No sooner was he in office than Lincoln received word that the Sumter garrison, unless supplied or withdrawn, would shortly be starved out. Still, for about a month, Lincoln delayed to act. He was beset by contradictory advice. On the one hand, General Scott, Secretary of State William H. Seward, and others urged him to abandon the fort; and Seward, through a go-between, gave a group of Confederate commissioners to understand that the fort would in fact be abandoned. On the other hand, many Republicans insisted that any show of weakness would bring disaster to their party and to the Union. Finally Lincoln ordered the preparation of two relief expeditions, one for Ft. Sumter and the other for Ft. Pickens, in Florida. (He afterward said he would have been willing to withdraw from Sumter if he could have been sure of holding Pickens.) Before the Sumter expedition, he sent a messenger to tell the South Carolina governor:

> I am directed by the President of the United States to notify you to expect an attempt will be made to supply Fort-Sumpter with provisions only; and that, if such attempt be not resisted, no effort to throw in men, arms, or ammunition, will be made, without further notice, or in case of an attack upon the Fort.

Without waiting for the arrival of Lincoln's expedition, the Confederate authorities presented to Major Anderson a demand for Sumter's prompt evacuation, which he refused. On April 12, 1861, at dawn, the Confederate batteries in the harbour opened fire.

Attack on
Ft. Sumter:
Civil War

"Then, and thereby," Lincoln informed Congress when it met on July 4, "the assailants of the Government, began the conflict of arms." The Confederates, however, accused him of being the real aggressor. They said he had cleverly manoeuvred them into firing the first shot so as to put upon them the onus of war guilt. Though some historians have repeated this charge, it appears to be a gross distortion of the facts. Lincoln was determined to preserve the Union; to do so he thought he must take a stand against the Confederacy, and he concluded he might as well take this stand at Sumter.

Lincoln's primary aim was neither to provoke war nor to maintain peace. In preserving the Union, he would have been glad to preserve the peace also, but he was ready to risk a war that he thought would be short.

After the firing on Ft. Sumter, Lincoln called upon the state governors for troops (Virginia and three other states of the upper South responded by joining the Confederacy). He then proclaimed a blockade of the Southern ports. These steps—the Sumter expedition, the call for volunteers, and the blockade—were the first important decisions of Lincoln as commander in chief of the army and navy. He still needed a strategic plan and a command system for carrying it out.

General Scott advised him to avoid battle with the Confederate forces in Virginia, to get control of the Mississippi River, and by tightening the blockade to hold the South in a gigantic squeeze. Lincoln had little confidence in Scott's comparatively passive and bloodless "Anaconda" plan. He believed the war must be actively fought if it ever was to be won. Overruling Scott, he ordered a direct advance on the Virginia front, which resulted in defeat and rout for the Federal forces at Bull Run (July 21, 1861). After a succession of more or less sleepless nights, Lincoln produced a set of memorandums on military policy. His basic thought was this: the armies should advance concurrently on several fronts and should move in such directions as to hold and use the support of Unionists in Missouri, Kentucky, western Virginia, and eastern Tennessee. He later explained:

Lincoln's
strategy
as
commander
in chief

> I state my general idea of this war to be that we have the *greater* numbers, and the enemy has the *greater* facility of concentrating forces upon points of collision; that we must fail, unless we can find some way of making *our* advantage an over-match for *his;* and that this can only be done by menacing him with superior forces at *different* points, at the *same* time.

This, with the naval blockade, comprised the essence of Lincoln's strategy.

From 1861 to 1864, while hesitating to impose his ideas upon his generals, Lincoln experimented with command personnel and organization. Accepting the resignation of Scott (November 1861), he put George B. McClellan in charge of the armies as a whole. After a few months, disgusted by the slowness of McClellan, he demoted him to the command of the Army of the Potomac alone. He questioned the soundness of McClellan's plans for the peninsular campaign, repeatedly compelled McClellan to alter them, and, after the Seven Days' Battles before Richmond (June–July 1862), ordered him to give them up. Then he tried a succession of commanders for the army in Virginia—John Pope, McClellan again, Ambrose E. Burnside, Joseph Hooker, and George Gordon Meade—but was disappointed with each of them in turn. Meanwhile, he had in Henry W. Halleck a general in chief who gave advice and served as a liaison with field officers but who shrank from making important decisions. For nearly two years the Federal armies had no very effective unity of command. President Lincoln, General Halleck, and War Secretary Edwin M. Stanton acted as an informal council of war. Lincoln, besides transmitting official orders through Halleck, also communicated directly with the generals, sending personal suggestions in his own name. To generals opposing Robert E. Lee, he suggested that the object was to destroy Lee's army, not to capture Richmond or to drive the invader from Northern soil.

Finally Lincoln looked to the West for a top general. He admired the Vicksburg Campaign of Ulysses S. Grant. Nine days after the Vicksburg surrender (which occurred on July 4, 1863), he sent Grant a "grateful acknowledgment for the almost inestimable service" he had done the country. Lincoln sent also an admission of his own error. He said he had expected Grant to bypass Vicksburg and go on down the Mississippi, instead of crossing the river and turning back to approach Vicksburg from the rear. "I feared it was a mistake," he wrote in his letter of congratulations. "I now wish to make the personal acknowledgment that you were right, and I was wrong."

Choice of
Grant as
commander
of Federal
armies

In March 1864 Lincoln promoted Grant to lieutenant general and gave him command of all the Federal armies. At last Lincoln had found a man who, with such able subordinates as William T. Sherman, Philip Sheridan, and George H. Thomas, could put into effect those parts of Lincoln's concept of a large-scale, coordinated offensive that still remained to be carried out. Grant was only a member, though an important one, of a top-command arrangement that Lincoln eventually had devised. Overseeing everything was Lincoln himself, the commander in chief. Taking the responsibility for men and supplies was Stanton, the secretary of war. Serving as a presidential adviser and as a liaison with military men was Halleck, the chief of staff. And directing all the armies, while accompanying Meade's Army of the Potomac, was Grant, the general in chief. Thus Lincoln pioneered in the creation of a high command, an organization for amassing all the energies and resources of a people in the grand strategy of total war. He combined statecraft and the overall direction of armies with an effectiveness that year by year increased. His achievement is all the more remarkable in view of his lack of training and experience in the art of warfare. This lack may have been an advantage as well as a handicap. Unhampered by outworn military dogma, Lincoln could all the better apply his practical insight and common sense—some would say his military genius—to the winning of the Civil War.

Lincoln and the slavery question

There can be no doubt of Lincoln's deep and sincere devotion to the cause of personal freedom. Before his election to the presidency he had spoken often and eloquently on the subject. In 1854, for example, he said he hated the Douglas attitude of indifference toward the possible spread of slavery to new areas. "I hate it because of the monstrous injustice of slavery itself," he declared. "I hate it because it deprives our republican example of its just influence in the world: enables the enemies of free institutions with plausibility to taunt us as hypocrites. . . ." In 1855, writing to his friend Joshua Speed, he recalled a steamboat trip the two had taken on the Ohio River 14 years earlier. "You may remember, as I well do," he said, "that from Louisville to the mouth of the Ohio there were, on board, ten or a dozen slaves, shackled together with irons. That sight was a continual torment to me; and I see something like it every time I touch the Ohio, or any other slave-border."

Yet, as president, Lincoln was at first reluctant to adopt an abolitionist policy. There were several reasons for his hesitancy, He had been elected on a platform pledging no interference with slavery within the states, and in any case he doubted the constitutionality of federal action under the circumstances. He was concerned about the possible difficulties of incorporating nearly 4,000,000 Negroes, once they had been freed, into the nation's social and political life. Above all, he felt that he must hold the border slave states in the Union, and he feared that an abolitionist program might impel them, in particular his native Kentucky, toward the Confederacy. So he held back while others went ahead. When Gen. John C. Frémont and Gen. David Hunter, within their respective military departments, proclaimed freedom for the slaves of disloyal masters, Lincoln revoked the proclamations. When Congress passed confiscation acts (in 1861 and 1862), he refrained from a full enforcement of the provisions authorizing him to seize slave property. And when Horace Greeley in the *New York Tribune* appealed to him to enforce these laws, Lincoln patiently replied (Aug. 22, 1862):

> My paramount object in this struggle is to save the Union, and is not either to save or to destroy slavery. If I could save the Union without freeing any slave I would do it; and if I could save it by freeing all the slaves I would do it; and if I could save it by freeing some and leaving others alone, I would also do that.

Meanwhile, in response to the rising antislavery sentiment, Lincoln came forth with an emancipation plan of his own. According to his proposal, the slaves were to be freed by state action, the slaveowners were to be compensated, the federal government was to share the financial burden, the emancipation process was to be gradual, and the freedmen were to be colonized abroad. Congress indicated its willingness to vote the necessary funds for the Lincoln plan, but none of the border slave states were willing to launch it, and in any case few Negro leaders desired to see their people sent abroad.

Emancipation Proclamation and the antislavery amendment

While still hoping for the eventual success of his gradual plan, Lincoln took a quite different step by issuing his preliminary (Sept. 22, 1862) and his final (Jan. 1, 1863) Emancipation Proclamation. This famous decree, which he justified as an exercise of the president's war powers, applied only to those parts of the country actually under Confederate control, not to the loyal slave states nor to the Federally occupied areas of the Confederacy. Directly or indirectly the proclamation brought freedom during the war to fewer than 200,000 slaves. Yet it had great significance as a symbol. It indicated that the Lincoln government had added freedom to reunion as a war aim, and it attracted liberal opinion in England and Europe to increased support of the Union cause.

Lincoln himself doubted the constitutionality of his step, except as a temporary war measure. After the war the slaves freed by the proclamation would have risked re-enslavement, had nothing else been done to confirm their liberty. Something else was done: the Thirteenth Amendment was added to the Constitution, and Lincoln played a large part in bringing about this change in the fundamental law. Through the chairman of the Republican National Committee he urged the party to include a plank for such an amendment in its platform of 1864. The plank, as adopted, stated that slavery was the cause of the rebellion, that the President's proclamation had aimed "a death blow at this gigantic evil," and that a constitutional amendment was necessary to "terminate and forever prohibit" it When Lincoln was re-elected on this platform and the Republican majority in Congress was increased, he was justified in feeling, as he apparently did, that he had a mandate from the people for the Thirteenth Amendment. The newly chosen Congress, with its overwhelming Republican majority, was not to meet until after the lame duck session of the old Congress during the winter of 1864–65. Lincoln did not wait. Using his resources of patronage and persuasion upon certain of the Democrats, he managed to get the necessary two-thirds vote before the session's end. He rejoiced as the amendment went out to the states for ratification, and he rejoiced again and again as his own Illinois led off and other states followed one by one in acting favourably upon it. (He did not live to rejoice in its ultimate adoption.)

Lincoln deserves his reputation as the Great Emancipator. His claim to that honour, if it rests uncertainly upon his famous proclamation, has a sound basis in the support he gave to the antislavery amendment. It is well founded also in his greatness as the war leader who carried the nation safely through the four-year struggle that brought freedom in its train. And, finally, it is strengthened by the practical demonstrations he gave of respect for human worth and dignity, regardless of colour. During the last two years of his life he welcomed Negroes as visitors and friends in a way no president had done before. One of his friends was the distinguished former slave Frederick Douglass. Afterward Douglass wrote: "In all my interviews with Mr. Lincoln I was impressed with his entire freedom from prejudice against the colored race."

Wartime politics

To win the war, President Lincoln had to have popular support. The reunion of North and South required, first of all, a certain degree of unity in the North. But the North contained various groups with special interests of their own. Lincoln faced the task of attracting to his administration the support of as many divergent groups and individuals as possible. So he gave much of his time and attention to politics, which in one of its aspects is the art of attracting such support. Fortunately for the Union cause, he was a president with rare political skill. He had the knack of appealing to fellow politicians and talking to them in their own language. He had a talent for smoothing over personal differences and holding the loyalty of men antagonistic to one another. Inheriting the spoils system, he made good use of it, disposing of government jobs in such a way as to strengthen his administration and further its official aims.

The opposition party remained alive and strong. Its membership included war Democrats and peace Democrats, often called "Copperheads," a few of whom collaborated with the enemy. Lincoln did what he could to cultivate the assistance of the war Democrats, as in securing from Congress the timely approval of the Thirteenth Amendment. So far as feasible, he conciliated the peace Democrats. He gave heed to the complaints of one of them, Gov. Horatio Seymour of New York, in regard to the draft quota for that state. He commuted the prison sentence of another, Congressman Clement L. Vallandigham of Ohio, to banishment within the Confederate lines. In dealing with persons suspected of treasonable intent, Lincoln at times authorized his generals to make arbitrary arrests. He justified this action on the ground that he had to allow some temporary sacrifice of parts of the Constitution in order to maintain the Union and thus preserve the Constitution as a whole. He let his generals suspend several newspapers, but only for short periods, and he promptly revoked a military order suppressing the hostile *Chicago Times*. In a letter to one of his generals he expressed his policy thus:

> You will only arrest individuals and suppress assemblies or newspapers when they may be working palpable injury to the

> military in your charge, and in no other case will you interfere with the expression of opinion in any form or allow it to be interfered with violently by others. In this you have a discretion to exercise with great caution, calmness, and forbearance.

Considering the dangers and provocations of the time, Lincoln was quite liberal in his treatment of political opponents and the opposition press. He was by no means the dictator critics often accused him of being.

Within his own party he confronted factional divisions and personal rivalries that caused him as much trouble as did the activities of the Democrats. True, he and most of his fellow partisans agreed fairly well upon their principal economic aims. With his approval, the Republicans enacted into law the essentials of the program he had advocated from his early Whig days—a protective tariff; a national banking system; and federal aid for internal improvements, in particular for the construction of a railroad to the Pacific Coast. The Republicans disagreed among themselves, however, on many matters regarding the conduct and purposes of the war. Two main factions arose: the "radicals" and the "conservatives." Lincoln himself inclined in spirit toward the conservatives, but he had friends among the radicals as well, and he strove to maintain his leadership over both. In appointing his cabinet, he chose his several rivals for the 1860 nomination and, all together, gave representation to every important party group. Wisely he included the outstanding conservative, Seward, and the outstanding radical, Salmon P. Chase. Cleverly he overcame cabinet crises and kept these two opposites among his oflicial advisers until Chase's resignation in 1864.

Reconstruction of the South

He had to deal with even more serious factional uprisings in Congress. The big issue was the "reconstruction" of the South. The seceded states of Louisiana, Arkansas, and Tennessee having been largely recovered by the Federal armies, Lincoln late in 1863 proposed his "ten percent plan," according to which new state governments might be formed when 10 percent of the qualified voters had taken an oath of future loyalty to the United States. The radicals rejected Lincoln's proposal as too lenient, and they carried through Congress the Wade–Davis Bill, which would have permitted the remaking and re-admission of states only after a majority had taken the loyalty oath. When Lincoln pocket vetoed that bill, its authors published a "manifesto" denouncing him.

Already he was the candidate of the "Union" (that is, the Republican) party for re-election to the presidency, and the Wade–Davis manifesto signalized a movement within the party to displace him as the party's nominee. He waited quietly and patiently for the movement to collapse, but even after it had done so, the party remained badly divided. A rival Republican candidate, John C. Frémont, nominated much earlier by a splinter group, was still in the field. Leading radicals promised to procure Frémont's withdrawal if Lincoln would obtain the resignation of his conservative postmaster general, Montgomery Blair. Eventually Frémont withdrew and Blair resigned. The party was reunited in time for the election of 1864.

In 1864, as in 1860, Lincoln was the chief strategist of his own electoral campaign. He took a hand in the management of the Republican Speakers' Bureau, advised state committees on campaign tactics, hired and fired government employees to strengthen party support, and did his best to enable as many soldiers and sailors as possible to vote. Most of the citizens in uniform voted Republican. He was re-elected with a large popular majority (55 percent) over his Democratic opponent, General McClellan.

Lincoln's second term

In 1864 the Democratic platform called for an armistice and a peace conference, and prominent Republicans as well as Democrats demanded that Lincoln give heed to Confederate peace offers, irregular and illusory though they were. In a public letter, he stated his own conditions:

> Any proposition which embraces the restoration of peace, the integrity of the whole Union, and the abandonment of slavery, and which comes by and with an authority that can control the armies now at war against the United States will he received and considered by the Executive government of the United States, and will be met by liberal terms on other substantial and collateral points.

When conservatives protested to him against the implication that the war must go on to free the slaves, even after reunion had been won, he explained: "To me it seems plain that saying reunion and abandonment of slavery would be considered, if offered,is not saying that nothing *else* or *less* would be considered, if offered." After his re-election, in his annual message to Congress, he said: "In stating a single condition of peace, I mean simply to say that the war will cease on the part of the government, whenever it shall have ceased on the part of those who began it." On Feb. 3, 1865, he met personally with Confederate commissioners on a steamship in Hampton Roads. He promised to be liberal with pardons if the South would quit the war, but he insisted on reunion as a precondition for any peace arrangement. In his second inaugural address he embodied the spirit of his policy in the famous words "with malice toward none; with charity for all." His terms satisfied neither the Confederate leaders nor the radical Republicans, and so no peace was possible until the final defeat of the Confederacy.

Policy toward the South

At the end of the war, Lincoln's policy for the defeated South was not clear in all its details, though he continued to believe that the main object should be to restore the "seceded States, so-called," to their "proper practical relation" with the Union as soon as possible. He possessed no fixed and uniform program for the region as a whole. As he said in the last public speech of his life (April 11, 1865), "so great peculiarities" pertained to each of the states, and "such important and sudden changes" occurred from time to time, and "so new and unprecedented" was the whole problem that "no exclusive and inflexible plan" could "safely be prescribed." With respect to states like Louisiana and Tennessee, he continued to urge acceptance of new governments set up under his 10 percent plan during the war. With respect to states like Virginia and North Carolina, he seemed willing to use the old rebel governments temporarily as a means of transition from war to peace. He was on record as opposing the appointment of "strangers" (carpetbaggers) to govern the South. He hoped that the Southerners themselves, in forming new state governments, would find some way by which whites and Negroes "could gradually live themselves out of their old relation to each other, and both come out better prepared for the new." A program of education for the freedmen, he thought, was essential for preparing them for their new status. He also suggested that the vote be given immediately to some Negroes—"as, for instance, the very intelligent, and especially those who have fought gallantly in our ranks."

On the question of reconstruction, however, Lincoln and the extremists of his own party stood even farther apart in early 1865 than a year before. Some of the radicals were beginning to demand a period of military occupation for the South, the confiscation of planter estates and their division among the freedmen, and the transfer of political power from the planters to their former slaves. In April 1865 Lincoln began to modify his own stand in some respects and thus to narrow the gap between himself and the radicals. He recalled the permission he had given for the assembling of the rebel legislature of Virginia, and he approved in principle—or at least did not disapprove—Stanton's scheme for the military occupation of southern states. After the cabinet meeting of April 14, Attorney General James Speed inferred that Lincoln was moving toward the radical position. "He never seemed so near our views," Speed believed. What Lincoln's reconstruction policy would have been, if he had lived to complete his second term, can only be guessed at.

Assassination

On the evening of April 14, 1865, John Wilkes Booth shot Lincoln as he sat in Ford's Theatre in Washington, and early the next morning Lincoln died.

REPUTATION AND CHARACTER

"Now he belongs to the ages," Stanton is supposed to have said as Lincoln breathed his last. Many thought of him as a martyr. The assassination had occurred on Good Friday, and on the following Sunday, memorable as "Black Easter," hundreds of speakers found a sermon in the event. Some of them saw more than mere chance in the fact that assassination day was also crucifixion day. One declared: "Jesus Christ died for the world; Abraham Lincoln died for his country." Thus the posthumous growth of his reputation was influenced by the timing and circumstances of his death, which won for him a kind of sainthood.

Among the many who remembered Lincoln from personal acquaintance, one was sure he had known him more intimately than any of the rest and influenced the world's conception of him more than all the others put together. That one was his former law partner Herndon. When Lincoln died, Herndon began a new career as Lincoln authority, collecting reminiscences wherever he could find them and adding his own store of memories. Though admiring Lincoln, he objected to the trend toward sanctifying the man. He saw, as the main feature of Lincoln's life, the far more than ordinary rise of a self-made man, a rise from the lowest depths to the greatest heights—"from a stagnant, putrid pool, like the gas which, set on fire by its own energy and self-combustible nature, rises in jets, blazing, clear, and bright." To emphasize this point, Herndon gave his most eager attention to evidences of the dismal and sordid in Lincoln's background. An extremely significant event in Lincoln's development, as Herndon viewed it, was a "romance of much reality" with Ann Rutledge. Lincoln loved no one but Ann and, after her death, never ceased to grieve for her. His memory of her both saddened and inspired him. As for his wife, Mary Todd, she married him out of spite, then devoted herself to making him miserable. So Herndon would have it, and after him countless biographers and novelists and playwrights elaborated upon his views, which persist as accepted knowledge about Lincoln despite their refutation by historical scholarship.

Lincoln has become a myth as well as a man. The legendary is to be found in imaginative literature and in folklore, in poems, plays, novels, anecdotes, and the like. It is also to be found in ostensibly factual productions, including footnoted biographies and histories.

The Lincoln of legend has grown into a protean god who can assume a shape to please almost any worshipper. He is Old Abe and at the same time a natural gentleman. He is Honest Abe and yet a being of superhuman shrewdness and cunning. He is also Father Abraham, the wielder of authority, the support of the weak; and he is an equal, a neighbour, and a friend.

The greatness of Lincoln

Lincoln the man has a reputation that may be considered apart from that of Lincoln the myth. While he was still alive, his reputation began to grow, and before his death his qualities of greatness already were widely recognized. In the midst of the Civil War, for instance, the *Washington Chronicle* found a resemblance between him and George Washington in their "sure judgment"; "perfect balance of thoroughly sound faculties"; and "great calmness of temper, great firmness of purpose, supreme moral principle, and intense patriotism." The *Buffalo Express* referred to his "remarkable moderation and freedom from passionate bitterness," then added: "We do not believe that Washington himself was less indifferent to the exercise of power for power's sake." An English newspaper, the *Liverpool Post,* suggested that "no leader in a great contest ever stood so little chance of being the subject of hero worship as Abraham Lincoln," if one were to judge only by the way he looked. His long arms and legs, his grotesque figure, made him too easy to caricature and ridicule. "Yet," this newspaper concluded, "a worshiper of human heroes might possibly travel a great deal farther and fare much worse for an idol than selecting this same lanky American." His inner qualities—his faithfulness, honesty, resolution, insight, humour, and courage–would "go a long way to make up a hero," whatever the man's personal appearance.

Lincoln's best ideas and finest phrases were considered and written and rewritten with meticulous revisions. Some resulted from a slow gestation of thought and phrase through many years. One of his recurring themes—probably his central theme––was the promise and the problem of self-government. As early as 1838, speaking to the Young Men's Lyceum of Springfield on "The Perpetuation of Our Political Institutions," he recalled the devotion of his Revolutionary forefathers to the cause, and went on to say:

> Their ambition aspired to display before an admiring world, a practical demonstration of the truth of a proposition, which had hitherto been considered, at best no better, than problematical; namely, *the capability of a people to govern themselves.*

Again and again he returned to this idea, especially after the coming of the Civil War, and he steadily improved his phrasing. In his first message to Congress after the fall of Ft. Sumter, he declared that the issue between North and South involved more than the future of the United States.

> It presents to the whole family of man, the question, whether a constitutional republic, or a democracy—a government of the people, by the same people—can, or cannot, maintain its territorial integrity, against its own domestic foes.

And finally at Gettysburg he made the culminating, the supreme statement, concluding with the words:

> . . . that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

BIBLIOGRAPHY. For the general reader, the most usable and valuable guide is PAUL M. ANGLE, A *Shelf of Lincoln Books* (1946). Practically all the known writings of Lincoln himself are available in *The Collected Works of Abraham Lincoln,* ed. by ROY P. BASLER, 9 vol. (1953–55). A judicious selection from these volumes is reprinted in *The Living Lincoln,* ed. by PAUL M. ANGLE and EARL S. MIERS (1955).

Classic multivolume biographies are JOHN G. NICOLAY and JOHN HAY, *Abraham Lincoln: A History,* 10 vol. (1890; abridged ed., 1 vol., 1966); ALBERT J. BEVERIDGE, *Abraham Lincoln, 1809–1850,* 2 vol. (1928); CARL SANDBURG, *Abraham Lincoln: The Prairie Years,* 2 vol. (1926), and *The War Years,* 4 vol. (1939); and JAMES G. RANDALL, *Lincoln the President,* 4 vol. (1945–55). The best of the one-volume lives is BENJAMIN P. THOMAS, *Abraham Lincoln* (1952).

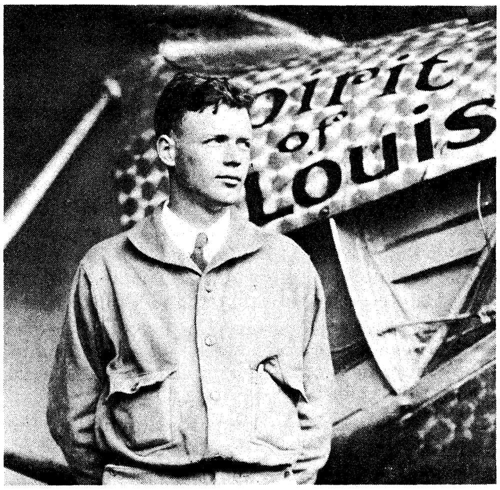
Matters of controversy and conflicting interpretation are discussed by DAVID DONALD in *Lincoln Reconsidered,* 2nd ed. (1961); and by RICHARD N. CURRENT in *The Lincoln Nobody Knows* (1958). Important works dealing with aspects of Lincoln's statesmanship are DON E. FEHRENBACHER, *Prelude to Greatness: Lincoln in the 1850's* (1962); DAVID M. POTTER, *Lincoln and His Party in the Secession Crisis* (1942); WILLIAM B. HESSELTINE, *Lincoln and the War Governors* (1948); KENNETH M. STAMPP, *And the War Came* (1950); and T. HARRY WILLIAMS, *Lincoln and the Radicals* (1941) and *Lincoln and His Generals* (1952). RUTH PAINTER RANDALL writes authoritatively of Lincoln's family life in *Mary Lincoln: Biography of a Marriage* (1953) and *Lincoln's Sons* (1955). Mrs. Randall effectively refutes the views of William H. Herndon, which may most conveniently be found in *Herndon's Life of Lincoln,* ed. by PAUL M. ANGLE (1930). An able and realistic treatment of Lincoln's legal career is JOHN J. DUFF, A. *Lincoln, Prairie Lawyer* (1960).

(R.N.C.)

Lindbergh, Charles A.

Probably the best known but least understood figure in aeronautical history, Charles Lindbergh heightened worldwide interest in aviation by his dramatic nonstop solo flight across the Atlantic in May 1927. That flight brought him fame, wealth, and a great personal tragedy. Its immediate aftermath also generated in him a great antipathy toward any form of personal publicity, and not until the mid-1960s did the self-generated screen covering his and his family's activities begin to dissipate.

Lindbergh was born in Detroit on February 4, 1902. His early years were spent mostly in Little Falls, Minnesota, and in Washington, D.C., where for ten years his father represented the 6th district of Minnesota in the Congress. His formal education ended during his second year at the University of Wisconsin, in Madison, when his growing interest in aviation led to enrollment in a flying school in Lincoln, Nebraska, and the purchase of a World War I Curtiss Jenny, with which he made stunt-flying tours through Southern and Midwestern states. After a year at the army flying schools in Texas (1924–25), he became an airmail pilot (1926), flying the route from St.



Lindbergh, 1927.
By courtesy of the Library of Congress, Washington. D.C.

Louis, Missouri, to Chicago. During this period he obtained financial backing from a group of St. Louis businessmen to compete for the $25,000 prize offered for the first nonstop flight between New York and Paris. In the monoplane "Spirit of St. Louis" he made the flight in 33½ hours on May 20–21, 1927. Overnight Lindbergh became a folk hero on both sides of the Atlantic and a well-known figure in most of the world. There followed a series of goodwill flights in Europe and America.

First nonstop New York–Paris flight

In Mexico, Lindbergh met Anne Morrow, daughter of the United States ambassador, Dwight Morrow. They were married in May 1929 and together made flights that took them to many countries of the world. During this period, Lindbergh acted as technical adviser to two airlines, Transcontinental Air Transport and Pan American Airways, personally pioneering many of their routes.

In March 1932 Lindbergh's two-year-old son, Charles Augustus, Jr., was kidnapped from the Lindberghs' home near Hopewell, New Jersey, and murdered. Partly because of Lindbergh's worldwide popularity, this became the most celebrated crime of the 1930s, and it was a major subject of newspaper attention until April 1936, when Bruno Richard Hauptmann was executed after being convicted of the kidnap-murder. The publicity was so distasteful to the Lindberghs that they took refuge in Europe. After 1936, when he visited German centres of aviation, Lindbergh repeatedly warned against the growing air power of Nazi Germany. His decoration by the German government in 1938 led to considerable criticism, as did the speeches advocating American neutrality in World War II he made in 1940–41 after his return to the United States. Criticism of his public statements by Pres. Franklin D. Roosevelt led Lindbergh to resign his Air Corps Reserve commission in April 1941.

When the United States entered the war, however, Lindbergh, as a civilian, threw himself unobtrusively into the war effort, serving as a consultant to the Ford Motor Company and to the United Aircraft Corporation. In the latter capacity he flew 50 combat missions during a tour of duty in the Pacific; and later, after the end of the war in Europe, he accompanied a navy technical mission in Europe investigating German aviation developments.

Following World War II, Lindbergh and his family lived quietly in Connecticut and then in Hawaii. He continued as consultant to Pan American World Airways and to the U.S. Department of Defense. He was a member of the National Advisory Committee for Aeronautics and served on a number of other aeronautical boards and committees. He received many honours and awards, in addition to the Medal of Honor that had been awarded to him by special act of Congress in 1927. For his services to the government he was appointed brigadier general in the Air Force Reserve by Pres. Dwight D. Eisenhower in 1954. His book *The Spirit of St. Louis,* describing the flight to Paris, was published in 1953 and gained him a Pulitzer Prize. He was also the author of *We* (1927), *Of Flight and Life* (1948), and, with the French surgeon and biologist Alexis Carrel, *The Culture of Organs* (1938), concerning researches on which he and Carrel had collaborated.

His *Wartime Journals* (1970) is a record (not initially intended for publication) that was handwritten as time permitted during the years 1938–45. Lindbergh died at his home on the island of Maui, Hawaii, on August 26, 1974.

BIBLIOGRAPHY. Lindbergh describes his nonstop airplane flight from New York to Paris in *The Spirit of St. Louis* (1953). An earlier volume in which he tells the story of his youth and his famous flight is *We* (1927). See also *The Wartime Journals of Charles A. Lindbergh* (1970), which covers the period from March 1938 to June 1945. Two biographies are KENNETH S. DAVIS, *The Hero* (1959); and WALTER S. ROSS, *The Last Hero: Charles A. Lindbergh* (1967).

Linguistics

Linguistics is the scientific study of language. The word was first used in the middle of the 19th century to emphasize the difference between a newer approach to the study of language that was then developing and the more traditional approach of philology. The differences were and are largely matters of attitude, emphasis, and purpose. The philologist is concerned primarily with the historical development of languages as it is manifest in written texts and in the context of the associated literature and culture. The linguist, though he may be interested in written texts and in the development of languages through time, tends to give priority to spoken languages and to the problems of analyzing them as they operate at a given point in time, without reference to their history. The linguist, in principle, is interested in all languages and not merely in the great literary languages of the world.

Synchronic and diachronic linguistics

The field of linguistics may be conveniently divided in terms of three dichotomies: synchronic versus diachronic, theoretical versus applied, microlinguistics versus macrolinguistics. A synchronic description of a language describes the language as it is at a given time; a diachronic description is concerned with the historical development of the language and the structural changes that have taken place in it between successive points in time. The goal of theoretical linguistics is the construction of a general theory of the structure of language or of a general theoretical framework for the description of languages (there is some controversy as to the difference between these two conceptions of the subject); the aim of applied linguistics is the application of the findings and techniques of the scientific study of language to a variety of practical tasks, especially to the elaboration of improved methods of language teaching. The terms microlinguistics and macrolinguistics are not yet well established, and they are, in fact, used here purely for convenience. The former refers to a narrower and the latter to a much broader view of the scope of linguistics. According to the microlinguistic view, languages should be analyzed for their own sake and without reference to their social function, to the manner in which they are acquired by children, to the psychological mechanisms that underlie the production and reception of speech, to the literary and the aesthetic or communicative function of language, and so on. In contrast, macrolinguistics embraces all of these aspects of language. A number of areas within macrolinguistics have been given terminological recognition: psycholinguistics, sociolinguistics, anthropological linguistics, dialectology, mathematical and computational linguistics, and stylistics. Macrolinguistics should not be identified with applied linguistics. The application of linguistic methods and concepts to language teaching may well involve other disciplines in a way that microlinguistics does not. But there is, in principle, a theoretical aspect to every part of macrolinguistics, no less than to microlinguistics.

About half of this article is devoted to theoretical, synchronic microlinguistics, which is generally acknowledged as the central part of the subject; it will be abbreviated henceforth as theoretical linguistics.

This article is divided into the following sections:

- I. History of linguistics
  - Earlier history
  - The 19th century
  - The 20th century
- II. Methods of synchronic linguistic analysis
  - Structural linguistics
  - Transformational–generative grammar
  - Tagmemics
  - Stratificational grammar
  - The Prague school
- III. Historical (diachronic) linguistics
  - Linguistic change
  - The comparative method
  - Language classification
- IV. Linguistics and other disciplines
  - Psycholinguistics
  - Sociolinguistics
  - Other relationships

I. History of linguistics

EARLIER HISTORY

Greek naturalists and conventionalists

Greek contributions. Western linguistic theory originated in Greece in the 5th century BC and was, at first, dominated by mainly philosophical issues, the most famous of which was whether language is of natural or conventional origin. According to the naturalists, the relationship between the form of a word and its meaning was, in the original state of language at least, of the kind exemplified by such onomatopoeic English words as "splash," "cuckoo," and so forth. Words whose form was no longer "naturally" related to their meaning in this way were traced back to what was assumed to have been their original "natural" state by the application of ad hoc and often quite fanciful etymological principles. The conventionalists argued that there were very few onomatopoeic forms, that they varied from language to language, and that they could be removed from the vocabulary without affecting the efficiency of the language as a vehicle of communication. The controversy endured for centuries and is still encountered in discussions of the origin of language. For the history of linguistics, its importance lies in the fact that it concentrated the attention of scholars upon the forms of words and grammatical patterns in their language.

The second great linguistic controversy of antiquity, that between the analogists and the anomalists, had a more direct and specific effect upon the development of traditional grammar. This concerned the grammatical structure of language—whether it is basically regular or irregular. Aristotle was an analogist, and the Alexandrian scholars followed him; the Stoics were, for the most part, anomalists. Both the Stoics and the Alexandrians contributed in roughly equal measure to the development of Greek grammatical theory, the former to the identification of the major grammatical categories, the latter to the establishment of the main patterns of inflection.

It was in Alexandria, one of the great centres of Hellenistic literary scholarship, that the grammar of Classical Greek was first codified. The so-called *Technē grammatikē* ("Art of Grammar") of Dionysius Thrax (c. 100 BC) was regarded as definitive for more than a thousand years, and it exerted a tremendous influence upon the grammars of the language written by Western scholars right down to modern times. The ***Technd*** of Dionysius dealt mainly with the accidence, or inflection, of Greek words. Greek syntax was described more than three centuries later, also in Alexandria, by Apollonius Dyscolus (2nd century AD). The grammar of Greek, as it is presented today in the standard handbooks, was fixed in all but relatively minor respects by the works of Dionysius and Apollonius.

Grammars of Donatus and Priscian

Roman practice. Roman grammarians were heavily dependent upon Greek models, and the fact that Greek and Latin were very similar in grammatical structure meant that Latin could be described in terms of the Greek grammatical categories without much difficulty. The Latin grammatical tradition, at first more strongly influenced by the Stoics than by the Alexandrians, later drew heavily from the grammars of Dionysius and Apollonius and, later, from those of Donatus (*c.* 400 AD) and Priscian (*c.* 500 AD). They were used not only for teaching Latin until well after the Renaissance, but they served as models for the description of many of the vernacular languages of Europe.

Medieval views on language. Though a number of advances were made in grammatical theory by medieval scholars, in general they were content to work with the grammar of Priscian. The most important development in the Middle Ages, as far as the history of linguistics is concerned, was the rise of what is usually called speculative grammar in the 13th and 14th centuries, when, dissatisfied with the mere description of the facts, scholastic grammarians set out to explain the structure not only of Latin but in principle of all languages in terms of the alleged congruence of grammatical, metaphysical, and epistemological categories. This interest in a universal grammar was revived in 17th-century France by the so-called rational grammar (*grammaire raisonnée*) of the Port Royal Abbey, on the basis, however, of Cartesian rather than scholastic philosophy.

Post-Renaissance work. Even during the Middle Ages there had been some sporadic concern with languages other than Latin. After the Renaissance this increased considerably and during the Reformation was further promoted by the Protestants' belief in the importance of translating the Bible into vernacular languages. With the expansion of Europe there evolved an increased appreciation of the range and diversity of the world's languages. By the 17th century, grammars of certain of the major indigenous languages of South America were being published, and some knowledge of Chinese, Vietnamese, Sanskrit, and other languages of the Far East, Southeast Asia, and India had been acquired, mainly as a result of the efforts of the Jesuit missionaries. All of this was to contribute to the formation of an intellectual climate in which linguistic research could set forth on a new and vigorous course of development in the 19th century.

THE 19TH CENTURY

Discovery of the Indo-European language family

Development of the comparative method. It is generally agreed that the most outstanding achievement of linguistic scholarship in the 19th century was the development of the comparative method, which comprised a set of principles whereby languages could be systematically compared with respect to their sound systems, grammatical structure, and vocabulary and shown to be "genealogically" related. As French, Italian, Portuguese, Romanian, Spanish, and the other Romance languages had evolved from Latin, so Latin, Greek, and Sanskrit as well as the Celtic, Germanic, and Slavic languages and many other languages of Europe and Asia had evolved from some earlier language, to which the name Indo-European or Proto-Indo-European is now customarily applied. That all the Romance languages were descended from Latin and thus constituted one "family" had been known for centuries; but the existence of the Indo-European family of languages and the nature of their genealogical relationship was first demonstrated by the 19th-century comparative philologists. (The term philology in this contest is not restricted to the study of literary languages.)

The main impetus for the development of comparative philology came toward the end of the 18th century, when it was discovered that Sanskrit bore a number of striking resemblances to Greek and Latin. An English orientalist, Sir William Jones, though he was not the first to observe these resemblances, is generally given the credit for bringing them to the attention of the scholarly world and putting forward the hypothesis, in 1786, that all three languages must have "sprung from some common source; which perhaps no longer exists." By this time, a number of texts and glossaries of the older Germanic languages (Gothic, Old High German, and Old Norse) had been published, and Jones realized that Germanic as well as Old Persian and perhaps Celtic had evolved from the same "common source." The next important step came in 1822, when the German scholar Jacob Grimm, following the Danish linguist Rasmus Rask (whose work, being written in Danish, was less accessible to most European scholars), pointed out in the second edition of his comparative grammar of Germanic that there were a number of systematic correspondences between the sounds of Germanic and the sounds of Greek, Latin. and Sanskrit in related words. Grimm noted, for example, that where Gothic (the oldest surviving Germanic language) had an *f*, Latin, Greek, and Sanskrit frequently had a *p* (*e.g.*, Gothic *fotus*, Latin *pedis*, Greek *podós*, Sanskrit *padás*, all meaning "foot"); when Gothic had a *p*, the non-Germanic languages had a *b;* when Gothic had a *b*, the non-Germanic languages had what Grimm called an "aspirate" (Latin *f*, Greek *ph*, Sanskrit *bh*). In order to account for these correspondences he postulated a cyclical "soundshift" (*Lautverschiebung*) in the prehistory of Germanic, in which the original "aspirates" became voiced unaspirated stops (*bh* became b, etc.), the original voiced unaspirated stops became voiceless (b became p, etc.), and the original voiceless (unaspirated) stops became "aspirates" (*p* became f). Grimm's term, "aspirate," it will be noted, covered such phonetically distinct categories as aspirated stops (*bh*, *ph*), produced with an accompanying audible puff of breath, and fricatives (f), produced with audible friction as a result of incomplete closure in the vocal tract.

The **Neogrammarians** and sound laws

In the work of the next 50 years the idea of sound change was made more precise, and, in the 1870s, a group of scholars known collectively as the *Jurzggramrnatiker* ("young grammarians," or Neogrammarians) put forward the thesis that all changes in the sound system of a language as it developed through time were subject to the operation of regular sound laws. Though the thesis that sound laws were absolutely regular in their operation (unless they were inhibited in particular instances by the influence of analogy) was at first regarded as most controversial, by the end of the 19th century it was quite generally accepted and had become the cornerstone of the comparative method. Using the principle of regular sound change, scholars were able to reconstruct "ancestral" common forms from which the later forms found in particular languages could be derived. By convention, such reconstructed forms are marked in the literature with an asterisk. Thus, from the reconstructed Proto-Indo-European word for "ten," **dekm*, it was possible to derive Sanskrit *daśa*, Greek *déka*, Latin *decem*, and Gothic *taihun* by postulating *a* number of different sound laws that operated independently in the different branches of the Indo-European family. The question of sound change is dealt with in greater detail in the section entitled *Historical Linguistics.*

The role of analogy. Analogy has been mentioned in connection with its inhibition of the regular operation of sound laws in particular word forms. This was how the Neogrammarians thought of it. In the course of the 20th century, however, it has come to be recognized that analogy, taken in its most general sense, plays a far more important role in the development of languages than simply that of sporadically preventing what would otherwise be a completely regular transformation of the sound system of a language. When a child learns to speak he tends to regularize the anomalous, or irregular, forms by analogy with the more regular and productive patterns of formation in the language; *e.g.*, he will tend to say "comed" rather than "came," "dived rather than "dove," and so on, just as he will say "talked," "loved," and so forth. The fact that the child does this is evidence that he has learned or is learning the regularities or rules of his language. He will go on to "unlearn" some of the analogical forms and substitute for them the anomalous forms current in the speech of the previous generation. But in some cases, he will keep a "new" analogical form (*e.g.*, "dived" rather than "dove"), and this may then become the recognized and accepted form.

Ideas of Wilhelm von Humboldt

Additional 19th-century theories **and development.** *Inner and outer form.* One of the most original, if not one of the most immediately influential, linguists of the 19th century was the learned Prussian statesman, Wilhelm von Humboldt (died 1835). His interests, unlike those of most of his contemporaries, were not exclusively historical. Following the German philosopher Johann Gottfried von Herder (1744–1803), he stressed the connection between national languages and national character: this was but a commonplace of romanticism. More original was Humboldt's theory of "inner" and "outer" form in language. The outer form of language was the raw material (the sounds) from which different languages were fashioned; the inner form was the pattern, or structure, of grammar and meaning that was imposed upon this raw material and differentiated one language from another. This "structural" conception of language was to become dominant, for a time at least, in many of the major centres of linguistics by the middle of the 20th century. Another of Humboldt's ideas was that language was something dynamic, rather than static, and was an activity itself rather than the product of activity. A language was not a set of actual utterances produced by speakers but the underlying principles or rules that made it possible for speakers to produce such utterances and, moreover, an unlimited number of them. This idea was taken up by a German philologist, Heymann Steinthal, and, what is more important, by the physiologist and psychologist Wilhelm Wundt, and thus influenced late 19th- and early 20th-century theories of the psychology of language. Its influence, like that of the distinction of inner and outer form, can also be seen in the thought of Ferdinand de Saussure, a Swiss linguist. But its full implications were probably not perceived and made precise until the middle of the 20th century, when the U.S. linguist Noam Chomsky re-emphasized it and made it one of the basic notions of generative grammar (see below).

Development of phonetics

Phonetics and dialectology. Many other interesting and important developments occurred in 19th-century linguistic research, among them work in the areas of phonetics and dialectology. Research in both these fields was promoted by the Neogrammarians' concern with sound change and by their insistence that prehistoric developments in languages were of the same kind as developments taking place in the languages and dialects currently spoken. The development of phonetics in the West was also strongly influenced at this period, as were many of the details of the more philological analysis of the Indo-European languages, by the discovery of the works of the Indian grammarians who, from the time of the Sanskrit grammarian Pāṇini (5th or 6th century BC), if not before, had arrived at a much more comprehensive and scientific theory of phonetics, phonology, and morphology than anything achieved in the West until the modern period.

THE 20TH CENTURY

Structuralism. The term structuralism has been used as a slogan and rallying cry by a number of different schools of linguistics, and it is necessary to realize that it has somewhat different implications according to the context in which it is employed. It is convenient to draw first a broad distinction between European and American structuralism and, then, to treat them separately.

Structural linguistics in Europe. Structural linguistics in Europe is generally said to have begun in 1916 with the posthumous publication of the *Cours de Linguistique Générale (Coume in General Linguistics)* of Ferdinand de Saussure. Much of what is now considered as Saussurean can be seen, though less clearly, in the earlier work of Humboldt, and the general structural principles that Saussure was to develop with respect to synchronic linguistics in the *Cours* had been applied almost 40 years before (1879) by Saussure himself in a reconstruction of the Indo-European vowel system. The full significance of the work was not appreciated at the time. Saussure's structuralism can be summed up in two dichotomies (which jointly cover what Humboldt referred to in terms of his own distinction of inner and outer form): (1) *langue* versus *parole* and (2) form versus substance. By *langue*, best translated in its technical Saussurean sense as language system, is meant the totality of regularities and patterns of formation that underlie the utterances of a language; by *parole*, which can be translated as language behaviour, is meant the actual utterances themselves. Just as two performances of a piece of music given by different orchestras on different occasions will differ in a variety of details and yet be identifiable as performances of the same piece, so two utterances may differ in various ways and yet be recognized as instances, in some sense, of the same utterance. What the two musical performances and the two utterances have in common is an identity of form, and this form, or structure, or pattern, is in principle independent of the substance, or "raw material," upon which it is imposed. "Structuralism," in the European sense then, refers to the view that there is an abstract relational structure that underlies and is to be distinguished from actual utterances—a system underlying actual behaviour—and that this is the primary object of study for the linguist.

European sense of structuralism

Two important points arise here: first, that the structural approach is not in principle restricted to synchronic linguistics; second, that the study of meaning, as well as the study of phonology and grammar, can be structural in orientation. In both cases "structuralism" is opposed to "atomism" in the European literature. It was Saussure who drew the terminological distinction between synchronic and diachronic linguistics in the *Cours*; despite the undoubtedly structural orientation of his own early work in the historical and comparative field, he maintained that, whereas synchronic linguistics should deal with the structure of a language system at a given point in time, diachronic linguistics should be concerned with the historical development of isolated elements—it should be atomistic. Whatever the reasons that led Saussure to take this rather paradoxical view, his teaching on this point was not generally accepted, and scholars soon began to apply structural concepts to the diachronic study of languages. The most important of the various schools of structural linguistics to be found in Europe in the first half of the 20th century have included the Prague school, most notably represented by Nikolay Sergeyevich Trubetskoy (died 1938) and Roman Jakobson (born 1896), both Russian émigrés, and the Copenhagen (or glossematic) school, centred around Louis Hjelmslev (died 1970). John Rupert Firth (died 1960) and his followers, sometimes referred to as the London school, were less Saussurean in their approach, but, in a general sense of the term, their approach may also be described appropriately as structural linguistics.

Structural linguistics in America. American and European structuralism shared a number of features. In insisting upon the necessity of treating each language as a more or less coherent and integrated system, both European and American linguists of this period tended to emphasize, if not to exaggerate, the structural uniqueness of individual languages. There was especially good reason to take this point of view given the conditions in which American linguistics developed from the end of the 19th century. There were hundreds of indigenous American Indian languages that had never been previously described. Many of these were spoken by only a handful of speakers and, if they were not recorded before they became extinct, would be permanently inaccessible. Under these circumstances, such linguists as Franz Boas (died 1942) were less concerned with the construction of a general theory of the structure of human language than they were with prescribing sound methodological principles for the analysis of unfamiliar languages. They were also fearful that the description of these languages would be distorted by analyzing them in terms of categories derived from the analysis of the more familiar Indo-European languages.

The work of Sapir, Bloomfield, and Whorf

After Boas, the two most influential American linguists were Edward Sapir (died 1939) and Leonard Bloomfield (died 1949). Like his teacher Boas, Sapir was equally at home in anthropology and linguistics, the alliance of which disciplines has endured to the present day in many American universities. Boas and Sapir were both attracted by the Humboldtian view of the relationship between language and thought, but it was left to one of Sapir's pupils, Benjamin Lee Whorf, to present it in a sufficiently challenging form to attract widespread scholarly attention. Since the republication of Whorf's more important papers in 1956, the thesis that language determines perception and thought has come to be known as the Whorfian hypothesis.

Sapir's work has always held an attraction for the more anthropologically inclined American linguists. But it was Bloomfield who prepared the way for the later phase of what is now thought of as the most distinctive manifestation of American "structuralism." When he published his first book in 1914, Bloomfield was strongly influenced by Wundt's psychology of language. In 1933, however, he published a drastically revised and expanded version with the new title Language; this book dominated the field for the next 30 years. In it Bloomfield explicitly adopted a behaviouristic approach to the study of language, eschewing in the name of scientific objectivity all reference to mental or conceptual categories. Of particular consequence was his adoption of the behaviouristic theory of semantics according to which meaning is simply the relationship between a stimulus and a verbal response. Because science was still a long way from being able to give a comprehensive account of most stimuli, no significant or interesting results could be expected from the study of meaning for some considerable time, and it was preferable, as far as possible, to avoid basing the grammatical analysis of a language on semantic considerations. Bloomfield's followers pushed even further the attempt to develop methods of linguistic analysis that were not based on meaning. One of the most characteristic features of "post-Bloomfieldian" American structuralism, then, was its almost complete neglect of semantics.

Another characteristic feature, one that was to be much criticized by Chomsky, was its attempt to formulate a set of "discovery procedures"—procedures that could be applied more or less mechanically to texts and could be guaranteed to yield an appropriate phonological and grammatical description of the language of the texts. Structuralism, in this narrower sense of the term, is represented, with differences of emphasis or detail, in the major American textbooks published during the 1950s.

More recent developments. The most significant development in linguistic theory and research in recent years was the rise of generative grammar, and, more especially, of transformational–generative grammar, or transformational grammar, as it came to be known. Two versions of transformational grammar were put forward in the mid-1950s, the first by Zellig S. Harris and the second by Noam Chomsky, his pupil. It is Chomsky's system that has attracted the most attention so far. As first presented by Chomsky in Syntactic Structures (1957), transformational grammar can be seen partly as a reaction against post-Bloomfieldian structuralism and partly as a continuation of it. What Chomsky reacted against most strongly was the post-Bloomfieldian concern with discovery procedures. In his opinion, linguistics should set itself the more modest and more realistic goal of formulating criteria for evaluating alternative descriptions of a language without regard to the question of how these descriptions had been arrived at. The statements made by linguists in describing a language should, however, be cast within the framework of a far more precise theory of grammar than had hitherto been the case, and this theory should be formalized in terms of modern mathematical notions. Within a few years, Chomsky had broken with the post-Bloomfieldians on a number of other points also. He had adopted what he called a "mentalistic" theory of language, by which term he implied that the linguist should be concerned with the speaker's creative linguistic competence and not his performance the actual utterances produced. He had challenged the post-Bloomfieldian concept of the phoneme (see below), which many scholars regarded as the most solid and enduring result of the previous generation's work. And he had challenged the structuralists' insistence upon the uniqueness of every language, claiming instead that all languages were, to a considerable degree, cut to the same pattern—they shared a certain number of formal and substantive universals.

The work of Harris and Chomsky

The effect of Chomsky's ideas has been phenomenal. It is hardly an exaggeration to say that there is no major theoretical issue in linguistics today that is debated in terms other than those in which he has chosen to define it, and every school of linguistics tends to define its position in relation to his. Among the rival schools are tagmemics, stratificational grammar, and the Prague school. Tagmemics is the system of linguistic analysis developed by the U.S. linguist Kenneth L. Pike and his associates in connection with their work as Bible translators. Its foundations were laid during the 1950s, when Pike differed from the post-Bloomfieldian structuralists on a number of principles, and it has been further elaborated since then. Tagmemic analysis has been used for analyzing a great many previously unrecorded languages, especially in Central and South America and in West Africa. Stratificational grammar, developed by a U.S. linguist, Sydney M. Lamb, has been seen by some linguists as an alternative to transformational grammar. Not yet fully expounded or widely exemplified in the analysis of different languages, stratificational grammar is perhaps best characterized as a radical modification of post-Bloomfieldian linguistics, but it has many features that link it with European structuralism. The Prague school has been mentioned above for its importance in the period immediately following the publication of Saussure's ***Cours.*** Many of its characteristic ideas (in particular, the notion of distinctive features in phonology) have been taken up by other schools. But there has been further development in Prague of the functional approach to syntax (see below). The work of M.A.K. Halliday in England derived much of its original inspiration from Firth (above), but Halliday provided a more systematic and comprehensive theory of the structure of language than Firth had, and it has been quite extensively illustrated.

Prague school

II. Methods of synchronic linguistic analysis

STRUCTURAL LINGUISTICS

This section is concerned mainly with a version of structuralism (which may also be called descriptive linguistics) developed by scholars working in a post-Bloomfieldian tradition.

Phonology. With the great progress made in phonetics in the late 19th century, it had become clear that the question whether two speech sounds were the same or not was more complex than might appear at first sight. Two utterances of what was taken to be the same word might differ quite perceptibly from one occasion of utterance to the next. Some of this variation could be attributed to a difference of dialect or accent and is of no concern here. But even two utterances of the same word by the same speaker might vary from one occasion to the next. Variation of this kind, though it is generally less obvious and would normally pass unnoticed, is often clear enough to the trained phonetician and is measurable instrumentally. It is known that the "same" word is being uttered, even if the physical signal produced is variable, in part, because the different pronunciations of the same word will cluster around some acoustically identifiable norm. But this is not the whole answer, because it is actually impossible to determine norms of pronunciation in purely acoustic terms. Once it has been decided what counts as "sameness" of sound from the linguistic point of view, the permissible range of variation for particular sounds in particular contexts can be measured, and, within certain limits, the acoustic cues for the identification of utterances as "the same" can be determined.

What is at issue is the difference between phonetic and phonological (or phonemic) identity, and for these purposes it will be sufficient to define phonetic identity in terms solely of acoustic "sameness." Absolute phonetic identity is a theoretical ideal never fully realized. From a purely phonetic point of view, sounds are more or less similar, rather than absolutely the same or absolutely different. Speech sounds considered as units of phonetic analysis in this article are called phones, and, following the normal convention, are represented by enclosing the appropriate alphabetic symbol in square brackets. Thus [p] will refer to a p sound (*i.e.,* what is described more technically as a voiceless, bilabial stop); and [pit] will refer to a complex of three phones—a p sound, followed by an *i* sound, followed by a *t* sound. A phonetic transcription may be relatively broad (omitting much of the acoustic detail) or relatively narrow (putting in rather more of the detail), according to the purpose for which it is intended. A very broad transcription will be used in this article except when finer phonetic differences must be shown.

Phones: speech sounds

Phonological, or phonemic, identity was referred to above as "sameness of sound from the linguistic point of view." Considered as phonological units—*i.e.*, from the point of view of their function in the language—sounds are described as phonemes and are distinguished from phones by enclosing their appropriate symbol (normally, but not necessarily, an alphabetic one) between two slash marks. Thus /p/ refers to a phoneme that may be realized on different occasions of utterance or in different contexts by a variety of more or less different phones. Phonological identity, unlike phonetic similarity, is absolute: two phonemes are either the same or different, they cannot be more or less similar. For example, the English words "bit" and "pit" differ phonemically in that the first has the phoneme /b/ and the second has the phoneme /p/ in initial position. As the words are normally pronounced, the phonetic realization of /b/ will differ from the phonetic realization of /p/ in a number of different ways: it will be at least partially voiced (*i.e.,* there will be some vibration of the vocal cords), it will be without aspiration (*i.e.*, there will be no accompanying slight puff of air, as there will be in the case of the phone realizing /p/), and it will be pronounced with less muscular tension. It is possible to vary any one or all of these contributory differences, making the phones in question more or less similar, and it is possible to reduce the phonetic differences to the point that the hearer cannot be certain which word, "bit" or "pit," has been uttered. But it must be either one or the other; there is no word with an initial sound formed in the same manner as /p/ or /b/ that is halfway between the two. This is what is meant by saying that phonemes are absolutely distinct from one another —they are discrete rather than continuously variable.

Classifying phones or phonemes

How it is known whether two phones realize the same phoneme or not is dealt with differently by different schools of linguists. The "orthodox" post-Bloomfieldian school regards the first criterion to be phonetic similarity. Two phones are not said to realize the same phoneme unless they are sufficiently similar. What is meant by "sufficiently similar" is rather vague, but it must be granted that for every phoneme there is a permissible range of variation in the phones that realize it. As far as occurrence in the same context goes, there are no serious problems. More critical is the question of whether two phones occurring in different contexts can be said to realize the same phoneme or not. To take a standard example from English: the phone that occurs at the beginning of the word "pit" differs from the phone that occurs after the initial /s/ of "spit." The "*p* sound" occurring after the /s/ is unaspirated (*i.e.*, it is pronounced without any accompanying slight puff of air). The aspirated and unaspirated "*p* sounds" may be symbolized rather more narrowly as [pʰ] and [p] respectively. The question then is whether [pʰ] and [p] realize the same phoneme /p/ or whether each realizes a different phoneme. They satisfy the criterion of phonetic similarity, but this, though a necessary condition of phonemic identity, is not a sufficient one.

The next question is whether there is any pair of words in which the two phones are in minimal contrast (or opposition); that is, whether there is any context in English in which the occurrence of the one rather than the other has the effect of distinguishing two or more words (in the way that [pʰ] versus [b] distinguishes the so-called minimal pairs "pit" and "bit," "pan" and "ban," and so on). If there is, it can be said that, despite their phonetic similarity, the two phones realize (or "belong to") different phonemes—that the difference between them is phonemic. If there is no context in which the two phones are in contrast (or opposition) in this sense, it can be said that they are variants of the same phoneme—that the difference between them is nonphonemic. Thus, the difference between [pʰ] and [p] in English is nonphonemic; the two sounds realize, or belong to, the same phoneme, namely /p/. In several other languages—e.g., Hindi—the contrast between such sounds as [pʰ] and [p] is phonemic, however. The question is rather more complicated than it has been represented here. In particular. it should be noted that [p] is phonetically similar to [b] as well as to [pʰ] and that, although [pʰ] and [b] are in contrast, [p] and [b] are not. It would thus be possible to regard [p] and [b] as variants of the same phoneme. Most linguists, however, have taken the alternative view, assigning [p] to the same phoneme as [pʰ]. Here it will suffice to note that the criteria of phonetic similarity and lack of contrast do not always uniquely determine the assignment of phones to phonemes. Various supplementary criteria may then be invoked.

Free variation and complementary distribution

Phones that can occur and do not contrast in the same contest are said to be in free variation in that context, and, as has been shown, there is a permissible range of variation for the phonetic realization of all phonemes. More important than free variation in the same context, however? is systematically determined variation according to the contest in which a given phoneme occurs. To return to the example used above: [p] and [pʰ], though they do not contrast, are not in free variation either. Each of them has its own characteristic positions of occurrence, and neither occurs, in normal English pronunciation, in any context characteristic for the other (e.g., only [pʰ] occurs at the beginning of a word, and only [p] occurs after s). This is expressed by saying that they are in complementary distribution. (The distribution of an element is the whole range of contexts in which it can occur.) Granted that [p] and [pʰ] are variants of the same phoneme /p/, it can be said that they are contextually, or positionally, determined variants of it. To use the technical term, they are allophones of /p/. The allophones of a phoneme, then, are its contextually determined variants and they are in complementary distribution.

The post-Bloomfieldians made the assignment of phones to phonemes subject to what is now generally referred to as the principle of bi-uniqueness. The phonemic specification of a word or utterance was held to determine uniquely its phonetic realization (except for free variation), and, conversely, the phonetic description of a word or utterance was held to determine uniquely its phonemic analysis. Thus, if two words or utterances are pronounced alike, then they must receive the same phonemic description; conversely, two words or utterances that have been given the same phonemic analysis must be pronounced alike. The principle of bi-uniqueness was also held to imply that, if a given phone was assigned to a particular phoneme in one position of occurrence, then it must be assigned to the same phoneme in all its other positions of occurrence; it could not be the allophone of one phoneme in one context and of another phoneme in other contexts.

A second important principle of the post-Bloomfieldian approach was its insistence that phonemic analysis should be carried out prior to and independently of grammatical analysis. Neither this principle nor that of bi-uniqueness was at all widely accepted outside the post-Bloomfieldian school, and they have been abandoned by the generative phonologists (see below).

Phonemes of the kind referred to so far are segmental; they are realized by consonantal or vocalic (vowel) segments of words, and they can be said to occur in a certain order relative to one another. For example, in the phonemic representation of the word "bit," the phoneme /b/ precedes /i/, which precedes /t/. But nonsegmental, or suprasegmental, aspects of the phonemic realization of words and utterances may also be functional in a language. In English, for example, the noun "import" differs from the verb "import" in that the former is accented on the first and the latter on the second syllable. This is called a stress accent: the accented syllable is pronounced with greater force or intensity. Many other languages distinguish words suprasegmentally by tone. For example, in Mandarin Chinese the words **haò** "day" and **haǒ** "good" are distinguished from one another in that the first has a falling tone and the second a falling–rising tone: these are realized, respectively, as (1) a fall in the pitch of the syllable from high to low and (2) a change in the pitch of the syllable from medium to low and back to medium. Stress and tone are suprasegmental in the sense that they are "superimposed" upon the sequence of segmental phonemes. The term tone is conventionally restricted by linguists to phonologically relevant variations of pitch at the level of words. Intonation, which is found in all languages, is the variation in the pitch contour or pitch pattern of whole utterances, of the kind that distinguishes (either of itself or in combination with some other difference) statements from questions or indicates the mood or attitude of the speaker (as hesitant, surprised, angry, and so forth). Stress, tone, and intonation do not exhaust the phonologically relevant suprasegmental features found in various languages, but they are among the most important.

Suprasegmental phonemes

A complete phonological description of a language includes all the segmental phonemes and specifies which allophones occur in which contexts. It also indicates which sequences of phonemes are possible in the language and which are not: it will indicate, for example, that the sequences /bl/ and /br/ are possible at the beginning of English words but not /bn/ or /bm/. A phonological description also identifies and states the distribution of the suprasegmental features. Just how this is to be done, however, has been rather more controversial in the post-Bloomfieldian tradition. Differences between the post-Bloomfieldian approach to phonology and approaches characteristic of other schools of structural linguistics will be treated below.

Morphology. The grammatical description of many, if not all, languages is conveniently divided into two complementary sections: morphology and syntax. The relationship between them, as generally stated, is as follows: morphology accounts for the internal structure of words, and syntax describes how words are combined to form phrases, clauses, and sentences.

Distinction between morphology and syntax

There are many words in English that are fairly obviously analyzable into smaller grammatical units. For example, the word "unacceptability" can be divided into un-, accept, abil-, and -ity (abil- being a variant of -able). Of these, at least three are minimal grammatical units, in the sense that they cannot be analyzed into yet smaller grammatical units—un-, abil-, and -ity. The status of accept, from this point of view, is somewhat uncertain. Given the existence of such forms as accede and accuse, on the one hand, and of except, excede, and excuse, on the other, one might be inclined to analyze accept into ac- (which might subsequently be recognized as a variant of ad-) and -cept. The question is left open. Minimal grammatical units like urt-, abil-, and -ity are what Bloomfield called morphemes; he defined them in terms of the "partial phonetic-semantic resemblance" holding within sets of words. For example, "unacceptable," "untrue," and "ungracious" are phonetically (or, phonologically) similar as far as the first syllable is concerned and are similar in meaning in that each of them is negative by contrast with a corresponding positive adjective ("acceptable," "true," "gracious"). This "partial phonetic-semantic resemblance" is accounted for by noting that the words in question contain the same morpheme (namely, un-) and that this morpheme has a certain phonological form and a certain meaning.

Bloomfield's definition of the morpheme in terms of "partial phonetic-semantic resemblance" was considerably modified and, eventually, abandoned entirely by some of his followers. Whereas Bloomfield took the morpheme to be an actual segment of a word, others defined it as being a purely abstract unit, and the term morph was introduced to refer to the actual word segments. The distinction between morpheme and morph (which is, in certain respects, parallel to the distinction between phoneme and phone) may be explained by means of an

Plural morphs in English

example. If a morpheme in English is posited with the function of accounting for the grammatical difference between singular and plural nouns, it may be symbolized by enclosing the term plural within brace brackets. Now the morpheme (plural) is represented in a number of different ways. Most plural nouns in English differ from the corresponding singular forms in that they have an additional final segment. In the written forms of these words, it is either -*s* or -es (*e.g.*, "cat" : "cats"; "dog" : "dogs"; "fish" : "fishes"). The word segments written -*s* or -es are morphs. So also is the word segment written -en in "oxen." All these morphs represent the same morpheme. But there are other plural nouns in English that differ from the corresponding singular forms in other ways (*e.g.*, "mouse" : "mice"; "criterion" : "criteria"; and so on) or not at all (*e.g.*, "this sheep" : "these sheep"). Within the post-Bloomfieldian framework no very satisfactory account of the formation of these nouns could be given. But it was clear that they contained (in some sense) the same morpheme as the more regular plurals.

Morphs that are in complementary distribution and represent the same morpheme are said to be allomorphs of that morpheme. For example, the regular plurals of English nouns are formed by adding one of three morphs on to the form of the singular: /s/, /z/, or /iz/ (in the corresponding written forms both /s/ and /z/ are written -*s* and /iz/ is written -*es*). Their distribution is determined by the following principle: if the morph to which they are to be added ends in a "sibilant" sound (*e.g.*, *s, z, sh, ch*), then the syllabic allomorph /iz/ is selected (*e.g.*, fish-es /fiš-iz/, *match-es* /mač-iz/); otherwise the nonsyllabic allomorphs are selected, the voiceless allomorph /s/ with morphs ending in a voiceless consonant (*e.g.*, cat-s /kat-s/) and the voiced allomorph /z/ with morphs ending in a vowel or voiced consonant (*e.g.*, flea-s /fli-z/, *dog-s* /dog-z/). These three allomorphs, it will be evident, are in complementary distribution, and the alternation between them is determined by the phonological structure of the preceding morph. Thus the choice is phonologically conditioned.

Past participle morphs in English

Very similar is the alternation between the three principal allomorphs of the past participle ending. /id/, /t/, and /d/, all of which correspond to the -*ed* of the written forms. If the preceding morph ends with /t/ or /d/, then the syllabic allomorph /id/ is selected (*e.g.*, wail-ed /weit-id/). Otherwise, if the preceding morph ends with a voiceless consonant, one of the nonsyllabic allomorphs is selected—the voiceless allomorph /t/ when the preceding morph ends with a voiceless consonant (*e.g., pack-ed* /pak-t/) and the voiced allomorph /d/ when the preceding morph ends with a vowel or voiced consonant (*e.g.,* row-ed /rou-d/; tarne-d /teim-d/). This is another instance of phonological conditioning. Phonological conditioning may be contrasted with the principle that determines the selection of yet another allomorph of the past participle morpheme. The final /n/ of show-n or *see-n* (which marks them as past participles) is not determined by the phonological structure of the morphs *show* and see. For each English word that is similar to "show" and "see" in this respect, it must be stated as a synchronically inexplicable fact that it selects the /n/ allomorph. This is called grammatical conditioning. There are various kinds of grammatical conditioning.

Alternation of the kind illustrated above for the allomorphs of the plural morpheme and the /id/, /d/, and /t/ allomorphs of the past participle is frequently referred to as morphophonemic. Some linguists have suggested that it should be accounted for not by setting up three allomorphs each with a distinct phonemic form but by setting up a single morph in an intermediate morphophonemic representation. Thus, the regular plural morph might be said to be composed of the morphophoneme /Z/ and the most common past-participle morph of the morphophoneme /D/. General rules of morphophonemic interpretation would then convert /Z/ and /D/ to their appropriate phonetic form according to context. This treatment of the question foreshadows, on the one hand, the stratificational treatment and, on the other, the generative approach, though they differ considerably in other respects.

An important concept in grammar and, more particularly, in morphology is that of free and bound forms. A bound form is one that cannot occur alone as a complete utterance (in some normal context of use). For example, -*ing* is bound in this sense, whereas wait is not, nor is waiting. Any form that is not bound is free. Bloomfield based his definition of the word on this distinction between bound and free forms. Any free form consisting entirely of two or more smaller free forms was said to be a phrase (*e.g.*, "poor John" or "ran away"), and phrases were to be handled within syntax. Any free form that was not a phrase was defined to be a word and to fall within the scope of morphology. One of the consequences of Bloomfield's definition of the word was that morphology became the study of constructions involving bound forms. The so-called isolating languages, which make no use of bound forms (*e.g.*, Vietnamese), would have no morphology.

Inflection and derivation

The principal division within morphology is between inflection and derivation (or word formation). Roughly speaking, inflectional constructions can be defined as yielding sets of forms that are all grammatically distinct forms of single vocabulary items, whereas derivational constructions yield distinct vocabulary items. For example, "sings," "singing," "sang," and "sung" are all inflectional forms of the vocabulary item traditionally referred to as "the verb to sing"; but "singer," which is formed from "sing" by the addition of the morph -er (just as "singing" is formed by the addition of -*ing*), is one of the forms of a different vocabulary item. When this rough distinction between derivation and inflection is made more precise, problems occur. The principal consideration, undoubtedly, is that inflection is more closely integrated with and determined by syntax. But the various formal criteria that have been proposed to give effect to this general principle are not uncommonly in conflict in particular instances, and it probably must be admitted that the distinction between derivation and inflection, though clear enough in most cases, is in the last resort somewhat arbitrary.

Bloomfield and most linguists have discussed morphological constructions in terms of processes. Of these, the most widespread throughout the languages of the world is affixation; *i.e.*, the attachment of an affix to a base. For example, the word "singing" can be described as resulting from the affixation of -*ing* to the base sing. (If the affix is put in front of the base, it is a prefix; if it is put after the base, it is a suffix; and if it is inserted within the base, splitting it into two discontinuous parts, it is an infix.) Other morphological processes recognized by linguists need not be mentioned here, but reference may be made to the fact that many of Bloomfield's followers from the mid-1940s were dissatisfied with the whole notion of morphological processes. Instead of saying that -*ing* was affixed to *sing* they preferred to say that sing and -*ing* co-occurred in a particular pattern or arrangement, thereby avoiding the implication that *sing* is in some sense prior to or more basic than -*ing*. The distinction of morpheme and morph (and the notion of allomorphs) was developed in order to make possible the description of the morphology and syntax of a language in terms of "arrangements" of items rather than in terms of "processes" operating upon more basic items. Nowadays, the opposition to "processes" is, except among the stratificationalists, almost extinct. It has proved to be cumbersome if not impossible to describe the relationship between certain linguistic forms without deriving one from the other or both from some common underlying form, and most linguists no longer feel that this is in any way reprehensible.

Bloomfield's theory of syntax

Syntax. Syntax, for Bloomfield, was the study of free forms that were composed entirely of free forms. Central to his theory of syntax were the notions of form classes and constituent structure. (These notions were also relevant though less central in the theory of morphology.) Bloomfield defined form classes, rather imprecisely, in terms of some common "recognizable phonetic or grammatical feature" shared by all the members. He gave as examples the form class consisting of "personal substantive expressions" in English (defined as "the forms that, when spoken with exclamatory final pitch, are calls for a person's presence or attention"; *e.g.*, "John," "Boy," "Mr. Smith"); the form class consisting of "infinitive expressions" (defined as "forms which. when spoken with exclamatory final pitch, have the meaning of a command"; *e.g.*, "run," "jump," "come here"); the form class of "nominative substantive expressions" (*e.g.*, "John," "the boys"); and so on. It should be clear from these examples that form classes are similar to though not identical with the traditional parts of speech, and that one and the same form can belong to more than one form class.

What Bloomfield had in mind as the criterion for form class membership (and therefore of syntactic equivalence) may best be expressed in terms of substitutability. Form classes are sets of forms (whether simple or complex, free or bound) any one of which may be substituted for any other in a given construction or set of constructions throughout the sentences of the language.

Constituents and immediate constituents

The smaller forms into which a larger form may be analyzed are its constituents, and the larger form is a construction. For example, the phrase "poor John" is a construction analyzable into, or composed of, the constituents "poor" and "John." Because there is no intermediate unit of which "poor" and "John" are constituents that is itself a constituent of the construction "poor John," the forms "poor" and "John" may be described not only as constituents but also as immediate constituents of "poor John." Similarly, the phrase "lost his watch" is composed of three word forms, "lost," "his," and "watch," all of which may be described as constituents of the construction. Not all of them, however, are its immediate constituents. The forms “his” and "watch combine to make the intermediate construction "his watch"; it is this intermediate unit that combines with "lost" to form the larger phrase "lost his watch." The immediate constituents of "lost his watch" are "lost" and "his watch; the immediate constituents of "his watch" are the forms "his" and "watch." By the constituent structure of a phrase or sentence is meant the hierarchical organization of the smallest forms of which it is composed (its ultimate constituents) into layers of successively more inclusive units. Viewed in this way, the sentence "Poor John lost his watch" is more than simply a sequence of five word forms associated with a particular intonation pattern. It is analyzable into the immediate constituents "poor John" and "lost his watch," and each of these phrases is analyzable into its own immediate constituents and so on, until, at the last stage of the analysis, the ultimate constituents of the sentence are reached. The constituent structure of the whole sentence is represented by means of a tree diagram in Figure 1.

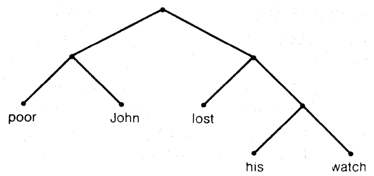


Figure 1: The constituent structure of a simple sentence (see text).

Each form, whether it is simple or composite, belongs to a certain form class. Using arbitrarily selected letters to denote the form classes of English, "poor" may be a member of the form class *A*, "John" of the class B, "lost" of the class *C*, "his" of the class *D*, and "watch" of the class E. Because "poor John" is syntactically equivalent to (*i.e.*, substitutable for) "John," it is to be classified as a member of A. So too, it can be assumed, is "his watch." In the case of "lost his watch" there is a problem. There are very many forms—including "lost," "ate," "stole"—that can occur, as here, in construction with a member of A and can also occur alone; for example, "lost" is substitutable for “stole the money," as "stole" is substitutable for either or for “lost his watch." This being so, one might decide to classify constructions like "lost his watch" as members of C. On the other hand, there are forms that—though they are substitutable for "lost," "ate," "stole," and so on when these forms occur alone—cannot be used in combination with a following member of A (cf. "died," "existed"): and there are forms that, though they may be used in combination with a following member of A, cannot occur alone (*cf.* "enjoyed"). The question is whether one respects the traditional distinction between transitive and intransitive verb forms. It may be decided, then, that "lost," "stole," "ate" and so forth belong to one class, *C* (the class to which "enjoyed" belongs), when they occur "transitively" (*i.e.*, with a following member of A as their object) but to a different class, *F* (the class to which "died" belongs), when they occur "intransitively." Finally, it can be said that the whole sentence “Poor John lost his watch" is a member of the form class *G*. Thus the constituent structure not only of "Poor John lost his watch" but of a whole set of English sentences can be represented by means of the tree diagram given in Figure 2. New sentences of the same type can be constructed by substituting actual forms for the class labels.

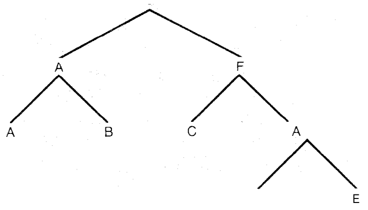


Figure 2: The constituent structure of a class of simple sentences with arbitrary letters used to represent the form class of each constituent (see text).

Endocentric and exocentric constructions

Any construction that belongs to the same form class as at least one of its immediate constituents is described as endocentric; the only endocentric construction in the model sentence above is "poor John." All the other constructions, according to the analysis, are exocentric. This is clear from the fact that in Figure 2 the letters at the nodes above every phrase other than the phrase A + B (*i.e.*, "poor John," "old Harry," and so on) are different from any of the letters at the ends of the lower branches connected directly to these nodes. For example, the phrase D + E (*i.e.*, "his watch," “the money," and so forth) has immediately above it a node labelled A, rather than either *D* or E. Endocentric constructions fall into two types: subordinating and coordinating. If attention is confined, for simplicity,-to constructions composed of no more than two immediate constituents. it can be said that subordinating constructions are those in which only one immediate constituent is of the same form class as the whole construction, whereas coordinating constructions are those in which both constituents are of the same form class as the whole construction. In a subordinating construction (*e.g.*, "poor John"), the constituent that is syntactically equivalent to the whole construction is described as the head, and its partner is described as the modifier: thus, in "poor John," the form "John" is the head, and "poor" is its modifier. An example of a coordinating construction is "men and women," in which, it may be assumed, the immediate constituents are the word “men” and the word "women," each of which is syntactically equivalent to "men and women." (It is here implied that the conjunction "and" is not a constituent, properly so called, but an element that, like the relative order of the constituents, indicates the nature of the construction involved. Not all linguists have held this view.)

Ambiguous constructions

One reason for giving theoretical recognition to the notion of constituent structure is that it helps to account for the ambiguity of certain constructions. A classic example is the phrase "old men and women," which may be interpreted in two different ways according to whether one associates "old" with "men and women" or just with "men." Under the first of the two interpretations, the immediate constituents are "old and "men and women"; under the second, they are "old men" and “women.” The difference in meaning cannot be attributed to any one of the ultimate constituents but results from a difference in the way in which they are associated with one another. Ambiguity of this kind is referred to as syntactic ambiguity. Not all syntactic ambiguity is satisfactorily accounted for in terms of constituent structure.

Semantics. Bloomfield thought that semantics, or the study of meaning, was the weak point in the scientific investigation of language and would necessarily remain so until the other sciences whose task it was to describe the universe and man's place in it had advanced beyond their present state. In his textbook *Language* (1933), he had himself adopted a behaviouristic theory of meaning, defining the meaning of a linguistic form as "the situation in which the speaker utters it and the response which it calls forth in the hearer." Furthermore, he subscribed, in principle at least, to a physicalist thesis, according to which all science should be modelled upon the so-called exact sciences and all scientific knowledge should be reducible, ultimately, to statements made about the properties of the physical world. The reason for his pessimism concerning the prospects for the study of meaning was his feeling that it would be a long time before a complete scientific description of the situations in which utterances were produced and the responses they called forth in their hearers would be available. At the time that Bloomfield was writing, physicalism was more widely held than it is today, and it was perhaps reasonable for him to believe that linguistics should eschew mentalism and concentrate upon the directly observable. As a result, for some 30 years after the publication of Bloomfield's textbook, the study of meaning was almost wholly neglected by his followers; most American linguists who received their training during this period had no knowledge of, still less any interest in, the work being done elsewhere in semantics.

Result of Bloomfield's physicalist tendencies

Two groups of scholars may be seen to have constituted an exception to this generalization: anthropologically minded linguists and linguists concerned with Bible translation. Much of the description of the indigenous languages of America has been carried out since the days of Boas and his most notable pupil Sapir by scholars who were equally proficient both in anthropology and in descriptive linguistics; such scholars have frequently added to their grammatical analyses of languages some discussion of the meaning of the grammatical categories and of the correlations between the structure of the vocabularies and the cultures in which the languages operated. It has already been pointed out that Boas and Sapir and, following them, Whorf were attracted by Humboldt's view of the interdependence of language and culture and of language and thought. This view was quite widely held by American anthropological linguists (athough many of them would not go as far as Whorf in asserting the dependence of thought and conceptualization upon language).

Also of considerable importance in the description of the indigenous languages of America has been the work of linguists trained by the American Bible Society and the Summer Institute of Linguistics, a group of Protestant missionary linguists. Because their principal aim is to produce translations of the Bible, they have necessarily been concerned with meaning as well as with grammar and phonology. This has tempered the otherwise fairly orthodox Bloomfieldian approach characteristic of the group.

The two most important developments evident in recent work in semantics are, first, the application of the structural approach to the study of meaning and, second, a better appreciation of the relationship between grammar and semantics. The second of these developments will be treated in the following section on *Transformational–Generative Grammar*. The first, structural semantics, goes back to the period preceding World War II and is exemplified in a large number of publications, mainly by German scholars—Jost Trier, Leo Weisgerber, and their collaborators.

The structural approach to semantics is best explained by contrasting it with the more traditional "atomistic" approach, according to which the meaning of each word in the language is described, in principle, independently of the meaning of all other words. The structuralist takes the view that the meaning of a word is a function of the relationships it contracts with other words in a particular lexical field, or subsystem, and that it cannot be adequately described except in terms of these relationships. For example, the colour terms in particular languages constitute a lexical field, and the meaning of each term depends upon the place it occupies in the field. Although the denotation of each of the words "green," "blue," and "yellow" in English is somewhat imprecise at the boundaries, the position that each of them occupies relative to the other terms in the system is fixed: "green" is between "blue" and "yellow," so that the phrases "greenish yellow" or "yellowish green" and "bluish green" or "greenish blue" are used to refer to the boundary areas. Knowing the meaning of the word "green" implies knowing what cannot as well as what can be properly described as green (and knowing of the borderline cases that they are borderline cases). Languages differ considerably as to the number of basic colour terms that they recognize, and they draw boundaries within the psychophysical continuum of colour at different places. Blue, green, yellow, and so on do not exist as distinct colours in nature, waiting to be labelled differently, as it were, by different languages; they come into existence, for the speakers of particular languages, by virtue of the fact that those languages impose structure upon the continuum of colour and assign to three of the areas thus recognized the words "blue," "green," "yellow."

Colour terms in different languages

The language of any society is an integral part of the culture of that society, and the meanings recognized within the vocabulary of the language are learned by the child as part of the procers of acquiring the culture of the society in which he is brought up. Many of the structural differences found in the vocabularies of different languages are to be accounted for in terms of cultural differences. This is especially clear in the vocabulary of kinship (to which a considerable amount of attention has been given by anthropologists and linguists), but it holds true of many other semantic fields also. A consequence of the structural differences that exist between the vocabularies of different languages is that, in many instances, it is in principle impossible to translate a sentence "literally" from one language to another.

It is important, nevertheless, not to overemphasize the semantic incommensurability of languages. Presumably, there are many physiological and psychological constraints that, in part at least, determine one's perception and categorization of the world. It may be assumed that, when one is learning the denotation of the more basic words in the vocabulary of one's native language, attention is drawn first to what might be called the naturally salient features of the environment and that one is, to this degree at least, predisposed to identify and group objects in one way rather than another. It may also be that human beings are genetically endowed with rather more specific and linguistically relevant principles of categorization. It is possible that, although languages differ in the number of basic colour categories that they distinguish, there is a limited number of hierarchically ordered basic colour categories from which each language makes its selection and that what counts as a typical instance, or focus, of these universal colour categories is fixed and does not vary from one language to another. If this hypothesis is correct, then it is false to say, as many structural semanticists have said, that languages divide the continuum of colour in a quite arbitrary manner. But the general thesis of structuralism is unaffected, for it still remains true that each language has its own unique semantic structure even though the total structure is, in each case, built upon a substructure of universal distinctions.

TRANSFORMATIONAL–GENERATIVE GRAMMAR

A generative grammar, in the sense in which Noam Chomsky uses the term generative, is a rules system formalized with mathematical precision that generates, with-

Generation of grammatical sentences

out drawing upon any information that is not represented explicitly in the system, the grammatical sentences of the language that it describes, or characterizes, and assigns to each sentence a structural description, or grammatical analysis. All the concepts introduced in this definition of "generative" grammar will be explained and exemplified in the course of this section. Generative grammars fall into several types; this exposition is concerned mainly with the type known as transformational (or, more fully, transformational–generative). Transformational grammar was initiated by Zellig S. Harris in the course of work on what he called discourse analysis (the formal analysis of the structure of continuous text). It was further developed and given a somewhat different theoretical basis by Chomsky.

Harris's grammar. Harris distinguished within the total set of grammatical sentences in a particular language (for example, English) two complementary subsets: kernel sentences (the set of kernel sentences being described as the kernel of the grammar) and nonkernel sentences. The difference between these two subsets lies in nonkernel sentences being derived from kernel sentences by means of transformational rules. For example, "The workers rejected the ultimatum" is a kernel sentence that may be transformed into the nonkernel sentences "The ultimatum was rejected by the workers" or "Did the workers reject the ultimatum?" Each of these may be described as a transform of the kernel sentence from which it is derived. The transformational relationship between corresponding active and passive sentences (*e.g.*, "The workers rejected the ultimatum" and "The ultimatum was rejected by the workers") is conventionally symbolized by the rule $N_1V\ N_2 \rightarrow N_2$ be V+en by N_1, in which N stands for any noun or noun phrase, V for any transitive verb, en for the past participle morpheme, and the arrow (→) instructs one to rewrite the construction to its left as the construction to the right. (There has been some simplification of the rule as it was formulated by Harris.) This rule may be taken as typical of the whole class of transformational rules in Harris's system: it rearranges constituents (what was the first nominal, or noun, N,, in the kernel sentence is moved to the end of the transform, and what was the second nominal, N_2, in the kernel sentence is moved to initial position in the transform), and it adds various elements in specified positions (be, en, and by). Other operations carried out by transformational rules include the deletion of constituents; *e.g.*, the entire phrase "by the workers" is removed from the sentence "The ultimatum was rejected by the workers" by a rule symbolized as N_2 be V+en by $N_1 \rightarrow N_2$ be V+en. This transforms the construction on the left side of the arrow (which resulted from the passive transformation) by dropping the by-phrase, thus producing "The ultimatum was rejected."

Functions of Harris's rules

Chomsky's grammar. Chomsky's system of transformational grammar, though it was developed on the basis of his work with Harris, differs from Harris' in a number of respects. It is Chomsky's system that has attracted the most attention and has received the most extensive exemplification and further development. As outlined in ***Syntactic Structures*** (1957), it comprised three sections, or components: the phrase-structure component, the transformational component, and the morphophonemic component. Each of these components consisted of a set of rules operating upon a certain "input" to yield a certain "output." The notion of phrase structure may be dealt with independently of its incorporation in the larger system. In the following system of rules, S stands for Sentence, NP for Noun Phrase, VP for Verb Phrase, Det for Determiner, Aux for Auxiliary (verb), N for Noun, and V for Verb stem.

(1) S → NP + VP
(2) VP → Verb + NP
(3) NP → Det + N
(4) Verb → Aux + V
(5) Det → *the, a, . . .*
(6) N → *man, ball, . . .*
(7) Aux → *will, can, . . .*
(8) V → *hit, see, . . .*

This is a simple phrase-structure grammar. It generates and thereby defines as grammatical such sentences as "The man will hit the ball," and it assigns to each sentence that it generates a structural description. The kind of structural description assigned by a phrase-structure grammar is, in fact, a constituent structure analysis of the sentence.

In these rules, the arrow can he interpreted as an instruction to rewrite (this is to be taken as a technical term) whatever symbol appears to the left of the arrow as the symbol or string of symbols that appears to the right of the arrow. For example, rule (2) rewrites the symbol VP as the string of symbols Verb + NP, and it thereby defines Verb + NP to be a construction of the type VP. Or, alternatively and equivalently, it says that constructions of the type VP may have as their immediate constituents constructions of the type Verb and NP (combined in that order). Rule (2) can be thought of as creating or being associated with the tree structure in Figure 3.

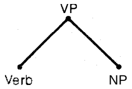


Figure 3: The constituent structure, or phrase structure, assigned by the rule VP→Verb + NP (see text).

Integrated system of rules

Rules (1)–(8) do not operate in isolation but constitute an integrated system. The symbol S (standing mnemonically for "sentence") is designated as the initial symbol. This information is not given in the rules (1)–(8), but it can be assumed either that it is given in a kind of protocol statement preceding the grammatical rules or that there is a universal convention according to which S is always the initial symbol. It is necessary to begin with a rule that has the initial symbol on the left. Thereafter any rule may he applied in any order until no further rule is applicable; in doing so, a derivation can be constructed of one of the sentences generated by the grammar. If the rules are applied in the following order: (1), (2), (3), (3), (4), (5), (5), (6), (6), (7), (8), then assuming that "the" is selected on both applications of (5), "man" on one application of (6), and "ball" on the other, "will" on the application of (7), and "hit" on the application of (8), the following derivation of the sentence "The man will hit the ball" will have been constructed:

(i)	*S*	
(ii)	NP + VP	by rule (1)
(iii)	NP + Verb + NP	by rule (2)
(iv)	Det + N + Verb + NP	by rule (3)
(v)	Det + N + Verb + Det + N	by rule (3)
(vi)	Det + N + Aux + V + Det + N	by rule (4)
(vii)	*the* + N + Aux + V + Det + N	by rule (5)
(viii)	*the* + N + Aux + V + *the* + *N*	by rule (5)
(ix)	*the* + *man* + Aux + V + *the* + N	by rule (6)
(x)	*the* + *man* + Aux + V + *the* + *ball*	by rule (6)
(xi)	*the* + *man* + *will* + V + *the* + *ball*	by rule (7)
(xii)	*the* + *man* + *will* + *hit* + *the* + *ball*	by rule (8)

Many other derivations of this sentence are possible, depending on the order in which the rules are applied. The important point is that all these different derivations are equivalent in that they can be reduced to the same tree diagram; namely, the one shown in Figure 4. If this is compared with the system of rules, it will be seen that each application of each rule creates or is associated with a portion (or subtree) of the tree. The tree diagram, or

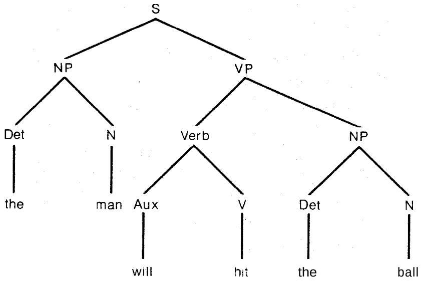


Figure 4: Structural description of the sentence "The man will hit the ball," assigned by the rules of a simple phrase-structure grammar (see text).

phrase marker, may now be considered as a structural description of the sentence "The man hit the ball." It is a description of the constituent structure, or phrase structure, of the sentence, and it is assigned by the rules that generate the sentence.

It is important to interpret the term generate in a static, rather than a dynamic, sense. The statement that the grammar generates a particular sentence means that the sentence is one of the totality of sentences that the grammar defines to be grammatical or well formed. All the sentences are generated, as it were, simultaneously. The notion of generation must be interpreted as would be a mathematical formula containing variables. For example, in evaluating the formula $y^2 + y$ for different values of y, one does not say that the formula itself generates these various resultant values (2, when $y = 1$; 5, when $y = 2$; etc.) one after another or at different times; one says that the formula generates them all simultaneously or, better still perhaps, timelessly. The situation is similar for a generative grammar. Although one sentence rather than another can be derived on some particular occasion by making one choice rather than another at particular places in the grammar, the grammar must be thought of as generating all sentences statically or timelessly.

Phrase-structure and transformational rules

It has been noted that, whereas a phrase-structure grammar is one that consists entirely of phrase-structure rules, a transformational grammar (as formalized by Chomsky) includes both phrase-structure and transformational rules (as well as morphophonemic rules). The transformational rules depend upon the prior application of the phrase-structure rules and have the effect of converting, or transforming, one phrase marker into another. What is meant by this statement may be clarified first with reference to a purely abstract and very simple transformational grammar, in which the letters stand for constituents of a sentence (and S stands for "sentence"):

PS rules
(1) $S \rightarrow A + B$
(2) $B \rightarrow C + D$
(3) $A \rightarrow a + b$
(4) $C \rightarrow c + e + f$
(5) $D \rightarrow d + g + h$
T rules
(6) $A + C + D \rightarrow D + A$

The first five rules are phrase-structure rules (PS rules); rule (6) is a transformational rule (T rule). The output of rules (1)–(5) is the terminal string $a + b + c + e + f + d + g + h$, which has associated with it the structural description indicated by the phrase marker shown in Figure 5 (left). Rule (6) applies to this terminal string of the PS rules and the associated phrase marker. It has the effect of deleting C (and the constituents of C) and permuting A and D (together with their constituents). The result is the string of symbols $d + g + h + a + b$, with the associated phrase marker shown in Figure 5 (right).

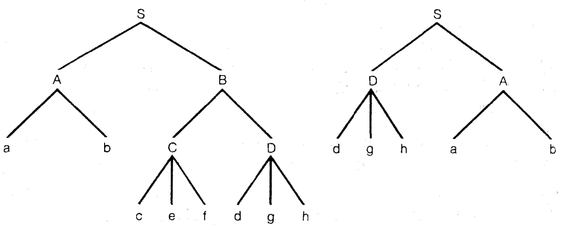


Figure 5: Phrase markers.
(Left) A phrase marker associated with the terminal string of a set of phrase-structure rules. (Right) The phrase marker resulting from the application to the phrase marker shown at left of the transformational rule $A + C + D \rightarrow D + A$ (see text).

The phrase marker shown in Figure 5 (left) may be described as underlying, and the phrase marker shown in Figure 5 (right) as derived with respect to rule (6). One of the principal characteristics of a transformational rule is its transformation of an underlying phrase marker into a derived phrase marker in this way. Transformational rules, in contrast with phrase-structure rules, are also formally more heterogeneous and may have more than one symbol on the left-hand side of the arrow. The linguistic importance of these abstract considerations may be explained with reference to the relationship that holds in English between active and passive sentences.

Chomsky's rule for relating active and passive sentences (as given in ***Syntactic Structures***) is very similar, at first sight, to Harris', discussed above. Chomsky's rule is:

$$\mathrm{NP,} - \mathbf{Aux} - \mathrm{V} - \mathbf{NP_2} \rightarrow \mathbf{NP_2} - \mathbf{Aux} + \mathrm{be} + \mathrm{en} - \mathrm{V} - \mathrm{by} + \mathrm{NP,}$$

Passive transformation

This rule, called the passive transformation, presupposes and depends upon the prior application of a set of phrase-structure rules. For simplicity, the passive transformation may first be considered in relation to the set of terminal strings generated by the phrase-structure rules (1)–(8) given earlier. The string "the + man + will + hit + the + ball" (with its associated phrase marker, as shown in Figure 4) can be treated not as an actual sentence but as the structure underlying both the active sentence "The man will hit the ball" and the corresponding passive "The ball will be hit by the man." The passive transformation is applicable under the condition that the underlying, or "input," string is analyzable in terms of its phrase structure as NP – Aux – V – NP (the use of subscript numerals to distinguish the two NPs in the formulation of the rule is an informal device for indicating the operation of permutation). In the phrase marker in Figure 4 "the" + "man" are constituents of NP, "will" is a constituent of Aux, "hit" is a constituent of V, and "the" + "ball" are constituents of NP. The whole string is therefore analyzable in the appropriate sense, and the passive transformation converts it into the string "the + ball + will + be + en + hit + by + the + man." A subsequent transformational rule will permute "en + hit" to yield "hit + en," and one of the morphophonemic rules will then convert "hit + en" to "hit" (as "ride + en" will be converted to "ridden"; "open + en" to "opened," and so on).

Every transformational rule has the effect of converting an underlying phrase marker into a derived phrase marker. The manner in which the transformational rules assign derived constituent structure to their input strings is one of the major theoretical problems in the formalization of transformational grammar. Here it can be assumed not only that "be + en" is attached to Aux and "by" to NP (as indicated by the plus signs in the rule as it has been formulated above) but also that the rest of the derived structure is as shown in Figure 6. The phrase marker in Figure 6 formalizes the fact, among others, that "the ball" is the subject of the passive sentence "The ball will be hit by the man," whereas "the man" is the subject of the corresponding active "The man will hit the ball" (cf. Figure 4).

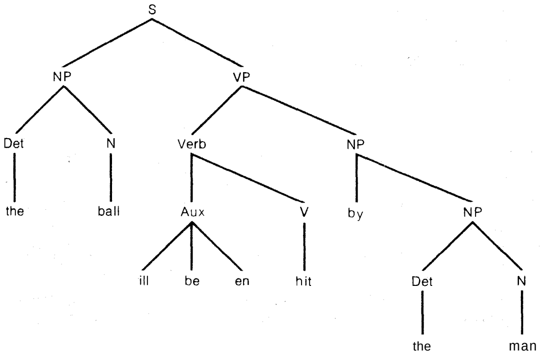


Figure 6: A possible derived phrase marker for a passive sentence (see text).

Although the example above is a very simple one, and only a single transformational rule has been considered independently of other transformational rules in the same system, the passive transformation must operate, not only upon simple noun phrases like "the man" or "the ball," but upon noun phrases that contain adjectives ("the old man"), modifying phrases ("the man in the corner"), relative clauses ("the man who checked in last night"), and so forth. The incorporation, or embedding, of these other structures within the noun phrase will be brought about by the prior application of other transformational rules. It should also be clear that the phrase-structure rules require extension to allow for the various forms of the verb ("is hitting," "hit," "was hitting," "has hit," "has been hitting," etc.) and for the distinction of singular and plural.

Operation of transformations on underlying structures

It is important to note that, unlike Harris', Chomsky's system of transformational grammar does not convert one sentence into another: the transformational rules operate upon the structures underlying sentences and not upon actual sentences. A further point is that even the simplest sentences (*i.e.*, kernel sentences) require the application of at least some transformational rules. Corresponding active and passive sentences, affirmative and negative sentences, declarative and interrogative sentences, and so on are formally related by deriving them from the same underlying terminal string of the phrase-structure component. The difference between kernel sentences and nonkernel sentences in Syntactic Structures (in Chomsky's later system the category of kernel sentences is not given formal recognition at all) resides in the fact that kernel sentences are generated without the application of any optional transformations. Nonkernel sentences require the application of both optional and obligatory transformations, and they differ one from another in that a different selection of optional transformations is made.

Modifications in Chomsky's grammar. Chomsky's system of transformational grammar was substantially modified in 1965. Perhaps the most important modification was the incorporation, within the system, of a semantic component, in addition to the syntactic component and phonological component. (The phonological component may be thought of as replacing the morphophonemic component of Syntactic Structures.) The rules of the syntactic component generate the sentences of the language and assign to each not one but two structural analyses: a deep structure analysis as represented by the underlying phrase marker, and a surface structure analysis, as represented by the final derived phrase marker. The underlying phrase marker is assigned by rules of the base (roughly equivalent to the PS [Phrase-Structure] rules of the earlier system); the derived phrase marker is assigned by the transformational rules. The interrelationship of the four sets of rules is shown diagrammatically in Figure 7. The meaning of the sentence is derived (mainly, if not wholly) from the deep structure by means of the rules of semantic interpretation; the phonetic realization of the sentence is derived from its surface structure by means of the rules of the phonological component. The grammar ("grammar" is now to be understood as covering semantics and phonology, as well as syntax) is thus an integrated system of rules for relating the pronunciation of a sentence to its meaning. The syntax, and more particularly the base, is at the "heart" of the system, as it were: it is the base component (as the arrows in the diagram indicate) that generates the infinite class of structures underlying the well-formed sentences of a language. These structures are then given a semantic and phonetic "interpretation" by the other components.

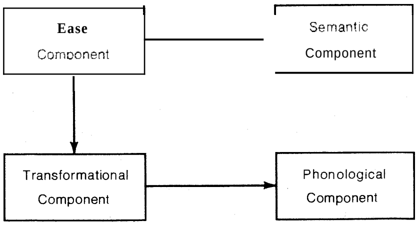


Figure 7: Diagrammatic representation of a transformational grammar (see text).

Base: categorial rules and a lexicon

The base consists of two parts: a set of categorial rules and a lexicon. Taken together, they fulfill a similar function to that fulfilled by the phrase-structure rules of the earlier system. But there are many differences of detail. Among the most important is that the lexicon (which may be thought of as a dictionary of the language cast in a particular form) lists, in principle, all the vocabulary words in the language and associates with each all the syntactic, semantic, and phonological information required for the correct operation of the rules. This information is represented in terms of what are called features. For example, the entry for "boy" might say that it has the syntactic features: [+ Noun], [+ Count], [+ Common], [+ Animate], and [+ Human]. The categorial rules generate a set of phrase markers that have in them, as it were, a number of "slots" to be filled with items from the lexicon. With each such "slot" there is associated a set of features that define the kind of item that can fill the "slot." If a phrase marker is generated with a "slot" for the head of a noun phrase specified as requiring an animate noun (*i.e.*, a noun having the feature [+ Animate]), the item "boy" would be recognized as being compatible with this specification and could be inserted in the "slot" by the rule of lexical substitution. Similarly, it could be inserted in "slots" specified as requiring a common noun, a human noun, or a countable noun, but it would be excluded from positions that require an abstract noun (*e.g.*, "sincerity") or an uncountable noun (*e.g.*, "water"). By drawing upon the syntactic information coded in feature notation in the lexicon, the categorial rules might permit such sentences as "The boy died," while excluding (and thereby defining as ungrammatical) such nonsentences as "The boy elapsed."

Generative semanticists

One further important aspect of generative phonology development of transformational grammar is the relationship between syntax and semantics. Scholars working in the field are now agreed that there is a considerable degree of interdependence between the two, and the problem is how to formalize this interdependence. One school of linguists, called generative semanticists, accept the general principles of transformational grammar but have challenged Chomsky's conception of deep structure as a separate and identifiable level of syntactic representation. In their opinion, the basic component of the grammar should consist of a set of rules for the generation of well-formed semantic representations. These would then be converted by a succession of transformational rules into strings of words with an assigned surface-structure syntactic analysis, there being no place in the passage from semantic representation to surface structure identifiable as Chomsky's deep structure. Chomsky himself has denied that there is any real difference between the two points of view and has maintained that the issue is purely one of notation. That this argument can be put forward by one party to the controversy and rejected by the other is perhaps a sufficient indication of the uncertainty of the evidence. Of greater importance than the overt issues, in so far as they are clear, is the fact that linguists are now studying much more intensively than they have in the past the complexities of the interdependence of syntax, on the one hand, and semantics and logic, on the other. Whether it will prove possible to handle all these complexities within a comprehensive generative grammar remains to be seen.

The role of the phonological component of a generative grammar of the type outlined by Chomsky is to assign a phonetic "interpretation" to the strings of words generated by the syntactic component. These strings of words are represented in a phonological notation (taken from the lexicon) and have been provided with a surface-structure analysis by the transformational rules (see Figure 7). The phonological elements out of which the word forms are composed are segments consisting of what are referred to technically as distinctive features (following the usage of the Prague school, see below). For example, the word form "man," represented phonologically, is composed of three segments: the first consists of the features [+ consonantal], [+ bilabial], [+ nasal], etc.; the second of the features [+ vocalic], [+ front], [+ open], etc.; and the third of the features [+ consonantal], [+ alveolar], [+ nasal], etc. (These features should be taken as purely illustrative; there is some doubt about the definitive list of distinctive features.) Although these segments may be referred to as the "phonemes" /m/, /a/, and /n/, they should not be identified theoretically with units of the kind discussed in the section on *Phonology* under *Structural linguistics.* They are closer to what many American structural linguists called "morphophonemes" or the Prague school linguists labelled "archiphonemes," being unspecified for any feature that is contextually redundant or predictable. For instance, the first segment of the phonological representation of "man" will not include the feature [+ voice]; because nasal consonants are always phonetically voiced in this position in English, the feature [+ voice] can be added to the phonetic specification by a rule of the phonological component.

Dependence of generative phonology on syntax

One further important aspect of generative phonology (*i.e.*, phonology carried out within the framework of an integrated generative grammar) should be mentioned: its dependence upon syntax. Most American structural phonologists made it a point of principle that the phonemic analysis of an utterance should be carried out without regard to its grammatical structure. This principle was controversial among American linguists and was not generally accepted outside America. Not only has the principle been rejected by the generative grammarians, but they have made the phonological description of a language much more dependent upon its syntactic analysis than has any other school of linguists. They have claimed, for example, that the phonological rules that assign different degrees of stress to the vowels in English words and phrases and alter the quality of the relatively unstressed vowel concomitantly must make reference to the derived constituent structure of sentences and not merely to the form class of the individual words or the places in which the word boundaries occur.

TAGMEMICS

The system of tagmemic analysis, as presented by Kenneth L. Pike, was developed for the analysis not only of language but of all of human behaviour that manifests the property of patterning. In the following treatment, only language will be discussed.

Modes of language. Every language is said to be trimodal—*i.e.*, structured in three modes: phonology, grammar, and lexicon. These modes are interrelated but have a considerable degree of independence and must be described in their own terms. Phonology and lexicon should not be seen as mere appendages to grammar, the former simply specifying which phonemes can combine to form morphemes (or morphs), and the latter simply listing the morphemes and other meaningful units with a description of their meaning. There are levels of structure in each of the modes, and the units of one level are not necessarily coterminous with those of another. Phonemes, for example, may combine to form syllables and syllables to form phonological words ("phonological word" is defined as the domain of some phonological process such as accentuation, assimilation, or dissimilation), but the morpheme (or morph) will not necessarily consist of an integral number of syllables, still less of a single syllable. Nor will the word as a grammatical unit necessarily coincide with the phonological word. Similarly, the units of lexical analysis, sometimes referred to as lexemes (in one sense of this term), are not necessarily identifiable as single grammatical units, whether as morphemes, words, or phrases. No priority, then, is ascribed to any one of the three modes.

Tagmeme defined

The originality of tagmemic analysis and the application of the term tagmeme is most clearly manifest in the domain of grammar. By a tagmeme is meant an element of a construction, the element in question being regarded as a composite unit, described in such terms as "slot-filler" or "function-class." For example, one of the tagmemes required for the analysis of English at the syntactic level might be noun-as-subject, in which "noun" refers to a class of substitutable, or paradigmatically related, morphemes or words capable of fulfilling a certain grammatical function, and "subject" refers to the function that may be fulfilled by one or more classes of elements. In the tagmeme noun-as-subject—which, using the customary tagmemic symbolism, may be represented as Subject:noun—the subject slot is filled by a noun. When a particular tagmeme is identified in the analysis of an actual utterance, it is said to be manifested by the particular member of the grammatical class that occurs in the appropriate slot in the utterance. For example, in the utterance "John is asleep," the subject tagmeme is manifested by the noun "John." Tagmemicists insist that tagmemes, despite their bipartite structure, are single units. In grammatical analysis, the distribution of tagmemes, not simply of classes, is stated throughout the sentences of the language. Subject:noun is a different tagmeme from Object:noun, as it is also a different tagmeme from Subject:pronoun.

Hierarchy of levels. Within the grammar of a language there is a hierarchy of levels, units of one level being composed of sequences of units of the level below. In many languages, five such levels are recognized, defined in terms of the following units: morpheme, word, phrase, clause, and sentence. (The term level is being used in a different sense from that in which it was used earlier to refer to phonology and grammar.) The difference between morphology and syntax is simply a difference between two of these five levels, no greater than the difference, for example, between the phrase level and the clause level. Normally, tagmemes at one level are manifested by units belonging to the level below: clause tagmemes by phrases, phrase tagmemes by words, and so on. Intermediate levels may, however, be skipped. For example, the subject tagmeme in a clause may be manifested by a single word in English (*e.g.*, "John," "water") and not necessarily by a phrase ("the young man").

It is also possible for there to be loop-backs in the grammatical hierarchy of a language. This means that a unit of higher level may be embedded within the structure of a unit of lower level; for example, a clause may fill a slot within a phrase (*e.g.*, "who arrived late," in "the man who arrived late").

Tagmemic notation

In regard to the notation of tagmemics, a construction is symbolized as a string of tagmemes (which commonly, though not necessarily, will be sequentially ordered according to the order in which elements manifesting the tagmemes occur in utterances). Each tagmeme is marked as obligatory or optional by having preposed to it a plus sign (+) or a plus-or-minus sign (±), respectively. For example, a formula representing the structure of a clause composed of an obligatory subject tagmeme, an obligatory predicate tagmeme, and an optional object tagmeme might be $\mathrm{Cl} = + \mathrm{S{:}n} + \mathrm{P{:}v} \pm \mathrm{O{:}n}$ (in which Cl stands for a clause of a certain type and n and v stand for the classes of nouns and verbs, respectively). This formula does not represent in any way the fact (if it is a fact) that the predicate tagmeme and object tagmeme together form a unit that is one of the two immediate constituents of the clause. It is one of the characteristic features of tagmemic grammar that it gives much less emphasis to the notion of constituent structure than other American approaches to grammatical analysis.

STRATIFICATIONAL GRAMMAR

Sememic, lexemic, morphemic, and phonemic strata

This system of analysis (whose principal advocate is Sydney M. Lamb, a U.S. linguist) is called stratificational because it is based upon the notion that every language comprises a restricted number of structural layers or strata, hierarchically related in such a way that units or combinations of units on one stratum realize units or combinations of units of the next higher stratum. The number of strata may vary from language to language. Four strata have been recognized for English, and it is probable that all languages may have at least these four: the sememic, the lexemic, the morphemic, and the phonemic strata. The sememic stratal system constitutes the semology of the language; the lexemic and morphemic stratal systems constitute the grammar (in the narrower sense of this term); and the phonemic system constitutes the phonology. In some later stratificational work, the term grammar covers the three higher stratal systems—the sememic, the lexemic, and the morphemic—and is opposed to "phonology." The deep structure of sentences is described on the sememic stratum and the surface structure on the morphemic. In the present account, "grammar" is used in the narrower sense and will be opposed to "semology" as well as "phonology."

The originality of stratificational grammar does not reside in the recognition of these three major components of a linguistic description, which are common to almost all schools of linguistics. The stratificational approach to linguistic description is distinguished from others in that it relates grammar to semology and phonology by means of the same notion of realization that it employs to relate the lexemic and the morphemic stratal systems within the grammatical component. Another distinguishing feature of stratificational grammar, in its later development at least, is its description of linguistic structure in terms of a network of relationships, rather than by means of a system of rules; linguistic units are said to be nothing more than points, or positions, in the relational network.

Technical terminology. Lamb has been very careful to make the terminology of stratificational grammar as consistent and perspicuous as possible; but, in fitting some of the more or less established terms into his own theoretical framework, he has reinterpreted them in a potentially confusing manner. Thus, the same terms have been used in different senses in different versions of the system. For example, "morpheme" in stratificational grammar corresponds neither to the unit to which Bloomfield applied the term (*i.e.*, to a word segment consisting of phonemes) nor to the more abstract grammatical unit that a Bloomfieldian morpheme might be described as representing (*e.g.*, the past-tense morpheme that might be variously represented by such allomorphs as /id/, /t/, /d/, etc.). Lamb describes the morpheme as a unit composed of morphons (roughly equivalent to what other linguists have called morphophonemes) that is related to a combination of one or more compositional units of the stratum above, lexons, by means of the relationship of realization. For example, the word form "hated" realizes (on the morphemic stratum) a combination of two lexons, one of which, the stem, realizes the lexeme HATE and the other, the suffix, realizes the PAST TENSE lexeme; each of these two lexons is realized on the stratum below by a morpheme. Another example brings out more clearly the difference between morphemes (the minimal grammatical elements) and lexemes (the minimal meaningful elements). The word form "understood" realizes a combination of three morphemes UNDER, STAND, and PAST. UNDER and STAND jointly realize the single lexeme UNDERSTAND (whose meaning cannot be described as a function of the meanings of UNDER and STAND), whereas the single PAST morpheme directly realizes the single lexeme PAST TENSE.

Morpheme defined in stratificational terms

The stratificational framework, presented in Lamb's work, consistently separates compositional and realizational units, the former being designated by terms ending in the suffix -on (semon, lexon, morphon, phonon), the latter by terms ending in the suffix *-eme* (sememe, lexeme, morpheme, phoneme). Ons are components or compounds of emes on the same stratum (semons are components of sememes, lexons are composed of lexemes, etc.) and emes realize ons of the stratum above (phonemes realize morphons, morphemes realize lexons, etc.). Each stratum has its own combinatorial pattern specifying the characteristic combinations of elements on that stratum. Syllable structure is specified on the phonemic stratum, the structure of word forms on the morphemic stratum, the structure of phrases on the lexemic stratum, and the structure of clauses and sentences on the sememic stratum. Phonons are roughly equivalent to phonological distinctive features and include such properties or components of phonemes as labial, nasal, and so on. Sememes are roughly equivalent to what other linguists have called semantic components or features and include such aspects of the meaning of the lexeme "man" as "male," "adult," "human," and so forth. Once again, however, compositional function is distinguished from interstratal realizational function, so that no direct equivalence can be established with nonstratificational terminology. In more recent work in stratificational grammar, the notion that emes are composed of ons has been abandoned, and greater emphasis is laid upon the fact that emes are points, or positions, in a relational network; they are connected to other points in the network but have themselves no internal structure.

Interstratal relationships. One of the principal characteristics of the stratificational approach is that it sets out to describe languages without making use of rules that convert one entity into another. (Reference has been made above to the antipathy many linguists have felt towards describing languages in terms of processes.) The stratificationalist would handle the phenomena in terms of the interstratal relationships of realization. Various kinds of interstratal relationships, other than that of one-to-one correspondence may be recognized: diversification, in which one higher unit has alternative realizations; zero realization, in which a higher unit has no overt realization on the lower stratum; neutralization, in which two or more higher units are mapped into the same lower level unit; and so on. All these interstratal one–many or many–one relations are then analyzed in terms of the logical notions of conjunction and disjunction (AND-relations versus OR-relations), of ordering (*x* precedes y in an AND-relationship, *x* is selected in preference to y in an OR-relationship), and the directionality ("upward towards meaning, or "downward" towards sound). Many of the phenomena that are described by other linguists in terms of processes that derive one unit from another can be described elegantly enough in terms of interstratal relationships of this kind.

Types of interstratal relationships

Critics, however, have objected to the proliferation of strata and theoretical constructs in stratificational grammar, arguing that they result from an a priori commitment to the notion of realization and that the only stratal distinction for which there is any independent evidence is the distinction of phonology and grammar. It has been suggested by Lamb that stratificational grammar provides a model for the way in which linguistic information is stored in the brain and activated during the production and reception of speech. But little is known yet about the neurology of language and speech, and it would be premature to draw any firm conclusions about this aspect of stratificational grammar (see below Psycholinguistics).

THE PRAGUE SCHOOL

What is now generally referred to as the Prague school comprises a fairly large group of scholars, mainly European, who, though they may not themselves have been members of the Linguistic Circle of Prague, derived their inspiration from the work of Vilém Mathesius, Nikolay Trubetskoy, Roman Jakobson and other scholars based in Prague in the decade preceding World War II.

Combination of structuralism and functionalism. The most characteristic feature of the Prague school approach is its combination of structuralism with functionalism. The latter term (like "structuralism") has been used in a variety of senses in linguistics. Here it is to be understood as implying an appreciation of the diversity of functions fulfilled by language and a theoretical recognition that the structure of languages is in large part determined by their characteristic functions. Functionalism, taken in this sense, manifests itself in many of the more particular tenets of Prague school doctrine.

One very famous functional analysis of language, which, though it did not originate in Prague, was very influential there, was that of the German psychologist Karl Biihler, who recognized three general kinds of function fulfilled by language: *Darstellungsfunktion*, Kundgabefunktion, and Appelfunktion. These terms may be translated, in the present context, as the cognitive, the expressive, and the conative (or instrumental) functions. The cognitive function of language refers to its employment for the transmission of factual information; by expressive function is meant the indication of the mood or attitude of the speaker (or writer); and by the conative function of language is meant its use for influencing the person one is addressing or for bringing about some practical effect. A number of scholars working in the

Various functions of language

Prague tradition have suggested that these three functions correlate in many languages, at least partly, with the grammatical categories of mood and person. The cognitive function is fulfilled characteristically by 3rd-person nonmodal utterances (*i.e.*, utterances in the indicative mood, making no use of modal verbs); the expressive function by 1st-person utterances in the subjunctive or optative mood; and the conative function by 2nd-person utterances in the imperative. The functional distinction of the cognitive and the expressive aspects of language has also been applied by Prague school linguists in their work on stylistics and literary criticism. One of their key principles is that language is being used poetically or aesthetically when the expressive aspect is predominant, and that it is typical of the expressive function of language that this should be manifest in the form of an utterance and not merely in the meanings of the component words.

Phonemes as distinctive features

Phonological contributions. The Prague school is best known for its work on phonology. Unlike the American phonologists, Trubetskoy and his followers did not take the phoneme to be the minimal unit of analysis. Instead, they defined phonemes as sets of distinctive features. For example, in English, /b/ differs from /p/ in the same way that /d/ differs from /t/ and /g/ from /k/. Just how they differ in terms of their articulation is a complex question. For simplicity, it may be said that there is just one feature, the presence of which distinguishes /b/, /d/, and /g/ from /p/, /t/, and /k/, and that this feature is voicing (vibration of the vocal cords). Similarly, the feature of labiality can be extracted from /p/ and /b/ by comparing them with /t/, /d/, /k/, and /g/; the feature of nasality from /n/ and /m/ by comparing them with /t/ and /d/, on the one hand, and with /p/ and /b/, on the other. Each phoneme, then, is composed of a number of articulatory features and is distinguished by the presence or absence of at least one feature from every other phoneme in the language. The distinctive function of phonemes, which depends upon and supports the principle of the duality of structure, can be related to the cognitive function of language. This distinctive feature analysis of Prague school phonology as developed by Jakobson has become part of the generally accepted framework for generative phonology (see above).

Two other kinds of phonologically relevant function are also recognized by linguists of the Prague school: expressive and demarcative. The former term is employed here in the sense in which it was employed above (*i.e.*, in opposition to "cognitive"); it is characteristic of stress, intonation, and other suprasegmental aspects of language that they are frequently expressive of the mood and attitude of the speaker in this sense. The term demarcative is applied to those elements or features that in particular languages serve to indicate the occurrence of the boundaries of words and phrases and, presumably, make it easier to identify such grammatical units in the stream of speech. There are, for example, many languages in which the set of phonemes that can occur at the beginning of a word differs from the set of phonemes that can occur at the end of a word. These and other devices are described by the Prague school phonologists as having demarcative function: they are boundary signals that reinforce the identity and syntagmatic unity of words and phrases.

Marked and unmarked features

Theory of markedness. The notion of markedness was first developed in Prague school phonology but was subsequently extended to morphology and syntax. When two phonemes are distinguished by the presence or absence of a single distinctive feature. one of them is said to be marked and the other unmarked for the feature in question. For example, /b/ is marked and /p/ unmarked with respect to voicing. Similarly, in morphology, the regular English verb can be said to be marked for past tense (by the suffixation of *-ed*) but to be unmarked in the present (cf. "jumped" versus "jump"). It is often the case that a morphologically unmarked form has a wider range of occurrences and a less definite meaning than a morphologically marked form. It can be argued, for example, that, whereas the past tense form in English (in simple sentences or the main clause of complex sentences) definitely refers to the past, the so-called present tense form is more neutral with respect to temporal reference: it is nonpast in the sense that it fails to mark the time as past, but it does not mark it as present. There is also a more abstract sense of markedness, which is independent of the presence or absence of an overt feature or affix. The words "dog" and "bitch" provide examples of markedness of this kind on the level of vocabulary. Whereas the use of the word "bitch" is restricted to females of the species, "dog" is applicable to both males and females. "Bitch" is the marked and "dog" the unmarked term, and, as is commonly the case, the unmarked term can be neutral or negative according to context (cf. "That dog over there is a bitch" versus "It's not a dog, it's a bitch"). The principle of markedness, understood in this more general or more abstract sense, is now quite widely accepted hy linguists of many different schools, and it is applied at all levels of linguistic analysis.

Recent contributions. Current Prague school work is still characteristically functional in the sense in which this term was interpreted in the pre-World War II period. The most valuable contribution made by the postwar Prague school is probably the distinction of theme and rheme and the notion of "functional sentence perspective" or "communicative dynamism." By the theme of a sentence is meant that part that refers to what is already known or given in the context (sometimes called, by the scholars, the topic or psychological subject); by the rheme, the part that conveys new information (the comment or psychological predicate). It has been pointed out that, in languages with a free word order (such as Czech or Latin), the theme tends to precede the rheme, regardless of whether the theme or the rheme is the grammatical subject and that this principle may still operate, in a more limited way, in languages, like English, with a relatively fixed word order (cf. "That book I haven't seen before"). But other devices may also be used to distinguish theme and rheme. The rheme may be stressed ("Jóhn saw Mary") or made the complement of the verb "to be" in the main clause of what is now commonly called a cleft sentence ("It's Jóhn who saw Mary").

Case grammar

The general principle that has guided research in "functional sentence perspective" is that the syntactic structure of a sentence is in part determined by the communicative function of its various constituents and the way in which they relate to the context of utterance. A somewhat different but related aspect of functionalism in syntax is seen in current work in what is called case grammar. Case grammar is based upon a small set of syntactic functions (agentive, locative, benefactive. instrumental, and so on) that are variously expressed in different languages but that are held to determine the grammatical structure of sentences. Although case grammar does not derive directly from the work of the Prague school, it is very similar in inspiration.

III. Historical (diachronic) linguistics

LINGUISTIC CHANGE

All languages change in the course of time. Written records make it clear that 15th-century English is quite noticeably different from 20th-century English, as is 15th-century French or German from modern French or German. It was the principal achievement of the 19th-century linguists not only to realize more clearly than their predecessors the ubiquity of linguistic change but also to put its scientific investigation on a sound footing by means of the comparative method (see the section on the History of *linguistics: The* 19th century). This will be treated in greater detail in the following section. Here various kinds, or categories, of linguistic change will be listed and exemplified.

Sound change. Since the beginning of the 19th century, when scholars observed that there were a number of systematic correspondences in related words between the sounds of the Germanic languages and the sounds of what were later recognized as other Indo-European languages, particular attention has been paid in diachronic linguistics to changes in the sound systems of languages.

Certain common types of sound change, most notably assimilation and dissimilation, can be explained, at least partially, in terms of syntagmatic, or contextual, conditioning. By assimilation is meant the process by which one sound is made similar in its place or manner of articulation to a neighbouring sound. For example, the word "cupboard" was presumably once pronounced, as the spelling indicates, with the consonant cluster pb in the middle. The p was assimilated to b in manner of articulation (*i.e.*, voicing was maintained throughout the cluster), and subsequently the resultant double consonant bb was simplified. With a single b in the middle and an unstressed second syllable, the word "cupboard," as it is pronounced nowadays, is no longer so evidently a compound of "cup" and "board" as its spelling still shows it to have been. The Italian words *notte* "night" and otto "eight" manifest assimilation of the first consonant to the second consonant of the cluster in place of articulation (cf. Latin *nocte*(*m*), octo). Assimilation is also responsible for the phenomenon referred to as umlaut in the Germanic languages. The high front vowel *i* of suffixes had the effect of fronting and raising preceding back vowels and, in particular, of converting an a sound into an *e* sound. In Modern German this is still a morphologically productive process (cf. *Mann* "man": Manner "men"). In English it has left its mark in such irregular forms as "men" (from **manniz*), "feet" (from **fotiz*), and "length" (from **langþa*).

Umlaut

Dissimilation refers to the process by which one sound becomes different from a neighbouring sound. For example, the word "pilgrim" (French *pèlerin*) derives ultimately from the Latin *peregrinus;* the *l* sound results from dissimilaton of the first r under the influence of the second r. A special case of dissimilation is haplology, in which the second of the two identical or similar syllables is dropped. Examples include the standard modern British pronunciations of "Worcester" and "Gloucester" with two syllables rather than three and the common pronunciation of "library" as if it were written "libry." Both assimilation and dissimilation are commonly subsumed under the principle of "ease of articulation." This is clearly applicable in typical instances of assimilation. It is less obvious how or why a succession of unlike sounds in contiguous syllables should be easier to pronounce than a succession of identical or similar sounds. But a better understanding of this phenomenon, as of other "slips of the tongue," may result from current work in the physiological and neurological aspects of speech production.

Not all sound change is to be accounted for in terms of syntagmatic conditioning. The change of p, t, and k to f, θ (the th sound in "thin"), and h or of b, *d*, g to p, t, and k in early Germanic cannot be explained in these terms. Nor can the so-called Great English Vowel Shift that, in the 15th century, modified the quality of all the long vowels (cf. "profane" : "profanity"; "divine" : "divinity"; and others). Attempts have been made to develop a general theory of sound change, notably by the French linguist André Martinet. But no such theory has yet won universal acceptance, and it is likely that the causes of sound change are multiple.

Phonetic vs. phonological sound change

Sound change is not necessarily phonological; it may be merely phonetic (see above Structural *linguistics*: Phonology). The pronunciation of one or more of the phones realizing a particular phoneme may change slightly without affecting any of the previously existing phonological distinctions; this no doubt happens quite frequently as a language is transmitted from one generation to the next. Two diachronically distinct states of the language would differ in this respect in the same way as two coexistent but geographically or socially distinct accents of the same language might differ. It is only when two previously distinct phonemes are merged or a unitary phoneme splits into two (typically when allophonic variation becomes phonemic) that sound change must definitely be considered as phonological. For example, the sound change of p to *f*, ***t*** to *θ*(*th*), and k to ***h***, on the one hand, and of b to p, *d* to *t*, and g to k, on the other, in early Germanic had the effect of changing the phonological system. The voiceless stops did not become fricatives in all positions; they remained as voiceless stops after s. Consequently, the *p* sound that was preserved after *s* merged with the p that derived by sound change from b. (It is here assumed that the aspirated p sound and the unaspirated p sound are to be regarded as allophones of the same phoneme). Prior to the Germanic sound shift the phoneme to be found at the beginning of the words for "five" or "father" also occurred after *s* in words for "spit" or "spew"; after the change this was no longer the case.

Grammatical change. A language can acquire a grammatical distinction that it did not have previously, as when English developed the progressive ("He is running") in contrast to the simple present ("He runs"). It can also lose a distinction; *e.g.*, modern spoken French has lost the distinction between the simple past (*Il marcha* "he walked") and the perfect (*Il* a *marché* "he has walked"). What was expressed by means of one grammatical device may come to be expressed by means of another. For example, in the older Indo-European languages the syntactic function of the nouns and noun phrases in a sentence was expressed primarily by means of case endings (the subject of the sentence being in the nominative case, the object in the accusative case, and so on); in most of the modern Indo-European languages these functions are expressed by means of word order and the use of prepositions. It is arguable, although it can hardly be said to have been satisfactorily demonstrated yet, that the grammatical changes that take place in a language in the course of time generally leave its deep structure unaffected and tend to modify the ways in which the deeper syntactic functions and distinctions are expressed (whether morphologically, by word order, by the use of prepositions and auxiliary verbs, or otherwise), without affecting the functions and distinctions themselves. Many grammatical changes, especially changes in the morphology of a language, are traditionally accounted for in terms of analogy (see above History of Linguistics).

Semantic change. Towards the end of the 19th century, a French scholar, Michel Bréal, set out to determine the laws that govern changes in the meaning of words. This was the task that dominated semantic research until the 1930s, when scholars began to turn their attention to the synchronic study of meaning. Many systems for the classification of changes of meaning have been proposed, and a variety of explanatory principles have been suggested. So far no "laws" of semantic change comparable to the phonologist's sound laws have been discovered. It seems that changes of meaning can be brought about by a variety of causes. Most important, perhaps, and the factor that has been emphasized particularly by the so-called words-and-things movement in historical semantics is the change undergone in the course of time by the objects or institutions that words denote. For example, the English word "car" goes back through Latin carrus to a Celtic word for a four-wheeled wagon. It now denotes a very different sort of vehicle; confronted with a model of a Celtic wagon in a museum, one would not describe it as a car. Semantic change of this kind is continuous.

Causes of changes in meaning

Some changes in the meaning of words are caused by their habitual use in particular contexts. The word "starve" once meant "to die" (cf. Old English steorfan, German *sterben*); in most dialects of English, it now has the more restricted meaning "to die of hunger," though in the north of England "He was starving" can also mean "He was very cold" (*i.e.*, "dying" of cold, rather than hunger). Similarly, the word "deer" has acquired a more specialized meaning than the meaning "wild animal" that it once bore (cf. German Tier); and "meat," which originally meant food in general (hence, "sweetmeats" and the archaic phrase "meat and drink") now denotes the flesh of an animal treated as food. In all such cases, the narrower meaning has developed from the constant use of the word in a more specialized context, and the contextual presuppositions of the word have in time become part of its meaning.

Borrowing. Languages borrow words freely from one another. Usually this happens when some new object or institution is developed for which the borrowing language has no word of its own. For example, the large number of words denoting financial institutions and operations borrowed from Italian by the other western European languages at the time of the Renaissance testifies to the importance of the Italian bankers in that period. (The word "bank" itself, in this sense, comes through French from the Italian banca). Words now pass from one language to another on a scale that is probably unprecedented, partly because of the enormous number of new inventions that have been made in the 20th century and partly because international communications are now so much more rapid and important. The vocabulary of modern science and technology is very largely international.

THE COMPARATIVE METHOD

Reconstruction of an earlier language

The comparative method in historical linguistics is concerned with the reconstruction of an earlier language or earlier state of a language on the basis of a comparison of related words and expressions in different languages or dialects derived from it. The comparative method was developed in the course of the 19th century for the reconstruction of Proto-Indo-European and was subsequently applied to the study of other language families. It depends upon the principle of regular sound change—a principle that, as explained above, met with violent opposition when it was introduced into linguistics by the Neogrammarians in the 1870s but by the end of the century had become part of what might be fairly described as the orthodox approach to historical linguistics (see above History of *linguistics:* The *19th* century). Changes in the phonological systems of languages through time were accounted for in terms of sound laws.

Grimm's law. The most famous of the sound laws is Grimm's law (though Grimm himself did not use the term law). Some of the correspondences accounted for by Grimm's law are given in Table 1. It will be observed that, when other Indo-European languages, including Latin and Greek, have a voiced unaspirated stop (*b*, d), Gothic has the corresponding voiceless unaspirated stop (p, t) and that, when other Indo-European languages have a voiceless unaspirated stop, Gothic has a voiceless fricative (f, *θ*). The simplest explanation would seem to be that, under the operation of what is now called Grimm's law, in some prehistoric period of Germanic (before the development of a number of distinct Germanic languages), the voiced stops inherited from Proto-Indo-European became voiceless and the voiceless stops became fricatives. The situation with respect to the sounds corresponding to the Germanic voiced stops is more complex. Here there is considerable disagreement between the other languages: Greek has voiceless aspirates (*ph*, th), Sanskrit has voiced aspirates (bh, dh), Latin has voiceless fricatives in word-initial position (f) and voiced stops in medial position (*b*, d), Slavic has voiced stops (*b*, *d*), and so on. The generally accepted hypothesis is that the Proto-Indo-European sounds from which the Germanic voiced stops developed were voiced aspirates and that they are preserved in Sanskrit but were changed in the other Indo-European languages by the loss of either voice or aspiration. (Latin, having lost the voice in initial position, subsequently changed both of the resultant voiceless aspirates into the fricative f, and it lost the aspiration in medial position.) It is easy to see that this hypothesis yields a simpler account of the correspondences than any of the alternatives. It is also in accord with the fact that voiced aspirates are rare in the languages of the world and, unless they are supported by the coexistence in the same language of phonologically distinct voiceless aspirates (as they are in Hindi and other north Indian languages), appear to be inherently unstable.

Table 1: Labial and Dental Stops in the Indo-European Languages

Greek	Latin	Gothic	Sanskrit	Slavic
p	p	f	P	P
b	b	p	b	b
ph	f/b	b	bh	b
t	t	θ	t	t
d	d	t	d	d
th	f/d	d	dh	d

Proto-Indo-European reconstruction. Reconstruction of the Proto-Indo-European labial stops (made with the lips) and dental stops (made with the tip of the tongue touching the teeth) is fairly straightforward. More controversial is the reconstruction of the Proto-Indo-European sounds underlying the correspondences shown in Table 2.

Table 2: Velar and Palatal Stops in the Indo-European Languages

Greek	Latin	Gothic	Sanskrit	Slavic
k	k	h	s	s
g	g	k	j	z
kh	h/g/f	g	h	z
p/t/k	qu	wh	k	k
b/d/g	v/gu	q	g	g
ph/th/kh	f/v/gu	w	gh	g

Controversy concerning Proto-Indo-European velar consonants

According to the most generally accepted hypothesis, there were in Proto-Indo-European at least two distinct series of velar (or "guttural") consonants: simple velars (or palatals), symbolized as *k, *g, and *gh, and labiovelars, symbolized as *k^w, *g^w, and *g^wh. The labiovelars may be thought of as velar stops articulated with simultaneous lip-rounding. In one group of languages, the labial component is assumed to have been lost, in another group the velar component; and it is only in the Latin reflex of the voiceless *k^w that both labiality and velarity are retained (cf. Latin quis from *k^wi-). It is notable that the languages that have a velar for the Proto-Indo-European labiovelar stops (*e.g.*, Sanskrit and Slavic) have a sibilant or palatal sound (s or 5) for the Proto-Indo-European simple velars. Earlier scholars attached great significance to this fact and thought that it represented a fundamental division of the Indo-European family into a western and an eastern group. The western group—comprising Celtic, Germanic, Italic, and Greek—is commonly referred to as the centum group; the eastern group—comprising Sanskrit, Iranian, Slavic, and others—is called the satem (*satam*) group. (The words centum and satem come from Latin and Iranian, respectively, and mean "hundred." They exemplify, with their initial consonant, the two different treatments of the Proto-Indo-European simple velar~.) Nowadays less importance is attached to the centum–satem distinction. But it is still generally held that in an early period of Indo-European, there was a sound law operative in the dialect or dialects from which Sanskrit, Iranian, Slavic and the other so-called satem languages developed that had the effect of palatalizing the original Proto-Indo-European velars and eventually converting them to sibilants.

Steps in the comparative method. The information given in the previous paragraphs is intended to illustrate what is meant by a sound law and to indicate the kind of considerations that are taken into account in the application of the comparative method. The first step is to find sets of cognate or putatively cognate forms in the languages or dialects being compared: for example, Latin *decem* = Greek deka = Sanskrit *daśa* = Gothic taihun, all meaning "ten." From sets of cognate forms such as these, sets of phonological correspondences can be extracted; *e.g.*, (1) Latin *d* = Greek *d* = Sanskrit d = Gothic t; (2) Latin *e* = Greek e = Sanskrit a = Gothic ai (in the Gothic orthography this represents an *e* sound); (3) Latin c (*i.e.*, a *k* sound) = Greek k = Sanskrit *ś* = Gothic *h*; (4) Latin em = Greek a = Sanskrit *a* = Gothic *un*. A set of "reconstructed" phonemes can be postulated (marked with an asterisk by the standard convention) to which the phonemes in the attested languages can be systematically related by means of sound laws. The reconstructed Proto-Indo-European word for "ten" is ***dekm***. From this form the Latin word can be derived by means of a single sound change, *m changes to em (usually symbolized as *m > em); the Greek by means of the sound change *m > u (*i.e.*, vocalization of the

Phonological correspondences in cognate forms syllabic nasal and loss of nasality); the Sanskrit by means of the palatalizing sound law, **k* > *ś* and the sound change **m* > *a* (whether this is assumed to be independent of the law operative in Greek or not); and the Gothic by means of Grimm's law (**d* > *t*, **k* > *h*) and the sound change **m* > *un*.

Most 19th-century linguists took it for granted that they were reconstructing the actual word forms of some earlier language, that **dekm*, for example, was a pronounceable Proto-Indo-European word. Many of their successors have been more skeptical about the phonetic reality of reconstructed starred forms like **dekm*. They have said that they are no more than formulae summarizing the correspondences observed to hold between attested forms in particular languages and that they are, in principle, unpronounceable. From this point of view, it would be a matter of arbitrary decision which letter is used to refer to the correspondences: Latin d = Greek d = Sanskrit *d* = Gothic *t*, and so on. Any symbol would do, provided that a distinct symbol is used for each distinct set of correspondences. The difficulty with this view of reconstruction is that it seems to deny the very raison d'être of historical and comparative linguistics. Linguists want to know, if possible, not only that Latin *decem*, Greek *deka*, and so on are related, but also the nature of their historical relationshiphow they have developed from common ancestral form. They also wish to construct, if feasible, some general theory of sound change. This can be done only if some kind of phonetic interpretation can be given to the starred forms. The important point is that the confidence with which a phonetic interpretation is assigned to the phonemes that are reconstructed will vary from one phoneme to another. It should be clear from the discussion above, for example, that the interpretation of **d* as a voiced dental or alveolar stop is more certain than the interpretation of **k* as a voiceless velar stop. The starred forms are not all on an equal footing from a phonetic point of view.

Model of the family tree

Criticisms of the comparative method. One of the criticisms directed against the comparative method is that it is based upon a misleading genealogical metaphor. In the mid-19th century, the German linguist August Schleicher introduced into comparative linguistics the model of the "family tree." There is obviously no point in time at which it can be said that new languages are "born" of a common parent language. Nor is it normally the case that the parent language "lives on" for a while, relatively unchanged, and then "dies." It is easy enough to recognize the inappropriateness of these biological expressions. No less misleading, however, is the assumption that languages descended from the same parent language will necessarily diverge, never to converge again, through time. This assumption is built into the comparative method as it is traditionally applied. And yet there are many clear cases of convergence in the development of well-documented languages. The dialects of England are fast disappearing and are far more similar in grammar and vocabulary today than they were even a generation ago. They have been strongly influenced by the standard language. The same phenomenon, the replacement of nonstandard or less prestigious forms with forms borrowed from the standard language or dialect, has taken place in many different places at many different times. It would seem, therefore, that one must reckon with both divergence and convergence in the diachronic development of languages: divergence when contact between two speech communities is reduced or broken and convergence when the two speech communities remain in contact and when one is politically or culturally dominant.

The comparative method presupposes linguistically uniform speech communities and independent development after sudden, sharp cleavage. Critics of the comparative method have pointed out that this situation does not generally hold. In 1872 a German scholar, Johannes Schmidt, criticized the family-tree theory and proposed instead what is referred to as the wave theory, according to which different linguistic changes will spread, like waves, from a politically, commercially, or culturally important centre along the main lines of communication, but successive innovations will not necessarily cover exactly the same area. Consequently, there will be no sharp distinction between contiguous dialects, but, in general, the further apart two speech communities are, the more linguistic features there will be that distinguish them (see below *Dialectology and linguistic geography*).

Use of irregular forms for reconstruction

Internal reconstruction. The comparative method is used to reconstruct earlier forms of a language by drawing upon the evidence provided by other related languages. It may be supplemented by what is called the method of internal reconstruction. This is based upon the existence of anomalous or irregular patterns of formation and the assumption that they must have developed, usually by sound change, from earlier regular patterns. For example, the existence of such patterns in early Latin as *honos* : *honoris* ("honor" : "of honor") and others in contrast with *orator* : *oratoris* ("orator" : "of the orator") and others might lead to the supposition that *honoris* developed from an earlier **honosis*. In this case, the evidence of other languages shows that **s* became *r* between vowels in an earlier period of Latin. But it would have been possible to reconstruct the earlier intervocalic **s* with a fair degree of confidence on the basis of the internal evidence alone. Clearly, internal reconstruction depends upon the structural approach to linguistics.

The most recent development in the field of historical and comparative linguistics has come from the theory of generative grammar (see above *Transformational-generative grammar*). If the grammar and phonology of a language are described from a synchronic point of view as an integrated system of rules, then the grammatical and phonological similarities and differences between two closely related languages, or dialects, or between two diachronically distinct states of the same language can be described in terms of the similarities and differences in two descriptive rule systems. One system may contain a rule that the other lacks (or may restrict its application more or less narrowly); one system may differ from the other in that the same set of rules will apply in a different order in the one system from the order in which they apply in the other. Language change may thus be accounted for in terms of changes introduced into the underlying system of phonological and grammatical rules (including the addition, loss, or reordering of rules) during the process of language acquisition. So far these principles have been applied principally to sound change. There has also been a little work done on diachronic syntax.

LANGUAGE CLASSIFICATION

There are two kinds of classification of languages practiced in linguistics: genetic (or genealogical) and typological. The purpose of genetic classification is to group languages into families according to their degree of diachronic relatedness. For example, within the Indo-European family, such subfamilies as Germanic or Celtic are recognized; these subfamilies comprise German, English, Dutch, Swedish, Norwegian, Danish, and others, on the one hand, and Irish, Welsh, Breton, and others, on the other. So far, most of the languages of the world have been grouped only tentatively into families, and many of the classificatory schemes that have been proposed will no doubt be radically revised as further progress is made.

Typological classifications

A typological classification groups languages into types according to their structural characteristics. The most famous typological classification is probably that of isolating, agglutinating, and inflecting (or fusional) languages, which was frequently invoked in the 19th century in support of an evolutionary theory of language development. Roughly speaking, an isolating language is one in which all the words are morphologically unanalyzable (*i.e.*, in which each word is composed of a single morph); Chinese and, even more strikingly, Vietnamese are highly isolating. An agglutinating language (*e.g.*, Turkish) is one in which the word forms can be segmented into morphs, each of which represents a single grammatical category. An inflecting language is one in which there is no one-to-one correspondence between particular word segments and particular grammatical categories.

The older Indo-European languages tend to be inflecting in this sense. For example, the Latin suffix -is represents the combination of categories "singular" and "genitive" in the word from *hominis* "of the man," but one part of the suffix cannot he assigned to "singular" and another to "genitive," and *-is* is only one of many suffixes that in different classes (or declensions) of words represent the combination of "singular" and "genitive."

There is, in principle, no limit to the variety of ways in which languages can be grouped typologically. One can distinguish languages with a relatively rich phonemic inventory from languages with a relatively poor phonemic inventory, languages with a high ratio of consonants to vowels from languages with a low ratio of consonants to vowels, languages with a fixed word order from languages with a free word order, prefixing languages from suffixing languages, and so on. The problem lies in deciding what significance should be attached to particular typological characteristics. Although there is, not surprisingly, a tendency for genetically related languages to be typologically similar in many ways, typological similarity of itself is no proof of genetic relationship. Nor does it appear true that languages of a particular type will be associated with cultures of a particular type or at a certain stage of development. What has emerged from recent work in typology is that certain logically unconnected features tend to occur together, so that the presence of feature A in a given language will tend to imply the presence of feature B. The discovery of unexpected implications of this kind calls for an explanation and gives a stimulus to research in many branches of linguistics.

IV. Linguistics and other disciplines

PSYCHOLINGUISTICS

The term psycholinguistics was coined in the 1940s and came into more general use after the publication of Charles E. Osgood and Thomas A. Sebeok's *Psycholinguistics: A Survey of Theory and Research Problems* (1954), which reported the proceedings of a seminar sponsored in the United States by the Social Science Research Council's Committee on Linguistics and Psychology.

The boundary between linguistics (in the narrower sense of the term: see the introduction of this article) and psycholinguistics is difficult, perhaps impossible, to draw. So too is the boundary between psycholinguistics and psychology. What characterizes psycholinguistics as it is practiced today as a more or less distinguishable field of research is its concentration upon a certain set of topics connected with language and its bringing to bear upon them the findings and theoretical principles of both linguistics and psychology. The range of topics that would be generally held to fall within the field of psycholinguistics nowadays is rather narrower, however, than that covered in the survey by Osgood and Sebeok.

Language learning by children. One of the topics most central to psycholinguistic research is the acquisition of language by children. The term acquisition is preferred to "learning," because "learning" tends to be used by psychologists in a narrowly technical sense, and many psycholinguists believe that no psychological theory of learning, as currently formulated, is capable of accounting for the process whereby children, in a relatively short time, come to achieve a fluent control of their native language. Since the beginning of the 1960s, research on language acquisition has been strongly influenced by Chomsky's theory of generative grammar, and the main problem to which it has addressed itself has been how it is possible for young childen to infer the grammatical rules underlying the speech they hear and then to use these rules for the construction of utterances that they have never heard before. It is Chomsky's conviction, shared by a number of psycholinguists, that children are born with a knowledge of the formal principles that determine the grammatical structure of all languages, and that it is this innate knowledge that explains the success and speed of language acquisition. Others have argued that it is not grammatical competence as such that is innate but more general cognitive principles and that the application of these to language utterances in particular situations ultimately yields grammatical competence. Many recent works have stressed that all children go through the same stages of language development regardless of the language they are acquiring. It has also been asserted that the same basic semantic categories and grammatical functions can be found in the earliest speech of children in a number of different languages operating in quite different cultures in various parts of the world.

Theories of language acquisition

Although Chomsky was careful to stress in his earliest writings that generative grammar does not provide a model for the production or reception of language utterances, there has been a good deal of psycholinguistic research directed toward validating the psychological reality of the units and processes postulated by generative grammarians in their descriptions of languages. Experimental work in the early 1960s appeared to show that nonkernel sentences took longer to process than kernel sentences and, even more interestingly, that the processing time increased proportionately with the number of optional transformations involved. More recent work has cast doubt on these findings, and most psycholinguists are now more cautious about using grammars produced by linguists as models of language processing. Nevertheless, generative grammar continues to be a valuable source of psycholinguistic experimentation, and the formal properties of language, discovered or more adequately discussed by generative grammarians than they have been by others, are generally recognized to have important implications for the investigation of short-term and long-term memory and perceptual strategies.

Speech perception. Another important area of psycholinguistic research that has been strongly influenced by recent theoretical advances in linguistics and, more especially, by the development of generative grammar is speech perception. It has long been realized that the identification of speech sounds and of the word forms composed of them depends upon the context in which they occur and upon the hearer's having mastered, usually as a child, the appropriate phonological and grammatical system. Throughout the 1950s, work on speech perception was dominated (as was psycholinguistics in general) by information theory, according to which the occurrence of each sound in a word and each word in an utterance is statistically determined by the preceding sounds and words. Information theory is no longer as generally accepted as it was a few years ago, and more recent research has shown that in speech perception the cues provided by the acoustic input are interpreted, unconsciously and very rapidly, with reference not only to the phonological structure of the language but also to the more abstract levels of grammatical organization.

Information theory and speech perception

Other areas of research. Other areas of psycholinguistics that should be briefly mentioned are the study of aphasia and neurolinguistics. The term aphasia is used to refer to various kinds of language disorders; recent work has sought to relate these, on the one hand, to particular kinds of brain injury and, on the other, to psychological theories of the storage and processing of different kinds of linguistic information. One linguist has put forward the theory that the most basic distinctions in language are those that are acquired first by children and are subsequently most resistant to disruption and loss in aphasia. This, though not disproved, is still regarded as controversial. Two kinds of aphasia are commonly distinguished. In motor aphasia the patient manifests difficulty in the articulation of speech or in writing and may produce utterances with a simplified grammatical structure, but his comprehension is not affected. In sensory aphasia the patient's fluency may be unaffected, but his comprehension will be impaired and his utterances will often be incoherent.

Neurolinguistics should perhaps be regarded as an independent field of research rather than as part of psycholinguistics. In 1864 it was shown that motor aphasia is produced by lesions in the third frontal convolution of the left hemisphere of the brain. Shortly after the connection had been established between motor aphasia and damage

Speech area in the brain

to this area (known as Broca's area), the source of sensory aphasia was localized in lesions of the posterior part of the left temporal lobe. More recent work has confirmed these findings. The technique of electrically stimulating the cortex in conscious patients has enabled brain surgeons to induce temporary aphasia and so to identify a "speech area" in the brain. It is no longer generally believed that there are highly specialized "centres" within the speech area, each with its own particular function; but the existence of such a speech area in the dominant hemisphere of the brain (which for most people is the left hemisphere) seems to be well established. The posterior part of this area is involved more in the comprehension of speech and the construction of grammatically and semantically coherent utterances, and the anterior part is concerned with the articulation of speech and with writing. Little is yet known about the operation of the neurological mechanisms underlying the storage and processing of language. (See also the articles entitled PERCEPTION; SPEECH, PHYSIOLOGY OF.)

SOCIOLINGUISTICS

Delineation of the field. Just as it is difficult to draw the boundary between linguistics and psycholinguistics and between psychology and psycholinguistics, so it is difficult to distinguish sharply between linguistics and sociolinguistics and between sociolinguistics and sociology. There is the further difficulty that, because the boundary between sociology and anthropology is also unclear, sociolinguistics merges with—and, some would say, subsumes—anthropological linguistics (see below).

It is frequently suggested that there is a conflict between the sociolinguistic and the psycholinguistic approach to the study of language, and it is certainly the case that two distinct points of view are discernible in the literature at the present time. Chomsky has described linguistics as a branch of cognitive psychology, and neither he nor most of his followers have yet shown much interest in the relationship between language and its social and cultural matrix. On the other hand, many modern schools of linguistics that have been very much concerned with the role of language in society would tend to relate linguistics more closely to sociology and anthropology than to any other discipline. It would seem that the opposition between the psycholinguistic and the sociolinguistic viewpoint must ultimately be transcended. The acquisition of language, a topic of central concern to psycholinguists, is in part dependent upon and in part itself determines the process of socialization; and the ability to use one's native language correctly in the numerous socially prescribed situations of daily life is as characteristic a feature of linguistic competence, in the broad sense of this term, as is the ability to produce grammatical utterances. Some of the most recent work in sociolinguistics and psycholinguistics has sought to widen the notion of linguistic competence in this way. So far, however, sociolinguistics and psycholinguistics tend to be regarded as relatively independent areas of research.

Language and socialization

Social dimensions. Language is probably the most important instrument of socialization that exists in all human societies and cultures. It is largely by means of language that one generation passes on to the next its myths, laws, customs, and beliefs, and it is largely by means of language that the child comes to appreciate the structure of the society into which he is born and his own place in that society.

As a social force, language serves both to strengthen the links that bind the members of the same group and to differentiate the members of one group from those of another. In many countries there are social dialects as well as regional dialects, so that it is possible to tell from a person's speech not only where he comes from but what class he belongs to. In some instances social dialects can transcend regional dialects. This is notable in England, where standard English in the so-called Received Pronunciation (RP) can be heard from members of the upper class and upper middle class in all parts of the country. The example of England is but an extreme manifestation of a tendency that is found in all countries: there is less regional variation in the speech of the higher than in that of the lower socioeconomic classes. In Britain and the United States and in most of the other English-speaking countries, people will almost always use the same dialect, regional or social, however formal or informal the situation and regardless of whether their listeners speak the same dialect or not. (Relatively minor adjustments of vocabulary may, however, be made: an Englishman speaking to an American may employ the word "elevator" rather than "lift" and so on.) In many communities throughout the world, it is common for members to speak two or more different dialects and to use one dialect rather than another in particular social situations. This is commonly referred to as code-switching. Code-switching may operate between two distinct languages (*e.g.*, Spanish and English among Puerto Ricans in New York) as well between two dialects of the same language. The term diglossia (rather than bilingualism) is frequently used by sociolinguists to refer to this by no means uncommon phenomenon.

Variables influencing language usage

In every situation, what one says and how one says it depends upon the nature of that situation, the social role being played at the time, one's status vis-à-vis that of the person addressed, one's attitude towards him, and so on. Language interacts with nonverbal behaviour in social situations and serves to clarify and reinforce the various roles and relationships important in a particular culture. Sociolinguistics is far from having satisfactorily analyzed or even identified all the factors involved in the selection of one language feature rather than another in particular situations. Among those that have been discussed in relation to various languages are: the formality or informality of the situation; power and solidarity relationships between the participants; differences of sex, age, occupation, socioeconomic class, and educational background; and personal or transactional situations. Terms such as style and register (as well as a variety of others) are employed by many linguists to refer to the socially relevant dimensions of phonological, grammatical, and lexical variation within one language. So far there is very little agreement as to the precise application of such terms. (For further treatment of sociolinguistics, see the article DIALECTS.)

OTHER RELATIONSHIPS

Anthropological linguistics. The fundamental concern of anthropological linguistics is to investigate the relationship between language and culture. To what extent the structure of a particular language is determined by or determines the form and content of the culture with which it is associated remains a controversial question. Vocabulary differences between languages correlate obviously enough with cultural differences, but even here the interdependence of language and culture is not so strong that one can argue from the presence or absence of a corresponding cultural difference. For example, from the fact that English—unlike French, German, Russian, and many other languages—distinguishes lexically between monkeys and apes, one cannot conclude that there is an associated difference in the cultural significance attached to these animals by English-speaking societies. Some of the major grammatical distinctions in certain languages may have originated in culturally important categories (*e.g.*, the distinction between an animate and an inanimate gender). But they seem to endure independently of any continuing cultural significance. The "Whorfian hypothesis" (the thesis that one's thought and even perception are determined by the language one happens to speak), in its strong form at least, is no longer debated as vigorously as it was a few years ago. Anthropologists continue to draw upon linguistics for the assistance it can give them in the analysis of such topics as the structure of kinship. A more recent development, but one that has not so far produced any very substantial results, is the application of notions derived from generative grammar to the analysis of ritual and other kinds of culturally prescribed behaviour (see also LANGUAGE).

The "Whorfian hypothesis"

Dialectology and linguistic geography. Dialectology, the scientific investigation of regional dialects, was stimulated towards the end of the 19th century by the desire to answer some of the questions posed by the findings of historical and comparative linguistics. It was expected and frequently found to be true that older forms that had disappeared from the standard language as a result of analogical levelling or borrowing were preserved in one or another of the local dialects. It was soon discovered, however, that nonstandard regional dialects could not be assumed to be invariably more conservative than the standard language in this respect or more impervious to borrowing. The investigation of regional dialects may be carried out from two rather different points of view. The speech of a local community may be analyzed without reference to the standard language or neighbouring dialects to produce a more or less complete phonological and grammatical description. Various dialects of the major European languages have been described in this way. Alternatively, a sample of the local dialects in a particular area may be systematically compared in order to discover in what respects they resemble or differ from one another. The latter kind of investigation is commonly called dialect geography; its results are usually published in the form of a dialect atlas that shows the distribution of particular phonological or grammatical features or particular words. Dialect geography has confirmed the wave theory (see above Historical [diachronic] Linguistics) and has demonstrated the impossibility of drawing a sharp distinction between contiguous dialects. Every feature has its own distribution. If a line, or isogloss, is drawn on the map between the localities that differ with respect to each of a set of selected features, the result will be a network of isoglosses; the fewer isoglosses that separate any two localities, the more similar are the dialects spoken in the two localities (see also the article DIALECTS).

Use of computers in linguistic research

Computational linguistics. By computational linguistics is meant no more than the use of electronic digital computers in linguistic research. At a theoretically trivial level, computers are employed to scan texts and to produce, more rapidly and more reliably than was possible in the past, such valuable aids to linguistic and stylistic research as word lists, frequency counts, and concordances. Theoretically more interesting, though much more difficult, is the automatic grammatical analysis of texts by computer. Considerable progress was made in this area by research groups working on machine translation and information retrieval in the United States, Great Britain, the Soviet Union, France, and a few other countries between the mid-1950s and the mid-1960s. But much of the original impetus for this work has disappeared, in part because of the realization that the theoretical problems involved in machine translation are much more difficult than they were at first thought to be and in part as a consequence of a loss of interest among linguists in the development of discovery procedures. Whether automatic syntactic analysis and fully automatic high-quality machine translation are even feasible in principle is a controversial question.

Mathematical linguistics. What is commonly referred to as mathematical linguistics comprises two areas of research: the study of the statistical structure of texts and the construction of mathematical models of the phonological and grammatical structure of languages. These two branches of mathematical linguistics, which may be termed statistical and algebraic linguistics, respectively, are typically distinct. Attempts have been made to derive the grammatical rules of languages from the statistical structure of texts written in those languages, but such attempts are generally thought not only to have been unsuccessful so far in practice but also to be in principle doomed to failure. That languages have a statistical structure is a fact well known to cryptographers. Within linguistics it is of considerable typological interest to compare languages from a statistical point of view (the ratio of consonants to vowels, of nouns to verbs, and so on). Statistical considerations are also of value in stylistics (see below).

Algebraic linguistics

Algebraic linguistics derives principally from the work of Noam Chomsky in the field of generative grammar (see above *Chomsky's* grammar). In his earliest work Chomsky described three different models of grammar—finite-state, phrase-structure, and transformational—and compared them in terms of their capacity to generate all and only the sentences of natural languages and, in doing so, to reflect in an intuitively satisfying manner the underlying formal principles and processes. Other models have also been investigated, and it has been shown that certain different models are equivalent in generative power to phrase-structure grammars. The problem is to construct a model that has all the formal properties required to handle the processes found to be operative in languages but that prohibits rules that are not required for linguistic description. It is an open question whether such a model, or one that approximates more closely to this ideal than current models do, will be a transformational grammar or a grammar of some radically different character.

Stylistics. The term stylistics is employed in a variety of senses by different linguists. In its widest interpretation it is understood to deal with every kind of synchronic variation in language other than what can be ascribed to differences of regional dialect. At its narrowest interpretation it refers to the linguistic analysis of literary texts. One of the aims of stylistics in this sense is to identify those features of a text that give it its individual stamp and mark it as the work of a particular author. Another is to identify the linguistic features of the text that produce a certain aesthetic response in the reader. The aims of stylistics are the traditional aims of literary criticism. What distinguishes stylistics as a branch of linguistics (for those who regard it as such) is the fact that it draws upon the methodological and theoretical principles of modern linguistics.

Philosophy and linguistics. The analysis of language has always been a subject of particular concern to philosophers, and traditional grammar was strongly influenced by the dominant philosophical attitudes of the day. Modern linguistics and modern philosophical theories have so far had little influence on one another. Some philosophers have shown an interest in Chomsky's controversial suggestion that work in generative grammar lends support to the rationalists in their long-standing dispute about the source of human knowledge. Potentially more fruitful, perhaps, is the interest shown by a number of linguists in philosophical treatments of reference, quantification, and presupposition, in systems of modal logic, and in the work of the so-called philosophers of ordinary language. (For a more detailed treatment of philosophy and linguistics, see SEMANTICS.)

Language teaching

Applied linguistics. In the sense in which the term applied linguistics is most commonly used nowadays it is restricted to the application of linguistics to language teaching. Much of the recent expansion of linguistics as a subject of teaching and research in the universities in many countries has come about because of its value, actual and potential, for writing better language textbooks and devising more efficient methods of teaching languages. Linguistics is also widely held to be relevant to the training of teachers of the deaf and speech therapists. Outside the field of education in the narrower sense, applied linguistics (and, more particularly, applied sociolinguistics) has an important part to play in what is called language planning; *i.e.*, in advising governments, especially in recently created states, which language or dialect should be made the official language of the country and how it should be standardized.

BIBLIOGRAPHY. ROBERT H. ROBINS, ***A Short History of Linguistics***, 2nd ed. (1979); ***General Linguistics: An Introductory Survey***, 3rd ed. (1980), a comprehensive and balanced treatment of the entire field; LEONARD BLOOMFIELD, ***Language*** (1933, reprinted 1963), a classic introduction to the subject and essential reading for an understanding of subsequent American work; CHARLES F. HOCKETT, ***A Course in Modern Linguistics*** (1958), a comprehensive, stimulating, though somewhat personal textbook, representative of the post-Bloomfieldian period in the United States; JOHN LYONS, ***Language and Linguistics: An Introduction*** (1981), a more recent textbook covering theoretical developments of the 1960s and 1970s, and (ed.), ***New Horizons in Linguistics*** (1970), contains previously unpublished chapters on developments in most areas of linguistics; MARTIN JOOS (ed.), ***Readings in Linguistics*** (1957), an excellent selection of key articles on structuralism

in the post-Bloomfieldian period; Z.S. HARRIS, *Methods in Structural Linguistics* (1951), the most extreme and consistent expression of the distributional approach to linguistic analysis—important for the development of generative grammar; NOAM CHOMSKY, *Syntactic Structures* (1957), the first generally accessible and relatively nontechnical treatment of generative grammar, widely recognized as one of the most revolutionary books on language to appear in the 20th century; J.P.B. ALLEN and S. PIT CORDER (eds.), *The Edinburgh Course in Applied Linguistics.* 4 vol. (1973–77), a collection of readings covering a wide range of views; STEPHEN ULLMANN, *Semantics: An Introduction to the Science of Meaning* (1962); J.A. FODOR, T.G. BEVER, and M.F. GARRETT, *The Psychology of Language* (1974), and JOSEPH F. KESS, *Psycholinguistics* (1976), two works in psycholinguistics; DELL HYMES (ed.), *Language in Culture and Society* (1964), an excellent selection of articles in sociolinguistics and anthropological linguistics; and M.A.K. HALLIDAY, *Language as a Social Semiotic: The Interpretation of Language and Meaning* (1978), a collection of essays covering various aspects of this theory.

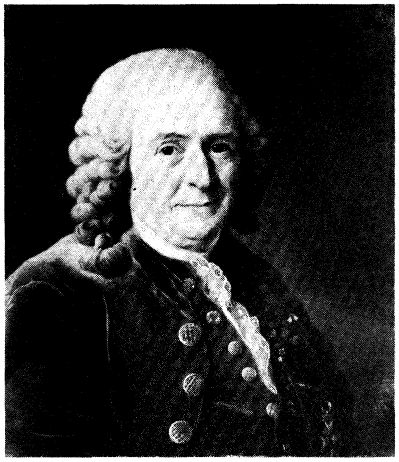
Periodicals: Language (quarterly), *Word* (3/yr.), *International Journal of American Linguistics* (quarterly) (United States); *Philological Society Transactions* (semiannual), *Journal of Linguistics* (semiannual) (Great Britain); *Lingua* (monthly), *Studies in Language* (3/yr.) (Holland); *Bulletin de la Société de Linguisrique de Paris* (semiannual) (France).

(J.Lyo.)

Linnaeus

Carolus Linnaeus (the Latinized form of Carl von Linne), Swedish botanist, was the first to enunciate the principles for defining genera and species and to adhere to a uniform use of two names (one for genus and one for species) for naming plants and animals. He was born on May 23, 1707, at South Råshult, where his father was a curate. His love of flowers developed at an early age; when only eight years old he was nicknamed "the little botanist." He studied at the universities of Lund and Uppsala, receiving his degree in medicine from the latter.

By courlesy of the Svenska Portrattarkivet Stockholm



Linnaeus. portrait by A. Roslin, 1775. In the Svenska Portrattarkivet, Stockholm

At Uppsala he met the veteran botanist Olof Celsius, who had a profound influence on Linnaeus' subsequent career. Linnaeus was appointed lecturer in botany in 1730 and two years later conducted explorations in Lapland for the Uppsala Academy of Sciences. The results of his journey were published in Amsterdam in 1737 as the *Flora Lapponica* and in English by Sir J.E. Smith as *Lachesis Lapponica* (1811). His reputation was firmly established by this work and, even more, by the appearance in 1735 of his *Systema Naturae* and of the *Genera Plantarum* two years later (the *Species Plantarum* was not published until 1753). For purposes of nomenclature of flowering plants and ferns, the first edition of the *Species Plantarum* has been internationally agreed upon as the starting point, together with the fifth edition of the *Genera Plantarum*, published in 1754.

Major work

The *Systema Naturne*, which Linnaeus had shown to the botanist Jan Fredrik Gronovius in manuscript, so impressed Gronovius that he published it at his own expense. Linnaeus' system was based mainly on flower parts, which tend to remain unchanged during the course of evolution. Although artificial, as Linnaeus himself recognized, such a system had the supreme merit of enabling students rapidly to place a plant in a named category. It came into use at a period when the richness of the world's vegetation was being discovered at a rate that outstripped more leisurely methods of investigation. So successful was his method in practice that its facile application was the greatest obstacle to its replacement by the more natural systems that superseded it.

In 1736 Linnaeus visited England, where he met the botanist and physician Sir Hans Sloane in London and Johann Jakob Dillenius, the first professor of botany at Oxford. He returned to Holland to complete his work on the *Hortus Cliffortianus* and in Paris visited the three Jussieu brothers, distinguished botanists with whom he established a close friendship. Soon afterward, he went once again to Sweden, and in 1738 he settled in Stockholm as a practicing physician, a profession in which he attained considerable success. In 1739 he married Sara Moraea, the daughter of a physician. Two years after his marriage he was appointed to the chair of medicine at Uppsala, but a year later he exchanged this for the chair of botany, his true calling. An inveterate classifier, he not only systematized the plant and animal kingdoms but even classified the mineral kingdom and drew up a treatise on the kinds of diseases known in his day.

His later years were taken up by teaching and the preparation of other works: *Flora Suecica* (1745) and *Fauna Suecica* (1746); two volumes of observations made during journeys in Sweden, *Vastgota resa* (1747) and *Skånska resa* (1751); *Hortus Upsaliensis* (1748); his *Philosophia Botanica* (1751); and the important *Species Plantarum* (1753), in which the specific names are fully set forth. In 1755 he declined an invitation from the King of Spain to settle in that country with a liberal salary and full liberty of conscience. In 1761 he was granted a patent of nobility, antedated to 1757, from which time he was styled Carl von Linne. An apoplectic attack in 1774 greatly weakened him, and he died on January 10, 1778, at Uppsala, in the cathedral in which he was interred. The Linnaean manuscripts and his herbarium and collections of insects and shells, purchased by Sir J.E. Smith in 1783, are carefully preserved by the Linnean Society at Burlington House, London.

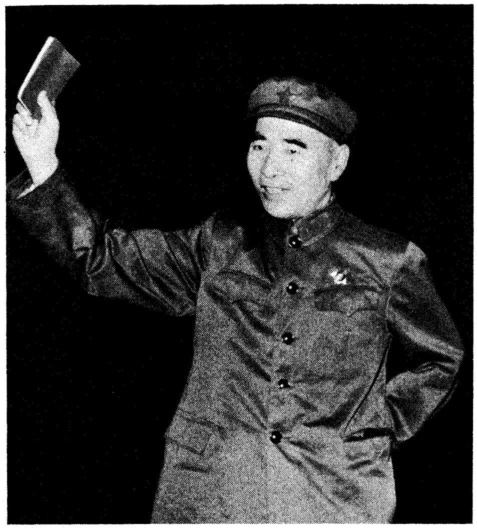
BIBLIOGRAPHY. The most comprehensive work on Linnaeus is T.M. FRIES, *Linné, Lefnadsteckning*, 2 vol. in 11 pt. (1903: Eng. trans., *Linnaeus*, 1923), including a selective but extensive bibliography. Other studies recommended are: OSCAR LEVERTIN, *Carl von Linné* (1906); J.M. HULTH, *Bibliographia Linnaeana* (1907); WILHELM JUNK, *Linné im Lichte neuerer Forschung* (1925); NORAN GOURLIE, *The Prince of Botanists: Carl Linnaeus* (1953); KNUT HAGBERG, *Carl Linnaeus* (1951; Eng. trans., 1952); W.T. STEARN, "The Background of Linnaeus's Contributions to the Nomenclature and Methods of Systematic Biology," *Syst. Zool.*, 8:4–22 (1959); B.D. JACKSON. *Linnaeus* (1923); GEORGE H.M. LAWRENCE, "Linnaeus," *Horticulture*, 42:22–23 (1964); WILFRED BLUNT. *The Complear Naturalist: A Life of Linnaeus* (1971); and JAMES L. LARSON, *Reason and Experience: The Representation of Natural Order in the Work of Carl von Linné* (1971).

(E.J.S.)

Lin Piao

One of modern China's outstanding military leaders, Lin Piao (in Pinyin romanization, Lin Biao), as a field commander of the Red Army, made substantial contributions to the Communists' 22-year struggle for power under the direction of Mao Tse-tung (Mao Ze-dong) and to their eventual triumph in 1949. He became, in 1966, Mao's heir apparent as ruler of the new China, but since his mysterious death in 1971 he has been vilified as a renegade and a traitor.

Lin Piao was born on December 5, 1907, to a family of modest means in Huang-kang (Huang-gang) county of Hupeh (Hu-bei) Province in Central China. He received his primary education in the village school and then entered middle school in Wu-ch'ang (Wu-chang), the provincial capital, in 1921. In middle school, he was deeply



Lin Piao, 1967
Easifoto

affected by the social and cultural upheaval then taking place, which some have called the Chinese Renaissance. He soon became interested in Socialism and Communism, and in 1925, after his graduation from middle school, he joined the Socialist Youth League.

Military career. Also in 1925, Lin Piao went south to Canton to enroll in the Whampoa Academy and there began his military career. China at this time suffered from the twin evils of warlordism and imperialism; *i.e.*, internal disunity and foreign encroachment. In order to fight the warlords and curb the imperialists, the Nationalists, led by Sun Yat-sen until his death in March 1925, had secured the assistance of the Soviet Union and the cooperation of the fledgling Chinese Communist Party and were then preparing a military expedition from their base in Canton. The Whampoa Academy, headed by Sun's successor, Chiang Kai-shek, was to train the officers for the revolutionary army. Lin Piao had been at the academy less than a year when Chiang launched the Northern Expedition in July 1926. Despite the brevity of his formal training, Lin quickly demonstrated his military prowess. A few months later, when the expedition reached the Yangtze River in Central China, he had risen from deputy platoon leader to battalion commander. But, when Chiang then turned savagely against his Communist allies in 1927, Lin forsook his mentor and fled with the Communists.

Northern Expedition of 1926

In the spring of 1928, Lin Piao joined Mao Tse-tung in the hills of south central China and established himself at once as one of the ablest and most active commanders in Mao's small but growing Red Army. From 1928 to 1934 he helped to enlarge the Communist-controlled territory in Kiangsi (Jiang-xi) Province and defended it against repeated attacks by the Nationalists. In 1932 he was promoted to corps commander. When the Communists were finally driven from their Kiangsi base in 1934 by the Nationalists, Lin's First Army Corps formed the vanguard of the epic retreat known as the Long March, which brought them a year later into the northern province of Shensi (Shen-xi), where they were able to regroup. By then the 28-year-old Lin was already a legendary figure with a reputation of never having lost a battle. In Shensi he became the president of the Red Army Academy and was among the handful of commanders who ranked just below the important military leaders Chu Teh (Zhu De) and P'eng Te-huai (Peng De-huai).

In 1937 the bitter civil war between the Nationalists and the Communists ended temporarily as the two sides formed a united front against the common foe, Japan. In September 1937, just after the outbreak of war, Lin Piao, in command of one of the Red Army's three divisions, scored an important early victory over the Japanese invaders. But he was wounded in battle the following spring and retired from the field for the rest of the war. When his wound did not heal, he went to the Soviet Union for medical treatment and stayed three years. After his return to China in 1942, he served briefly as a member of the Communist liaison team with the Nationalists. In 1943 he resumed the presidency of the Military and Political Academy in the Communist capital of Yen-an (Yan-an). As the war with Japan neared its end in 1945, he was elected for the first time to the Communist Party's 44-member Central Committee.

Conquest of Manchuria

When World War II ended, the civil war in China resumed. Returning to the field once more, Lin Piao went to Manchuria, in northeastern China, as commander of what later became the Fourth Field Army. In a brilliant display of Maoist strategy, he first abandoned the cities of Manchuria to the Nationalists and concentrated instead on securing the support of the peasants in the countryside. Using guerrilla warfare, he then patiently and methodically whittled away at his numerically superior enemy. Gradually, he isolated the Nationalists in the cities and eventually forced garrison after garrison to surrender. By the end of 1948, his army, which had originally numbered 100,000, had grown to 800,000, and he had captured all of Manchuria. Lin's victory in Manchuria ensured the rapid collapse of Chiang Kai-shek's Nationalists in the rest of China. His own forces, moving south, captured Peking (Bei-jing) in January 1949, Wu-han in May, and Canton in October.

Positions in the People's Republic. With the establishment of the People's Republic in October 1949, Lin Piao was appointed to many high posts in the government and the party. At first he was both administrative head and party chief of the six-province "Central-South" region of China. In 1954, when the central government was reorganized, he was named a vice premier of the State Council (or Cabinet) and a vice chairman of the National Defense Council. Among army officers he ranked only behind the aging Chu Teh and P'eng Te-huai, then the minister of defense. In 1955 he was elevated within the party to the Central Committee's 13-man Politburo and then in May 1958 to the Politburo's seven-man Standing Committee. But throughout these early years of Communist rule Lin, who may have been in chronic ill health, seldom appeared in public and was probably only occasionally active at his various posts.

Minister of defense

In late 1958, Lin Piao suddenly began to assume a more active and important role in the army and the party. In September 1959 he succeeded P'eng Te-huai as minister of defense, after P'eng was ousted for opposing Mao's economic and defense policies. Lin then inaugurated a reformation of the army that both intensified the political education of its soldiers and upgraded their military training. As a result, Lin's army became in the early 1960s both an example of how, according to Mao's teachings, professional expertise may be combined with political consciousness and a model for the rest of society, including the party itself, to emulate. This movement to "learn from the People's Liberation Army" eventually developed in 1965 into the extensive purge of the party known as the Great Proletarian Cultural Revolution, whose principal casualty was Liu Shao-ch'i (Liu Shao-qi), the party organizer who for more than 20 years had been Mao's second in command. In August 1966, the 58-year-old Lin Piao replaced Liu as the future successor to Mao; this position was formalized in April 1969, when Lin was so designated by the new constitution. From 1966 to 1971, the army took over the role previously played by the party in ruling China, By 1971, however, Lin and the army may have amassed more political authority than Mao thought desirable. In a desperate move to avoid being purged, Lin and others of the military high command plotted a coup that failed. It was announced that Lin Piao was killed on September 13, 1971, in an airplane crash in Mongolia as he was fleeing to the Soviet Union. Since then he has been posthumously criticized as a rightist reactionary and a traitor to the cause of Chinese Communism.

Assessment. Throughout his life, Lin Piao was more a doer than a thinker. His writings are few and uninspiring. They deal primarily with questions of military strategy and tactics, especially the latter (of which he was a master), or with the importance of political indoctrination.

As a leader, Lin lacked Mao's wit and charisma and Chou En-lai's charm and urbanity. In contrast to these other two members of the ruling triumvirate in the late 1960s, Lin seemed almost colourless. Even as a military commander, he was characterized more by caution and deliberation than by dash and flamboyance. Lin Piao was married at least twice. His last wife, Yeh Ch'iin (Ye Qun), rose to political prominence herself during the Cultural Revolution; in April 1969 she joined the 21-member Politburo. It was reported that she died with her husband and their son, Lin Li-guo, in the airplane crash.

BIBLIOGRAPHY. The best accounts of Lin's life are two studies by THOMAS W. ROBINSON, *A Politico-Military Biography of Lin Piao, Part I,* 1907–1949 (1971) and "Lin Piao as an Elite Type," in ROBERT A SCALAPINO (ed.), *Elites in the People's Republic of China* (1972). Other biographic accounts are the brief entries in DONALD W. KLEIN and ANNE B. CLARK, *Biographic Dictionary of Chinese Communism,* 1921–1965, 2 vol. (1971); and HOWARD L. BOORMAN (ed.), *Biographical Dictionary of Republican China,* vol. 2 (1968). MICHAEL Y.M. KAU (ed.), *The Lin Piao Affair: Power Politics and Military Coup* (1975), discusses the circumstances of Lin's death and contains a compendium of documents including a translation of Lin's official biography as of 1969.

(E.J.M.R.)

Lin Tse-hsii

A leading Chinese scholar and official of the Ch'ing (Manchu) dynasty, Lin Tse-hsii (in Pinyin romanization, Lin Ze-xu) was a proponent of the revitalization of traditional Chinese thought and institutions, a movement known as the "Statecraft school." He could, however, neither comprehend the implications of the European challenge nor overcome the weakness and conservative opposition of his contemporaries. Later, the so-called Self-strengthening Movement adopted Lin's program of reform; still later generations of revolutionaries abandoned Chinese culture in order to save China but accepted Lin as a national hero because of his courage and example in opposing the British before the Opium War (1839–42).

Rise as administrator

Lin was born in Hou-kuan (Hou-guan) in Fukien Province in 1785; his father was a teacher, who, though poor, was determined that his sons should have the grounding in the Confucian Classics that alone could advance them in the governmental bureaucracy. Lin Tse-hsii, the second son, proved immensely capable and passed the initial examinations in 1804. He then was selected as an aide to the governor of his native province, an informal apprenticeship that served to balance the abstract, moral, and largely literary content of his early education. In 1811 Lin passed the highest of the examinations, the *chin-shih,* and joined the Hanlin Academy, which advised the Emperor and helped him to draft documents. In 1820 he took up his first regular administrative post and rose through a number of the most responsible offices in the bureaucracy. After starting in the salt monopoly, he supervised water-control systems in several localities, administered the collection of taxes, and served a term as a local judge, during which he earned the respectful nickname "Lin the Clear Sky." Lin's quick rise showed him to be an effective organizer and ambitious bureaucrat.

Role in the Opium War

Following the traditional period of mourning and retirement at the death of his father, a time that also served for reflection and literary activity, Lin returned to official life in the upper reaches of the government. When, in the middle of the 1830s, the Tao-kuang (Dao-guang) emperor became alarmed over the growth of the opium trade by British and Chinese smugglers—both for the obvious moral reasons and for the more practical one that even illegal imports had to be paid for with the export of Chinese silver—Lin submitted a memorial condemning a suggestion that the trade be legalized. In support of his position he cited the measures by which he had suppressed the drug traffic in the provinces of which he was then governor general. The Emperor, who for almost two decades had vainly attempted to enforce the ban on the importation of opium, responded by appointing Lin Imperial Commissioner in late 1838, vesting him with extraordinary powers. After an unusual 19 personal audiences with the Emperor, Lin proceeded to Canton, the hub of the trade. His diary for this period survives and conveys a vivid picture of a Chinese official of the time at work: making the arduous journey from Peking (Bei-jing); perspiring in the heat of Canton's subtropical climate as he kowtows before the very written instructions of the Emperor; peremptorily summoning the British merchants and officials; vainly trying to make the corrupt Chinese officials, grown soft on the profits and use of opium, perform their duties; and composing an ode of apology to the god of the sea for defiling his ocean with confiscated opium.

Lin was only too successful. He forced foreign merchants to surrender their stocks of opium for destruction and put pressure on them to guarantee that they would cease importing the cargo. Yet, when the British retaliated by ravaging large parts of South China, the Emperor, who had personally approved Lin's tough policies, quickly dismissed him. Although exiled to the northwest frontier, Lin served quietly and loyally, was soon called back to important service, and was rewarded with the title of Grand Guardian of the Heir Apparent for pacifying rebel Muslims in the province of Yunnan (Yün-nan). He died in 1850, on his way to help suppress the Taiping Rebellion.

Assessment

Lin's significance goes beyond his career. He belonged to a small but later influential group of reform officials and scholars whose slogan was "find in antiquity the sanction for present-day reform." This "Statecraft school" pioneered in the compilation of practical information for use in governing on many subjects, including geography and the knowledge of the history of foreign countries. Yet when Lin, one of the most experienced and well-informed men of his day, went to Canton, he had no idea that his success in stopping the opium trade would only open up his country to the humiliating and ruinous penetration by foreign interests that were to hasten its downfall. He simply drew on the precedents of generations of officials whose policy against the Central Asian tribes had been to play them off against each other and to whom commercial considerations were somewhat petty matters; he did not comprehend the significance of the British demands for free trade and international equality, which were based on their concept of a commercial empire. This concept was a radical challenge to the Chinese world order, which knew only an empire and subject peoples; at his arrival in Canton, Lin still thought of the British as dependent on the Chinese and was convinced that they would perish from constipation without Chinese tea and rhubarb. In a famous letter to Queen Victoria, written when he arrived in Canton, Lin asked whether she would allow the importation of such a poisonous substance into her own country, and requested her to forbid her subjects to bring it into his. Lin relied on aggressive moral tone, meanwhile proceeding relentlessly against British merchants in a manner that could only insult their government. The only lesson Lin drew from China's humiliation was that it was necessary to learn more about these "barbarians" and to import their technology. The task of fathoming the relation between the technology of the "barbarians" and their politics and economics was left to later generations.

BIBLIOGRAPHY. HSIN–PAO CHANG, *Commissioner Lin and the Opium War* (1964), is the most up-to-date study, including an extensive Chinese bibliography; see also ARTHUR WALEY, *The Opium War Through Chinese Eyes* (1958).

(C.W.H.)

Lipid

Lipids, or fats, are best defined as a diverse group of organic compounds found in plants, in animals, and in microorganisms. Characteristically, they are greasy to the touch and insoluble in water but soluble in alcohol, ether, and other organic solvents. Lipids comprise one of the three large classes of foods and, with proteins and carbohydrates, are components of all living cells. The proportion of lipids in foodstuffs varies; it is 0.2 percent in white potatoes and 70 percent in some nut kernels. Fats such as olive oil and cod-liver oil contain a mixture of fatty substances, lipids called triglycerides. Triglycerides comprise almost 90 percent of the adipose, or fat, tissue of animals. Contrary to popular opinion, adipose tissue is an energy source and can be utilized when needed. In fact, triglycerides are sometimes called nature's storehouse of energy because, on a weight basis, they contain more than twice as much energy as do carbohydrates and proteins.

The lipids made by cells are important not only because they serve as an energy source but also because they form structural components. Lipids such as lecithin and cephalin, which are soluble in both water and fats, serve a vital role in the cell by binding water-soluble compounds such as proteins to lipid-soluble substances. Lecithin is an important structural component of the cell membrane, where it maintains continuity between the water and lipid phases inside and outside the cell. The function of certain enzymes (biological catalysts) depends upon their attachment to lipids such as lecithin.

The structure of lipids varies from simple chainlike molecules consisting of hydrogen, carbon, and oxygen to complex ring, or cyclic, structures with side chains of varying composition and complexity. Many naturally occurring lipids are associated with proteins, in combinations called lipoproteins. The type of lipids dealt with in this article include neutral lipids (or triglycerides; *i.e.*, fatty-acid esters of the alcohol glycerol), phosphoglycerides (or phospholipids; *i.e.*, fatty-acid esters of glycerol and phosphoric acid or one of its derivatives), and sphingolipids (complex lipids containing compounds—sphingosine or phytosphingosine—other than glycerol). The characteristics of fatty acids, which are important components of many lipids, also are described. Lipoproteins and the preparation and analysis of lipids are dealt with briefly. Sterols and carotenoids, also included here as lipids, are considered in detail in two related articles—STEROID and ISOPRENOID. Other related articles include METABOLISM; OILS, FATS, AND WAXES; and MEMBRANE, BIOLOGICAL.

IMPORTANCE

Lipids as food reserves in cells. The most important role of the fatty-acid components of neutral lipids in plant and animal tissues is to provide a fuel supply for cells; *i.e.*, neutral lipids comprise a reserve supply of potential energy and are broken down, when needed, in such a manner that the energy liberated is employed to make an energy-rich compound called adenosine triphosphate (ATP), which in turn is utilized in energy-requiring cellular processes such as muscle contraction and the synthesis of cell constituents. The energy in a fatty-acid molecule is transformed into ATP by a process known as fatty-acid oxidation (or beta oxidation). For a detailed discussion of the mechanism of fatty-acid oxidation, see METABOLISM.

Triglycerides and fatty acids are formed during digestive processes in animals. After a mammal ingests a fatty meal, the fats are acted upon by digestive secretions containing the enzyme lipase, which breaks down at least part of the triglycerides. The breakdown products and the remaining intact triglycerides then are absorbed through the intestinal cell wall and are recombined, at least in part, to form triglycerides and phospholipids. These lipids, in the form of very small droplets (chylomicrons), are transported in blood and in chyle (a milky fluid from the small intestine) to points of utilization or storage in the body.

One function of bile salts in digestion is to promote the linkage of (*i.e.*, emulsify) lipid-soluble groups with water-soluble ones (such as those in enzymes) and also to increase the solubility of lipids in water. Both emulsification and solubilization are necessary because lipids are completely metabolized only at the lipid–water interface created by bile salts and by the salts of fatty acids (soaps), which are formed during the partial breakdown of lipids (see also DIGESTION AND DIGESTIVE SYSTEMS).

If an animal ingests more energy-rich substances (*e.g.*, fats, carbohydrates) than it can utilize in energy-requiring functions, excess fatty acids combine with glycerol to form neutral lipids, which are stored in the animal; *e.g.*, in adipose tissue in mammals. If the energy requirements of the animal increase, the stored neutral lipids may then be broken down, each molecule forming three molecules of fatty acid and one molecule of glycerol. The three molecules of fatty acid combine with a protein (albumin) in mammalian blood plasma and are carried in the bloodstream to various tissues and organs that require energy. It is probably also as depots of concentrated energy that neutral lipids function in plant reproductive structures such as pollen grains and seeds; *i.e.*, the lipids serve as a food reserve for the developing embryo.

Neutral
lipids as
depot
material

The types of neutral lipids in an individual animal may vary according to the animal species and the composition of fats in the food it consumes. Fats used by or stored in animal tissues come from two sources---diet and enzymatic synthesis. The lipids synthesized from carbohydrates or proteins are characteristic of the animal species, whereas those resynthesized from dietary fats are characteristic of the food ingested. Many animals require some lipids containing one or more specific fatty acids, usually linoleic, linolenic, and arachidonic, to prevent the development of an essential fatty-acid deficiency, which is manifested by skin lesions, scaliness, poor hair growth, and low growth rates. The so-called essential fatty acids cannot be synthesized by the animal and must, therefore, be supplied in the diet (see also NUTRITIONAL DISEASES AND DISORDERS).

Lipids as structural components of cells. Neutral lipids. Subcutaneous deposits of neutral lipids insulate animals against cold because of the low rate of heat transfer in fats, a property especially important to animals commonly found in cold waters or cold climates; *e.g.*, whales, walruses, and bears. Neutral lipids also provide structural support or padding for organs.

Phosphoglycerides. Phosphoglycerides, or phospholipids, are important constituents of cell membranes and of the membranes of cell components such as mitochondria; they function in the conduction of nerve impulses, in the insulation of nerve cells (see NERVE IMPULSE), in certain enzyme-catalyzed reactions within cells, and in blood coagulation and the transport of lipids from the liver to other tissues and organs in mammals.

Phosphoglycerides contain acidic (negatively charged) and basic (positively charged) groups as well as fatty-acid groups. Because they have both charged, water-attracting (hydrophilic) groups and lipid-attracting, water-repelling (hydrophobic) groups, phosphoglycerides are moderately soluble in both water and lipids and serve, therefore, an important role in the cell in binding both types of compounds together; lecithin is a naturally occurring phosphoglyceride of special importance in the cell membrane because its hydrophilic groups maintain continuity between the water outside and that inside the cell, and its hydrophobic groups dissolve lipid materials, allowing them to enter the cell (see MEMBRANE, BIOLOGICAL).

Phosphoglycerides in blood bodies called platelets function in the process of blood-clot formation. Blood coagulation in mammals involves the conversion of a soluble protein (fibrinogen) into an insoluble derivative (fibrin clot). This reaction in part is catalyzed by the enzyme thrombin, which is derived from another protein (prothrombin) by the action of the enzyme prothrombinase. Prothrombinase in turn is formed by the interaction of two blood-plasma proteins, a phosphoglyceride, and calcium. The exact role of the phosphoglyceride in this reaction, however, has not yet been established with certainty.

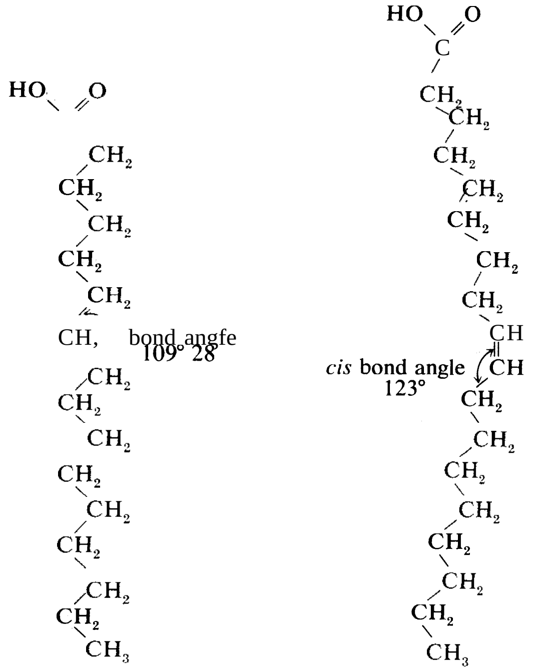
Sphingolipids. Lipids found in nervous tissue (especially the brain) and necessary for its normal function include the sphingolipids, among them sphingomyelins, cerebrosides, sulfatides, and gangliosides. Schwann cells, the source for the membranous sheath surrounding certain nerve fibres, release a lipid material (myelin) that contains sphingomyelins. Myelination of nerve fibres is necessary for the normal development of nerve tissue, and absence or alteration of the myelination process may result in severe mental retardation in man. Sphingolipids have not been studied as exhaustively as some of the other lipids because they are difficult to isolate and separate into homogeneous components.

Sterols and carotenoids. Sterols include cholesterol in vertebrates and saponins and digitaloids in plants; cholesterol, with phospholipids, plays a role in membrane structure and an important role in the synthesis of numerous other biologically active sterols, commonly referred to as steroids—*e.g.*, bile acids, certain hormones, vitamin D. Carotenoids, widely distributed in living things, are pigments (coloured molecules) and function in the photosynthetic process in plants. These substances are discussed in detail in STEROIDS; ISOPRENOIDS.

FATTY ACIDS

The fatty acids of the naturally occurring lipids have an even number of carbon (C) atoms because they are synthesized from acetyl groups, each of which contains two carbon atoms. Fatty acids with 16 (palmitic acid) and 18 (stearic acid) carbon atoms are most commonly found in nature, but the reasons for their abundance have not yet been established. Important constituents of lipids in plants, animals, and micro-organisms, fatty acids usually are not found in free form but, instead, are bound to other compounds to form fatty-acid-containing lipids; *e.g.*, neutral lipids (tryglycerides), sterols, phosphoglycerides such as lecithin, and sphingolipids such as sphingomyelin. Two typical fatty acids are palmitic and oleic.

Fatty-acid-containing lipids



palmitic acid

oleic acid

The most stable arrangement of methylene groups ($-CH_2-$ groups), which comprise the hydrocarbon moiety of the molecule, is represented in the formula for palmitic acid; the double bond ($-CH=CH-$) in the oleic acid molecule, whose structure differs from that of palmitic acid only in the double bond between carbon atoms 9 and 10 (the carbon in the carboxyl, or $-COOH$, group is number 1), changes the shape of the molecule from that of palmitic acid. The double bond, which indicates that the acid contains two fewer hydrogen atoms than its counterpart (palmitic acid), makes the oleic acid molecule almost symmetrical because it is located exactly in the middle of the hydrocarbon chain. An additional change in the shape of fatty-acid molecules occurs in those acids that contain two double bonds; *e.g.*, linoleic acid. Fatty-acid molecules containing one or more double bonds ($-CH=CH-$) are called unsaturated fatty acids; those without double bonds are saturated acids. Double bonds give rise to a phenomenon called geometrical isomerism; *i.e.*, the positions of certain atoms in the acid molecule may lie on the same side of an imaginary plane through the molecule, in which case a bond is said to be cis, or on opposite sides of the plane, in which case the bond is said to be trans. Unsaturated fatty acids of mammalian tissues usually contain cis double bonds (see structural formulas for palmitic and oleic acids); in addition, fatty acids found in mammals usually consist of long, chainlike arrangements of the hydrocarbon portion of the molecule (*e.g.*, as in palmitic acid) rather than branched or cyclic structures. On the other hand, fatty acids with cyclic and branched structures occur in plants and bacteria.

Classes of fatty acids. Saturated fatty acids. Typical naturally occurring saturated fatty acids are chainlike (nonbranched) compounds with an even number of carbon atoms; *e.g.*, palmitic acid, stearic acid. Names given to fatty acids other than the more commonly used, or trivial, names are called systematic names (*n*-hexadecanoic acid for palmitic and *n*-octadecanoic acid for stearic); they provide information about either the correct geometric form of an unsaturated fatty acid or the exact location of substituted groups in the molecule.

Although palmitic acid and stearic acid are the major saturated fatty acids found in animal and plant tissues, significant amounts of other saturated fatty acids, such as myristic acid and lauric acid, occur in certain tissues, and lignoceric acid and behenic acid are found in high concentrations in brain sphingolipids. Small amounts of fatty acids with an odd number of carbon atoms also are known; *e.g.*, pentadecanoic acid and heptadecanoic acid. A list of the most common saturated chainlike fatty acids (*i.e.*, those containing 12 or more carbon atoms), including their usual sources, is presented in Table 1. Several important short-chain fatty acids include butyric acid (four carbon atoms, or C_4) and caproic acid (C_6), which are important constituents of milk lipids, and octanoic acid (C_8) and decanoic acid (C_{10}), which are present in palm oil.

Table 1: Some Naturally Occurring Saturated Fatty Acids

trivial name	systematic name	chain length*	typical sources
Lauric acid	*n*-dodecanoic acid	12	palm-kernel oil, nutmeg
Myristic acid	*n*-tetradecanoic acid	14	palm-kernel oil, nutmeg
Palmitic acid	n-hexadecanoic acid	16	olive oil, animal lipids
Stearic acid	n-octadecanoic acid	18	cocoa butter, animal lipids
Behenic acid	n-docosanoic acid	22	brain tissue, radish oil
Lignoceric acid	*n*-tetracosanoic acid	24	brain tissue, carnauba wax

*In carbon atoms.

Unsaturated fatty acids. Unsaturated fatty acids have structures similar to those of saturated ones but contain at least one or more olefinic, or double bond. Fatty acids with one double bond are called monounsaturated fatty acids; those with two or more are called polyunsaturated ones. The naturally occuring unsaturated fatty acids may be distinguished solely by the degree of unsaturation and usually contain one or two double bonds; fatty acids containing acetylenic, or triple bonds ($-C\equiv C-$), also have been found in nature.

Mono- and polyunsaturated fatty acids

Monounsaturated fatty acids (also called monoethenoic, monoenoic, ethylenic, olefinic, or alkenoic acids) contain two fewer hydrogen atoms (H) than do corresponding saturated acids and thus have one double bond (*i.e.*, $-CH_2-CH_2$ becomes $-CH=CH-$); monounsaturated acids comprise the largest group of unsaturated fatty acids. The major and most representative monounsaturated acids in animal and plant tissues, oleic acid and palmitoleic acid, have the following formulas:

$$CH_3(CH_2)_7CH=CH(CH_2)_7COOH$$

oleic acid

$$CH_3(CH_2)_5CH=CH(CH_2)_7COOH.$$

palmitoleic acid

These two widely distributed fatty acids exist in abundant quantities in nature and have several common features; *i.e.*, a double bond between carbon atoms 9 and 10 (counting from the —COOH, or carboxyl, end) and a cis configuration at this double bond. Fatty acids with a double bond at other positions are listed in Table 2; *e.g.*, nervonic acid has a double bond between carbon atoms 15 and 16, as indicated by the systematic name, cis-15-tetracosenoic acid.

Table 2: Types of Monounsaturated Long-Chain Fatty Acids

trivial name	systematic name	chain length*	typical source(s)
Palmitoleic acid	cis-9-hexadecenoic acid	16	marine algae, pine oil
Oleic acid	cis-9-octadecenoic acid	18	animal tissues, olive oil
Gadoleic acid	cis-9-eicosenoic acid	20	fish oils (cod, sardine)
Erucic acid	cis-13-docosenoic acid	22	rapeseed oil
Nervonic acid	cis-15-tetracosenoic acid	24	elasmobranch fishes, brain

*In carbons.

Polyunsaturated acids (also called polyethenoic, polyenoic, alkapolyenoic, alkadienoic, or alkatrienoic acids) are found in much smaller amounts in naturally occurring lipids than are monounsaturated ones. The formulas of the three common polyunsaturated fatty acids are given below.

$$CH_3(CH_2)_4CH=CHCH_2CH=CH(CH_2)_7COOH$$

linoleic acid

$$CH_3CH_2CH=CHCH_2CH=CHCH_2CH=CH(CH_2)_7COOH$$

linolenic acid

$$CH_3(CH_2)_4CH=CHCH_2CH=CHCH_2CH=$$
$$CHCH_2CH=CH(CH_2)_3COOH.$$

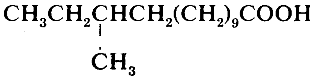
arachidonic acid

Linoleic acid and arachidonic acid comprise most of the polyunsaturated fatty acids found in animal tissues. The double bonds usually are located at specific positions, as in linoleic and arachidonic acids. The most important polyunsaturated acids, which contain 18 to 22 carbon atoms, are listed in Table 3.

Most mammals cannot synthesize linoleic acid, even though they are able to convert it to other unsaturated acids; *e.g.*, arachidonic acid and the eicosapolyenoic acids. The polyunsaturated fatty acids in mammals therefore are derived largely from the diet, *i.e.*, they are essential fatty acids.

Alkyl-, hydroxy-, and alicyclic-substituted fatty acids

Fatty acids *with* substituted groups. Alkyl-substituted, or alkylalkanoic, fatty acids have one alkyl group on the molecule. An alkyl group is a hydrocarbon group with the general formula C_nH_{2n+1}; the methyl group ($-CH_3$) is the alkyl group in the example below. Methyltetradecanoic acid is found in butterfat.

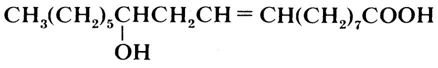


$$CH_3CH_2\underset{\displaystyle CH_3}{\underset{|}{C}}HCH_2(CH_2)_9COOH$$

12-methyltetradecanoic acid (anteiso)

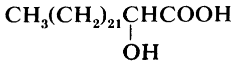
An excellent source of alkyl-substituted fatty acids is the bacterium that causes human tuberculosis; the acid is called tuberculostearic acid, or D-(1)-10-methyloctadecanoic acid. Other saturated acids isolated from tubercle bacilli include phytomonic acid (a 10- or 11-methylnonadecanoic acid), phthioic acid (a mixture containing fatty acids with 23 to 31 carbon atoms), mycocerosic acid (a polymethylated fatty acid with 31 carbon atoms), and a number of alkyl branched-chain, unsaturated fatty acids.

Several hydroxy fatty acids (*i.e.*, containing one — OH group) occur in various lipid sources. Ricinoleic acid and cerebronic acid are typical examples.



$$CH_3(CH_2)_5\underset{\displaystyle OH}{\underset{|}{C}}HCH_2CH=CH(CH_2)_7COOH$$

ricinoleic acid
(*d*-12-hydroxy-cis-9-octadecanoic acid)

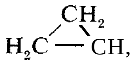


$$CH_3(CH_2)_{21}\underset{\displaystyle OH}{\underset{|}{C}}HCOOH$$

cerebronic acid
(2-hydroxytetracosanoic acid)

Ricinoleic acid occurs in castor oil, and cerebronic acid is present in brain tissue; a-hydroxynervonic acid (an unsaturated fatty acid with 24 carbon atoms) also occurs in brain tissue. Very small amounts of other hydroxy fatty acids are found in animal and bacterial lipids.

Alicyclic-substituted fatty acids, which contain a cyclopropane group



attached at or near the centre of the molecule, occur in certain bacteria. One such acid, lactobaciilic acid, comprises about 30 percent of the total fatty acids of the bacterium Lactobacillus *arabinosus*. Fatty acids containing cyclopropane groups have not yet been found in animal tissues. A series of compounds similar in structure has, however, been found in the chaulmoogra oils.

Several alicyclic acids, called prostaglandins, occur in seminal fluid and in certain glands of sheep and man. Hormone-like substances with widespread physiological effects, prostaglandins lower blood pressure, and stimulate the contraction of smooth muscles.

Physical and chemical properties of fatty acids. Solubility. Long-chain fatty acids (*i.e.*, with ten or more carbon atoms) are insoluble in water; short-chain fatty acids (two to eight carbon atoms), however, mix easily with water. The reason for the low solubility of long-chain fatty-acid molecules is that the methylene groups ($-CH_2-$) comprising the hydrocarbon chain are not ionized (*i.e.*, lack charge) and thus are nonpolar; they are more important in determining solubility properties of the acids than is the carboxyl (— COOH) component, which is ionized and therefore polar. Common table salt is about 1,000 times more soluble in water than is stearic acid; glucose is about 3,300 times more soluble. Fatty acids dissolve in nonpolar solvents or solvents that are less polar than water.

Reactivity of the carboxyl *group*. The low water solubility of fatty acids, which are members of a large group of organic acids known as carboxylic acids, makes difficult the determination of their properties as acids. Small amounts of fatty acids, which are weak acids compared with mineral acids such as hydrochloric and nitric, do not completely dissociate into ionized (charged) moieties in water. The carboxyl group (— COOH) is responsible for the acidic properties of fatty acids (see also CARBOXYLIC ACIDS AND THEIR DERIVATIVES).

Chlorides, esters, and amides

Most reactions at the carboxyl group of fatty acids involve the hydroxyl moiety (— OH) but are influenced by the carbonyl moiety ($-C=O$). Several reactions are typical of fatty acids; *e.g.*, formation of chlorides, esters, and amides. Fatty-acid chlorides, called acyl chlorides, are formed by the reaction of a fatty acid with the compound thionyl chloride ($SOCl_2$). Salts of fatty acids are formed in reactions with bases (*e.g.*, sodium hydroxide), in which case the products are called soaps, or with alcohols to form esters. If palmitic acid reacts with the alcohol methanol (CH_3OH), the product, called methyl palmitate, is the methyl ester of palmitic acid; the methyl esters of fatty acids are often used to determine the properties of the acids. The reaction of a fatty acid with ammonia in the presence of heat results in the formation of a product called an amide; the amides of fatty acids also are useful in characterizing them. Amides are found in nature.

Reactivity of the hydrocarbon chain. The most chemically reactive part of a hydrocarbon chain of a fatty acid is the double bond. Halogens (*e.g.*, chlorine, bromine, iodine) and their derivatives, called halides, react with unsaturated fatty acids to remove the double bonds, thereby forming saturated acids; this is known as an addition reaction. Halides such as hydriodic acid (HI) readily attack a double bond to form derivative compounds.

In the presence of a suitable catalyst (*e.g.*, platinum oxide, palladium on charcoal), hydrogen undergoes an addition reaction with unsaturated fatty acids; that is, hydrogen is added to the positions at which double bonds are found. This hydrogenation reaction results in the formation of saturated fatty acids. The double bonds of polyunsaturated fatty acids are hydrogenated in a specific order; linolenic acid (a fatty acid with 18 carbon atoms and three double bonds; Table 3), for example, is hydrogenated first at the bond between carbon atoms 12 and 13 to form a 9,15-octadecadienoic acid, which can then be further hydrogenated to form a saturated acid. For a discussion of the commercial use of the hydrogenation process, see OILS, FATS, AND WAXES.

Table 3: Types of Polyunsaturated Fatty Acids

trivial name	systematic name	chain length*	typical source(s)
Linoleic acid	cis-9-, cis-12-octadecadienoic acid	18	corn oil, animal tissues, bacteria
Linolenic acid	cis-9-, cis-12-, cis-15-octadecatrienoic acid	18	animal tissues
	5,8,11-eicosatrienoic acid	20	
	8,11,14-eicosatrienoic acid	20	brain tissue
	7,10,13-docosatrienoic acid	22	phospholipids
	8,11,14-docosatrienoic acid	22	
Arachidonic acid	5,8,11,14-eicosatetraenoic acid	20	liver, brain tissue
	4,7,10,13-docosatetraenoic acid	22	brain tissue
	4,7,10,13,16,19-docosahexaenoic acid	22	brain tissue

*In carbon atoms.

Bond alternation, which is defined as any change involving double bonds such as those found in unsaturated fatty acids, also occurs in living cells. The first type of bond alternation, called autoxidation, is a nonenzymatic process; the second, lipoxidation, is an enzyme-catalyzed reaction. Both involve oxygen.

Autoxi-
dation

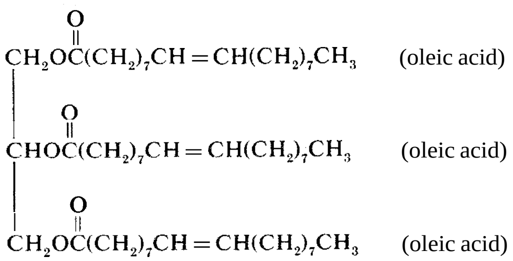
Autoxidation, or rancidification, which occurs slowly and spontaneously in air, involves the absorption of oxygen by an unsaturated fatty-acid molecule, with the formation of a compound called a hydroperoxide, which decomposes. Autoxidation may occur during the drying or handling of certain oils, under adverse conditions of fat storage, and in certain mammalian deficiencies; *i.e.*, vitamin E deficiency. The acids that undergo autoxidation usually are the polyunsaturated ones; *e.g.*, linoleic acid and arachidonic acid.

Lipoxidation contrasts with autoxidation in that an enzyme (lipoxidase) catalyzes the addition of oxygen to a double bond in linoleic acid. Lipoxidase is found in legume seeds (for example, soybeans) and in adipose (fat) tissue.

DERIVATIVES OF FATTY ACIDS

As was previously emphasized, naturally occurring fatty acids are normally found in cells bound to other substances to form triglycerides (neutral lipids), phosphoglycerides (phospholipids), and sphingolipids. Neutral lipids are found almost exclusively in the cytoplasmic compartment of cells, phosphoglycerides and sphingolipids almost exclusively in the membranous structures. For metabolic aspects of these lipids, see METABOLISM.

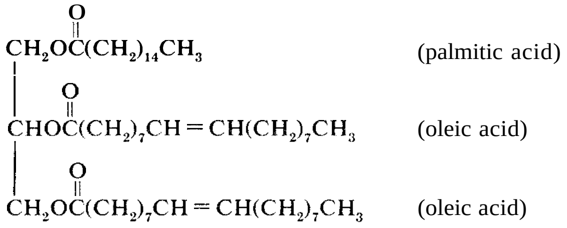
Neutral lipids. *General features.* Triglycerides, or neutral lipids, are compounds consisting of glycerol and three fatty acids; the structural formula of a typical triglyceride, triolein, is shown below.



$$\begin{array}{ll}
\mathrm{CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3} & \text{(oleic acid)}\\
\mathrm{CHO\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3} & \text{(oleic acid)}\\
\mathrm{CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3} & \text{(oleic acid)}
\end{array}$$

Triolein is called a simple triglyceride since it contains only one type of fatty acid (oleic acid). Naturally occurring triglycerides, however, usually are mixed triglycerides; *i.e.*, they contain more than one type of fatty acid. An example of a mixed triglyceride is palmitodiolein, the fatty-acid composition of which is, as the name indicates, one molecule of palmitic acid and two molecules of oleic acid.

This triglyceride may have structural arrangements other than the one shown; *i.e.*. the fatty-acid molecules may be arranged with palmitic acid occupying any of the two possible different positions. (There are only two chemically different positions since the top and the bottom positions in the above formula are undistinguishable.)



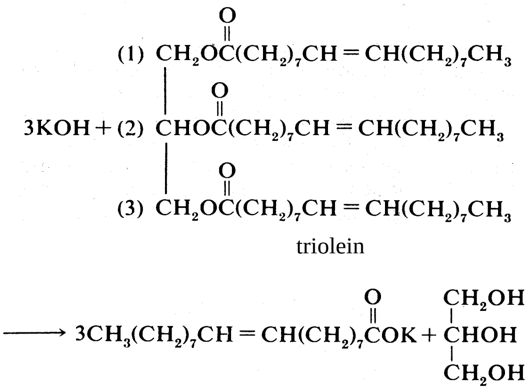
$$\begin{array}{ll}
\mathrm{CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_{14}CH_3} & \text{(palmitic acid)}\\
\mathrm{CHO\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3} & \text{(oleic acid)}\\
\mathrm{CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3} & \text{(oleic acid)}
\end{array}$$

palmitodiolein

Physical and chemical characteristics. Triglycerides, called neutral lipids because they possess no charged groups, do not migrate in an electrical field and are insoluble in water. Information about the structure of naturally occurring mixed triglycerides containing different fatty acids may be obtained by chemical, enzymatic, and physical methods. Chemical and enzymatic techniques are summarized below.

Saponifi-
cation

Triglyceride molecules undergo reactions with bases such as potassium hydroxide (KOH). The reaction, shown below, often called a saponification reaction, results in the formation of an alkali soap (potassium oleate) and glycerol.

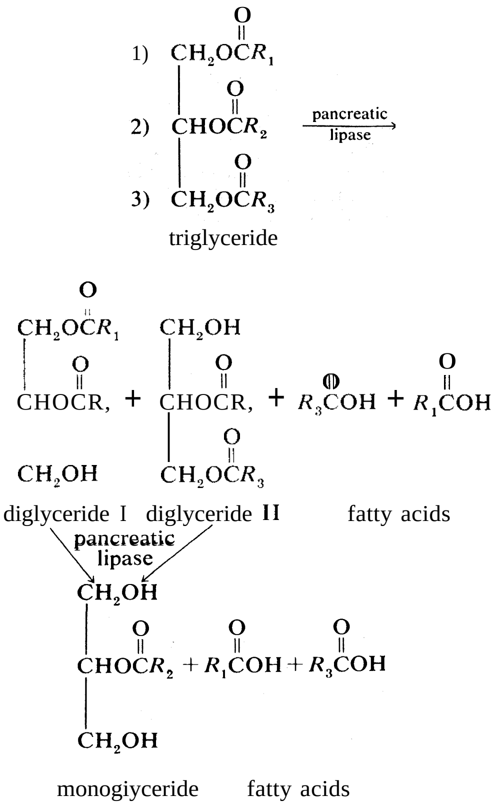


$$\begin{array}{l}
\mathrm{(1)\ CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3}\\
\mathrm{3KOH+(2)\ CHO\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3}\\
\mathrm{(3)\ CH_2O\overset{O}{\overset{\|}{C}}(CH_2)_7CH=CH(CH_2)_7CH_3}\\
\text{triolein}\\
\longrightarrow \mathrm{3CH_3(CH_2)_7CH=CH(CH_2)_7\overset{O}{\overset{\|}{C}}OK + \begin{array}{c}CH_2OH\\ |\\ CHOH\\ |\\ CH_2OH\end{array}}
\end{array}$$

potassium oleate glycerol

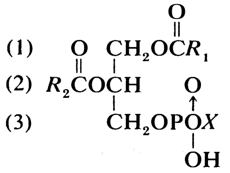
If acid is added to a solution containing the soap, free fatty acids are formed. Although the saponification reaction is useful in establishing the general nature of the fatty acids associated with a specific triglyceride, it provides no information on their specific locations on the glycerol molecule [positions (1), (2), or (3) in the triolein formula above].

The distribution of fatty acids on the glycerol molecule is determined by using an enzyme (pancreatic lipase) that is found in mammals and plays a role in lipid digestion. Pancreatic lipase acts only at positions (1) and (3) of the triglyceride molecule in the sequence outlined below; the overall reaction occurs in two steps. The first products of the lipase reaction are two molecules of glycerol, each containing two fatty acids (labeled diglycerides I and II) and two molecules of free fatty acids (represented by the general formula RCOOH) from positions (1) and (3) of the triglyceride. The lipase next reacts with the two diglycerides. A monoglyceride that contains one fatty acid at position (2), and free fatty acids from positions (1) and (3) of the diglycerides are the products of this second reaction. R_1, R_2, and R_3 in the reaction sequence represent hydrocarbon chains of fatty acid ester groups.



The various products of the enzyme-catalyzed reactions may be isolated and analyzed to provide information about the chemical nature of the fatty acids at each of the three positions.

Phosphoglycerides (phospholipids). The most simple phosphoglyceride contains fatty acids combined with glycerophosphate; *i.e.*, glycerol, containing a molecule of phosphoric acid at position (3), as shown below in the general formula. R_1 and R_2 represent hydrocarbon chains of fatty-acid ester groups at positions (1) and (2) of glycerol; X at position (3) may be a hydrogen atom or an organic compound (*e.g.*, serine, inositol, choline, glycerol, glycerophosphoric acid, or a diglyceride).



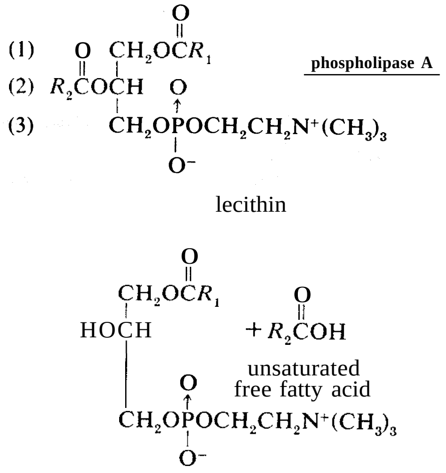
Variations from the general structure occur; phospholipids called plasmalogens, for example, contain an unsaturated ether group ($-OCH=CHR_1$) in position (1) instead of a fatty-acid ester group, and other phospholipids called glyceryl ether phospholipids contain a saturated ether group ($-OCH_2CH_2R$) in position (1). The properties and characteristics of several phosphoglycerides are summarized in the following sections.

Phosphatidyl moiety

The term phosphatidyl is used to describe the portion of the phosphoglyceride molecule containing the fatty-acid ester groups and glycerophosphate.

Phosphatidyl choline. This phospholipid, also known as lecithin, is the difatty-acid derivative of glyceryl-3-phosphorylcholine and is perhaps the most representative of all the phosphoglycerides. A lecithin molecule has a nonpolar portion (the two fatty-acid ester groups), which is insoluble in water, and a polar portion (phosphorylcholine), which is soluble in water. Since lecithin molecules contain both polar and nonpolar components, they can orient themselves in a specific way at oil–water or water–air interfaces. The nitrogen-containing component of lecithin (choline) is a strongly basic compound and has a positive charge; since phosphoric acid has a negative charge, lecithin has both acidic and basic groups and behaves as a zwitterion (*i.e.*, a dipolar ion); it does not, therefore, migrate in an electrical field.

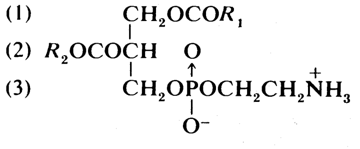
In animal tissues, the fatty-acid ester groups of lecithin are specific with regard to type and position; *e.g.*, reactions (see example below) using the enzyme phospholipase A (from snake venom) as a catalyst show that lecithins from different sources have saturated fatty acids predominately at position (1) and unsaturated fatty acids at position (2). The action of this enzyme on lecithin results in the liberation of the unsaturated fatty acid at position (2); the remainder of the molecule is called a lysolecithin. R_1 and R_2 represent hydrocarbon chains of fatty-acid ester groups in the equation below.



lysolecithin

The reaction proceeds most efficiently in a diethyl ether solution, which removes liberated fatty acids from the surfaces of the enzyme molecules. The free fatty acid represents the unsaturated fatty-acid ester group at position (2) in lecithin; the lysolecithin has a saturated fatty acid ester group at position (1). The specificity of the fatty-acid positions may act in metabolic utilization.

Phosphatidyl ethanolamine. The structure of this phosphoglyceride, found in most sources containing phosphatidyl choline, is represented below; R_1 and R_2 represent hydrocarbon chains of fatty-acid ester groups. The fatty acids in phosphatidyl ethanolamine, as in phosphatidyl choline, usually occupy specific positions; saturated fatty acids are located primarily at position (1), unsaturated fatty acids at position (2). Ethanolamine ($\overset{+}{N}H_3CH_2CH_2OH$) is a nitrogen-containing basic compound derived from the amino acid serine. This phosphoglyceride, unlike phosphatidyl choline, is not a zwitterion and exists in various ionic, or charged, forms.

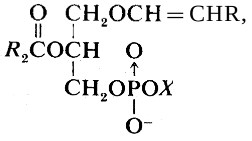


phosphatidyl ethanolamine

Phosphatidyl serine. Although amino acids other than serine may occur in natural phospholipids, phosphatidyl serine is the most representative one. It is a minor chemical component of cells, compared with phosphatidyl choline and phosphatidyl ethanolamine.

Plasmalogens. These naturally occurring phosphoglycerides, under certain conditions, release an organic compound called an aldehyde. The substituent at the (1) position in the general formula below is called a vinyl ether ($-OCH=CHR_1$); that on position (2) is a fatty-acid ester, usually of an unsaturated fatty acid. The nitrogen base (X) in most plasmalogens is ethanolamine; choline-containing plasmalogens occur in lesser amounts.

Another phospholipid, similar in general structure to plasmalogens, is found in eggs, brain tissue, and red blood cells of cows; it contains a saturated ether $(-OCH_2CH_2R)$ substituent instead of a vinyl ether.



$$\begin{array}{l} \overset{O}{\overset{\|}{R_2C}}O\overset{CH_2OCH=CHR_1}{\overset{|}{C}}H \\ \qquad\;\; \overset{|}{C}H_2O\underset{O^-}{\overset{\overset{O}{\uparrow}}{P}}OX \end{array}$$

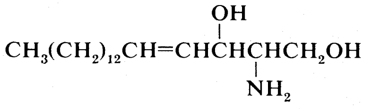
general structure of plasmalogens

Phosphatidyl inositol. Phosphatidyl inositol, which is not present in large amounts in animals or plants, contains an alcohol (inositol) with a ring structure. Several similar compounds have been isolated from brain tissue, including phosphogylcerides containing inositol polyphosphates; *e.g.,* L-myo-inositol-1,4,5-triphosphate and L-myo-inositol-2,4,5-triphosphate. More complex inositol-containing phospholipids may contain ethanolamine, tartaric acid, or the sugar galactose.

Phosphatidic acids

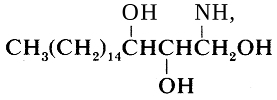
Minor phosphoglycerides. Not all of the glycerol-containing phospholipids are represented by the structures already described. Minor phosphoglycerides (*e.g.*, phosphatidic acids, phosphatidylglycerol, bisphosphatidic acid, and diphosphatidylglycerol, or cardiolipin) usually are present in cells in very small amounts, less than 2 or 3 percent, and often occur in tissues as salts. Phosphatidic acids, which are glycerol phosphate molecules containing two molecules of fatty acid, are important in the biosynthetic pathways that result in the formation of phosphoglycerides and neutral lipids. Small amounts of phosphatidic acids occur in mammals, but they are abundant in plants (*e.g.*, cabbage). Cardiolipin, or diphosphatidylglycerol, which is found in heart muscle, usually contains only unsaturated fatty acids; it has been used in the diagnosis of syphilis. Phosphatidylglycerol, which occurs in a variety of sources, may be involved in both the synthesis and breakdown of phosphatidic acids.

Sphingolipids. The sphingolipids, a group of complex lipids, contain either sphingosine or the closely related compound phytosphingosine instead of glycerol.



$$CH_3(CH_2)_{12}CH{=}CH\overset{OH}{\overset{|}{C}}H\underset{NH_2}{\underset{|}{C}}HCH_2OH$$

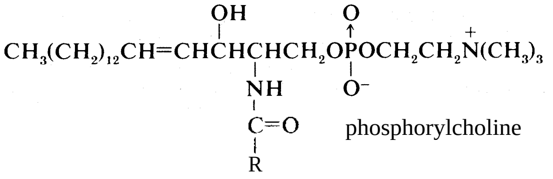
sphingosine



$$CH_3(CH_2)_{14}\overset{OH}{\overset{|}{C}}H\underset{OH}{\underset{|}{\overset{NH_2}{\overset{|}{C}}}}HCH_2OH$$

phytosphingosine

Sphingosine is found in the sphingolipids of animals, phytosphingosine in those of plants. Animal sphingolipids occur in high concentration in the brain, in the peripheral nerves, and in the myelin sheaths surrounding nerve fibres; they occur in lower amounts in several other tissues; *e.g.*, liver, blood plasma, kidney. Sphingomyelins are the most frequently encountered sphingolipids in animals: R represents the hydrocarbon chain of a fatty-acid ester group in the formula below.



$$CH_3(CH_2)_{12}CH{=}CH\overset{OH}{\overset{|}{C}}H\underset{\underset{\underset{R}{|}}{\underset{C=O}{\underset{|}{NH}}}}{\underset{|}{C}}HCH_2O\underset{O^-}{\underset{|}{\overset{\overset{O}{\uparrow}}{P}}}OCH_2CH_2\overset{+}{N}(CH_3)_3$$

a sphingomyelin

Other sphingolipids include glycosphingolipids, or glycolipids, which contain a sugar instead of the phosphoryl choline unit. A typical glycosphingolipid, present primarily in brain tissue, is the *O*-galactosido derivative of N-acyl sphingosine; it is generally called a cerebroside. The sugar component of the molecule is galactose, a hexose sugar (*i.e.*, containing six carbon atoms). The fatty acids found in cerebrosides usually contain 24 carbon atoms and have either a double bond or a hydroxyl group. Some cerebrosides contain sulfuric acid linked to galactose; these molecules are called sulfatides. Other glycosphingolipids are called gangliosides; they differ from the cerebrosides in the more complex chemical nature of the sugar group on the sphingosine molecule. The various ganglioside molecules usually differ in the number of sugar molecules.

Sialic acids

The neuraminic acid component, usually present as N-acetyl-*O*-*N*-diacetyl-, or N-glycolylneuraminic acid, is commonly known as a sialic acid. Naturally occurring sialic acids, which are found as components of gangliosides and other complex molecules, are widely distributed in animal tissues and in bacteria.

STEROLS AND CAROTENOIDS

Widely distributed in nature, sterols and carotenoids usually are closely associated with the fatty-acid-containing lipids in cells. The most commonly occurring natural sterol in vertebrates is cholesterol, which is the starting substance, or precursor, for many other important biologically active sterols (often referred to as steroids). The carotenoids, which are sometimes referred to as tetraterpenes (C_{40}), are widely distributed in plants and micro-organisms and to a lesser extent in animals, in which they probably originate in dietary sources, since animals cannot synthesize them. A frequently encountered caroteniod is β-carotene, a precursor of vitamin A alcohol, which functions in mammalian vision. Other carotenoids function in photosynthesis in plants. These lipids are discussed in detail in ISOPRENOIDS; STEROIDS.

LIPOPROTEINS

Lipids in living cells normally are associated with proteins. Lipid–protein complexes, or lipoproteins, occur either as soluble components (*e.g.*, those in the blood plasma of mammals or in egg yolk) or as more insoluble types (*e.g.*, those in membranes of cells). The discussion that follows is concerned primarily with composition, structure, and biological behaviour of the soluble lipoproteins in blood plasma, which are relatively easy to isolate and purify and thus have been studied extensively.

One outstanding feature of the association between lipids and proteins is the lack of significant chemical bonds between them. Since lipids can be separated from the protein component, it is assumed that lipoproteins are held together by weak physical forces.

Lipoproteins isolated from various sources usually fall into one of two general groups. In one group are the highly ordered lipoproteins, which are characterized by high protein and low lipid content; the other group consists of disorganized lipoproteins, which are characterized by low protein and high lipid content. A large number of different lipoproteins have been isolated; their molecular weights and lipid contents vary over a wide range. A suitable characterization of a lipoprotein includes several criteria; *e.g.*, solubility characteristics, behaviour when spun in a centrifuge, and chemical composition.

Physical characteristics. The physical character of lipoproteins resembles that of proteins. Lipoproteins that differ with respect to physical properties can be separated by differences in solubility, in electrophoretic behaviour (migration of a charged particle in an electrical field), and in centrifugal behaviour. A well-established method for separating proteins (salt fractionation) is also useful for separating lipoproteins; it has been replaced to a great extent, however, by a more effective method, centrifugation, mentioned below. Another method now largely replaced by centrifugation involves the separation, using ethanol, or protein components in human blood plasma to obtain two major lipoprotein fractions, designated as α_1-lipoprotein and β-lipoprotein.

Centrifugation

Lipoproteins were first identified in human blood when it was spun in a centrifuge at extremely high speeds (ultracentrifugation). Lipoproteins separate from other serum constituents after 20 to 24 hours of centrifugation. The lipoproteins gather as a uniform layer at the top of the centrifuge tube. If the density (mass per unit volume) of a solution containing the serum lipoproteins is adjusted by the addition of a soluble material such as a salt, so that the solution at the bottom of the tube has a higher density than that at the top, and if the mixture then is spun in an ultracentrifuge, the lipoproteins become distributed according to the densities of the components. The degree of separation of various lipoproteins, therefore, depends on the density adjustment of the solution before ultracentrifugation.

The physical characterization of low-density lipoproteins (*i.e.*, high lipid content) is often expressed in S_f values; S_f indicates low density lipoprotein flotation rate in a sodium chloride solution of density 1.063 grams per cubic centimetre at 26° C (79" F) in a unit centrifugal field and is described in Svedberg units (10^{-13} centimetres/second/dyne/gram).

Chemical nature. The chemical nature of the three groups of human-blood-plasma lipoproteins, the very low density lipoproteins (VLDL), the low-density lipoproteins (LDL_1), and the high-density lipoproteins (HDL_3), which differ from each other in various ways (*e.g.*, types and amounts of both lipids and proteins, immunochemical properties), are summarized in Table 4. Variations in age and sex of individuals is reflected by differences in distribution of low-density lipoproteins (LDL_1) and high-density lipoproteins (HDL_3) in their blood serums. The chylomicrons occur as transient components produced during the intestinal absorption of lipids.

Table 4: Characteristics of Human Blood Plasma Lipoproteins
(a composite picture)

	HDL_3	LDL_1	VLDL
Lipid, in percent	37	48	90
Component lipids, in percent of total lipid			
Phosphatidyl choline Sphingomyelin	44	24	20
Triglyceride	20	16	56
Cholesterol ester	24	12	8
Cholesterol	6	12	8
Free fatty acid	6	1	1
Protein, in percent	45	15	10
Water, in percent	18	37	...
Diameter, in angstrom units	84	185	500
Density, hydrated (grams per cubic centimetre)	1.13	1.04	0.91
Molecular weight (approx.)	300,000	2×10^6	8×10^6

PREPARATION AND ANALYSIS OF LIPIDS

The isolation and identification of lipids is a challenging task. The procedures involve the separation of lipids from other cellular constituents using organic solvents (*e.g.*, chloroform, methanol), in which the lipids dissolve. Two solvents usually are used to separate lipids from other cellular constituents. One solvent, usually methanol, separates the lipoproteins into lipid and protein components; the other, usually chloroform, dissolves the lipids. After this mixture is treated to remove low-molecular-weight compounds, such as inorganic salts, amino acids, and sugars, the lipids are separated by column chromatography into individual components. A typical chromatographic separation of the total lipids, however, does not necessarily result in separation of the components into pure compounds. Additional analytical techniques, such as thin-layer chromatography, mass spectrometry, infrared spectrometry, and optical activity, are needed to establish the identity of each component.

BIBLIOGRAPHY. COMMISSION ON BIOCHEMICAL NOMENCLATURE, "The Nomenclature of Lipids," *Biochemistry*, 6: 3287–3292 (1967), nomenclature as approved by this international commission.

Physiological significance of lipids: E.J. MASORO, *Physiological Chemistry of Lipids in Mammals* (1968), a well-written introduction to lipid chemistry and metabolism in mammals.

Chemistry of fatty acids: K.S. MARKLEY (ed.), *Fatty Acids*, 2nd rev. ed., pt. 1–3 (1960–64), an excellent and extensive discussion of fatty acid chemistry.

Derivatives of fatty acids: L.L.M. VAN DENNEN, "Phospholipids and Biomembranes," *Progress in the Chernistry of Fats and Other Lipids*, vol. 8, pt. 1 (1965), detailed treatment of the chemistry and biochemistry of phospholipids as related to cell membranes; D. CHAPMAN, *Introduction to Lipids* (1969), a brief account of the chemistry of simple and complex lipids, with particular attention to fatty acids, glycerides, phosphoglycerides, and sphingolipids.

Lipid-protein complexes: E. TRIA and A.M. SCANU (eds.), *Structural and Functional Aspects of Lipoproteins in Living Systems* (1969), an excellent treatise describing and summarizing the status of the physical and chemical nature of lipid-protein complexes in naturally occurring systems.

Preparation and analysis of lipids: "Chromatography," *Progress in the Chemistry of Fats and Other Lipids*, vol. 8, pr. 3 (1969), three excellent articles on thin-layer, paper, and column chromatography.

(D.J.H./Ed.)

Lippi, Fra Filippo

The painter Fra Filippo Lippi lived and worked in mid-15th-century Italy, in the generation immediately following that of Masaccio, one of the great initiators of the Renaissance, whose austere influence Lippi combined with the more medieval tradition of the International Gothic style. But Lippi was more than an eclectic or transitional artist. Instead, he is seen to have had a rich artistic personality of his own, a clear artistic vision, and an openness toward innovating experience, indicated not only by his pictures but also by his colourful adventures.

Filippo Lippi was born in Florence, probably in 1406. After the death of both his father and mother, he stayed with an aunt for some years, and in 1421 he pronounced the vows of a Carmelite monk at Sta. Maria del Carmine, in Florence. The Brancacci chapel of this monastery was at this time being decorated with frescoes by Masaccio. These frescoes, which were to be among the most glorious and influential paintings of the Renaissance, were Lippi's first important contact with art.

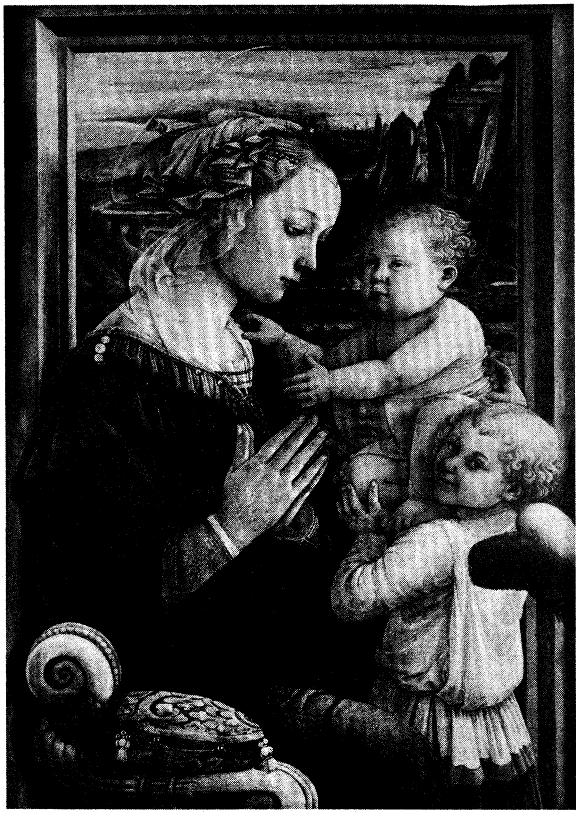
In 1432 Lippi left the monastery after having painted some frescoes in the church and in the cloister. According to the Renaissance biographer Giorgio Vasari, who wrote a lively and fanciful profile of the painter, Lippi was abducted with some companions by the Moors on the Adriatic, held as a slave for 18 months, and then freed after he painted a portrait of his owner. It is known that in 1434 the artist was at Padua. None of the works executed in the period at Padua are known, but the effect of his presence may be recognized in the paintings of others there, such as Mantegna.

Association with the Medicis

In 1437 Lippi returned to Florence, protected by the powerful Medici family, and was commissioned to execute several works for convents and churches.

The qualities he acquired during his years of travel are affirmed with clarity in two works of 1437, immediately after he returned from Padua: "The Virgin and Child Between SS. Frediano and Augustin" and the "Madonna and Child." In both of these altarpieces, the influence of Masaccio is still evident, but it is absorbed into a different style, having the pictorial effect of bas-relief, rendered more evident by lines, so that it resembles the reliefs of the sculptors Donatello and Jacopo della Quercia. In these works, the colour is warm, toned down with shadings, approaching the limpid chromatics of his great contemporary Fra Angelico. Still further testimony of Lippi's development is the painting "The Annunciation" (Galleria Nazionale d'Arte Antica), formerly believed to be a late work but now dated between 1441 and 1443. It is composed in a new way, using the newly discovered effects of perspective and skillful contrasts between colour and form; the suggested movement of the light garments of the two frightened young girls at the door is rendered with such sensitivity as to anticipate Botticelli.

A famous altarpiece of the same time, Lippi's well-known "Coronation of the Virgin," is a complex work crowded with figures. The celebrated altarpiece is so sumptuous in appearance that it seems to have been painted in competition with Angelico; it marks a historic



"The Madonna and Child with Two Angels," by Fra Filippo Lippi, c. 1465. In the Uffizi, Florence.
Alinari

point in Florentine painting in its success in uniting as one scene the various panels of a polyptych.

The altarpieces are characterized by a solemnity of composition that is absent from the paintings in which he developed a typical motive of 15th-century Florentine art: the Madonna with the Child at her breast. The masterpiece of these is "Madonna with Child and Scenes from the Life of Mary," a circular painting now in the Pitti Palace in Florence; it is a clear and realistic mirror of life, transfigured in a most intimate way, and it had a great effect on Renaissance art.

A second "Coronation of the Virgin," executed about 1445, displays a marked change in the style of Lippi—from the plastic values suggested by his study of Masaccio to the serene chromatics of Angelico.

Elopement with Lucrezia Buti

In 1442 Lippi had been made rector of the church of S. Quirico at Legnaia. His life, however, became constantly more eventful, and tradition has given him the reputation (borne out in great part by documents) of a man dominated by love affairs and impatient of methodical or tranquil conduct. His adventures culminated in 1456 in his romantic flight from Prato, where he was painting in the convent of the nuns of Sta. Margherita, with a young woman of the convent, Lucrezia Buti. The Pope later gave permission to the former priest-painter to marry her, and from this union was born a son, Filippo, called Filippino, who was to be one of the most noted Florentine painters of the second half of the 15th century.

The bright and active city of Prato, a short distance from Florence, was the second home of Filippo Lippi. He returned to Prato often, staying there for long periods, painting frescoes and altarpieces. Accompanied by Fra Diamante, who had been his companion and collaborator since he was a young man, Lippi began to redecorate the walls of the choir of the cathedral there in 1452. He returned in 1463 and again in 1464, remaining in the city this time until 1467. At the centre of his activity in Prato stand the frescoes of the cathedral, with the four Evangelists and scenes from the lives of St. John the Baptist and St. Stephen. Perhaps the most solemn scene of the life and death of St. Stephen is the burial; at the sides of the funeral bed of the Saint stand a crowd of prelates and illustrious persons in mourning, among them Cardinal Carlo de' Medici, Fra Diamante, and the artist himself.

In 1467 Lippi and Fra Diamante left for Spoleto, where he had received a commission, through the Medici family, for another vast undertaking: the decorations and frescoes of the choir of the cathedral, which included the "Nativity," the "Annunciation," the "Death of Mary," and—in the centre of the vault of the apse—the "Coronation." These frescoes were Lippi's final work; they were interrupted by his death, for which there are two documented dates—according to the monks' necrology of Sta. Maria del Carmine in Florence it is October 8, 1469, and in the archives of Spoleto it is recorded as October 10, 1469. The Medici had a splendid sepulchre, designed by his son, erected for him in the Cathedral of Spoleto.

Assessment

Posthumous judgments of Filippo Lippi were often coloured by the traditions of his adventurous life. Moreover, his works have been criticized from time to time for their borrowings from other painters; nevertheless, it has also been recognized that his art was not diminished but rather enriched and rendered more balanced by what he took from Masaccio and Fra Angelico. He was constantly seeking the techniques to realize his artistic vision and the new ideas that made him one of the most appreciated artists of his time.

The 20th-century critic Bernard Berenson, who maintained that Lippi's true place as an artist was among the "painters of genius," also described him as "a high-class illustrator," intending by this to underline the importance of expressive content and the presentation of reality in his works. Later critics have recognized in Lippi a "narrative" spirit that reflected the life of his time and translated into everyday terms the ideals of the early Renaissance.

MAJOR WORKS

"The Reform of the Carmelite Rule" (fresco, 1432; Sta. Maria del Carmine, Florence); "Madonna of Humility" (c. 1432; Castello Sforzesco, Milan); "Madonna and Child (1437; Galleria Nazionale d'Arte Antica, Rome); "The Virgin and Child Between SS. Frediano and Augustin" (1437; Louvre, Paris); "Madonna and Child Enthroned (*c*. 1437; Metropolitan Museum of Art, New York); "Annunciation" (*c*. 1437–40; S. Lorenzo, Florence); "The Coronation of the Virgin" (1441–47; Uffizi, Florence); "The Madonna Enthroned with Saints" (1441–42; Uffizi); "Virgin and Child (1441–42; Museo Mediceo, Florence); "The Annunciation" (1441–43; Galleria Nazionale d'Arte Antica, Rome); "Annunciation" (*c*. 1442–47; Alte Pinakothek, Munich); "St. Lawrence Enthroned with SS. Cosmo and Damian and Alessandro Alessandri and Two of His Sons; St. Benedict; St. Anthony Abbott" ("Alessandri Altarpiece"; *c*. 1442–47; Metropolitan Museum of Art, New York); "Adoration of the Magi" (*c*. 1445–47; National Gallery of Art, Washington, D.C.); "Coronation of the Virgin" (*c*. 1445; Vatican Museum, Rome); "St. Bernard's Vision of the Virgin" (1447; National Gallery, London); choir frescoes (1452–64; cathedral, Prato); "Madonna with Child and Scenes from the Life of Mary" (1452; Pitti Palace, Florence); "The Virgin Enthroned and Two Saints" (1453; Museo Comunale, Prato); "The Adoration of the Child with St. Hilary" (*c*. 1455; Uffizi); "Seven Saints" (1458–60; National Gallery, London); "The Virgin Adoring the Child" (*c*. 1460; Staatliche Museen Preussischer Kulturbesitz, Berlin); "The Adoration of the Child with St. Bernard" (*c*. 1463; Uffizi); "The Madonna and Child with Two Angels" (*c*. 1465; Uffizi); choir and apse frescoes (commissioned 1466, begun 1467–69; cathedral, Spoleto).

BIBLIOGRAPHY. MARY PITTALUGA, "Fra Filippo Lippi," in the *World Encyclopedia of Art*, vol. 9, col. 257–261 (1964), a biographical and critical synopsis with modern bibliography; and *Filippo Lippi* (1949), includes a critical evaluation (in Italian) of the artist and precise attributions; EDWARD C. STRUTT, *Fra Filippo Lippi* (1901), the first systematic biography with comments on the principal works; BERNARD BERENSON, *The Study and Criticism of Italian Art*, 2nd Series (1902) and *Italian Pictures of the Renaissance* (1932), contain important critical evaluations of the artist; I.B. SUPINO, *Fra Filippo Lippi* (1902), in Italian; HENRIETTE MENDELSOHN, *Fra Filippo Lippi* (1909), in German; GIUSEPPE FIOCCO, *La pittura toscana del Quattrocento* (1945); ROBERT OERTEL, *Fra Filippo Lippi* (1942), a critical study (in German) with attributions; MARIO SALMI (ed.), *Filippo Lippi: Gli Affreschi nel duomo di Prato* (1944), with illustrations of particular frescoes and comments and notes.

(V.M.)

Liquid State

Most simple substances can exist in any of the three states of matter—solid, liquid, and gas—within certain ranges of temperature. A solid is almost rigid; any stress that a solid object is subjected to is proportional to the strain set up in it, and stresses that are sufficiently small are withstood by the body for indefinite periods of time. Two kinds of stress are important in this definition: tensile stress tends to pull the body apart, while shear stress tends to press one part past another. On a microscopic level, a crystalline solid is characterized by a high degree of spatial order in the arrangement of the molecules (see SOLID STATE OF MATTER). In the liquid and gas states (see GASEOUS STATE), collectively called the fluid states, stress is proportional to the rate of strain, and a shearing stress produces a viscous flow of parts. The spatial order of the molecules is less obvious than that of a solid; in a gas of low density it is almost negligible, while in a dense gas or in a liquid it can be described only in terms of averages over many different configurations. Such averages are called probability distribution functions and are treated below.

Types of liquid state

General considerations. Clear distinctions between the one solid and the two fluid phases, gas and liquid, hold only for those substances whose molecules are composed of a small number of atoms. When the number exceeds about 20, the liquid may often be cooled below the true melting point to form a glass (see GLASSY STATE): which has many of the mechanical properties of a solid but lacks crystalline order. If the number of atoms in the molecule exceeds about 100–200, then the classification into solids, liquids, and gases ceases to be useful. At low temperatures such substances are usually glasses or amorphous solids, and their rigidity falls with increasing temperature—*i.e.*, they do not have fixed melting points; some may, however, form true liquids. With these large molecules, the gaseous state is usually not attainable, because they decompose chemically before the temperature is high enough for the liquid to evaporate. Synthetic and natural high polymers, such as nylon and rubber, behave in this way.

If the molecules are large, rigid, and either roughly planar or linear, as in cholesteryl acetate or *p*-azoxyanisole, the solid may melt to an anisotropic liquid (not uniform in all directions) in which the molecules are free to move about but have great difficulty in rotating. Such a state is called liquid crystal, and the anisotropy produces changes of the refractive index (a measure of the change in direction of light when it passes from one medium into another) with the direction of the incident light and hence leads to unusual optical effects; however, no inorganic compounds and fewer than one in 200 of the known organic compounds form liquid crystals.

The theory of normal liquids is. therefore, predominantly the theory of the behaviour of substances consisting of simple molecules.

Thus, a liquid lacks both the strong spatial order of a solid, though it has the high density of solids, and the absence of order of a gas that results from the low density of gases; *i.e.*, gas molecules are relatively free of each other's influence. The combination of high density and of partial order in liquids has led to difficulties in developing quantitatively acceptable theories of liquids. Understanding of the liquid state, as of all states of matter, came with the kinetic molecular theory, which stated that matter consisted of particles in constant motion and that this motion was the manifestation of thermal energy. The greater the thermal energy of the particle, the faster it moved.

Transition to the liquid state

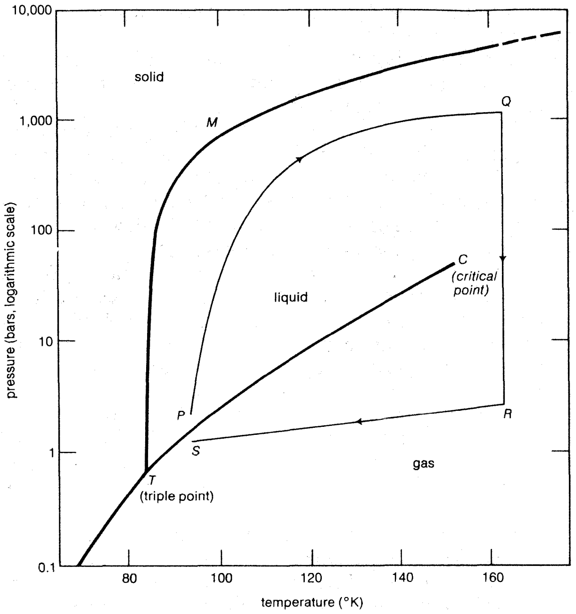
In very general terms, the particles may be molecules, atoms, charged atoms called ions, or electrons. In a gas, the particles are far enough away from one another and are moving fast enough to escape each other's influence, which may be of various kinds, such as the attraction or repulsion due to electrical charges, and specific forces of attraction that involve the electrons orbiting around atomic nuclei. The motion of particles is in a straight line, and the collisions that result occur with no resulting loss of energy although an exchange of energies may result between colliding particles. When a gas is cooled, its particles move more slowly, and those that are slow enough to linger in each other's vicinity will coalesce, because a force of attraction will overcome their lowered kinetic energy (energy of motion) and, by definition, thermal energy. Each particle, when it joins the liquid state with others, gives up a measure of heat called the latent heat of liquefaction, but each continues to move at the same speed within the liquid as long as the temperature remains at the condensation point. The distances the particles can travel in a liquid without colliding are of the order of molecular diameters. As the liquid is cooled, the particles move more slowly still, until at the freezing temperature the attractive energy produces so high a density that the liquid freezes into the solid state. They continue to vibrate, however, at the same speed, as long as the temperature remains at the freezing point, and the freezing process releases their latent heat of fusion. Heating a solid will provide the particles with the heat of fusion necessary to escape one another's influence enough to move about in the liquid state, and further heating will provide the liquid particles with their heat of evaporation, which will enable them to escape one another altogether and form the vapour of the liquid, or the gaseous state. Of course, this starkly simplified view of the states of matter ignores a great many complicating factors, most important being the fact that no two particles need be moving at the same speed in a gas or liquid or solid and the related fact that even in a solid some particles may have acquired all the energy they need to be gas particles, while even in a gas some particles may be practically motionless for a brief span of time. It is the average kinetic energy of the particles that has to be considered, together with the fact that the motion is random. At the interface between liquid and gas and between liquid and solid, an exchange of particles is always taking place, slow gas molecules condensing at the liquid surface and fast liquid molecules escaping into the gas. An equilibrium state is reached in any closed system, so that the number of exchanges in either direction is the same. Because the kinetic energy of particles in the liquid state can be defined only in statistical terms (*i.e.*, every possible value can be found), discussion of the liquid state, as of the gaseous state, involves formulations in terms of probability functions.

THE NATURE AND STRUCTURE OF LIQUIDS AT EQUILIBRIUM

Gibbs phase rule

Behaviour of liquids at equilibrium. The number of phases (or distinct states of a substance) that are present in a system at equilibrium is governed by a natural law called the phase rule by J. Willard Gibbs. All the properties of a homogeneous phase of a single substance (*e.g.*, a pure liquid) are determined if, in the absence of external forces, the two variables of pressure and temperature are specified. Such a system is, therefore, said to have two degrees of freedom. If there is more than one phase (*e.g.*, ice and water) or more than one substance present (*e.g.*, water and alcohol) or both, then the number of degrees of freedom (symbolized by F) is related to the number of phases (symbolized P) and the number of substances or components (symbolized C), by Gibbs phase rule, which states that the sum of the number of phases and degrees of freedom equals the sum of the number of components and the number two. This is written $P + F = C + 2$. Hence, for a single substance (when $C = 1$), the number of coexisting phases cannot exceed three, since F cannot be negative, though it can be zero. Solid, liquid, and gas can exist together in equilibrium only for one fixed pair of values of pressure and temperature, and at this point $F = 0$, since, being fixed, there is no freedom to choose the values of these variables. With two phases—solid and liquid or liquid and gas—the equation becomes $2 + F = 1 + 2$; hence, $F = 1$; *i.e.*, the system has one degree of freedom: either the pressure or the temperature can be chosen at will (within certain ranges), and the other is then fixed by the nature of the substance.

These relations are summarized in the Figure, the phase diagram of argon, in which one phase is shown as an



Phase diagram of argon.
Heavy lines show the boundaries between liquid and gas (*TC*) and between liquid and solid (TM). Line PQRS describes path from typically liquid state (P) to typically gaseous state (*S*) (see text).

area, two as a line, and three as the meeting of the lines at the triple point, T. The diagram shows that the liquid phase is merely a part of a single fluid state. There are abrupt changes in physical properties (*e.g.*, density, heat capacity, viscosity, dielectric constant) on crossing the lines that separate the liquid from the solid on one side and from the gas on the other. The second line, the vapour-pressure curve, ***TC*** (obtained by measuring the pressure at every temperature), does not continue to arbitrarily high temperatures but ends at the temperature and pressure that describe the critical point, C. At the critical point, two phases in equilibrium become identical and form one phase. A liquid at the pressure and temperature indicated at point *P* can be heated without changing its density along the line *PQ;* at *Q*, the pressure is lowered, while the temperature is kept constant, and the volume of the liquid is expanded along *QR;* at R, the gas is cooled along *RS*, to *S*, where it is changed from a liquid to a gas without undergoing that abrupt change of properties associated with boiling or evaporation. The path *PQRS* demonstrates the essential continuity of state between liquid and gas, which differ in degree but which comprise together the one fluid state. Strictly speaking, the term liquid should be applied only to the more dense of the two phases on the line TC, but it is generally extended to any dense fluid state at low temperatures; that is, to the area lying within the angle ***CTM***.

Phases in equilibrium

The line ***TM*** is the melting curve and is initially much steeper than the vapour-pressure curve; hence, as the pressure is changed, the temperature does not change much, and the melting temperature is little affected by pressure. No substance has been found to have a critical point on this line, and there are theoretical reasons for supposing that it continues indefinitely to high temperatures and pressures, until the substance is so compressed that the molecules break up into atoms, ions, and electrons. At pressures above 10^6 bars (the unit of pressure; one bar is approximately equal to the pressure of the atmosphere at sea level), it is believed that most substances pass into a metallic state, but steady pressures of this magnitude are not yet obtainable in the laboratory.

There are a few substances, such as water and bismuth, for which the density of the solid is less than that of the liquid at the triple point. For these two substances, the volume change for solid to liquid is negative; hence, the melting temperatures fall with increasing pressure. With water, this negative volume change (*i.e.*, shrinking) persists to 2.1 kilobars and —22" C (—8° F), at which point the normal form of ice changes to a more dense form, and, thereafter, the change in volume is positive.

It is often possible to cool a gas at constant pressure to a temperature lower than that of the vapour-pressure line without producing immediate condensation, since the liquid phase forms readily only in the presence of suitable nuclei, generally particles of dust or ions, about which the drops can grow. Unless the gas is scrupulously cleaned, these nuclei are always present; a supercooled vapour state is unstable, and liquid will ultimately condense. It is similarly possible to superheat a liquid to a temperature where, though still a liquid, the gas is the stable phase. Again, this occurs most readily with clean liquids heated in smooth vessels, because gas bubbles will collect around foreign particles or sharp points. When the superheated liquid changes to gas, it does so with almost explosive violence, a phenomenon familiar as the "bumping" noise of a heated liquid. A liquid also may be supercooled to below its freezing temperature; this is quite easy to do under carefully controlled conditions. It is not possible, however, to superheat a solid.

Triple-point temperature

Triple-point temperatures range from 14" K (0° K equals —273.15" C [—460" F]) for hydrogen to temperatures too high for accurate measurement; *e.g.*, around 4,500" K for diamond. Triple-point pressures are generally low, that of carbon dioxide at 5.2 bars being one of the highest. Most are around 10^{-3} bar, and those of some hydrocarbons are as low as 10^{-7} bar. The normal melting point of a substance is defined as the melting temperature at a pressure of one atmosphere (exactly equivalent to 1.01325 bars); it differs little from the triple point because of the steepness of melting lines (TM in the Figure). Critical temperatures (maximum temperature at which a gas can be liquefied by pressure) range from helium, at 5.2° K, to temperatures too high to measure. Critical pressures (at which liquid and vapour phases are identical) are generally about 40–100 bars. The normal boiling point is the temperature at which the vapour pressure reaches one atmosphere. The normal liquid range is defined as the temperature interval between the normal melting point and the normal boiling point, but such a restriction is artificial, the true liquid range being from triple point to critical point. Substances whose triple-point pressures are above atmospheric (*e.g.*, carbon dioxide) have no normal liquid range but sublime (*i.e.*, pass from solid phase to gas phase) at atmospheric pressure.

Many of the properties of a liquid near its triple point are closer to those of the solid than to those of the gas. It has a high density (typically 0.5–1.5 grams per cubic centimetre), a high refractive index (a measure of the change in direction of light rays passing from one medium into another; for liquids the index varies from 1.3 to 1.8), a high heat capacity at constant pressure (two to four joules per gram per °K, one joule being equal to 0.239 calorie), and a low compressibility ($0.5–1 \times 10^{-4}$ per bar). The compressibility falls to values characteristic of a solid (0.1×10^{-4} per bar, or less) as the pressure increases. A simple and widely used equation describes the change of specific volume with pressure. If $V(p)$ is the volume at pressure p, $V(0)$ is volume at zero pressure, and A and B are positive parameters (constants whose values may be arbitrarily assigned), then the difference in volume resulting from a change in pressure equals the product of A, the pressure, and the volume at zero pressure, divided by the sum of B and the pressure. This is written:

$$V(0) - V(p) = \frac{ApV(0)}{(B+p)}.$$

The pressure B is close to the pressure at which the compressibility has fallen to half its initial value and is generally about 500 bars for liquids near their triple points. It falls rapidly with increasing temperature.

As a liquid is heated along its vapour-pressure curve, TC, its density falls and its compressibility rises. Conversely, the density of the saturated vapour in equilibrium with the liquid rises; *i.e.*, the number of gas molecules in a fixed space above the liquid increases. Liquid and gas states approach each other with increasing rapidity as the temperature approaches C, until at this point they become identical and have a density about one-third of that of the liquid at T. The change of saturated-gas density (ρ_g) and liquid density (ρ_l) can be expressed by a simple equation when the temperature is close to the critical. If ρ_c is the density at the critical temperature T_c, then the difference between densities equals the difference between temperatures raised to a factor called beta, β. This is written:

$$(\rho_l - \rho_c) = (\rho_c - \rho_g) = (T_c - T)^{\beta},$$

and β is about 0.34. The compressibility and the heat capacity of the gas at constant pressure (C_p) become infinite as T approaches T_c from above along the path of constant density. The infinite compressibility implies that the pressure no longer restrains local fluctuations of density. The fluctuations grow to such an extent that their size is comparable with the wavelength of light, which is therefore strongly scattered. Hence, at the critical point, a normally transparent liquid is almost opaque and usually dark brown in colour. The classical description of the critical point and the results of modern measurement do not agree in detail, but recent considerations of thermodynamic stability show that there are certain regularities in behaviour that are common to all substances.

Shape of drops

Between a liquid and its vapour there is a dividing surface that has a measurable tension; work must be done to increase the area of the surface at constant temperature. Hence, in the absence of gravity or during free fall, the equilibrium shape of a volume of liquid is one that has a minimum area; *i.e.*, a sphere. In the Earth's field this shape is found only for small drops, for which the gravitational forces, since they are proportional to the volume, are negligible compared with surface forces, which are proportional to the area. The surface tension falls with rising temperature and vanishes at the critical point. There is a similar dividing surface between two immiscible liquids, but this usually has lower tension. It is believed that there is a tension also between a liquid and a solid, though it is not directly measurable because of the rigidity of the solid; it may be inferred, however, under certain assumptions, from the angle of contact between the liquid and the solid; *i.e.*, the angle at which the liquid's surface meets the solid. If this angle is zero, the liquid surface is parallel to the solid surface and is said to wet the solid completely.

Structure. A liquid lacks regular arrangement of its molecules; *i.e.*, it lacks the regular crystalline order of a solid. The molecular structure of a liquid, therefore, can be described and measured only on the basis of probability. Probability is computed by a mathematical operation that yields distribution functions; a distribution function measures the probable distribution of some property of molecules through the liquid. Distribution functions that describe liquid structure are formulated as follows.

Distribution functions for liquids

If the volume of a minute quantity of liquid that may be called an element of the liquid, symbolized by dr, is small compared with the volume of a molecule and is in a fixed position symbolized by r, sometimes it will contain the centre of a molecule, and sometimes it will not. The probability that it will do so, in a liquid at equilibrium, is the product of the number of molecules in the total volume, symbolized by N, and of the ratio of dr to the total volume, V. Written out, this equals Ndr/V. The ratio N/V is called the number density and is written as n. Thus, probability is the product ndr. If two elements of volume are considered, dr_1, or dr_2, and if they are separated by a distance symbolized as r_{12}, which is large compared with the size of a molecule, the probability that each such element of volume contains the centre of a molecule is given by the product of the probabilities of independent occupancy; that is, by $ndr_1 \times ndr_2$, or $n^2dr_1dr_2$. If r_{12}, the distance between r_1 and r_2, is comparable with the size of a molecule, the probabilities are not independent, and the joint probability can be written as the product of the probabilities of independent occupancy (of a particular state) and a function of r_{12} denoted by $g^{(2)}(r_{12})$. This function is called the pair distribution function and is the simplest measure of the structure of a liquid.

If the pair distribution function equals one—that is, if $g^{(2)} = 1$—this value corresponds to no order or structure, since the positions of the molecules are mutually independent. Such a limit of no structure at all is [illegible] in a gas in which the molecules are far apart; *i.e.*, in effect, a hypothetically perfect gas. In a more dense gas, where molecules do affect one another, it is found that a useful approximation to $g^{(2)}$ is a value called the Boltzmann factor of the energy of interaction of a pair of molecules, symbolized as $u(r)$. The energy can be found from the properties of a dilute gas, from calculations using quantum mechanics, and in some cases from molecular-beam-scattering experiments. For the noble gas argon, which is highly unreactive, $u(r)$ is large and positive (*i.e.*, the molecules repel each other) if r is less than 3.3 angstroms (one angstrom equals 10^{-7} of a millimetre), and $u(r)$ is negative (*i.e.*, the molecules attract one another) if the distance is greater than 3.3 angstroms; the value reduces to zero at large separations of the molecules.

More detailed mathematical calculations in terms of molecular forces are beyond the scope of this article.

In a liquid, the pair distribution function $g^{(2)}$ is a more complex function of r and depends on both the number density, n, and the absolute temperature, T. When r is less than the collision diameter (3.3 angstroms for argon), $g^{(2)}$ is essentially zero, since two molecules cannot "overlap." At slightly larger separations—when $g^{(2)}$ is greater than one—there is a more than average chance of finding a pair of molecules. This enhancement arises in part from the Boltzmann factor of $u(r)$ but, more importantly, from the geometric necessity of there being many pairs at such separations in a liquid whose density is little less than that of a solid. Minor maxima and minima at larger separations are the remnants in the liquid of the much stronger extrema found in the solid (for which the pair distribution function $g^{(2)}$ along a crystal axis consists of a set of many narrow peaks separated by regions in which it is zero). Thus, a liquid lacks the long-range order of a solid but shows appreciable order at short separations.

Importance of distribution function

A knowledge of the pair distribution function $g^{(2)}$ is the minimum information needed for the discussion of the thermodynamic properties of a liquid, since molecules interact predominantly in pairs, through the potential energy u (r), and since $g^{(2)}$ describes the statistics of the distribution of such pairs. The macroscopic properties of internal or cohesive energy of a liquid, u, and of the departure of the pressure from its value in a perfect gas can be calculated from averages over $g^{(2)}$. These results show that a knowledge of $g^{(2)}$ is the link between the microscopic energies and forces and the macroscopic properties.

The theory of liquids at equilibrium reduces to a large degree to the measurement and calculation of the pair distribution function $g^{(2)}$. 'Finer details of the structure can be codified formally in higher distribution functions that specify the probabilities of there being molecules at three specified positions, but little is known about these functions.

Pair distribution function. There are two methods of measuring the pair distribution function $g(r)$ (from which the superscript may be omitted): first, by X-ray or neutron diffraction from simple fluids and, second, by computer simulation of the molecular structure and motions in a liquid. In the first, the liquid is exposed to a specific, single type (monochromatic) radiation and the observed results are then subjected to a mathematical treatment known as a Fourier transform.

The second method of obtaining the pair distribution function $g(r)$ supposes that the energy of interaction, $u(r)$, for the liquid under study is known. A computer model of a liquid is set up, in which between 100 and 1,000 molecules are contained within a cube. There are now two methods of proceeding, by Monte Carlo calculation or by what is called molecular dynamics; only the latter need be discussed here. Each molecule is assigned a

Use of a computer model

random position and velocity, and Newton's equations of motion are solved to calculate the path of each molecule in the changing field of all the others. A molecule that leaves the cell is deemed to be replaced by a new one entering through the corresponding point on the opposite wall with equal velocity. After a few collisions per molecule, the distribution of velocities conforms with equations worked out by the British physicist James Clerk Maxwell, and after a longer time the mean positions are those appropriate to the density and mean kinetic energy (*i.e.*, temperature) of the liquid. Functions such as the pair distribution function $g(r)$ can now be evaluated by taking suitable averages as the system evolves in time. Since 1958 such computer experiments have added more to the knowledge of the molecular structure of simple liquids than all the theoretical work of the previous century.

The theory of liquids at equilibrium cannot be discussed without extensive mathematical analysis. It is, to a large degree, the theory of the calculation of the total correlation function, $h(r)$, for which $h(r)$ is $g(r) - 1$, in terms of the pair potential $u(r)$, which, for the purposes of this theory, is assumed to be known. Equations can be derived that can be interpreted by saying that the total function h is composed of a direct term, c, and of chains of direct terms through all other molecules in the liquid. The direct function c can be determined from X-ray-diffraction experiments (by means of a Fourier transform) and also from other calculations, including computer results based on the assumption that a liquid behaves to a first approximation as a hard-sphere fluid.

The structure of a simple liquid (*e.g.*, argon) is determined, to a large degree, by the problem of packing at high density molecules with strong repulsive forces between them. The above approach is, of course, not a complete theory of liquids because of its basis in computer experiments. In any event, molecules as polar as water, for example, cannot be included quantitatively in any general scheme—such liquids must be discussed individually. No simple theory can account for the fact that water has a maximum density at 4° C (39° F) and expands on cooling below that temperature.

MECHANICAL AND THERMAL PROPERTIES OF LIQUIDS

Transport properties. A liquid in which there exists a gradient of the concentration, temperature, or velocity is not at equilibrium and responds to these variations, which may be considered "forces," by an appropriate movement of a quantity of matter, energy, or momentum through a unit area, called a flux. Unless the gradients are extremely large (as in a shock wave), the flux is proportional to the force, and this relation is embodied in several laws of diffusion and conduction, for which the mathematical formulations are beyond the scope of this article.

Velocity is a vector property, and so the laws governing the transport of momentum and, hence, of viscous flow are more complicated than those of diffusion and conduction. The simplest case is the fluid that obeys Newton's laws, for which stress is proportional to rate of strain (see above).

Diffusion

Diffusion coefficients are about 10^{-9} m^2/s near the triple point and rise with increasing temperature and decreasing density. The thermal conductivity of a liquid is about ten times that of the gas. It generally rises with increasing temperature and appears to become infinitely large at the critical point in the same way as the heat capacity at constant volume. The thermal conductivities of liquid metals are unusually high, since most of the energy is carried by the free electrons. Hence, the thermal conductivity is closely related to the electrical conductivity.

The shear viscosity varies with the molecular complexity of the liquid; thus, in argon and liquid air it is a fraction of that in water, and in heavy oils it rises to hundreds of times that. The bulk viscosity was not detected in a monatomic liquid until 1965 and is discussed below in connection with the speed of sound.

The theory of transport properties (that is, of the relation between certain properties such as diffusion and viscosity and the mass, structure, and interaction potential of the molecules) is in a less developed state than that of the equilibrium properties. In a gas, the mechanism of transport is by convection on a molecular scale; that is, matter, energy, and momentum are carried by the actual movement of molecules. In a liquid, this is true also for diffusion, but the the major contribution to viscosity and thermal conductivity is through collisions. When a pair of molecules collide there is virtually instantaneous transfer of energy and momentum over a distance equal to the molecular diameter. In a gas, this distance is negligible compared with the mean free path of molecular travel, and so such transfer is unimportant; in a liquid with the molecules touching, it is dominant. The molecules spend much of their time moving in "cells" to which they are confined by their neighbours, but this restriction on their movement does not hinder transfer by collisions. Hence, diffusion is slow in a liquid, but the coefficients of viscosity and thermal conductivity are large compared with those of a gas. This simple division into transfer by convection and transfer by collision has been put into more quantitative form by making simplifying assumptions about the movement of molecules within their cells and about the activation energy needed to escape to a neighbouring cell. The resulting theory, however, although pictorially attractive, is not a satisfactory route to a theory of transport processes.

A proper theory of transport properties must proceed through the distribution functions. Since the liquid is not at equilibrium, these will be functions not only of the separations of the molecules but also of their velocities; moreover, they will change with time.

Some of the most useful results have come not from a direct attack on the central theoretical problem but from the use of theory to interrelate different properties of the liquid. One approach, started by Albert Einstein and continued by others, introduces a correlation between a function at time zero and its value at a later time, t.

Transport properties of dense fluids

A theory of the transport properties of dense fluids was developed by assuming (1) that the molecules were hard spheres, so that only binary collisions occur, and (2) that the velocities of two colliding molecules are uncorrelated before collision. Velocity correlation is neglected in the description of the fluid. This approach is satisfactory at moderate densities but less so as the true liquid density is reached. The cells in which molecules are effectively confined at high densities mean that two neighbouring molecules may well collide more than once before they diffuse apart. Hence, their velocities are not uncorrelated at the second and subsequent collisions. Attempts to remove this defect lead to mathematical difficulties that do not lend themselves to ready summary; these are now an active field of research.

Propagation of sound in liquids. A sound wave is a series of longitudinal compressions and expansions that travels through a liquid at a speed of about one kilometre per second, or about three times the speed of sound in air. The compressions and expansions are adiabatic (changes that take place without transfer of heat) and reversible if the frequency is not too high. Conduction of energy from the hot (compressed) to the cold (expanded) regions of the liquid introduces irreversible effects, which are dissipative, and so such conduction leads to the absorption of the sound. A longitudinal compression (in the direction of the wave) is a combination of a uniform compression and a shearing stress (a force that causes one plane of a substance to glide past an adjacent plane). Hence, both bulk and shear viscosity also govern the propagation of sound in a liquid.

These dissipative effects are small in simple liquids, which may be virtually "transparent" to sound at frequencies up to 100 megahertz (10^9 cycles per second). At these high frequencies, that elusive coefficient, the bulk viscosity, can be observed. More complex liquids, such as hydrocarbons, esters, and even so simple a liquid as carbon disulfide, show strong absorption of sound below one megahertz. The absorption is accompanied by dispersion—that is, by the change of speed with frequency—and that can be described formally in terms of a frequency-dependent bulk viscosity. The cause is either a failure of the vibrational modes of the molecule to acquire and lose energy, if the time of heating and cooling is less than 10^{-6} of a second, or a structural change of the liquid on compression, which needs a similar time to occur. The first is certainly the cause in carbon disulfide, and the latter is exemplified by the association and dissociation of acetic acid into paired molecules and hence the strong absorption of sound of high frequency in this liquid. Hydrocarbons may combine both effects.

Electrical and magnetic properties. If a liquid is placed in a static electric field, the field exerts a force on any free carriers of electric charge in the liquid, and the liquid, therefore, conducts electricity. Such carriers are of two kinds, mobile electrons and ions; *i.e.*, atoms or molecules that have lost or gained electrons. The former are present in abundance in liquid metals, which have conductivities that are generally about one-third of the conductivity of the solid. The fall in conductivity on melting arises from the greater disorder of the positive ions in the liquid and hence their greater ability to scatter electrons. The contribution of the ions is small, less than 5 percent in most liquid metals, but is the sole cause of conductivity in molten salts (consisting of negative and positive ions) and in their aqueous solutions. Such conductivities vary widely but are much lower than those of liquid metals.

Non-ionic liquids

Non-ionic liquids (composed of molecules that do not dissociate into ions) have negligible conductivities, but they are polarized by an electric field; that is, the liquid develops positive and negative poles and also a dipole moment (pole strength times distance between poles) that is oriented against the field, from which the liquid therefore acquires energy. This polarization is of three kinds; in the first, called electron polarization, the electrons in each atom are displaced from their usual positions, thus giving each molecule a small dipole moment. The molecules act independently, and their reaction to the static electric field is the same as to the fluctuating electrical component of the field of electromagnetic radiation. The contribution of electron polarization to the dielectric constant (a measure of the electric stress) of the liquid is numerically equal to the square root of its refractive index (the measure of the change in direction when light enters the liquid). A second effect of an electric field is to produce atomic polarization, in which there is a relative change in the mean positions of the atomic nuclei within one molecule. This generally small effect is observed at radio frequencies but not at optical, and so it is missing from the refractive index. The third effect is that of orientation polarization. If the molecules have permanent dipole moments, they are partially aligned by the field and contribute heavily to the polarization. Thus, the dielectric constant of a nonpolar liquid, such as a hydrocarbon, is about 2, that of a weakly polar liquid, such as chloroform or ethyl ether, about 5, while those of highly polar liquids, such as ethanol and water, lie between 25 and 80.

The partial alignment occurs in a static electrical field or in one whose frequency is less than the reciprocal of the time it takes the polar molecule to rotate. Hence, orientation polarization does not contribute to the refractive index. The degree of alignment is less than might be expected from independent dipoles because of the polar intermolecular forces. The full effects of the independent dipoles can be observed only in the gas phase or, in a reasonable measure, in a dilute solution of a polar substance in a nonpolar solvent.

Most liquids are diamagnetic (opposing the magnetizing force). The exceptions are those whose molecules have unpaired electrons (*e.g.*, nitric oxide) and the liquid metals. Even those metals that are ferromagnetic (strongly magnetic) in the solid state are only paramagnetic (slightly magnetic) in the liquid state.

Absorption and scattering of radiation. Knowledge of the structure of liquids is derived principally from the study of their interaction with radiation. One example, the determination of $g(r)$ at equilibrium from X-ray or neutron diffraction, has been touched upon above under ***Pair distribution function.*** In the more general case of the scattering of radiation, the cause of the scattering is different for different kinds of radiation; thus, light is scattered by the outer electrons of an atom, X-rays by all the electrons, and neutrons by the nuclei. If the atoms in liquids move as entities, however, then all types of scattering can be described by the atomic or molecular distribution functions.

BIBLIOGRAPHY. P.A. EGELSTAFF, ***An Introduction to the Liquid State*** (1967), a work that serves both as an introduction to liquid structure and as a more advanced account of the scattering of radiation; D.S. EISENBERG and W. KAUZMANN, ***The Structure and Properties of Water*** (1969), covers the internal structure of the water molecule as well as its behaviour in gaseous, liquid, and solid states; G.W. GRAY, ***Molecular Structure and the Properties of Liquid Crystals*** (1962), a comprehensive review of work up to 1962; R.M. MAZO, ***Statistical Mechanical Theories of Transport Processes*** (1967), a short book that deals primarily with liquids; J.S. ROWLINSON, ***Liquids and Liquid Mixtures,*** 2nd ed. (1969), covers the classical and statistical thermodynamics of liquids at equilibrium, with some emphasis on the properties at high pressures; H.N.V. TEMPERLEY, J.S. ROWLINSON, and G.S. RUSHBROOKE (eds.), ***Physics of Simple Liquids*** (1968), an advanced treatise that covers equilibrium and nonequilibrium statistical thermodynamics and computer calculations for molecularly simple liquids.

(Jo.S.R.)

Lisbon

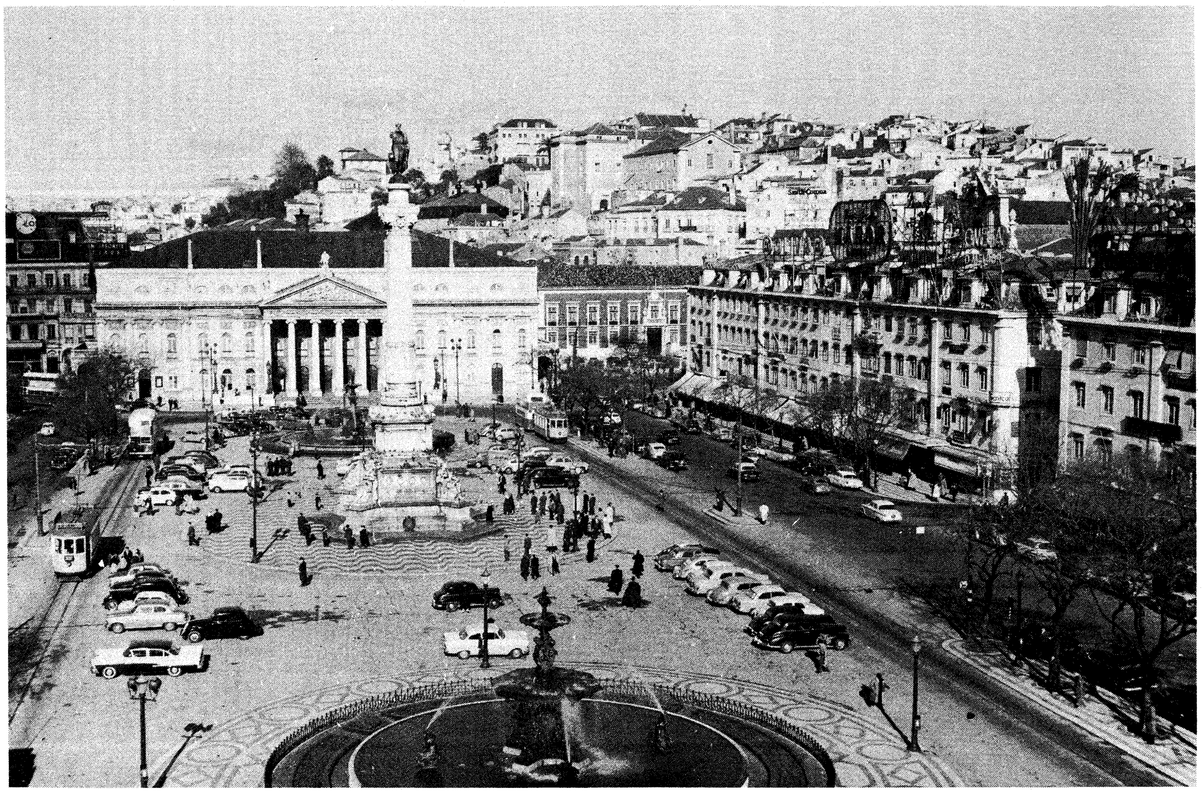
The capital, chief port, and largest city of Portugal—with, at more than 780,000 inhabitants, 10 percent of the national population—Lisbon (Portuguese, Lisboa) stands on the westernmost point of land of the European continent, where the Tagus River flows into the Atlantic Ocean. In the eight miles from the sea upstream to the city, the river is almost straight and about two miles wide. Just beyond the soaring gateway to Lisbon, the Salazar suspension bridge, which is the longest in western Europe, the waters suddenly broaden into a bay seven miles wide called the ***mar de palha,*** the straw sea—named for the sheen of its water, not its pollution. It is one of the continent's most beautiful harbours.

Scenically spectacular though it may be, the hill-cradled basin of burnished water also serves as a modern port, handling 9,000,000 tons of cargo a year. Amid the freighters, warships, liners, and ferryboats, a persistently picturesque note is struck by the ***fragatas*** with their Phoenician silhouettes, black hulls, and sepia sails; they perform most of the harbour's lighterage. The port maintains an intimacy with its city that was common in the days before steam. Vessels tie up at quays open to the everyday life of the town, where the clang of the trolley cars blends with the sound of ships' bells. At dawn, fishing smacks deposit their catch at the town's front doorstep for noisy auction to Lisbon dealers; while the fishwives, often barefoot, wait to fill the baskets they will peddle through the streets. Later the fish market gives way to the equally colourful and clamorous fruit and vegetable market.

The role of the river

It is traditional for Lisbon's poets to refer to the entwining Tagus as the city's lover. The river is indeed an ever-present part of the city's decor, visible from the many heights on which the town is built. The official entrance to Lisbon is a broad marble stair mounting from the water to a vast, arcaded square, the Praça do Comércio. The three landward sides of the square are surrounded by uniform 18th-century buildings, sea-green and white. This formal, baroque-flavoured composition is pierced by a monumental archway, built a century later, marking the entry into the central city. In the middle of the square, King Joseph (José) I, on horseback, is attended by a regiment of parked automobiles.

To the east, on the hill where Lisbon was first founded, the Castle of St. George (São Jorge in Portuguese; named for England's patron saint, in honour of an alliance made in 1386) still watches over the city. Just below it, the Church of St. Vincent-Outside-the-Walls guards the remains of the saint, which were—as depicted on the Lisbon coat of arms—miraculously brought to the city guided by two ravens. (Lisbon has a large number of local saints. Somewhat to the dismay of fellow townsmen, its most famous, St. Anthony, is known as St. Anthony of



The Rossio, in the commercial heart (Baixa) of Lisbon. In the centre, between twin fountains circled by mosaic, stands a statue of Pedro IV. The National Theatre of Dona Maria II overlooks the square in the background.

H. Armstrong Roberts

Padua. A church stands over his birthplace near the cathedral.) The towers of the dozen other old churches can still be picked out of the mass of buildings, but they have relinquished the skyline to the high-rise buildings of Lisbon's latest rebirth.

THE CITY'S ASPECT AND CHARACTER

Despite architectural change, the general outlines of the city remain as they were. It is still a city of balconies and vistas. On 17 of its prominences (many Lisbonese profess to see only seven traditional hills, as in Rome) the *miradouros,* garden balconies maintained by the city, are still frequented by citizens of all ages who gaze out from them down onto their city and river.

For centuries the citizens have been discussing the symptoms of an affliction they say is endemic in their strip of the Iberian Peninsula: *saudade,* very approximately translated as "melancholy," a variety of a state of anxiety tempered by fatalism. *Saudade* is said to be reflected in the music called *fado*, which is sung in its original form only in Lisbon, and in Lisbon only in two hillside plebeian precincts, Alfama and Bairro Alto. The word *fado* means "fate," usually unkind in the songs throbbingly sung to a penetrating but melodic music.

Saudade and dignity

One aspect of the Lisbonese character readily noticeable on the streets is self-respect, a nonaggressive esteem for human dignity. It can be seen in the waiter's pride in his deftness, in the cleanliness of the white shirt worn by the beggar gently soliciting alms among the café tables. (Lisbon, judged by net median family income, is the capital of western Europe's poorest people.) The pride is visible in the bearing and enthusiasm of the newspaper vendor, though he may be among the 40 percent illiterate in the Portuguese population.

Among the more affluent citizenry there are some indications that *saudade* may have soured into a sense approaching inferiority, and that the pride behind the facade of courtesy prickles with a dueller's sensitivity. Their baleful stares at tourists can be a reminder that much of Lisbon's history is a list of wars, uprisings, and assassinations.

A cross section of the population is visible during the late afternoon thronging of men in the Chiado neighbourhood, on the heights, and on the Rossio (city square), just below. Originally, the strolling began around three o'clock as a sort of peripatetic club for the professional men who could take time off to visit the intellectual coffeehouses. Today the coffeehouses have almost all gone, and the listless crowd scenes begin after office hours with a less distinguished cast.

The old, red brick bullring, with its Moorish arches and cupolas, still finds a public for its spectacle in which the bull is first tired either by a *cavaliero* who implants his darts from horseback or by a *toureiro* on foot who puts in the *farpas.* Next, a squad of eight acrobatic men stop the bull-eventually — with their bare hands. The bull is not slain. Near the bullring is the Arco en Cego housing development, most of whose occupants prefer to travel several miles north or west to find an outlet for their sporting enthusiasm at one of the three major football stadiums.

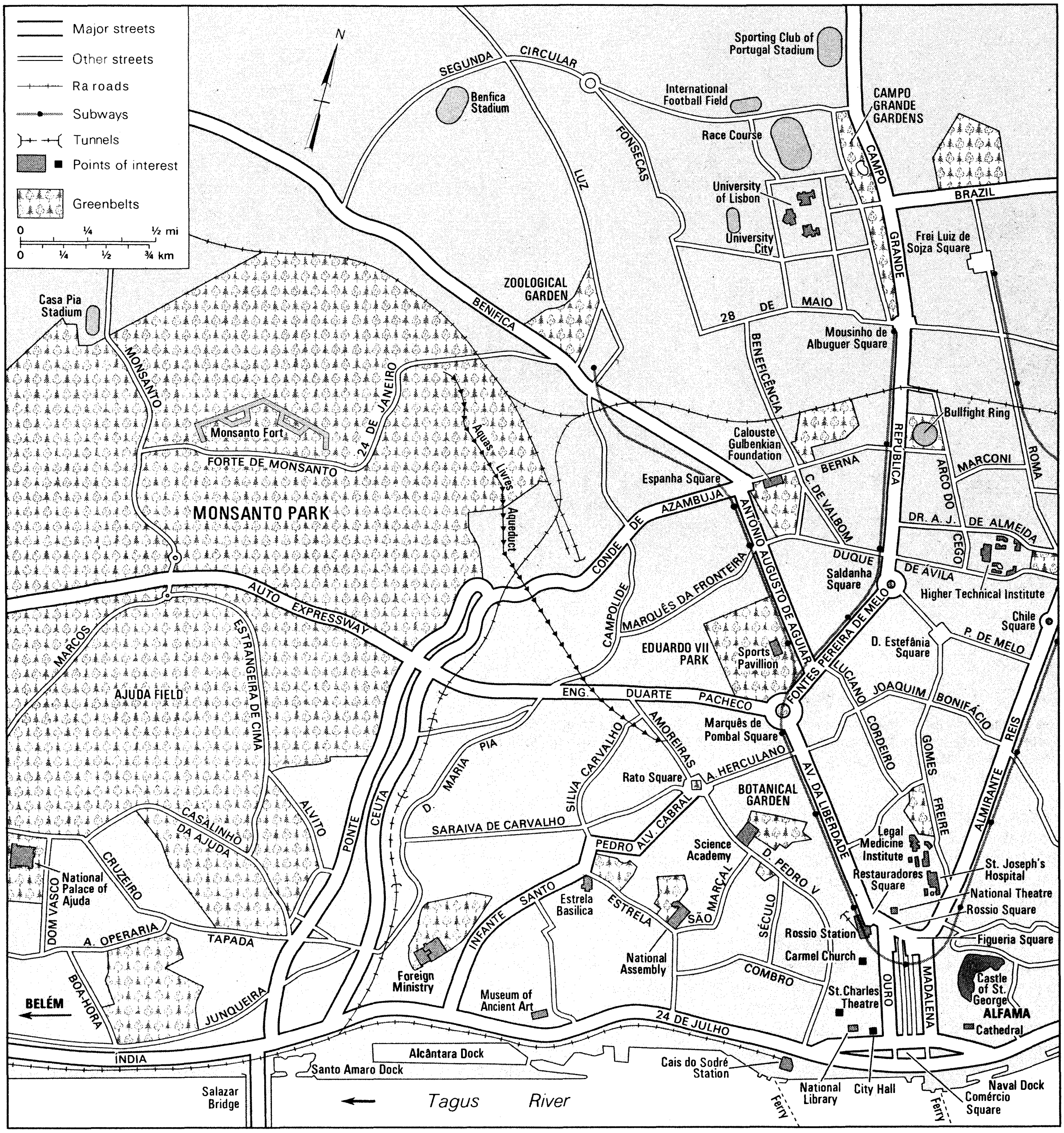
Finally, it may be observed that two aspects of life in Lisbon remain intact among the confusions of the city's sudden awakening from its long slumber: prejudice — religious and racial — and vandalism are virtually unknown.

THE HISTORY OF THE CITY

Prehistoric to Moorish times. The valley in which the heart of Lisbon now lies was, in prehistoric times, the bed of a forked branch of the Tagus. (The subway now forks at the same spot.) No evidence has yet been uncovered to show who were the first residents on the hills surrounding the valley, and there is no unassailable proof of the story favoured in Lisbon that the city was founded *c.* 1200 BC as a trading station by the far-ranging Phoenicians. In the city's ancient — and possibly Phoenician — name, Olisipo, lie the origins of the legend that the city's founder was Ulysses.

Origins of the city

During the Roman occupation (250 BC to c. AD 409), Julius Caesar raised the settlement to the dignity of a *municipium* and named it Felicitas Julia. A few inscribed stones testify that the Romans were there. They lost the city to the migratory peoples known as the Alani, who were driven out by the Suebi, who in turn were conquered



The city of Lisbon.

by the Visigoths. The base plan of the original fortifications is thought to be Visigothic and, if so, is the sole vestige of their reign.

The Muslims of North Africa took Lisbon when they overran the Iberian Peninsula in the 8th century; they stayed for **433** years, despite incursions by the Normans in **844** and by Alfonso VI of León in **1093.** Under the Moors the city was known under variations of "Lisbon": Luzbona, Lixbuna, Ulixbone, and Olissibona. Some authorities contend that the Muslims took this name from the conquered Roman castle, but Lisbon historians suggest that it derives from *água* ***boa*** ("good water").

The Portuguese conquest. Behind their walls, the Moors were able to hold out for months when the city was assailed by crusader forces—English, Flemish, Norman, and Portuguese under Afonso Henriques, the Portuguese king. The city finally fell in **1147** and successfully resisted Moorish attempts to win it back. The mosque was rebuilt as a church (presently the Sé [Cathedral], which has been restored in the styles of many periods, after fires and earthquakes), and the Moorish alcazar was transformed into a Portuguese royal palace.

Although **1,400** years of non-Portuguese occupation and invasion have left almost no trace among the stones of the capital, the presence of the outlander is still visible in the faces of the inhabitants, which range in cast from the Scandinavian to the Mauritanian.

After winning Lisbon, King Afonso established his court **105** miles to the north-northeast, atop a cliff at Coimbra. Lisbon did not become the national capital until more than a century later, in **1256.** Within its Moorish walls, of which some traces still remain, medieval Lisbon measured **1,443** feet at its widest and **1,984** feet at its longest, descending the hill below the castle. Even before the Portuguese conquest, some houses had already been built outside the walls down toward the river. The site of this first Lisbon is occupied by the lively Alfama quarter, which has kept the labyrinthine medieval street plan.

The medieval core

King Dinis I **(1279–1325)** decreed that Portuguese, the dialect of the Porto region, was to be the national language and founded **(1290)** the university. During his reign, other hilltops around the central valley were crowned with convents and churches.

In **1375** Lisbon was besieged and burned by the Castilians, after which King Ferdinand I swiftly erected new defenses. His wall—more than three miles long, with **77** towers and **38** gates and enclosing more than **247** acres—withstood the renewed Castilian attack of **1384.**

The Age of Discovery. When the Portuguese Age of Discovery (1420–1578) began, a city census showed 65,000 inhabitants occupying 23 parishes. A considerable number of them became rich, and the city was endowed with bigger and more luxurious buildings. African slaves became a familiar Lisbon sight, the trade in "black ivory" being another discovery in which Portugal led the rest of Renaissance Europe. After the great explorer Vasco da Gama led a Portuguese fleet to India in 1498, the Venetian monopoly on Oriental trade was broken; and colonies of German, Flemish, Dutch, English, and French traders set up in Lisbon. Greeks, Lombards, and Genoese who had lost their trading enclaves in Constantinople when that city fell to the Turks in 1453 also came to Lisbon.

King Manuel I (1495–1521) dominated this epoch. He was on the throne when the discovery of Brazil was announced, da Gama voyaged under his aegis, and in his name Albuquerque took Goa, in India. It was under Manuel that Portugal developed its sole contribution to European architecture, a style of decoration that one authority has described as "the high-water mark of late Gothic frenzy."

Manueline architecture

The elements of Manueline decoration celebrated the voyages of discovery, Manuel, and God. The cross topped the arms and the armillary spheres of the king, which crowned an assortment of hawsers, cordage, cork, seaweed, coral, shells, flowers, leaves, and fruit. Around balustrades, windows, arches, and columns these elements writhe and wreathe and knot themselves. The source of this overripe sculpture—unmistakable in the ivory and mother-of-pearl marquetry of the period's furniture—would seem to be Portuguese India, a unique reversal of the cultural current between western Europe and the colonies.

The prime examples of Manueline at Lisbon, the Tower of Belém and the Jerónimos Monastery, about four miles downstream from the city centre, are far less exuberant than those at the rival Portuguese cities of Batalha and Tomar. The tower and the monastery are nevertheless the most important architectural monuments in the capital. The five-story Tower of Belém on the riverbank was built in 1515 as a fort in the middle of the Tagus, which subsequently altered course. Girt by a cable carved in the stone, it has a stern Gothic interior but exhibits its North African touches on its turrets and crenellations and presents rounded Renaissance arches for the windows.

The monastery with its church and cloisters was begun in 1502 by an architect of French origin, Boytac (Boitaca), and was not finished until the end of the century. The work of four other architects succeeded one another, their styles passing from the Gothic through the Renaissance to the Baroque. Smoothed by time, the ensemble is harmonious and proudly Portuguese.

Manuel I promoted the urbanization of the central valley between Lisbon's hills, creating a city square, the Rossio, which at once became, and remains, Lisbon's "living room." Down by the Tagus he constructed a new palace, the Paços da Ribeira, with a large square laid out along its eastern flank—the Palace Terrace (Terreiro do Paço). Today the Praça do Comércio, constructed on the site after the 1775 earthquake, is often called by its 16th-century name. The area between the Rossio and the Palace Terrace was soon strewn with streets wandering among the new shops, churches, and hospitals of what had become a phenomenally prosperous city.

Decline in the 16th century

The prosperity was chimeric. Even as Magellan's ships returned from around the world in 1522, the realities had begun to rasp through the gilding of Portugal's Golden Age. John (João) the Pious, who had succeeded Manuel, invited the Jesuits and the Inquisition to come to Portugal to counter the ungodly materialism of Lisbon. He transferred (1537) the university to the royal palace at Coimbra, far from the capital's excesses, and invited professors from France, Italy, and Spain, thus creating one of the intellectual centres of Europe. Lisbon was visited with plagues and earthquakes during this time, but they proved easier to meet than the bill for 50 years of glory. Literally half the nation's population had vanished in pursuit of wealth in the new colonies. With farms deserted, food was imported from other European countries at crippling prices, and with so many skilled men absent, wages rose sharply, as did the cost of building and manufacturing materials. The colonial treasures, which had made Lisbon such a sybaritic queen of the seas, in the end cost more than they could fetch.

In 1578 King Sebastian of Portugal was killed while losing his entire army in an invasion of Morocco: two years later, the Spanish pushed into Portugal, and Philip II of Spain became king of both countries. In 1588 it was from Lisbon that the Invincible Armada sailed against England, Portugal's oldest ally. In the half century that followed, Lisbon lived relatively well as a port for the riches of the Spanish Main. In 1640 a conspiracy of Lisbon nobles struck for freedom and drove out the Spaniards. The square just north of the Rossio, Praça dos Restauradores, is named for them.

Disaster and reconstruction. In the first half of the 18th century the profits from the plantations and the gold and diamond deposits of Brazil brought a new flurry of optimism and excitement to Lisbon. This ended on the morning of November 1, 1755. The churches were crowded to honour the dead on All Saints' Day when the city was devastated by one of the greatest earthquakes ever recorded. There were two shocks, 40 minutes apart; and the waters of the Tagus, lifted from their bed, roared through the city. Then came fire. It is believed that 30,000 lives were lost, and more than 9,000 buildings were destroyed.

The great earthquake of 1755

Physically, Lisbon recovered with a celerity astonishing for the time, but the shock left its mark upon the thinking of generations to come. The reconstruction—a good deal of foreign aid was forthcoming—was achieved by Joseph I's prime minister, Sebastiaõ José de Carvalho, the virtual ruler of the nation. He put Manuel da Maia, engineer in chief of the realm, in charge of five architects and soon had a plan for remaking the totally devastated centre of the town, the Baixa. The riverside palace had been destroyed, and its terrace was expanded to create the new arcaded Praça do Comércio. Northward from there, a grid work of 48 streets led inland to the Rossio and a neighbouring new square, Figueria.

The two-story, uniform buildings were topped by two tiers of dormers projecting from a tiled roof. The corners of the eaves, in the Lisbon tradition, turned up curiously, in faint echo of a pagoda. The building style, evolved for fast, cheap construction, was Baroque, but virtually stripped of decoration. After the minister was rewarded with the title of marquês de Pombal, the style became known as the ***estilo*** pombalino.

The Sé and most of the churches were repaired or rebuilt, but the 14th-century Carmo (Carmel) Church was left as it was. Looming from its hilltops over the Baixa, the roofless Gothic shell now serves as an archaeological museum, while its cloister serves as a police barracks. The Palace of the Inquisition, utterly flattened, was not rebuilt when Pombal enlarged and realigned the Rossio, and on its site, 75 years later, the National Theatre of Dona Maria II was erected. Pombal banished the Jesuit order and transformed their establishment into St. Joseph's Hospital to replace the destroyed All Saints Hospital. The medical school scrambled for room at St. Joseph's until it acquired a new building of its own late in the 19th century. The Jesuit novice house was converted to serve as the Nobles' School. Later governments expelled more religious orders, whose buildings became barracks, hospitals, royal academies, and government offices. The modern National Assembly convenes in a transformed convent.

Nineteenth-century expansion. During the Peninsular War of the early 1800s, Lisbon alternated between French and British control, and after Napoleon's defeat, it was embroiled in civil war until 1834. That conflict was followed by ten years of revolutionary outbursts. Nineteenth-century Lisbon nevertheless continued to expand and, by 1885, embraced 20,378 acres while the population had doubled in 100 years to reach 300,000. Public buildings, such as the new city hall and the Ajuda Royal Palace, had been built, the harbour had been modernized and quays constructed on land reclaimed from the river. The railway had appeared, and with it the ornate neo-Manueline station in the Rossio. A system of horse cars served the lower town.

The opening of Avenida da Liberdade

The greatest change in the city, and the one most important for its future growth, was the opening of a new main street in 1880—the 460-foot-long and 27-foot-wide Avenida da Liberdade. (The 20-acre formal garden at the upper end was added early in the 20th century and named Eduardo VII in honour of a state visit by the British monarch.) The lower end, including the Praça dos Restauradores, had been the leafy gardens of the Passeio Publico (Public Promenade), and the paving operation was greeted by hostility on the part of the public. The municipality bordered the central six-lane carriageway with wide blue mosaic sidewalks graced with palms and shade trees, fountains, and ornamental waters stocked with goldfish and swans. So the street remains today, with the addition of outdoor cafés beneath the trees.

In conjunction with the new thoroughfare, a series of new streets, the "Avenidas Novas," expanded the city northward, and new neighbourhoods sprang up as well on the borders of the Avenida da Liberdade. In 1901 the electric streetcar made its appearance, enabling more people to live farther away from their employment in the Baixa. Three cable cars shuttled up and down the adjacent hills, and Alexandre-Gustave Eiffel designed the giant elevator that hisses grandly between the town's upper and lower levels.

New water supplies, augmenting those of the 1748 aqueduct of Águas Livres, of which the Lisbonese are still proud, were introduced from Alviela. Consequently, water was piped directly into houses, eliminating the ancient calling of *galego*, or water porter.

Salazar to the 1960s. In 1908 Portugal's king and crown prince were assassinated on a corner of the Terreiro do Paço. Two years later the new king, Manuel II, abdicated. A republic was declared, and a period of national instability ensued. When António Salazar took control of the near-bankrupt nation in 1932, he erected a corporate state of which he alone determined the policies until his retirement in 1968. As the president of the municipal council of Lisbon is appointed by the national executive, as is his colleague, the civil governor of the Lisbon District, the national oligarchy controls the administration of the capital. A group of career professionals with the title of director formulate projects and regulations for the municipal departments, and these are passed to the municipal council for discussion and to the council president for decision. With the economy controlled in minute detail, strikes illegal, and press, radio, and television censored, nothing happens in the capital that is not approved on a high level. Few changes have been approved, and these as much out of national pride as out of local need: the new University City, the huge new teaching hospital attached to modern buildings for the medical school, the large Higher Technical Institute, several express highways, a model housing development, and completion of the first leg of the subway system.

A new era of expansion

In the 1960s national policy began to change, allowing some economic expansion. The 30-year-old austerity program of stability and self-sufficiency (at an admittedly low level of investment and consumption) was somewhat softened. The flood of international tourists and international corporations began to be accommodated. In 1966, well ahead of schedule (and with $77,000,000 in aid from the United States), the Salazar Bridge was completed. New hotels and office buildings began changing the face of the city, and major improvements were made in the port.

THE CONTEMPORARY CITY

The rage to build has continued into the 1970s. The pastel-tinted and somewhat sleepy Lisbon that offered a neutral, 19th-century haven to 200,000 war refugees in the 1940s is disappearing in the din and dust of new construction. Lisbon is emerging as a bustling modern metropolis. Pombal's Baixa remains rigorously protected from change, but the four-story buildings of the Avenida da Liberdade and its ancillary streets have been almost totally replaced by ten-story buildings in an anonymous style sometimes categorized as airport modern. The new construction has also gained the hills, even the Alfama, whose established residents continue to hang laundry across the narrow streets and grill sardines on doorstep braziers.

The municipality has built whole new neighbourhoods in the north and northwest sectors of town, around the airport (the airport will be replaced in 1978 by a mammoth installation 50 miles south across the Tagus). Other building is pushing westward toward Belém as replacement for decayed neighbourhoods. These new structures, some of them 14 stories high, are designed by Lisbon architects who produced handsome, colourful, contemporary buildings. Usually about 40 percent of the new lodgings are reserved for the poorest families (a certain number pay nothing at all); 30 percent for people of moderate means; 20 percent for those with higher incomes; and 10 percent for the most prosperous. The newest project, in the Chelas District, implanted on heathland previously considered too difficult to build upon, housed 10 percent of the 1971 Lisbonese population. The mobile and moneyed in the late 1960s began to move out of town, and more and more villas were appearing in the countryside of the "Portuguese Riviera," which lies between Lisbon and the town of Estoril, 16 miles to the west.

Allocation of new housing by income groups

Parks, museums, and cultural life. The new housing developments are planted with trees and grass and given small parks to add to Lisbon's collection of more than 40 public gardens. The biggest, Monsanto, covers about eight square miles. Planted in the 1920s as a windbreak for the town, its rolling hills are now thickly forested, and recreational facilities (camping, swimming pool) are being enlarged. There are also two botanical gardens and a zoological garden within the city. The Castle of St. George is more garden than castle, richly planted within and without the ramparts and with superb views over the city.

Lisbon's newest cultural centre (October 1969), consisting of low-built modem structures set in a garden, was built by the Calouste Gulbenkian Foundation, about half a mile north of Eduardo VII Park. The centre presents music and ballet as well as exhibits of the other fine arts. It also houses the broad-ranging personal collection that Calouste Gulbenkian, an Armenian oil-lease negotiator, willed to the Portugese nation.

The city has more than a dozen other museums, including those of modern, antique, sacred, decorative, and folk art. In the latter, the visitor may ponder the differences between the innocent beauty of native artifacts and the ugliness of the corrupt versions in souvenir shops.

Theatres and theatrical churches

There are two principal theatres, the St. Charles (São Carlos) and the National Theatre of Dona Maria II. The St. Charles, constructed in the late 18th century, has a beautiful elliptical interior, and the National, built about 1840, displays a facade of six giant columns saved from the earthquake-destroyed convent church of St. Francisco. The interior, gutted by fire in 1966, has been restored.

Neither of these edifices is as theatrical as the interiors of some of the churches built or restored after the 1755 earthquake. In gold, marble, carved wood, rare tiles, these interiors are decorated in Baroque, Rococo, or rocaille. Lisbon is further distinguished in religion as one of the three places in the world whose chief Roman Catholic clergyman bears the title of patriarch.

Commerce and industry. Two of the forces acting to uproot 18th- and 19th-century Lisbon and transplant it into the 20th century were tourism and commerce. Lisbon's five largest banks increased the number of their branches from 50, in 1964, to 75, five years later. Their deposits rose from more than 12,000,000,000 escudos to more than 75,000,000,000 during the same period. To the long established local industries of soapmaking, munitions, and steel manufacture were added glassmaking, electronics, margarine manufacture and—with the raw gems coming from Portuguese Angola—diamond cutting. The petroleum refinery, largely state supported, was greatly expanded.

The greatest development has come just outside the city limits on the south bank of the Tagus, now readily accessible because of the new bridge; that district has become Portugal's most important manufacturing centre. The industrial belt is expected to continue growing from the river down to Setdbal, 25 miles south. One of the world's largest cement plants is found on the far bank, along with grain elevators, a steelmaking complex, a cork factory, and a plastics plant.

The port of Lisbon, with 19 miles of docks, now has special facilities for handling container-ship cargoes, and a new facility for car ferries. On the far bank two dry docks handle tankers of 326,000 tons and 100,000 tons, and a third is planned for 1,000,000-ton supertankers. The fishing port has a specialized dock attached to its refrigerating plants. In 1970 the port handled 400,000 passengers; more than five times that number arrived by air. The Salazar Bridge is the first in Lisbon's long history to span the Tagus. Built between the end of 1962 and mid-1966, it is 694 feet long, with a central span of 309 feet suspended 21 feet above mean water level. There is space under the roadway to carry two railroad tracks.

To the southeast of the bridge rises a gigantic figure of Christ. Visitors may climb to the top of it and avail themselves of a panoramic view of the city.

BIBLIOGRAPHY

General: PATRICK SWIFT and DAVID WRIGHT, *Lisbon* (1971), gives thorough coverage of all quarters of the city, its history, monuments, cultural institutions, and contemporary life, and also includes excursions outside Lisbon; CAROL WRIGHT, *Lisbon* (1971), on the city and the life of its inhabitants at various times of day, information for the visitor to Lisbon and its environs; VIVIAN ROWE, *The Road to Lisbon* (1962), mainly concerned with the journey from France to Lisbon but includes some material on the attractions of the city itself; in Portuguese, CAMARA MUNICIPAL DE LISBOA, *Grades de Lisbon* (1947).

The past: T.D. KENDRICK, *The Lisbon Earthquake* (1956), a study of the catastrophic earthquake of November 1, 1755, and its impact on European philosophers and theologians; JANET SCHAW, *Journal of a Lady of Quality*, 3rd ed. by EVANGELINE WALKER ANDREWS in collaboration with CHARLES MCLEAN ANDREWS (1939), part of which tells of the sojourn of the Scottish author in 18th-century Lisbon.

(B.E.)

Lister, Joseph

Celebrated British surgeon and medical scientist, Joseph Lister was the pioneer of antisepsis—the use of chemicals to prevent surgical infections—which in the latter half of the 19th century reduced surgical mortality to one-third of its previous levels. While his method, based on the use of antiseptics, is no longer employed, his principle—that bacteria must never gain entry to an operation wound—remains the basis of surgery to this day.

Joseph Lister was born in Essex on April 5, 1827, the second son of Joseph Jackson Lister and his wife Isabella Harris, members of the Society of Friends, or Quakers. J.J. Lister, a wine merchant and amateur physicist and microscopist, was elected a fellow of the Royal Society for his discovery that led to the modern achromatic (non-colour-distorting) microscope.

While both parents took an active part in Lister's education, his father instructing him in natural history and the use of the microscope, Lister received his formal schooling in two Quaker institutions, which laid far more emphasis upon natural history and science than did other schools. He became interested in comparative anatomy, and, before his 16th birthday, he had decided upon a surgical career.

After taking an arts course at University College, London, he enrolled in the faculty of medical science in October 1848. A brilliant student, he was graduated a bachelor of medicine with honours in 1852; in the same year he became a fellow of the Royal College of Surgeons and house surgeon at University College Hospital. A visit to Edinburgh in the fall of 1853 led to Lister's appointment as assistant to James Syme, the greatest surgical teacher of his day, and in October 1856 he was appointed surgeon to the Edinburgh Royal Infirmary. In April he had married Syme's eldest daughter. Lister, a deeply religious man, joined the Scottish Episcopal Church. The marriage, although childless, was a happy one, his wife entering fully into Lister's professional life.

Lister, 1857.
BY courtesy of the Wellcome Trustees, London

When three years later the Regius Professorship of Surgery at Glasgow University fell vacant, Lister was elected from seven applicants. In August 1861 he was appointed surgeon to the Glasgow Royal Infirmary, where he was in charge of wards in the new surgical block. The managers hoped that hospital disease (now known as operative sepsis—infection of the blood by disease-producing micro-organisms) would be greatly decreased in their new building. The hope proved vain, however. Lister reported that, in his Male Accident Ward, between 45 percent and 50 percent of his amputation cases died from sepsis between 1861 and 1865.

Beginning of antisepsis

In this ward Lister began his experiments with antisepsis. Much of his earlier published work had dealt with the mechanism of coagulation of the blood and role of the blood vessels in the first stages of inflammation. Both researches depended upon the microscope and were directly connected with the healing of wounds. Lister had already tried out methods to encourage clean healing and had formed theories to account for the prevalence of sepsis. Discarding the popular concept of miasma—direct infection by bad air—he postulated that sepsis might be caused by a pollen-like dust. There is no evidence that he believed this dust to be living matter, but he had come close to the truth. It is therefore all the more surprising that he became acquainted with the work of the bacteriologist Louis Pasteur only in 1865.

Pasteur had arrived at his theory that micro-organisms cause fermentation and disease by experiments on fermentation and putrefaction. Lister's education and his familiarity with the microscope, the process of fermentation, and the natural phenomena of inflammation and coagulation of the blood impelled him to accept Pasteur's theory as the full revelation of a half-suspected truth. At the start he believed the germs were carried solely by the air. This incorrect opinion proved useful, for it obliged him to adopt the only feasible method of surgically clean treatment. In his attempt to interpose an antiseptic barrier between the wound and the air, he protected the site of operation from infection by the surgeon's hands and instruments. He found an effective antiseptic in carbolic acid, which had already been used as a means of cleansing foul-smelling sewers and had been empirically advised as a wound dressing in 1863. Lister first successfully used his new method on August 12, 1865; in March 1867 he published a series of cases. The results were dramatic. Between 1865 and 1869, surgical mortality fell from 45 percent to 15 percent in his Male Accident Ward.

In 1869, Lister succeeded Syme in the chair of Clinical Surgery at Edinburgh. There followed the seven happiest years of his life when, largely as the result of German experiments with antisepsis during the Franco-German War, his clinics were crowded by foreign visitors and eager students. In 1875 Lister made a triumphal tour of the leading surgical centres in Germany. Next year he visited America but was received with little enthusiasm except in Boston and New York.

Opposition and vindication

Lister's work had been largely misunderstood in England and the United States. Opposition was directed against his germ theory rather than against his "carbolic treatment." The majority of practicing surgeons were unconvinced; while not antagonistic, they awaited clear proof that antisepsis constituted a major advance. Lister was not a spectacular operative surgeon and refused to publish statistics. Edinburgh, despite the ancient fame of its medical school, was regarded as a provincial centre. Lister understood that he must convince London before the usefulness of his work would be generally accepted.

His chance came in 1877, when he was offered the chair of Clinical Surgery at King's College. On October 26, 1877, Lister, at King's College Hospital, for the first time performed the then-revolutionary operation of wiring a fractured patella, or kneecap. It entailed the deliberate conversion of a simple fracture, carrying no risk to life, into a compound fracture, which often resulted in generalized infection and death. Lister's proposal was widely publicized and aroused much opposition. Thus, the entire success of his operation carried out under antiseptic conditions forced surgical opinion throughout the world to accept that his method had added greatly to the safety of operative surgery.

More fortunate than many pioneers, Joseph Lister saw the almost universal acceptance of his principle during his working life. He retired from surgical practice in 1893, after the death of his wife in the previous year. Many honours came to him. Created a baronet in 1883, he was made Baron Lister of Lyme Regis in 1897 and appointed one of the 12 original members of the Order of Merit in 1902. Throughout his life, he remained a gentle, shy, unassuming man, firm in his purpose because he humbly believed himself to be directed by God. He was uninterested in social success or financial reward. In person he was handsome, with a fine athletic figure, fresh complexion, hazel eyes, and silver hair. After some years of almost complete blindness and deafness, he died at Walmer, Kent, on February 10, 1912.

BIBLIOGRAPHY. Lister wrote no books but contributed many papers to professional journals. These are contained in *The Collected Papers of Joseph, Baron Lister, 2* vol. (1909). The authoritative biography is SIR RICKMAN J. GODLEE, *Lord Lister* (1917). Simpler biographies are FREDERICK F. CARTWRIGHT, *Joseph Lister* (1963); and RHODA TRUAX, *Joseph Lister* (1944). For personal reminiscences, see J.R. LEESON, *Lister As I Knew Him* (1927).

(F.F.C.)

Liszt, Franz

Hungarian composer and the greatest piano virtuoso of his time, Franz Liszt revolutionized the technique of piano playing and invented the "piano recital" as it is known today. He discovered new methods of writing for the piano that laid the foundations for the modern technique of piano composition; he also invented the symphonic poem for orchestra and the method of "transformation of themes," by which one or two themes in different forms can provide the basis for an entire work—a principle from which Richard Wagner derived his system of so-called leitmotifs in his operas. Liszt also considerably expanded the harmonic language of his time, and some of his later works anticipate 20th-century harmonic methods. He was in contact with all the leading artistic and social personalities of his time and did a great deal to encourage the performance of music of previously neglected masters such as Johann Sebastian Bach, Ludwig van Beethoven, and Franz Schubert and music of his contemporaries Hector Berlioz, Robert Schumann, and Wagner. He also helped younger composers, including the Norwegian Edvard Grieg, the Russians Mily Balakirev and Aleksandr Borodin, and the Frenchman Claude Debussy; and he taught a number of pupils who themselves became famous virtuosi.

By courtesy of the Museo Teatrale alla Scala, Milan

Liszt, lithograph by Joseph Kriehuber, 1846.

Youth and early training. Liszt was born at Raiding, Hungary, on October 22, 1811. His father, Ádám Liszt, was an official in the service of Prince Nicolas Eszterházy, whose palace in Eisenstadt was frequented by many celebrated musicians, including Joseph Haydn, Luigi Cherubini, and Johann Hummel. Ádám Liszt was a talented amateur musician who played the cello in the court concerts. By the time Franz was five years old he was already attracted to the piano and was soon given lessons by his father. He began to show interest in church and Gypsy music. He also developed into a very religious child, again because of the influence of his father, who, during his youth, had spent two years in the Franciscan order.

First public appearance as a pianist

Franz began to compose at the age of eight. When only nine he made his first public appearance as a concert pianist at Sopron and Pozsony (now Bratislava, Czechoslovakia). His playing so impressed the local Hungarian magnates that they put up a sum of money to pay for his musical education for the next six years. Ádám obtained leave of absence from his post and took Franz to Vienna, where he had piano lessons with Karl Czerny, a composer and pianist who had been a pupil of Beethoven, and studied composition with Antonio Salieri, the musical director at the Viennese court. He gave several concerts in Vienna, with great success. Beethoven was persuaded to attend one of them, and at the end he mounted the platform and kissed the boy on the forehead. Liszt was also invited to contribute a variation on a waltz by the publisher Anton Diabelli to a symposium in which 50 Austrian composers, including Schubert, Czerny, Hummel, and Ignaz Moscheles took part. Beethoven refused to subscribe to the symposium but instead wrote his own famous set of 33 variations on the theme.

Liszt moved with his family to Paris in 1823, giving concerts in Germany on the way. He was refused admission to the Paris Conservatoire because he was a foreigner; instead, he studied with Anton Reicha, a theorist who had been a pupil of Haydn's brother Michael, and Ferdinando Paer, the director of the Théâtre-Italien in Paris and a composer of light operas. Liszt's Paris debut on March 7, 1824, was sensational. Other concerts quickly followed, as well as a visit to London in June. He toured England again the following year, playing for George IV at Windsor Castle and also visiting Manchester, where his *New Grand Overture* was performed for the first time. This piece was used as the overture to his one-act opera *Don Sanche*, which was performed at the Paris Opéra on October 17, 1825. Liszt also composed a number of piano works at this time, including the *Etude en 48 exercices,* which was later transformed into the *Etudes d'exécution transcendante.*

In 1826 he toured France and Switzerland, returning to England again in the following year. Suffering from nervous exhaustion, Liszt expressed a desire to become a priest. His father took him to Boulogne to take sea-baths to improve his health; there Ádám died of typhoid fever. Liszt returned to Paris, and sent for his mother to join him; she had gone back to the Austrian province of Styria during his tours.

Despondency and period of doubt

Liszt now earned his living mainly as a piano teacher, and in 1828 he fell in love with one of his pupils, the daughter of the Comte de Saint-Cricq, who was the minister of commerce to Charles X of France. When the count insisted that their attachment be broken off, Liszt again became extremely ill; he was considered so close to death that his obituary appeared in a Paris newspaper. After his illness he underwent a long period of depression and doubt about his career. For over a year he did not touch the piano and was dissuaded from joining the church only through the efforts of his mother. He experienced much religious pessimism. He also conversed with the violinist Chrdtien Urhan, who introduced him to the Saint-Simonians, a socio-religious movement espousing a kind of universal brotherhood of men.

During this period Liszt took an active dislike to the career of a virtuoso. He made up for his previous lack of education by reading widely, and he came into contact with many of the leading artists of the day, including the writer and statesman Alphonse de Lamartine, Victor Hugo, and Heinrich Heine, the German poet and essayist. With the July Revolution of 1830 resulting in the abdication of Charles X and the coronation of Louis-Philippe, he sketched out a *Revolutionary Symphony.*

Between 1830 and 1832 he met the three men who were to have considerable influence on his artistic life. At the end of 1830 he first met Berlioz and heard the first performance of his *Symphorzie fantastique.* From Berlioz he inherited the command of the romantic orchestra and also the diabolic quality that remained with him for the rest of his life. He achieved the seemingly impossible feat of transcribing Berlioz' *Symphonie fantastique* for the piano in 1833, and he helped Berlioz considerably by transcribing other works of his and playing them in concert. In March 1831 he heard Paganini play for the first time. He again became interested in virtuoso technique and resolved to transfer some of Paganini's fantastic violin effects to the piano, writing a fantasia on his *Campanella.* At this time he also met Frédéric Chopin, whose poetical style of music exerted a profound influence on Liszt.

Years with Marie d'Agoult. In 1834 Liszt emerged as a mature composer with the single piece *Harmonies poétiques et religieuses*, based on a collection of poems by Lamartine, and the set of three *Apparitions.* The lyrical style of these works is in marked contrast to his youthful compositions, which reflected the style of his teacher Czerny. In the same year, through the poet and dramatist Alfred de Musset, he met the novelist George Sand and also Marie de Flavigny, comtesse d'Agoult, with whom he began an affair. In 1835 she 'left her husband and family to join Liszt in Switzerland; their first daughter, Blandine, was born in Geneva on December 18. Liszt and Madame d'Agoult lived together for four years, mainly in Switzerland and Italy, though Liszt made occasional visits to Paris. He also taught at the newly founded Geneva Conservatory and published a series of essays "On the Position of Artists," in which he endeavored to raise the status of the artist—who up to then had been regarded as a kind of superior servant—to that of a respected member of the community.

Composition of *Années de pèlerinage*

Liszt commemorated his years with Madame d'Agoult in the first two books of the *Années de pèlerinage (Years of Pilgrimage),* which are poetical evocations of Swiss and Italian scenes. He also wrote the first mature version of the *Transcendental Studies,* based on his youthful *Étude en 48 exercices,* but here transformed into pieces of terrifying virtuosity. He transcribed for the piano six of Paganini's pieces—five studies and *La campanella*—and also three Beethoven symphonies, some Schubert songs, and further works of Berlioz. He made these transcriptions to make the work of these men more available and thus spread the appreciation of their music, which was still greatly neglected at that time. Liszt also wrote a number of fantasias on popular operas of the day and dazzled audiences with them at his concerts.

His second daughter, Cosima, was born in 1837 and his son Daniel in 1839, but toward the end of that year his relations with Madame d'Agoult became strained and she returned to Paris with the children. Liszt then returned to his career as a virtuoso to raise money for the Beethoven Memorial Committee in Bonn for the completion of its Beethoven monument.

For the next eight years Liszt travelled all over Europe, giving concerts in countries as far apart as Ireland, Portugal, Turkey, and Russia. He continued to spend his summer holidays with Madame d'Agoult and the children on the island of Nonnenwerth in the Rhine River until 1844; then they finally parted and Liszt took the children to Paris. Liszt's brilliance and success were at their peak during these years as a virtuoso. Everywhere he was received with great adulation; gifts and decorations were showered on him; he had numerous mistresses, including the dancer Lola Montez and Marie Duplessis, the prototype of *La Dame aux camélias (The Lady of the Camellias),* a novel by Alexandre Dumas *(fils).* Nevertheless, he still continued to compose, writing songs as well as piano works.

His visit to Hungary in 1839–40, the first since his boyhood, was an important event. His renewed interest in the music of the Gypsies laid the foundations for his *Hungarian Rhapsodies* and other works composed in the Hungarian style. He also wrote a cantata for the Beethoven Festival of 1845, his first work for choius and orchestra, and some smaller choral works.

Affair with Carolyne Sayn-Wittgenstein

Compositionsat Weimar. In February 1847 Liszt met the princess Carolyne Sayn-Wittgenstein at Kiev and later spent some time at her estate in Poland. She quickly persuaded him to give up his career as a virtuoso and to concentrate on composition. He gave his final concert at Yelizavetgrad (modern Kirovograd) in September of that year. Having been director of music extraordinary to the Weimar court in Germany since 1843, and having conducted concerts there since 1844, Liszt decided to settle there permanently in 1848. He was later joined by the Princess, who had unsuccessfully tried to obtain a divorce from her husband. They resided together in the Altenburg, and Liszt now had ample time to compose, as well as to conduct the court orchestra in operas and concerts. This was the period of his greatest production: the first 12 symphonic poems, the Faust and Dante symphonies, the piano sonata, the two piano concerti, the *Totentanz* for piano and orchestra, revised versions of the *Transcendental* and *Paganini Studies* and of the first two books of the *Annkes de pèlerinage*, choral works, and numerous others. Some of these works had been sketched out in the 1840s or earlier, but even so, his productivity in this period remains astonishing. He also conducted a number of operas by modern composers, including Wagner's *Tannhäuser* in 1849 and the world premiere of *Lohengrin* in 1850; Schumann's *Genoveva* and his incidental music to *Manfred;* and Berlioz' *Benvenuto Cellini*, as well as operas by Giuseppe Verdi and Gaetano Donizetti and a number of works by lesser composers. In addition, he performed his own works with the Weimar orchestra and made numerous revisions as a result. For a time the composers August Conradi and Joachim Raff helped him with the preparation of his orchestral scores, but the final versions of all of them were by Liszt himself.

Hostility toward Liszt

The avant-garde composers of the day regarded Weimar as the one city where modern composers could be heard, and many of them came to Liszt as pupils. The so-called New German school hoisted the banner of modernism, which naturally annoyed the more academic musicians. Some members of the Weimar court also were upset by Liszt's continued support of Wagner, who had had to flee in 1849 with Liszt's help from Germany to Switzerland because of his political activism. The straitlaced citizens of Weimar also objected strongly to the Princess openly living with Liszt, and the Grand Duchess of Weimar was under pressure from her brother, Nicholas I of Russia, to ban Princess Sayn-Wittgenstein from all court functions. Furthermore, the Grand Duke who originally appointed Liszt died in 1853, and his successor took little interest in music. As a result, the director of the theatre was able to intrigue against Liszt and decrease the number of music productions in favour of inferior plays. Things came to a climax in 1858 at the first performance of *The Barber of Bagdad* by Liszt's pupil Peter Cornelius. Liszt conducted the work himself, and when the audience mounted a demonstration against the music, Liszt took it to be directed against him personally and handed in his resignation to the Grand Duke.

Though Liszt remained in Weimar until 1861, his position there became more and more difficult. His son Daniel had died in 1859 at the age of 20. Liszt was deeply distressed and wrote the oration for orchestra *Les Morts* in his son's memory. In May 1860 the Princess had left Weimar for Rome in the hope of having her divorce sanctioned by the Pope, and in September, in a troubled state of mind, Liszt had made his will. He left Weimar in August of the following year, and after travelling to Berlin and Paris, where he saw Marie d'Agoult, he arrived in Rome. He and the Princess hoped to be married on his 50th birthday. At the last moment, however, the Pope revoked his sanction of the Princess's divorce; they both remained in Rome in separate establishments.

Attempts to create a new kind of religious music

Eight years in Rome. For the next eight years Liszt lived mainly in Rome and occupied himself more and more with religious music. He completed the oratorios *Die Legende von der heiligen Elisabeth* and *Christus* and a number of smaller works. He hoped to create a new kind of religious music that would be more direct and moving than the rather sentimental style popular at the time. Liszt was one of the few 19th-century musicians to be interested in Gregorian plainsong, but his efforts were frowned on by the ecclesiastical authorities, and much of his sacred music remained unpublished until many years after his death.

In 1862 his daughter Blandine died at the age of 26. Liszt wrote his variations on a theme from the Bach cantata *Weinen, Klagen (Weeping, Mourning)* ending with the chorale *Was Gott tut das ist wohlgetan (What God Does Is Well Done)*, which must have been inspired by this event. The Princess's husband died in 1864, but there was no more talk of marriage; and in 1865 Liszt took the four minor orders of the Roman Catholic Church, though he never became a priest. In 1867 he wrote the *Hungarian Coronatiort Mass* for the coronation of the emperor Francis Joseph I of Austria as king of Hungary. This commission renewed his links with his native country. Meanwhile, his younger daughter Cosima, who, at the age of 19, had married Liszt's favourite pupil, Hans von Bülow, was having an affair with Wagner. She had an illegitimate child by Wagner, which led to a quarrel between the two composers that lasted until 1872.

Last years. In 1869 Liszt was invited to return to Weimar by the Grand Duke to give master classes in piano playing, and two years later he was asked to do the same in Budapest. From then until the end of his life he divided his time between Rome, Weimar, and Budapest. Numerous pupils and composers visited him, including Anton Rubinstein, Isaac Albéniz, Aleksandr Borodin, Camille Saint-Saens, and Gabriel-Urbain Fauré. After reconciliation with Wagner in 1872, he regularly attended the Bayreuth festivals. He appeared occasionally as a pianist in charity concerts and continued to compose. His later works anticipate the harmonic style of Claude Debussy. One late work called *Bagatelle Without Tonality* anticipates Béla Bartók and even Arnold Schoenberg. In the winter of 1882–83 he stayed with Wagner and Cosima in Venice. There a sudden premonition caused him to interrupt his work on the oratorio *Die Legende vom heiligen Stanislaus* and write two pieces called *La lugubre gondola (The Gondola of Mourning)*, inspired by the funeral processions he saw on the canals. Only two months later Wagner died in Venice, and his body was carried in a similar procession.

In 1886 Liszt left Rome for the last time. He attended concerts of his works in Budapest, Liège, and Paris and then went to London—his first visit there in 45 years. Several concerts of his works were given, including a performance of St. *Elisabeth*. He then went on to Antwerp, Paris, and Weimar. He played for the last time at a concert in Luxembourg on July 19. Two days later he arrived in Bayreuth for the festival. His health had not been good for some months, and he went to bed with a high fever, though he still managed to attend two Wagner performances. His illness developed into pneumonia, and his condition was not helped by the callous behaviour of Cosima, who left him alone in order to supervise the running of the festival. He died on July 31.

Assessment. Apart from his more than 700 compositions, Liszt wrote books on Chopin, Hungarian Gypsy music, Wagner's *Lohengrin* and *Tannhäuser*, John Field's nocturnes, the lieder of Robert Franz, and the Goethe Foundation in Weimar. His published essays and correspondence fill many volumes. A controversial figure in his time, he was attacked for his innovations, and his rivals were jealous of his brilliance and panache. For a long time he was regarded merely as a superficial composer of brilliant trifles, but in recent years his true stature has been seen more clearly as that of a composer who revolutionized the music of his time and anticipated many later developments. As Princess Carolyne said, "Liszt has flung his spear far into the future."

MAJOR WORKS

Orchestral works

TONE POEMS: Thirteen, also *A Faust Symphony* in three charactcr-pictures (after Goethe, 1854, rev. 1857–61); *A Symphony to Dante's Divina Commedia* (1855–56); *Trois Odes Funèbres* (1860–66).

PIANO AND ORCHESTRA: *Concerto No.* I *in E Flat Major* (1849, rev. 1853 and 1856); *Concerto No.* 2 *in A Major* (1839, rev. 1849, 1853, 1857, and 1861); *Fantasia on Hungarian Folk Tunes* (*c.* 1852); *Totentanz* (1849, rev. 1853 and 1859).

Solo piano

Apparitions (1834); *Album d'un voyageur* (1835–36); *Années ile pèlerinage* (1835–77); *Etudes d'exdcution transcendante d'après Paganini* (1838); *Harmonies poétiques et religieuses* (1845–52); *Concert Studies* (*c.* 1848 and *c.* 1862–63); *Études d'exe'cution transcendante* (1851); *Grandes Études de Paganini* (1851); *Sonata in B Minor* (1852–53); *Ldgendes* (*c.* 1863); *Historische ungarische Biidnisse* (1885).

Vocal music

SACRED CHORAL WORKS: *Missa solennis zur Einweihung der Basilika irt Gran* (called the *Gran Mass;* 1855, rev. 1857–58); *Christus*, oratorio (1855–66); *Die Legende von der heiligen Elisabeth,* oratorio (1857–62); *Requiem*, two tenors, two bass soloists, male chorus, organ, and brass (1867–68); *Die heilige Cäcilia*, mezzo-soprano, chorus, orchestra, or piano (*c.* 1874); 70 songs.

BIBLIOGRAPHY. The most detailed catalog of works in English is in *Grove's Dictionary of Music and Musicians,* 5th ed., vol. 5 (1954), updated in HUMPHREY **SEARLE**, *The Music of Liszt,* 2nd rev. ed. (1966); another catalog with different numbering is in the symposium, edited by ALAN WALKER, *Franz Liszt: The Man and His Music* (1970). A detailed catalog in German is in PETER **RAABE**, *Franz Liszt,* vol. 2 (1968). The most complete bibliography is LUDWIG KOCH, *Franz Liszt: Ein bibliographischer Versuch* (1936). The principal collections of Liszt manuscripts are in the Liszt Museum, housed in the Goethe-und-Schiller Archiv, Weimar, East Germany; the Hungarian National Museum, Budapest; the Library of Congress, Washington, D.C.; the Bibliothèque du Conservatoire National, Paris; and the British Museum. The chief editions of Liszt's letters in English are *Franz Liszts Briefe* (1893; Eng. trans. of vol. 1–2, *Letters* of *Franz*

Liszt, 1894 and 1968) and *Briefwechsel zwischen Franz Liszt und Hans von Bülow* (1898; Eng. trans., *Letters of Liszt and von Bülow,* 1898), both trans. by CONSTANCE BACHE; *Briefwechsel zwischen Wagner und Liszt,* 2 vol. (1887; Eng. trans. by FRANCIS HUEFPER, *Correspondence of Wagner and Liszt,* 2 vol., 1897, new ed. 1969); *Alexandre Borodine, d'après la biographie et la correspondance* (1893; Eng. trans. of pt. 2 by ROSA NEWMARCH, *Liszt, As Sketched in the Letters of Borodin,* 1895); and *The Letters of Franz Liszt to Marie zu Sayn-Wittgenstein,* trans. and ed. by HOWARD E. HUGO (1953). The best biography in English is SACHEVERELL SITWELL, *Liszt,* rev. ed. (1955, reprinted 1966). The official biography is by LINA RAMANN, *Franz Liszt: Als Künstler und Mensch, 2* vol. (1890–94), only the first volume of which has been translated into English: *Franz Liszt: Artist and Man,* 2 vol. (1882). Other useful biographical sources in English are AMY FAY, *Music Study in Germany* (1890, reprinted 1965); the *Life and Letters of Sir Charles Halle* (1896); the *Memoirs of Count Apponyi* (1935); and ALAN WALKER, *Liszt* (1971). The principal biographies in other languages are EMILE HARASZTI, *Franz Liszt* (1967), in French; and PETER RAABE, *Franz Liszt,* vol. *1* (1968), in German. The chief critical studies in English of Liszt's music are BENCE SZABOLCSI, *Liszt Ferenc estéje* (1956; Eng. trans., *The Tivilight of Ferenc Liszt,* 1959), concentrating on Liszt's later works; ALAN WALKER (*op. cit.*); and HUMPHREY SEARLE, *The Music of Liszt,* 2nd rev. ed. (1966). Other major studies are PETER RAABE, *Frani Liszt,* vol. *2* (1968), in German; FRIEDRICH SCHNAPP, "Verschollene Kompositionen Franz Liszts," in *Von Deutscher Tonkunst* (1942), dealing with doubtful or unknown works; and the Liszt-Bartók volume of *Studia Musicologica* (1963), essays, in various languages, on Liszt's life and works.

(H.Se.)

Literary Criticism

Criticism
and
scholarship

Construed loosely, literary criticism is the reasoned consideration of literary works and issues. It applies, as a term, to any argumentation about literature, whether or not specific works are analyzed. Plato's cautions against the risky consequences of poetic inspiration in general in his *Republic* are thus often taken as the earliest important example of literary criticism. More strictly construed, the term covers only what has been called "practical criticism," the interpretation of meaning and the judgment of quality. Criticism in this narrow sense can be distinguished not only from aesthetics (the philosophy of artistic value) but also from other matters that may concern the student of literature: biographical questions, bibliography, historical knowledge, sources and influences, and problems of method. Thus, especially in academic studies, "criticism" is often considered to be separate from "scholarship." In practice, however, this distinction often proves artificial, and even the most single-minded concentration on a text may be informed by outside knowledge, while many notable works of criticism combine discussion of texts with broad arguments about the nature of literature and the principles of assessing it. Criticism will here be taken to cover all phases of literary understanding, though the emphasis will be on the evaluation of literary works and of their authors' places in literary history. The development of literary criticism will be treated briefly, focussing on what is distinctive to literary criticism. A longer and more theoretical discussion of the role of criticism will be found in the article ARTS, CRITICISM OF THE, and a discussion of criticism as a form in which works of literature are sometimes cast, in the article LITERATURE, NONFICTIONAL PROSE. One particular aspect of literary criticism is covered in the article TEXTUAL CRITICISM. The following article deals with criticism largely in the context of Western literature. Its role in other literatures, along with the history of the literatures themselves, is treated in LITERATURE, EAST ASIAN, and in articles on the arts of various peoples, such as SOUTH ASIAN PEOPLES, ARTS OF; and AFRICAN PEOPLES, ARTS OF. Criticism and theory associated with particular genres of literature, such as tragedy and comedy, will be found in separate articles on such subjects.

FUNCTIONS OF LITERARY CRITICISM

The functions of literary criticism vary widely, ranging from the reviewing of books as they are published to systematic theoretical discussion. Though reviews may sometimes determine whether a given book will be widely sold, many works succeed commercially despite negative reviews, and many classic works, including Herman Melville's *Moby Dick* (1851), have acquired appreciative publics long after being unfavourably reviewed and at first neglected. One of criticism's principal functions is to express the shifts in sensibility that make such revaluations possible. The minimal condition for such a new appraisal is, of course, that the original text survive. The literary critic is sometimes cast in the role of scholarly detective, unearthing, authenticating, and editing unknown manuscripts. Thus, even rarefied scholarly skills may be put to criticism's most elementary use, the bringing of literary works to a public's attention.

The variety of criticism's functions is reflected in the range of publications in which it appears. Criticism in the daily press rarely displays sustained acts of analysis and may sometimes do little more than summarize a publisher's claims for a book's interest. Weekly and biweekly magazines serve to introduce new books but are often more discriminating in their judgments, and some of these magazines, such as *The* (London) *Times Literary Supplement* and *The New York Review of Books,* are far from indulgent toward popular works. Sustained criticism can also be found in monthlies and quarterlies with a broad circulation, in "little magazines" for specialized audiences, and in scholarly journals and books.

Criticism
as a
target of
criticism

Because critics often try to be lawgivers, implicitly or explicitly declaring which works deserve respect and presuming to say what they are "really" about, criticism is a perennial target of resentment. Misguided or malicious critics can discourage an author who has been feeling his way toward a new mode that offends received taste. Pedantic critics can actually obstruct a serious engagement with literature by deflecting attention toward inessential matters. As the contemporary French philosopher-critic Jean-Paul Sartre observed, the critic may announce that French thought is a perpetual colloquy between Pascal and Montaigne not in order to make those thinkers more alive but to make thinkers of his own time more dead. Criticism can antagonize authors even when it performs its function well. Authors who customarily regard literature as needing no advocates or investigators are less than grateful when told that their works possess unintended meaning or are imitative or incomplete.

What such authors may tend to forget is that their works, once published, belong to them only in a legal sense; their true owner is the public, which will appropriate them for its own concerns whether or not the critic urges them to. The critic's responsibility is not to the author's self-esteem but to the public and to his own standards of judgment, which are usually more exacting than the public's. The justification for his role rests on the premise that literary works are not in fact self-explanatory. A critic is socially useful to the extent that his society wants, and receives from him, a fuller understanding of literature than it could have achieved without him. In filling this appetite the critic whets it further, helping to create a public that cares about artistic quality. Without sensing the presence of such a public an author may either prostitute his talent or squander it in sterile acts of defiance. In this sense the critic is not a parasite but, potentially, someone who is responsible in part for the existence of good writing in his own time and afterward.

Although some critics believe that literature should be discussed in isolation from other matters, criticism usually seems to be openly or covertly involved with social and political debate. Since literature itself is often partisan, is always rooted to some degree in local circumstances, and has a way of calling forth affirmations of ultimate values, it is not surprising that the finest critics have never paid much attention to the alleged boundaries between criticism and other types of discourse. Especially in modern Europe, literary criticism has occupied a central place in debate about cultural and political issues. Sartre's own *What Is Literature?* (1947) is typical in its wide-ranging attempt to prescribe the literary intellectual's idea! relation to the development of his society and to literature as a manifestation of human freedom. Similarly, some prominent American critics, including Alfred Kazin, Lionel Trilling, Kenneth Burke, Philip Rahv, and Irving Howe, began as political radicals in the 1930s and sharpened their concern for literature on the dilemmas and disillusionments of that era. Trilling's influential *The* Liberal Imagination (1950) is simultaneously a collection of literary essays and an attempt to reconcile the claims of politics and art.

Such a reconciliation is bound to be tentative and problematic if the critic believes, as Trilling does, that literature possesses an independent value and a deeper faithfulness to reality than is contained in any political formula. In Marxist states, however, literature has usually been considered a means to social ends, and criticism has been cast in forthrightly partisan terms. Dialectical materialism does not necessarily turn the critic into a mere guardian of party doctrine, but it does forbid him to treat literature as a cause in itself, apart from the working class's needs as interpreted by the party. Where this utilitarian view prevails, the function of criticism is taken to be continuous with that of the state itself, namely, furtherance of the social revolution. The critic's main obligation is not to his texts but to the masses of people whose consciousness must be advanced in the designated direction. In periods of severe orthodoxy the practice of literary criticism has not always been distinguishable from that of censorship.

HISTORICAL DEVELOPMENT

Antiquity. Although almost all of the criticism ever written dates from the 20th century, questions first posed by Plato and Aristotle are still of prime concern, and every critic who has attempted to justify the social value of literature has had to come to terms with the opposing argument made by Plato in *The* Republic. The poet as a man and poetry as a form of statement both seemed untrustworthy to Plato, who depicted the physical world as an imperfect copy of transcendent ideas and poetry as a mere copy of the copy. Thus, literature could only mislead the seeker of truth. Plato credited the poet with divine inspiration, but this, too, was cause for worry: a man possessed by such madness would subvert the interests of a rational polity. Poets were therefore to be banished from the hypothetical republic.

In his Poetics—still the most respected of all discussions of literature—Aristotle countered Plato's indictment by stressing what is normal and useful about literary art. The tragic poet is not so much divinely inspired as he is motivated by a universal human need to imitate, and what he imitates is not something like a bed (Plato's example) but a noble action. Such imitation presumably has a civilizing value for those who empathize with it. Tragedy does arouse emotions of pity and terror in its audience, but these emotions are purged in the process (*katharsis*). In this fashion Aristotle succeeded in portraying literature as satisfying and regulating human passions instead of inflaming them.

Although Plato and Aristotle are regarded as antagonists, the narrowness of their disagreement is noteworthy. Both maintain that poetry is mimetic, both treat the arousing of emotion in the perceiver, and both feel that poetry takes its justification, if any, from its service to the state. It was obvious to both men that poets wielded great power over others. Unlike many modern critics who have tried to show that poetry is more than a pastime, Aristotle had to offer reassurance that it was not socially explosive.

Aristotle's practical contribution to criticism, as opposed to his ethical defense of literature, lies in his inductive treatment of the elements and kinds of poetry. Poetic modes are identified according to their means of imitation, the actions they imitate, the manner of imitation, and its effects. These distinctions assist the critic in judging each mode according to its proper ends instead of regarding beauty as a fixed entity. The ends of tragedy, as Aristotle conceived them, are best served by the harmonious disposition of six elements: plot, character, diction, thought, spectacle, and song. Thanks to Aristotle's insight into universal aspects of audience psychology, many of his dicta have proved to be adaptable to genres developed long after his time.

Later Greek and Roman criticism offers no parallel to Aristotle's originality. Much ancient criticism, such as that of Cicero, Horace, and Quintilian in Rome, was absorbed in technical rules of exegesis and advice to aspiring rhetoricians. Horace's verse epistle *The* Art of Poetry is an urbane amplification of Aristotle's emphasis on the decorum or internal propriety of each genre, now including lyric, pastoral, satire, elegy, and epigram, as well as Aristotle's epic, tragedy, and comedy. This work was later to be prized by Neoclassicists of the 17th century not only for its rules but also for its humour, common sense, and appeal to educated taste. On the *Sublime*, by the Roman-Greek known as "Longinus," was to become influential in the 18th century but for a contrary reason: when decorum began to lose its sway encouragement could be found in Longinus for arousing elevated and ecstatic feeling in the reader. Horace and Longinus developed, respectively, the rhetorical and the affective sides of Aristotle's thought, but Longinus effectively reversed the Aristotelian concern with regulation of the passions.

Horatian
and
Longinian
tendencies

Medieval period. In the Christian Middle Ages criticism suffered from the loss of nearly all the ancient critical texts and from an antipagan distrust of the literary imagination. Such Church Fathers as Tertullian, Augustine, and Jerome renewed, in churchly guise, the Platonic argument against poetry. But both the ancient gods and the surviving classics reasserted their fascination, entering medieval culture in theologically allegorized form. Encyclopaedists and textual commentators explained the supposed Christian content of pre-Christian works and the Old Testament. Although there was no lack of rhetoricians to dictate the correct use of literary figures, no attempt was made to derive critical principles from emergent genres such as the fabliau and the chivalric romance. Criticism was in fact inhibited by the very coherence of the theologically explained universe. When nature is conceived as endlessly and purposefully symbolic of revealed truth, specifically literary problems of form and meaning are bound to be neglected. Even such an original vernacular poet of the 14th century as Dante appears to have expected his Divine Comedy to be interpreted according to the rules of scriptural exegesis.

The Renaissance. Renaissance criticism grew directly from the recovery of classic texts and notably from Giorgio Valla's translation of Aristotle's Poetics into Latin in 1498. By 1549 the Poetics had been rendered into Italian as well. From this period until the later part of the 18th century Aristotle was once again the most imposing presence behind literary theory. Critics looked to ancient poems and plays for insight into the permanent laws of art. The most influential of Renaissance critics was probably Lodovico Castelvetro, whose 1570 commentary on Aristotle's Poetics encouraged the writing of tightly structured plays by extending and codifying Aristotle's idea of the dramatic unities. It is difficult today to appreciate that this obeisance to antique models had a liberating effect; one must recall that imitation of the ancients entailed rejecting scriptural allegory and asserting the individual author's ambition to create works that would be unashamedly great and beautiful. Classicism, individualism, and national pride joined forces against literary asceticism. Thus, a group of 16th-century French writers known as the Pléiade—notably Pierre de Ronsard and Joachim du Bellay—were simultaneously classicists, poetic innovators, and advocates of a purified vernacular tongue.

Classicism,
individualism, and
national
pride
as
literary
forces

The ideas of the Italian and French Renaissance were transmitted to England by Roger Ascham, George Gascoigne, Sir Philip Sidney, and others. Gascoigne's "Certayne notes of Instruction" (1575), the first English manual of versification, had a considerable effect on poetic practice in the Elizabethan Age. Sidney's Defence of Poe*sie* (1595) vigorously argued the poet's superiority to the philosopher and the historian on the grounds that his imagination is chained neither to lifeless abstractions nor to dull actualities. The poet "doth not only show the way, but giveth so sweet a prospect into the way, as will entice any man to enter into it." While still honouring the traditional conception of poetry's role as bestowing pleasure and instruction, Sidney's essay presages the Romantic claim that the poetic mind is a law unto itself.

Neoclassicism and its decline. The Renaissance in general could be regarded as a neoclassical period, in that ancient works were considered the surest models for modern greatness. Neoclassicism, however, usually connotes narrower attitudes that are at once literary and social: a worldly-wise tempering of enthusiasm, a fondness for proved ways, a gentlemanly sense of propriety and balance. Criticism of the 17th and 18th centuries, particularly in France, was dominated by these Horatian norms. French critics such as Pierre Corneille and Nicolas Boileau urged a strict orthodoxy regarding the dramatic unities and the requirements of each distinct genre, as if to disregard them were to lapse into barbarity. The poet was not to imagine that his genius exempted him from the established laws of craftsmanship.

Neoclassicism had a lesser impact in England, partly because English Puritanism had kept alive some of the original Christian hostility to secular art, partly because English authors were on the whole closer to plebeian taste than were the court-oriented French, and partly because of the difficult example of Shakespeare, who magnificently broke all of the rules. Not even the relatively severe classicist Ben Jonson could bring himself to deny Shakespeare's greatness, and the theme of Shakespearean genius triumphing over formal imperfections is echoed by major British critics from John Dryden and Alexander Pope through Samuel Johnson. The science of Newton and the psychology of Locke also worked subtle changes on neoclassical themes. Pope's *Essay on Criticism* (1711) is a Horatian compendium of maxims, but Pope feels obliged to defend the poetic rules as "Nature methodiz'd"—a portent of quite different literary inferences from Nature. Dr. Johnson, too, though he respected precedent, was above all a champion of moral sentiment and "mediocrity," the appeal to generally shared traits. His preference for forthright sincerity left him impatient with such intricate conventions as those of the pastoral elegy.

Progress of a cult of “feeling”

The decline of Neoclassicism is hardly surprising; literary theory had developed very little during two centuries of artistic, political, and scientific ferment. The 18th century's important new genre, the novel, drew most of its readers from a bourgeoisie that had little use for aristocratic dicta. A Longinian cult of "feeling" gradually made headway, in various European countries, against Neoclassical canons of proportion and moderation. Emphasis shifted from concern for meeting fixed criteria to the subjective state of the reader and then of the author himself. The spirit of nationalism entered criticism as a concern for the origins and growth of one's own native literature and as an esteem for such non-Aristotelian factors as "the spirit of the age." Historical consciousness produced by turns theories of literary progress and primitivistic theories affirming, as one critic put it, that "barbarous" times are the most favourable to the poetic spirit. The new recognition of strangeness and strong feeling as literary virtues yielded various fashions of taste for misty sublimity, graveyard sentiments, medievalism, Norse epics (and forgeries), Oriental tales, and the verse of plowboys. Perhaps the most eminent foes of Neoclassicism before the 19th century were Denis Diderot in France and, in Germany, Gotthold Lessing, Johann von Herder, Johann Wolfgang von Goethe, and Friedrich Schiller.

Romanticism. Romanticism, an amorphous movement that began in Germany and England at the turn of the 19th century, and somewhat later in France, Italy, and the United States, found spokesmen as diverse as Goethe and August and Friedrich von Schlegel in Germany, William Wordsworth and Samuel Taylor Coleridge in England, Madame de Stael and Victor Hugo in France, Alessandro Manzoni in Italy, and Ralph Waldo Emerson and Edgar Allan Poe in the United States. Romantics tended to regard the writing of poetry as a transcendentally important activity, closely related to the creative perception of meaning in the world. The poet was credited with the godlike power that Plato had feared in him; Transcendental philosophy was, indeed, a derivative of Plato's metaphysical Idealism. In the typical view of Percy Bysshe Shelley, poetry "strips the veil of familiarity from the world, and lays bare the naked and sleeping beauty, which is the spirit of its forms."

Wordsworth's preface to *Lyrical Ballads* (1800), with its definition of poetry as the spontaneous overflow of powerful feelings and its attack on Neoclassical diction, is regarded as the opening statement of English Romanticism. In England, however, only Coleridge in his *Biographia Literaria* (1817) embraced the whole complex of Romantic doctrines emanating from Germany; the British empiricist tradition was too firmly rooted to be totally washed aside by the new metaphysics. Most of those who were later called Romantics did share an emphasis on individual passion and inspiration, a taste for symbolism and historical awareness, and a conception of art works as internally whole structures in which feelings are dialectically merged with their contraries. Romantic criticism coincided with the emergence of aesthetics as a separate branch of philosophy, and both signalled a weakening in ethical demands upon literature. The lasting achievement of Romantic theory is its recognition that artistic creations are justified, not by their promotion of virtue, but by their own coherence and intensity.

Late 19th century. The Romantic movement had been spurred not only by German philosophy but also by the universalistic and utopian hopes that accompanied the French Revolution. Some of those hopes were thwarted by political reaction, while others were blunted by industrial capitalism and the accession to power of the class that had demanded general liberty. Advocates of the literary imagination now began to think of themselves as enemies or gadflies of the newly entrenched bourgeoisie. In some hands the idea of creative freedom dwindled to a bohemianism pitting "art for its own sake" against commerce and respectability. Aestheticism characterized both the Symbolist criticism of Charles Baudelaire in France and the self-conscious decadence of Algernon Swinburne, Walter Pater, and Oscar Wilde in England. At an opposite extreme, realistic and naturalistic views of literature as an exact record of social truth were developed by Vissarion Belinsky in Russia, Gustave Flaubert and Émile Zola in France, and William Dean Howells in the United States. Zola's program, however, was no less anti-bourgeois than that of the Symbolists; he wanted novels to document conditions so as to expose their injustice. Post-Romantic disillusion was epitomized in Britain in the criticism of Matthew Arnold, who thought of critical taste as a substitute for religion and for the unsatisfactory values embodied in every social class.

Merging intellectual currents

Toward the end of the 19th century, especially in Germany, England, and the United States, literary study became an academic discipline "at the doctoral level." Philology, linguistics, folklore study, and the textual principles that had been devised for biblical criticism provided curricular guidelines, while academic taste mirrored the prevailing impressionistic concern for the quality of the author's spirit. Several intellectual currents joined to make possible the writing of systematic and ambitious literary histories. Primitivism and Medievalism had awakened interest in neglected early texts; scientific Positivism encouraged a scrupulous regard for facts; and the German idea that each country's literature had sprung from a unique national consciousness provided a conceptual framework. The French critic Hippolyte Taine's *History of English Literature* (published in French, 1863–69) reflected the prevailing determinism of scientific thought; for him a work could be explained in terms of the race, milieu, and moment that produced it. For other critics of comparable stature, such as Charles Sainte-Beuve in France, Benedetto Croce in Italy, and George Saintsbury in England, historical learning only threw into relief the expressive uniqueness of each artistic temperament.

The 20th century. The ideal of objective research has continued to guide Anglo-American literary scholarship and criticism and has prompted work of unprecedented accuracy. Bibliographic procedures have been revolutionized; historical scholars, biographers, and historians of theory have placed criticism on a sounder basis of factuality. Important contributions to literary understanding have meanwhile been drawn from anthropology, linguistics, philosophy, and psychoanalysis. Impressionistic method has given way to systematic inquiry from which gratuitous assumptions are, if possible, excluded. Yet demands for a more ethically committed criticism have repeatedly been made, from the New Humanism of Paul Elmer More and Irving Babbitt in the United States in the 1920s, through the moralizing criticism of the Cambridge don F.R. Leavis and of the American poet Yvor Winters, to the most recent demands for "relevance."

No sharp line can be drawn between academic criticism and criticism produced by authors and men of letters. Many of the latter are now associated with universities, and the main shift of academic emphasis, from impressionism to formalism, originated outside the academy in the writings of Ezra Pound, T.S. Eliot, and T.E. Hulme, largely in London around 1910. Only subsequently did such academics as I.A. Richards and William Empson in England and John Crowe Ransom and Cleanth Brooks in the United States adapt the New Criticism to reform of the literary curriculum—in the 1940s. New Criticism has been the methodological counterpart to the strain of modernist literature characterized by allusive difficulty, paradox, and indifference or outright hostility to the democratic ethos. In certain respects the hegemony of New Criticism has been political as well as literary; and anti-Romantic insistence on irony, convention, and aesthetic distance has been accompanied by scorn for all revolutionary hopes. In Hulme conservatism and classicism were explicitly linked. Romanticism struck him as "spilt religion," a dangerous exaggeration of human freedom. In reality, however, New Criticism owed much to Romantic theory, especially to Coleridge's idea of organic form, and some of its notable practitioners have been left of centre in their social thought.

Multiplicity and factionalism

The totality of Western criticism in the 20th century defies summary except in terms of its restless multiplicity and factionalism. Schools of literary practice, such as Imagism, Futurism, Dadaism, and Surrealism, have found no want of defenders and explicators. Ideological groupings, psychological dogmas, and philosophical trends have generated polemics and analysis, and literary materials have been taken as primary data by sociologists and historians. Literary creators themselves have continued to write illuminating commentary on their own principles and aims. In poetry, Paul Valéry, Ezra Pound, Wallace Stevens; in the theatre, George Bernard Shaw, Antonin Artaud, Bertolt Brecht; and in fiction, Marcel Proust, D.H. Lawrence, and Thomas Mann have contributed to criticism in the act of justifying their art.

PRESENT STATE AND PROSPECTS

Most of the issues debated in 20th-century criticism appear to be strictly empirical, even technical, in nature. By what means can the most precise and complete knowledge of a literary work be arrived at? Should its social and biographical context be studied or only the words themselves as an aesthetic structure? Should the author's avowed intention be trusted, or merely taken into account, or disregarded as irrelevant? How is conscious irony to be distinguished from mere ambivalence, or allusiveness from allegory? Which among many approaches—linguistic, generic, formal, sociological, psychoanalytic, and so forth—is best adapted to making full sense of a text? Would a synthesis of all these methods yield a total theory of literature? Such questions presuppose that literature is valuable and that objective knowledge of its workings is a desirable end. These assumptions are, indeed, so deeply buried in most critical discourse that they customarily remain hidden from critics themselves, who imagine that they are merely solving problems of intrinsic interest.

What separates modern criticism from earlier work is its catholicity of scope and method, its borrowing of procedures from the social sciences, and its unprecedented attention to detail. As literature's place in society has become more problematic and peripheral, and as humanistic education has grown into a virtual industry with a large group of professionals serving as one another's judges, criticism has evolved into a complex discipline, increasingly refined in its procedures but often lacking a sense of contact with the general social will. Major modern critics, to be sure, have not allowed their "close reading" to distract them from certain perennial questions about poetic truth. the nature of literary satisfaction, and literature's social utility, but even these matters have sometimes been cast in "value-free" empirical terms.

Science as the outstanding influence

Recourse to scientific authority and method, then, is the outstanding trait of 20th-century criticism. The sociology of Marx, Max Weber, and Karl Mannheim, the mythological investigations of Sir James George Frazer and his followers, Edmund Husserl's phenomenology, Claude Levi-Strauss's anthropological structuralism, and the psychological models proposed by Sigmund Freud and C.G. Jung have all found their way into criticism. The result has been not simply an abundance of technical terms and rules, but a widespread belief that literature's governing principles can be located outside literature. Jungian "archetypal" criticism, for example, regularly identifies literary power with the presence of certain themes that are alleged to inhabit the myths and beliefs of all cultures, while psychoanalytic exegetes interpret poems in exactly the manner that Freud interpreted dreams. Such procedures may encourage the critic, wisely or unwisely, to discount traditional boundaries between genres, national literatures, and levels of culture; the critical enterprise begins to seem continuous with a general study of man. The impetus toward universalism can be discerned even in those critics who are most skeptical of it, the so-called historical relativists who attempt to reconstruct each epoch's outlook and to understand works as they appeared to their first readers. Historical relativism does undermine cross-cultural notions of beauty, but it reduces the record of any given period to data from which inferences can be systematically drawn. Here, too, in other words, uniform methodology tends to replace the intuitive connoisseurship that formerly typified the critic's sense of his role.

The debate over poetic truth may illustrate how modern discussion is beholden to extraliterary knowledge. Critics have never ceased disputing whether literature depicts the world correctly, incorrectly, or not at all, and the dispute has often had more to do with the support or condemnation of specific authors than with ascertainable facts about mimesis. Today it may be almost impossible to take a stand regarding poetic truth without also coming to terms with positivism as a total epistemology. The spectacular achievements of physical science have (with logic questioned by some) downgraded intuition and placed a premium on concrete, testable statements very different from those found in poems. Some of the most influential modern critics, notably I.A. Richards in his early works, have accepted this value order and have confined themselves to behavioristic study of how literature stimulates the reader's feelings. A work of literature, for them, is no longer something that captures an external or internal reality, but is merely a locus for psychological operations; it can only be judged as eliciting or failing to elicit a desired response.

Literature as a form of knowing

Other critics, however, have renewed the Shelleyan and Coleridgean contention that literary experience involves a complex and profound form of knowing. In order to do so they have had to challenge Positivism in general. Such a challenge cannot be convincingly mounted within the province of criticism itself and must depend rather on the authority of antipositivist epistemologists such as Alfred North Whitehead, Ernst Cassirer, and Michael Polanyi. If it is now respectable to maintain, with Wallace Stevens and others, that the world is known through imaginative apprehensions of the sort that poetry celebrates and employs, this is attributable to developments far outside the normal competence of critics.

The pervasive influence of science is most apparent in modern criticism's passion for total explanation of the texts it brings under its microscope. Even formalist schools, which take for granted an author's freedom to shape his work according to the demands of art, treat individual lines of verse with a dogged minuteness that was previously unknown, hoping thereby to demonstrate the "organic" coherence of the poem. The spirit of explanation is also apparent in those schools that argue from the circumstances surrounding a work's origin to the work itself, leaving an implication that the former have caused the latter. The determinism is rarely as explicit or relentless as it was in Taine's scheme of race, milieu, and moment, but this may reflect the fact that causality in general is now handled with more sophistication than in Taine's day.

Whether criticism will continue to aim at empirical exactitude or will turn in some new direction cannot be readily predicted, for the empiricist ideal and its sanctuary, the university, are not themselves secure from attack. The history of criticism is one of oscillation between periods of relative advance, when the imaginative freedom of great writers prompts critics to extend their former conceptions, and periods when stringent moral and formal presciptions are laid upon literature. In times of social upheaval criticism may more or less deliberately abandon the ideal of disinterested knowledge and be mobilized for a practical end. Revolutionary movements provide obvious instances of such redirection, whether or not they identify their pragmatic goals with the cause of science. It should be evident that the future of criticism depends on factors that lie outside criticism itself as a rationally evolving discipline. When a whole society shifts its attitudes toward pleasure, unorthodox behaviour, or the meaning of existence, criticism must follow along.

Literature
as a
source
of order

As Matthew Arnold foresaw, the waning of religious certainty has encouraged critics to invest their faith in literature, taking it as the one remaining source of value and order. This development has stimulated critical activity, yet, paradoxically, it may also be responsible in part for a growing impatience with criticism. What Arnold could not have anticipated is that the faith of some moderns would be apocalyptic and Dionysian rather than a sober and attentuated derivative of Victorian Christianity. Thought in the 20th century has yielded a strong undercurrent of anarchism which celebrates libidinous energy and self-expression at the expense of all social constraint, including that of literary form. In the critical writings of D.H. Lawrence, for example, fiction is cherished as an instrument of unconscious revelation and liberation. A widespread insistence upon prophetic and ecstatic power in literature seems at present to be undermining the complex, irony-minded formalism that has dominated modern discourse. As literary scholarship has acquired an ever-larger arsenal of weapons for attacking problems of meaning, it has met with increasing resentment from people who wish to be nourished by whatever is elemental and mysterious in literary experience.

An awareness of critical history suggests that the development is not altogether new, for criticism stands now approximately where it did in the later 18th century, when the Longinian spirit of expressiveness contested the sway of Boileau and Pope. To the extent that modern textual analysis has become what Hulme predicted, "a classical revival," it may not be welcomed by those who want a direct and intense rapport with literature. What is resisted now is not Neoclassical decorum but impersonal methodology, which is thought to deaden commitment. Such resistance may prove beneficial if it reminds critics that rationalized procedures are indeed no substitute for engagement. Excellent work continues to be written, not because a definitive method or synthesis of methods has been found, but on the contrary because the best critics still understand that criticism is an exercise of private sympathy, discrimination, and moral and cultural reflection.

BIBLIOGRAPHY. A useful compilation of essential texts is MARK SCHORER, JOSEPHINE MILES, and GORDON MCKENZIE (eds.), *Criticism,* rev. ed. (1958). The best survey of critical history is WILLIAM K. WIMSATT, JR., and CLEANTH BROOKS, *Literary Criticism: A Short History* (1957); G.M.A. GRUBE, *The Greek and Roman Critics* (1965); JOEL E. SPINGARN, *A History of Literary Criticism in the Renaissance,* 5th ed. (1925, paperback edition 1963); WALTER J. BATE, *From Classic to Romantic* (1946, reprinted 1961); and RENE WELLEK, *A History of Modern Criticism, 1750–1950*, 4 vol. (1955–65), are more specialized historical studies. Important theoretical statements are I.A. RICHARDS, *Principles of Literary Criticism* (1924, reprinted 1961); M.H. ABRAMS, *The Mirror and the Lamp* (1953); RENE WELLEK and AUSTIN WARREN, *Theory of Literature,* 3rd rev. ed. (1966); NORTHROP FRYE, *Anatomy of Criticism* (1957); and WAYNE C. BOOTH, *The Rhetoric of Fiction* (1961). WILLIAM EMPSON, *Seven Types of Ambiguity,* 3rd ed. (1956, reprinted 1963); ERICH AUERBACH, *Mimesis* (1946; Eng. trans., 1953); and LIONEL TRILLING, *The Liberal Imagination* (1950), are representative examples of modern criticism, combining theory with analysis of a wide variety of texts.

(F.C.C.)

Literature, Art of

Definitions of the word literature tend to be circular. *The Concise Oxford Dictionary* says it is "writings whose value lies in the beauty of form or emotional effect." The 19th-century critic Walter Pater referred to "the matter of imaginative or artistic literature" as a "transcript, not of mere fact, but of fact in its infinitely varied forms." But such definitions really assume that the reader already knows what literature is. And indeed its central meaning, at least, is clear enough. Deriving from the Latin *littera*, "a letter of the alphabet," literature is first and foremost mankind's entire body of writing; after that it is the body of writing belonging to a given language or people; then it is individual pieces of writing.

But already it is necessary to qualify these statements. To use the word writing when describing literature is itself misleading, for one may rightly speak of "oral literature" or "the literature of preliterate peoples." The art of literature is not reducible to the words on the page; they are there because of the craft of writing. As an art, literature is the organization of words to give pleasure; through them it elevates and transforms experience; through them it functions in society as a continuing symbolic criticism of values.

THE SCOPE OF LITERATURE

Literature is a form of human expression. But not everything expressed in words—even when organized and written down—is counted as literature. Those writings that are primarily informative—technical, scholarly, journalistic—would be excluded from the rank of literature by most, though not all, critics. Certain forms of writing, however, are universally regarded as belonging to literature as an art. Individual attempts within these forms are said to succeed if they possess something called artistic merit and to fail if they do not. The nature of artistic merit is less easy to define than to recognize. The writer need not even pursue it to attain it. On the contrary, a scientific exposition might be of great literary value and a pedestrian poem of none at all.

The purest (or, at least, the most intense) literary form is the lyric poem, and after it comes elegiac, epic, dramatic, narrative, and expository verse. Most theories of literary criticism base rhemselves on an analysis of poetry, because the aesthetic problems of literature are there presented in their simplest and purest form. Poetry that fails as literature is not called poetry at all but verse. Many novels—certainly all the world's great novels—are literature, but there are thousands that are not so considered. Most great dramas are considered literature (although the Chinese, possessors of one of the world's greatest dramatic traditions, consider their plays, with few exceptions, to possess no literary merit whatsoever).

Distinction
between
poetry
and
verse

The Greeks thought of history as one of the seven arts, inspired by a goddess, the muse Clio. All of the world's classic surveys of history can stand as noble examples of the art of literature, but most historical works and studies today are not written primarily with literary excellence in mind, though they may possess it, as it were, by accident.

The essay was once written deliberately as a piece of literature: its subject matter was of comparatively minor importance. Today most essays are written as expository, informative journalism, although there are still essayists in the great tradition who think of themselves as artists. Now, as in the past, some of the greatest essayists are critics of literature, drama, and the arts.

Some personal documents (autobiographies, diaries, memoirs, and letters) rank among the world's greatest literature. Some examples of this biographical literature were written with posterity in mind, others with no thought of their being read by anyone but the writer. Some are in a highly polished literary style; others, couched in a privately evolved language, win their standing as literature because of their cogency, insight, depth, and scope.

Many works of philosophy are classed as literature. The *Dialogues* of Plato (4th century BC) are written with great narrative skill and in the finest prose; the *Meditations* of the 2nd-century Roman emperor Marcus Aurelius are a collection of apparently random thoughts, and the Greek in which they are written is eccentric. Yet both are classed as literature, while the speculations of other philosophers, ancient and modern, are not. Certain scientific works endure as literature long after their scientific content has become outdated. This is particularly true of books of natural history, where the element of personal observation is of special importance. An excellent example is Gilbert White's *Natural History and Antiquities of Selbourne* (1789).

Role of oratory in literature

Oratory, the art of persuasion, was long considered a great literary art. The oratory of the American Indian, for instance, is famous, while in classical Greece, Polymnia was the muse sacred to poetry and oratory. Rome's great orator Cicero was to have a decisive influence on the development of English prose style. Abraham Lincoln's Gettysburg Address is known to every American schoolchild. Today, however, oratory is more usually thought of as a craft than as an art. Most critics would not admit advertising copywriting, purely commercial fiction, or cinema and television scripts as accepted forms of literary expression, although others would hotly dispute their exclusion. The test in individual cases would seem to be one of enduring satisfaction and, of course, truth. Indeed, it becomes more and more difficult to categorize literature, for in modern civilization words are everywhere. Man is subject to a continuous flood of communication. Most of it is fugitive, but here and there—in high-level journalism, in television, in the cinema, in commercial fiction, in westerns and detective stories, and in plain, expository prose—some writing, almost by accident, achieves an aesthetic satisfaction, a depth and relevance that entitle it to stand with other examples of the art of literature.

LITERARY COMPOSITION

Critical theories. *Western.* If the early Egyptians or Sumerians had critical theories about the writing of literature, these have not survived. From the time of classical Greece until the present day, however, Western criticism has been dominated by two opposing theories of the literary art, which might conveniently be called the expressive and constructive theories of composition.

The Greek philosopher and scholar Aristotle (384–322 BC) is the first great representative of the constructive school of thought. His *Poetics* (the surviving fragment of which is limited to an analysis of tragedy and epic poetry) has sometimes been dismissed as a recipe book for the writing of potboilers. Certainly, Aristotle is primarily interested in the theoretical construction of tragedy, much as an architect might analyze the construction of a temple, but he is not exclusively objective and matter of fact. He does, however, regard the expressive elements in literature as of secondary importance, and the terms he uses to describe them have been open to interpretation and a matter of controversy ever since.

The 1st-century Greek treatise On *the Sublime* (conventionally attributed to the 3rd-century Longinus) deals with the question left unanswered by Aristotle—what makes great literature "great"? Its standards are almost entirely expressive. Where Aristotle is analytical and states general principles, the pseudo-Longinus is more specific and gives many quotations: even so, his critical theories are confined largely to impressionistic generalities.

Constructive and expressive theories of criticism

Thus, at the beginning of Western literary criticism, the controversy already exists. Is the artist or writer a technician, like a cook or an engineer, who designs and constructs a sort of machine that will elicit an aesthetic response from his audience? Or is he a virtuoso who above all else expresses himself and, because he gives voice to the deepest realities of his own personality, generates a response from his readers because they admit some profound identification with him? This antithesis endures throughout Western European history—Scholasticism versus Humanism, Classicism versus Romanticism, Cubism versus Expressionism—and survives to this day in the common judgment of our contemporary artists and writers. It is surprising how few critics have declared that the antithesis is unreal, that a work of literary or plastic art is at once constructive and expressive, and that it must in fact be both.

Eastern. Critical theories of literature in the Orient, however, have been more varied. There is an immense amount of highly technical, critical literature in India. Some works are recipe books, vast collections of tropes and stylistic devices; others are philosophical and general. In the best period of Indian literature, the cultural climax of Sanskrit (c. 320–490), it is assumed by writers that expressive and constructive factors are twin aspects of one reality. The same could be said of the Chinese, whose literary manuals and books on prosody and rhetoric are, as with the West, relegated to the class of technical handbooks, while their literary criticism is concerned rather with subjective, expressive factors—and so aligns itself with the pseudo-Longinus' "sublime." In Japan, technical, stylistic elements are certainly important (Japanese discrimination in these matters is perhaps the most refined in the world), but both writer and reader above all seek qualities of subtlety and poignancy and look for intimations of profundity often so evanescent as to escape entirely the uninitiated reader.

Broad and narrow conceptions of poetry. Far Eastern literary tradition has raised the question of the broad and narrow definitions of poetry (a question familiar in the West from Edgar Allan Poe's advocacy of the short poem in his "Poetic Principle" [1850]). There are no long epic poems in Chinese, no verse novels of the sort written in England by Robert Browning or Alfred Lord Tennyson in the 19th century. In Chinese drama, apart from a very few of the songs, the verse as such is considered doggerel. The versified treatises on astronomy, agriculture, or fishing, of the sort written in Greek and Roman times and during the 18th century in the West, are almost unknown in the Far East. Chinese poetry is almost exclusively lyric, meditative, and elegiac, and rarely does any poem exceed 100 lines—most are little longer than Western sonnets; many are only quatrains. In Japan this tendency to limit length was carried even further. The ballad survives in folk poetry, as it did in China, but the "long poem" of very moderate length disappeared early from literature. For the Japanese, the tanka is a "long poem": in its common form it has 31 syllables; the *sedōka* has 38; the *dodoitsu*, imitating folk song, has 26. From the 17th century and onward, the most popular poetic form was the haiku, which has only 17 syllables.

This development is relevant to the West because it spotlights the ever-increasing emphasis which has been laid on intensity of communication, a characteristic of Western poetry (and of literature generally) as it has evolved since the late 19th century. In the Far East all cultivated people were supposed to be able to write suitable occasional poetry, and so those qualities that distinguished a poem from the mass consequently came to be valued above all others. Similarly, as modern readers in the West struggle with a "communication avalanche" of words, they seek in literature those forms, ideas, values, vicarious experiences, and styles that transcend the verbiage to be had on every hand.

Literary language. In some literatures (notably classical Chinese, Old Norse, Old Irish), the language employed is quite different from that spoken or used in ordinary writing. This marks off the reading of literature as a special experience. In the Western tradition, it is only in comparatively modern times that literature has been written in the common speech of cultivated men. The Elizabethans did not talk like Shakespeare nor 18th-century people in the stately prose of Samuel Johnson or Edward Gibbon (the so-called Augustan plain style in literature became popular in the late 17th century and flourished throughout the 18th, but it was really a special form of rhetoric with antecedent models in Greek and Latin). The first person to write major works of literature in the ordinary English language of the educated man was Daniel Defoe (1660?–1731), and it is remarkable how little the language has changed since. *Robinson Crusor* (1719) is much more contemporary in tone than the elaborate prose of 19th-century writers like Thomas De Quincey or Walter Pater. (Defoe's language is not, in fact, so very simple: simplicity is itself one form of artifice.)

Ambiguity. Other writers have sought to use language for its most subtle and complex effects and have deliberately cultivated the ambiguity inherent in the multiple or shaded meanings of words. Between the two world wars, "ambiguity" became very fashionable in English and American poetry and the ferreting out of ambiguities—from even the simplest poem—was a favourite critical sport. T.S. Eliot in his literary essays is usually considered the founder of this movement. Actually, the platform of his critical attitudes is largely moral, but his two disciples, I.A. Richards in *Principles of Literary Criticism* (1924) and William Empson in *Seven Types of Ambiguity* (1930), carried his method to extreme lengths. The basic document of the movement is Charles Kay Ogden and I.A. Richards' *The Meaning of Meaning* (1923), a work of enormous importance in its time. Only a generation later, however, their ideas were somewhat at a discount.

Translation. Certainly, William Blake or Thomas Campion, when they were writing their simple lyrics, were unaware of the ambiguities and multiple meanings that future critics would find in them. Nevertheless, language is complex. Words do have overtones; they do stir up complicated reverberations in the mind that are ignored in their dictionary definitions. Great stylists, and most especially great poets, work with at least a half-conscious, or subliminal, awareness of the infinite potentialities of language. This is one reason why the essence of most poetry and great prose is so resistant to translation (quite apart from the radically different sound patterns that are created in other-language versions). The translator must project himself into the mind of the original author; he must transport himself into an entirely different world of relationships between sounds and meanings, and at the same time he must establish an equivalence between one infinitely complex system and another. Since no two languages are truly equivalent in anything except the simplest terms, this is a most difficult accomplishment. Certain writers are exceptionally difficult to translate. There are no satisfactory English versions, for example, of the Latin of Catullus, the French of Baudelaire, the Russian of Pushkin, or of the majority of Persian and Arabic poetry. The splendour of Sophocles' Greek, of Plato at his best, is barely suggested even in the finest English versions. On the other hand, the Germans insist that Shakespeare is better in German than he is in English, a humorous exaggeration perhaps. But again, Shakespeare is resistant to translation into French. His English seems to lack equivalents in that language.

The very greatest translations may become classics in their own right, of enduring literary excellence (the King James Version of the Bible, appearing in 1611, is the outstanding example), but on the whole the approximate equivalence of most translations to their originals seems to have a very short life. The original work remains the same, of lasting value to its own people, but the translation becomes out of date with each succeeding generation as the language and criteria of literary taste change. Nothing demonstrates the complexity of literary language more vividly. An analogous process takes place when a reader experiences a literary work in his own language; each generation gets a "new version" from its own classics.

Yet the values of great literature are more fundamental than complexity and subtleties of meaning arising from language alone. Works far removed from contemporary man in time and in cultural background, composed in a variety of languages utterly different from one another in structure, have nevertheless been translated successfully enough to be deeply moving. The 20th century has seen an immense mass of the oral literature of preliterate peoples and of the writings of all the great civilizations translated into modern languages. Understanding the growth of literature and its forms in other civilizations has greatly enriched the understanding of our own.

Craftsmanship. *Prosody.* Literature, like music, is an art of time, or "tempo": it takes time to read or listen to, and it usually presents events or the development of ideas or the succession of images or all these together in time. The craft of literature, indeed, can be said to be in part the manipulation of a structure in time, and so the simplest element of marking time, rhythm, is therefore of basic importance in both poetry and prose. Prosody, which is the science of versification, has for its subject the materials of poetry and is concerned almost entirely with the laws of metre, or rhythm in the narrowest sense. It deals with the patterning of sound in time; the number, length, accent, and pitch of syllables; and the modifications of rhythm by vowels and consonants. In most poetry, certain basic rhythms are repeated with modifications (that is to say, the poem rhymes or scans or both) but not in all. It most obviously does neither in the case of the "free forms" of modern poetry; but neither does it in the entire poetry of whole cultures. Since lyric poetry is either the actual text of song or else is immediately derived from song, it is regular in structure nearly everywhere in the world, although the elements of patterning that go into producing its rhythm may vary. The most important of these elements in English poetry, for example, have been accent, grouping of syllables (called feet), number of syllables in the line, and rhyme at the end of a line (and sometimes within it). Other elements such as pitch, resonance, repetition of vowels (assonance), repetition of consonants (alliteration), and breath pauses (cadence) have also been of great importance in distinguishing successful poetry from doggerel verse, but on the whole they are not as important as the former, and poets have not always been fully conscious of their use of them. Greek and Latin poetry was consciously patterned on the length of syllables (long or short) rather than on their accent; but all the considerations of "sound" (such as assonance and alliteration) entered into the aesthetically satisfactory structure of a poem. Similarly, both the French and Japanese were content simply to count the syllables in a line—but again, they also looked to all the "sound" elements.

Elements of prosody

The rhythms of prose are more complicated, though not necessarily more complex, than those of poetry. The rules of prose patterning are less fixed; patterns evolve and shift indefinitely and are seldom repeated except for special emphasis. So the analysis of prose rhythm is more difficult to make than, at least, the superficial analysis of poetry.

Structure. The craft of writing involves more than mere rules of prosody. The work's structure must be manipulated to attract the reader. First, the literary situation has to be established. The reader must be directly related to the work, placed in it—given enough information on who, what, when, or why—so that his attention is caught and held (or, on the other hand, he must be deliberately mystified, to the same end).

Aristotle gave a formula for dramatic structure that can be generalized to apply to most literature: presentation, development, complication, crisis, and resolution. Even lyric poems can possess plot in this sense, but by no means are all literary works so structured, nor does such structure ensure their merit—it can be safely said that westerns, detective stories, and cheap melodramas are more likely to follow strictly the rules of Aristotle's *Poetics* than are great novels. Nevertheless, the scheme does provide a norm from which there is infinite variation. Neoclassical dramatists and critics, especially in 17th-century France, derived from Aristotle what they called the unities of time, action, and place. This meant that the action of a play should not spread beyond the events of one day and, best of all, should be confined within the actual time of performance. Nor should the action move about too much from place to place—best only to go from indoors to outdoors and back. There should be only one plot line, which might be relieved by a subplot, usually comic. These three unities—of time, place, and action—do not occur in Aristotle and are certainly not observed in Classical Greek tragedy. They are an invention of Renaissance critics, some of whom went even further, insisting also on what might be called a unity of mood. To this day there are those who, working on this principle, object to Shakespeare's use of comic relief within the tragic action of his plays—to the porter in *Macbeth,* for instance, or the gravediggers in *Hamlet.*

Unities of
time, place,
and action

Assiduous critics have found elaborate architectural structures in quite diffuse works—including Miguel de Cervantes' *Don Quixote* (1605–15), Sterne's *Tristram Shandy* (1759–67), Casanova's *Icosameron* (1788; 1928). But their "discoveries" are too often put there after the event. Great early novels such as the Chinese *Dream of the Red Chamber* (1754; first published in English 1929) and the Japanese *Tale of Genji* (early 11th century) usually develop organically rather than according to geometrical formulas, one incident or image spinning off another. Probably the most tightly structured work, in the Neoclassicists' sense, is the Icelandic *Njál's saga.*

The 19th century was the golden age of the novel, and most of the more famous examples of the form were systematically plotted, even where the plot structure simply traced the growth in personality of an individual hero or heroine. This kind of novel, of which in their very diverse ways Stendhal's *The Red and the Black* (1830) and Dickens' *David Copperfield* (1850) are great examples, is known as *Bildungsroman.* Gustave Flaubert's *Madame Bovary* (1857) is as rigorously classicist in form as the 17th-century plays of Racine and Corneille, which were the high point of the French classical theatre, although Flaubert obeys laws more complex than those of the Aristotelians. Novels such as Tolstoy's *War and Peace* (1865–69), Dostoyevsky's *Brothers Karamazov* (1880), and the works of Balzac owe much of their power to their ability to overwhelm the reader with a massive sense of reality. The latter 19th and early 20th centuries witnessed an attack on old forms, but what the new writers evolved was simply a new architecture. A novel like James Joyce's *Ulysses* (1922), which takes place in a day and an evening, is one of the most highly structured ever written. Novelists such as Joseph Conrad, Ford Madox Ford, Virginia Woolf, and, to a lesser extent, Henry James developed a multiple-aspect narrative, sometimes by using time shifts and flashbacks and by writing from different points of view, sometimes by using the device (dating back to Classical Greek romances) of having one or more narrators as characters within the story. (This technique, which was first perfected in the verse novels of Robert Browning, in fact reached its most extreme development in the English language in poetry: in Ezra Pound's *Cantos,* T.S. Eliot's *The Waste Land,* William Carlos Williams' *Paterson,* and the many long poems influenced by them.)

Narrative
devices in
the novel

CONTENT OF LITERATURE

The word as symbol. The content of literature is as limitless as the desire of human beings to communicate with one another. The thousands of years, perhaps hundreds of thousands, since the human species first developed speech have seen built up the almost infinite systems of relationships called languages. A language is not just a collection of words in an unabridged dictionary but the individual and social possession of living men, an inexhaustible system of equivalents, of sounds to objects and to one another. Its most primitive elements are those words that express direct experiences of objective reality, and its most sophisticated are concepts on a high level of abstraction. Words are not only equivalent to things, they have varying degrees of equivalence to one another. A symbol, says the dictionary, is something that stands for something else or a sign used to represent something, "as the lion is the symbol of courage, the cross the symbol of Christianity." In this sense all words can be called symbols, but the examples given—the lion and the cross—are really metaphors: that is, symbols that represent a complex of other symbols, and which are generally negotiable in a given society (just as money is a symbol for goods or labour). Eventually a language comes to be, among other things, a huge sea of implicit metaphors, an endless web of interrelated symbols. As literature, especially poetry, grows more and more sophisticated, it begins to manipulate this field of suspended metaphors as a material in itself, often as an end in itself. Thus, there emerge forms of poetry (and prose, too) with endless ramifications of reference, as in Japanese waka and haiku, some ancient Irish and Norse verse, and much of the poetry written in western Europe since the time of Baudelaire that is called modernist. It might be supposed that, at its most extreme, this development would be objective, constructive—aligning it with the critical theories stemming from Aristotle's *Poetics.* On the contrary, it is romantic, subjective art, primarily because the writer handles such material instinctively and subjectively, approaches it as the "collective unconscious," to use the term of the psychologist Carl Jung, rather than with deliberate rationality.

Themes and their sources. By the time literature appears in the development of a culture, the society has already come to share a whole system of stereotypes and archetypes: major symbols standing for the fundamental realities of the human condition, including the kind of symbolic realities that are enshrined in religion and myth. Literature may use such symbols directly, but all great works of literary art are, as it were, original and unique myths. The world's great classics evoke and organize the archetypes of universal human experience. This does not mean, however, that all literature is an endless repetition of a few myths and motives, endlessly retelling the first stories of civilized man, repeating the Sumerian *Epic of Gilgamesh* or Sophocles' *Oedipus the King.* The subject matter of literature is as wide as human experience itself. Myths, legends, and folktales lie at the beginning of literature. and their plots, situations, and allegorical (metaphorical narrative) judgments of life represent a constant source of literary inspiration that never fails. This is so because mankind is constant—men share a common physiology. Even social structures, after the development of cities, remain much alike. Whole civilizations have a life pattern that repeats itself through history. Jung's term "collective unconscious" really means that mankind is one species, with a common fund of general experience. Egyptian scribes, Soviet bureaucrats, and junior executives in New York City live and respond to life in the same ways; the lives of farmers or miners or hunters vary only within narrow limits. Love is love and death is death, for a South African Bushman and a French Surrealist alike. So the themes of literature have at once an infinite variety and an abiding constancy. They can be taken from myth, from history, or from contemporary occurrence, or they can be pure invention (but even if they are invented, they are nonetheless constructed from the constant materials of real experience, no matter how fantastic the invention).

Use of
symbols
and myths
in
literature

The writer's personal involvement. As time goes on, literature tends to concern itself more and more with the interior meanings of its narrative, with problems of human personality and human relationships. Many novels are fictional, psychological biographies which tell of the slowly achieved integration of the hero's personality or of his disintegration, of the conflict between self-realization and the flow of events and the demands of other people. This can be presented explicitly, where the characters talk about what is going on in their heads; either ambiguously and with reserve, as in the novels of Henry James, or

Presentation of
characters

overtly, as in those of Dostoyevsky. Alternatively, it can be presented by a careful arrangement of objective facts, where psychological development is described purely in terms of behaviour, and where the reader's subjective response is elicited by the minute descriptions of physical reality, as in the novels of Stendhal and the greatest Chinese novels like the *Dream of the Red Chamber*, which convince the reader that through the novel he is seeing reality itself, rather than an artfully contrived semblance of reality.

Literature, however, is not solely concerned with the concrete, with objective reality, with individual psychology, or with subjective emotion. Some deal with abstract ideas or philosophical conceptions. Much purely abstract writing is considered literature only in the widest sense of the term, and the philosophical works that are ranked as great literature are usually presented with more or less of a sensuous garment. Thus, Plato's *Dialogues* rank as great literature because the philosophical material is presented in dramatic form, as the dialectical outcome of the interchange of ideas between clearly drawn, vital personalities, and because the descriptive passages are of great lyric beauty. Karl Marx's *Das Kapital* (1867–95) approaches great literature in certain passages in which he expresses the social passion he shares with the Hebrew prophets of the Old Testament. Euclid's *Elements* and St. Thomas Aquinas' *Summa theologica* give literary, aesthetic satisfaction to some people because of their purity of style and beauty of architectonic construction. In short, most philosophical works that rank as great literature do so because they are intensely human. The reader responds to Pascal's *Pensées*, to Montaigne's *Essays,* and to Marcus Aurelius' *Meditations* as he would to living men. Sometimes the pretense of purely abstract intellectual rigour is in fact a literary device. The writings of the 20th-century philosopher Ludwig Wittgenstein, for example, owed much of their impact to this approach, while the poetry of Paul Valtry borrows the language of philosophy and science for its rhetorical and evocative power.

Relation of form to content. Throughout literary history, many great critics have pointed out that it is artificial to make a distinction between form and content, except for purposes of analytical discussion. Form determines content. Content determines form. The issue is, indeed, usually only raised at all by those critics who are more interested in politics, religion, or ideology than in literature; thus, they object to writers who they feel sacrifice ideological orthodoxy for formal perfection, message for style.

Style. But style cannot really be said to exist on paper at all; it is the way the mind of the author expresses itself in words. Since words represent ideas, there cannot be abstract literature unless a collection of nonsense syllables can be admitted as literature. Even the most avant-garde writers associated with the Cubist or Non-Objective painters used language, and language is meaning, though the meaning may be incomprehensible. Oscar Wilde and Walter Pater, the great 19th-century exponents of "art for art's sake," were in fact tireless propagandists for their views, which dominate their most flowery prose. It is true that great style depends on the perfect matching of content and form, so that the literary expression perfectly reflects the writer's intention; "poor style" reveals the inability of a writer to match the two—in other words, reveals his inability to express himself. This is why we say that "style expresses the man." The veiled style of Henry James, with its subtleties, equivocations, and qualifications, perfectly reflects his complicated and subtle mind and his abiding awareness of ambiguity in human motives. At the other extreme, the style of the early-20th-century U.S. novelist Theodore Dreiser—bumbling, clumsy, dogged, troubled—perfectly embodies his own attitudes toward life and is, in fact, his constant judgment of his subject matter. Sometimes an author, under the impression that he is simply polishing his style, may completely alter his content. As Flaubert worked over the drafts of *Madame Bovary,* seeking always the apposite word that would precisely convey his meaning, he lifted his novel from a level of sentimental romance to make it one of the great ironic tragedies of literature. Yet, to judge from his correspondence, he seems never to have been completely aware of what he had done, of the severity of his own irony.

Coordination of content and form in good style

Literature may be an art, but writing is a craft, and a craft must be learned. Talent, special ability in the arts, may appear at an early age; the special personality called genius may indeed be born, not made. But skill in matching intention and expression comes with practice. Naive writers, "naturals" like the 17th-century English diarist Samuel Pepys, the late 18th-century French naïf Restif de la Bretonne, the 20th-century American novelist Henry Miller, are all deservedly called stylists, although their styles are far removed from the deliberate, painstaking practice of a Flaubert or a Turgenev. They wrote spontaneously whatever came into their heads; but they wrote constantly, voluminously, and were, by their own standards, skilled practitioners.

Objective–subjective expression. There are certain forms of literature that do not permit such highly personal behaviour—for instance, formal lyric poetry and classic drama. In these cases the word "form" is used to mean a predetermined structure within whose mold the content must be fitted. These structures are, however, quite simple and so cannot be said to determine the content. Racine and Corneille were contemporaries; both were Neoclassic French dramatists; both abided by all the artificial rules—usually observing the "unities" and following the same strict rules of prosody. Yet their plays, and the poetry in which they are written, differ completely. Corneille is intellectually and emotionally a Neoclassicist—clear and hard, a true objectivist, sure of both his verse and the motivations of his characters. Racine was a great romantic long before the age of Romanticism. His characters are confused and tortured; his verse throbs like the heartbeats of his desperate heroines. He is a great sentimentalist in the best and deepest meaning of that word. His later influence on poets like Baudelaire and Paul Valtry is due to his mastery of sentimental expression, not, as they supposed, to his mastery of Neoclassic form.

Predetermined structure of classic drama

Verse on any subject matter can of course be written purely according to formula. The 18th century in England saw all sorts of prose treatises cast in rhyme and metre, but this was simply applied patterning. (Works such as *The Botanic Garden* [2 vol., 1794–95] by Erasmus Darwin should be sharply distinguished from James Thomson's *The Seasons* [1726–30], which is true poetry, not versified natural history—just as Virgil's *Georgics* is not an agricultural handbook.) Neoclassicism, especially in its 18th-century developments, confused—for ordinary minds, at any rate—formula with form and so led to the revolt called Romanticism. The leading theorists of that revolt, the poets William Wordsworth and Samuel Taylor Coleridge, in the "Preface" (1800) to *Lyrical Ballads* urged the observance of a few simple rules basic to all great poetry and demanded a return to the integrity of expressive form. A similar revolution in taste was taking place all over Europe and also in China (where the narrow pursuit of formula had almost destroyed poetry). The Romantic taste could enjoy the "formlessness" of William Blake's prophetic books, or Walt Whitman's *Leaves of Grass,* or the loose imagination of Shelley—but careful study reveals that these writers were not formless at all. Each had his own personal form.

Form and formlessness in modern literature

Time passes and the pendulum of taste swings. In the mid-20th century, Paul Valéry, T.S. Eliot, and Yvor Winters would attack what the latter called "the fallacy of expressive form," but this is itself a fallacy. All form in literature is expressive. All expression has its own form, even when the form is a deliberate quest of formlessness. (The automatic writing cultivated by the surrealists, for instance, suffers from the excessive formalism of the unconscious mind and is far more stereotyped than the poetry of the Neoclassicist Alexander Pope.) Form simply refers to organization, and critics who attack form do not seem always to remember that a writer organizes more than words. He organizes experience. Thus, his organization stretches far back in his mental process. Form is the other face of content, the outward, visible sign of inner spiritual reality.

LITERATURE AND ITS AUDIENCE

Folk and elite literatures. In preliterate societies oral literature was widely shared; it saturated the society and was as much a part of living as food, clothing, shelter, or religion. In barbaric societies, the minstrel might be a courtier of the king or chieftain, and the poet who composed liturgies might be a priest. But the oral performance itself was accessible to the whole community. As society evolved its various social layers, or classes, an "elite" literature began to be distinguishable from the "folk" literature of the people. With the invention of writing this separation was accelerated until finally literature was being experienced individually by the elite (reading a book), while folklore and folk song were experienced orally and more or less collectively by the illiterate common people.

Elite literature continuously refreshes itself with materials drawn from the popular. Almost all poetic revivals, for instance, include in their programs a new appreciation of folk song, together with a demand for greater objectivity. On the other hand folk literature borrows themes and, very rarely, patterns from elite literature. Many of the English and Scottish ballads that date from the end of the Middle Ages and have been preserved by oral tradition share plots and even turns of phrase with written literature. A very large percentage of these ballads contain elements that are common to folk ballads from all over western Europe; central themes of folklore, indeed, are found all over the world. Whether these common elements are the result of diffusion is a matter for dispute. They do, however, represent great psychological constants, archetypes of experience common to the human species, and so these constants are used again and again by elite literature as it discovers them in folklore.

Modern popular literature. There is a marked difference between true popular literature, that of folklore and folk song, and the popular literature of modern times. Popular literature today is produced either to be read by a literate audience or to be enacted on television or in the cinema; it is produced by writers who are members, however lowly, of an elite corps of professional literates. Thus, popular literature no longer springs from the people; it is handed to them. Their role is passive. At the best they are permitted a limited selectivity as consumers.

Certain theorists once believed that folk songs and even long, narrative ballads were produced collectively, as has been said in mockery "by the tribe sitting around the fire and grunting in unison." This idea is very much out of date. Folk songs and folk tales began somewhere in one human mind. They were developed and shaped into the forms in which they are now found by hundreds of other minds as they were passed down through the centuries. Only in this sense were they "collectively" produced. During the 20th century, folklore and folk speech have had a great influence on elite literature-on writers as different as Franz Kafka and Carl Sandburg, Selma Lagerlöf and Kawabata Yasunari, Martin Buber and Isaac Bashevis Singer. Folk song has always been popular with bohemian intellectuals, especially political radicals (who certainly are an elite). Since World War II the influence of folk song upon popular song has not just been great; it has been determinative. Almost all "hit" songs since the mid-century have been imitation folk songs; and some authentic folk singers attract immense audiences.

Influence of folklore on modem literature

Popular fiction and drama, westerns and detective stories, films and television serials, all deal with the same great archetypal themes as folktales and ballads, though this is seldom due to direct influence; these are simply the limits within which the human mind works. The number of people who have elevated the formulas of popular fiction to a higher literary level is surprisingly small. Examples are H.G. Wells's early science fiction, the western stories of Gordon Young and Ernest Haycox, the detective stories of Sir Arthur Conan Doyle, Georges Simenon, and Raymond Chandler.

The latter half of the 20th century has seen an even greater change in popular literature. Writing is a static medium: that is to say, a book is read by one person at a time; it permits recollection and anticipation; the reader can go back to check a point or move ahead to find out how the story ends. In radio, television, and the cinema the medium is fluent; the audience is a collectivity and is at the mercy of time. It cannot pause to reflect or to understand more fully without missing another part of the action, nor can it go back or forward. Marshall McLuhan in his book ***Understanding Media*** (1964) became famous for erecting a whole structure of aesthetic, sociological, and philosophical theory upon this fact. But it remains to be seen whether the new, fluent materials of communication are going to make so very many changes in civilization, let alone in the human mind—mankind has, after all, been influenced for thousands of years by the popular, fluent arts of music and drama. Even the most transitory television serial was written down before it was performed, and the script can be consulted in the files. Before the invention of writing, all literature was fluent because it was contained in people's memory. In a sense it was more fluent than music, because it was harder to remember. Man in mass society becomes increasingly a creature of the moment, but the reasons for this are undoubtedly more fundamental than his forms of entertainment.

LITERATURE AND ITS ENVIRONMENT

Social and economic conditions. Literature, like all other human activities, necessarily reflects current social and economic conditions. Class stratification was reflected in literature as soon as it had appeared in life. Among the American Indians, for instance, the chants of the shaman, or medicine man, differ from the secret, personal songs of the individual, and these likewise differ from the group songs of ritual or entertainment sung in community. In the Heroic Age, the epic tales of kings and chiefs that were sung or told in their barbaric courts differed from the folktales that were told in peasant cottages.

Reflection of class distinction in literature

The more cohesive a society, the more the elements—and even attitudes–evolved in the different class strata are interchangeable at all levels. In the tight clan organization that existed in late medieval times at the Scottish border, for example, heroic ballads telling of the deeds of lords and ladies were preserved in the songs of the common people. But where class divisions are unbridgeable, elite literature is liable to be totally separated from popular culture. An extreme example is the classic literature of the Roman Empire. Its forms and its sources were largely Greek—it even adopted its laws of verse patterning from Greek models, even though these were antagonistic to the natural patterns of the Latin language—and most of the sophisticated works of the major Latin authors were completely closed to the overwhelming majority of people of the Roman Empire.

Printing has made all the difference in the negotiability of ideas. The writings of the 18th-century French writers Voltaire, Rousseau, and Diderot were produced from and for almost as narrow a caste as the Roman elite, but they were printed. Within a generation they had penetrated the entire society and were of vital importance in revolutionizing it.

Class distinctions in the literature of modem times exist more in the works themselves than in their audience. Although Henry James wrote about the upper classes and Émile Zola about workingmen, both were, in fact, members of an elite and were read by members of an elite—moreover, in their day, those who read Zola certainly considered themselves more of an elite than did the readers of Henry James. The ordinary people, if they read at all, preferred sentimental romances and "penny dreadfuls." Popular literature had already become commercially produced entertainment literature, a type which today is also provided by television scripts.

The elite who read serious literature are not necessarily members of a social or economic upper class. It has been said of the most ethereal French poet, Stéphane Mallarmé, that in every French small town there was a youth who carried his poems in his heart. These poems are perhaps the most "elite" product of western European civilization, but the "youths" referred to were hardly the sons of dukes or millionaires. (It is a curious phenomenon that, since the middle of the 18th century in Europe and in the United States, the majority of readers of serious literature—as well as of entertainment literature—have been women. The extent of the influence that this audience has exerted on literature itself must be immense.)

Hippolyte Taine's ecological theory of literature

National and group literature. Hippolyte Taine, the 19th-century French critic, evolved an ecological theory of literature. He looked first and foremost to the national characteristics of western European literatures, and he found the source of these characteristics in the climate and soil of each respective nation. His *History of English Literature* (5 vol., 1863–69) is an extensive elaboration of these ideas. It is doubtful that anyone today would agree with the simplistic terms in which Taine states his thesis. It is obvious that Russian literature differs from English or French from German. English books are written by Englishmen, their scenes are commonly laid in England, they are usually about Englishmen and they are designed to be read by Englishmen—at least in the first instance. But modern civilization becomes more and more a world civilization, wherein works of all peoples flow into a general fund of literature. It is not unusual to read a novel by a Japanese author one week and one by a black writer from West Africa the next. Writers are themselves affected by this cross-fertilization. Certainly, the work of the great 19th-century Russian novelists has had more influence on 20th-century American writers than has the work of their own literary ancestors. Poetry does not circulate so readily, because catching its true significance in translation is so very difficult to accomplish. Nevertheless, for the past 100 years or so, the influence of French poetry upon all the literatures of the civilized world has not just been important, it has been pre-eminent. The tendentious elements of literature—propaganda for race, nation, or religion—have been more and more eroded in this process of wholesale cultural exchange.

Popular literature on the other hand is habitually tendentious both deliberately and unconsciously. It reflects and stimulates the prejudices and parochialism of its audience. Most of the literary conflicts that have seized the totalitarian countries during the 20th century stem directly from relentless efforts by the state to reduce elite literature to the level of the popular. The great proletarian novels of our time have been produced, not by Russians, but by American Negroes, Japanese, Germans, and—most proletarian of all—a German-American living in Mexico, B. Traven. Government control and censorship can inhibit literary development, perhaps deform it a little, and can destroy authors outright; but, whether in the France of Louis XIV or in the Soviet Union of the 20th century, it cannot be said to have a fundamental effect upon the course of literature.

Effect of censorship on literary development

The writer's position in society. A distinguishing characteristic of modern literature is the peculiar elite which it has itself evolved. In earlier cultures the artist, though he may have been neurotic at times, thought of himself as part of his society and shared its values and attitudes. Usually the clerkly caste played a personal, important role in society. In the modern industrial civilization, however, "scribes" became simply a category of skilled hired hands. The writer shared few of the values of the merchant or the entrepreneur or manager. And so the literary and artistic world came to have a subculture of its own. The antagonism between the two resultant sets of values is the source of what we call alienation—among the intellectuals at least (the alienation of the common man in urban, industrial civilization from his work, from himself, and from his fellows is another matter, although its results are reflected and intensified in the alienation of the elite). For about 200 years now, the artistic environment of the writer has not usually been shared with the general populace. The subculture known as bohemia and the literary and artistic movements generated in its little special society have often been more important—at least in the minds of many writers—than the historical, social, and economic movements of the culture as a whole. Even massive historical change is translated into these terms—the Russian Revolution, for instance, into Communist-Futurism, Constructivism, Socialist Realism. Western European literature could be viewed as a parade of movements—Romanticism, Realism, Naturalism, Futurism, Structuralism, and so on indefinitely. Some of the more journalistic critics, indeed, have delighted to regard it in such a way. But after the manifestos have been swept away, the meetings adjourned, the literary cafés of the moment lost their popularity, the turmoil is seen not to have made so very much difference. The Romantic Théophile Gautier (1811–72) and the Naturalist Emile Zola (1840–1902) have more in common than they have differences, and their differences are rather because of changes in society as a whole than because of conflicting literary principles.

Literary subcultures and movements

At first, changes in literary values are appreciated only at the upper levels of the literary elite itself, but often, within a generation, works once thought esoteric are being taught as part of a school syllabus. Most cultivated people once thought James Joyce's *Ulysses* incomprehensible or, where it was not, obscene. Today his methods and subject matter are commonplace in the commercial fiction of the mass culture. A few writers remain confined to the elite. Mallarmé is a good example—but he would have been just as ethereal had he written in the simplest French of direct communication. His subtleties are ultimately grounded in his personality.

Literature and the other arts. Literature has an obvious kinship with the other arts. Presented, a play is drama; read, a play is literature. Most important films have been based upon written literature, usually novels, although all the great epics and most of the great plays have been filmed at some time and thus have stimulated the younger medium's growth. Conversely, the techniques required in writing for film have influenced many writers in structuring their novels and have affected their style. Most popular fiction is written with "movie rights" in mind, and these are certainly a consideration with most modern publishers. Literature provides the libretto for operas, the theme for tone poems—even so anomalous a form as Nietzsche's *Thus Spake Zarathustra* was interpreted in music by Richard Strauss—and of course it provides the lyrics of songs. Many ballets and modern dances are based on stories or poems. Sometimes, music and dance are accompanied by a text read by a speaker or chanted by a chorus. The mid-19th century was the heyday of literary, historical, and anecdotal painting, though, aside from the Surrealists, this sort of thing died out in the 20th century. Cross-fertilization of literature and the arts now takes place more subtly, mostly in the use of parallel techniques—the rational dissociation of the Cubists or the spontaneous action painting of the Abstract Expressionists, for example, which flourished at the same time as the free-flowing uncorrected narratives of some novelists in the 1950s and '60s.

LITERARY GENRES

Critics have invented a variety of systems for treating literature as a collection of genres. Often these genres are artificial, invented after the fact with the aim of making literature less sprawling, more tidy. Theories of literature must be based upon direct experience of the living texts and so be flexible enough to contain their individuality and variety. Perhaps the best approach is historical, or genetic. What actually happened, and in what way did literature evolve up to the present day?

Oral literature

There is a surprising variety of oral literature among surviving preliterate peoples, and, as the written word emerges in history, the indications are that the important literary genres all existed at the beginning of civilized societies: heroic epic; songs in praise of priests and kings; stories of mystery and the supernatural; love lyrics; personal songs (the result of intense meditation); love stories; tales of adventure and heroism (of common peoples, as distinct from the heroic epics of the upper classes); satire (which was dreaded by barbaric chieftains); satirical combats (in which two poets or two personifications abused one another and praised themselves); ballads and folktales of tragedy and murder; folk stories, such as the tale of the clever boy who performs impossible tasks, outwits all his adversaries, and usually wins the hand of the king's daughter; animal fables like those attributed to Aesop (the special delight of Black Africa and Indian America); riddles, proverbs, and philosophical observations; hymns, incantations, and mysterious songs of priests; and finally actual mythology—stories of the origin of the world and the human race, of the great dead, and of the gods and demigods.

Epic. The true heroic epic never evolved far from its preliterate origins, and it arose only in the Heroic Age which preceded a settled civilization. The conditions reflected in, say, the *Iliad* and *Odyssey* are much the same as those of the Anglo-Saxon *Beowulf,* the German *Nibelungenlied,* or the Irish stories of Cdchulainn. The literary epic is another matter altogether. Virgil's *Aeneid,* for instance, or Milton's *Paradise Lost* are products of highly sophisticated literary cultures. Many long poems sometimes classified as epic literature are no such thing—Dante's *La divina Commedia,* for example, is a long theological, philosophical, political, moral, and mystical poem. Dante considered it to be a kind of drama which obeyed the rules of Aristotle's *Poetics.* Goethe's *Faust* is in dramatic form and is sometimes even staged—but it is really a philosophical poetic novel. Modern critics have described long poems such as T.S. Eliot's *Waste Land* and Ezra Pound's *Cantos* as "philosophical epics." There is nothing epic about them; they are reveries, more or less philosophical.

Lyric poetry. Lyric poetry never gets far from its origins, except that some of its finest examples—Medieval Latin, Provençal, Middle High German, Middle French, Renaissance—which today are only read, were actually written to be sung. In the 20th century, however, popular songs of great literary merit have become increasingly common—for example, the songs of Bertolt Brecht and Kurt Weill in German, of Georges Brassens and Anne Sylvestre in French, and of Leonard Cohen, Bob Dylan, and Joni Mitchell. It is interesting to note that, in periods when the culture values artificiality, the lyric becomes stereotyped. Then, after a while, the poets revolt and, usually turning to folk origins, restore to lyric poetry at least the appearance of naturalness and spontaneity.

Diverse forms of satire

Satire. The forms of satire are as manifold as those of literature itself—from those of the mock epic to the biting epigram. A great many social and political novels of today would have been regarded as satire by the ancients. Many of the great works of all time are satires, but in each case they have risen far above their immediate satirical objectives. The 16th-century medieval satire on civilization, the *Gargantua and Pantagruel* of Rabelais, grew under the hand of its author into a great archetypal myth of the lust for life. Cervantes' *Don Quixote,* often called the greatest work of prose fiction in the West, is superficially a satire of the sentimental romance of knightly adventure. But, again, it is an archetypal myth, telling the adventures of the soul of man—of the individual—in the long struggle with what is called the human condition. *The Tale of Genji* by Murasaki Shikibu has sometimes been considered by obtuse critics as no more than a satire on the sexual promiscuity of the Heian court. In fact, it is a profoundly philosophical, religious, and mystical novel.

The earliest prose fiction

Prose fiction. Extended prose fiction is the latest of the literary forms to develop. We have romances from classical Greek times that are as long as short novels; but they are really tales of adventure—vastly extended anecdotes. The first prose fiction of any psychological depth is the *Satyricon,* almost certainly attributed to Petronius Arbiter (died AD 65/66). Though it survives only in fragments, supposedly one-eleventh of the whole, even these would indicate that it is one of the greatest picaresque novels, composed of loosely connected episodes of robust and often erotic adventure. The other great surviving fiction of classical times is the *Metamorphoses* (known as *The Golden Ass)* by Apuleius (2nd century AD). In addition to being a picaresque adventure story, it is a criticism of Roman society, a celebration of the religion of Isis, and an allegory of the progress of the soul. It contains the justly celebrated story of Cupid and Psyche, a myth retold with psychological subtlety. Style has much to do with the value and hence the survival of these two works. They are written in prose of extraordinary beauty, although it is by no means of "classical" purity. The prose romances of the Middle Ages are closely related to earlier heroic literature. Some, like Sir Thomas Malory's 15th-century *Le Morte Darthur,* are retellings of heroic legend in terms of the romantic chivalry of the early Renaissance, a combination of barbaric, medieval, and Renaissance sensibility which, in the tales of Tristram and Isolt and Launcelot and Guinevere, produced something not unlike modem novels of tragic love.

The Western novel is a product of modem civilization, although in the Far East novels began a separate development as early as the 10th century. Extended prose works of complex interpersonal relations and motivations begin in 17th-century France with *The Princess of Cleves* (1678) by Madame de La Fayette. Eighteenth-century France produced an immense number of novels dealing with love analysis but none to compare with Madame de La Fayette's until Pierre Choderlos de Laclos wrote *Les Liaisons dangereuses* (1782). This was, in form, an exchange of letters between two corrupters of youth; but, in intent, it was a savage satire of the *ancien régime* and a heart-rending psychological study. The English novel of the 18th century was less subtle, more robust—vulgar in the best sense—and is exemplified by Henry Fielding's *Torn Jones* (1749) and Laurence Sterne's *Tristram Shandy.* The 19th century was the golden age of the novel. It became ever more profound, complex, and subtle (or, on the other hand, more popular, eventful, and sentimental). By the beginning of the 20th century it had become the most common form of thoughtful reading matter and had replaced, for most educated people, religious, philosophical, and scientific works as a medium for the interpretation of life. By the late 1920s the novel had begun to show signs of decay as a form, and no works have since been produced to compare with the recent past. This may prove to be a temporarily barren period, or else the novel may be losing its energy as a narrative art form and in this sense giving way to the medium of film.

Current state of the novel

Drama. Like lyric poetry, drama has been an exceptionally stable literary form. Given a little leeway, most plays written by the beginning of the 20th century could be adjusted to the rules of Aristotle's *Poetics.* Before World War I, however, all traditional art forms, led by painting, began to disintegrate, and new forms evolved to take their place. In drama the most radical innovator was August Strindberg (1849–1912), and from that day to this, drama (forced to compete with the cinema) has become ever more experimental, constantly striving for new methods, materials, and, especially, ways to establish a close relationship with the audience. All this activity has profoundly modified drama as literature.

Future developments. In the 20th century the methods of poetry have also changed drastically, although the "innovator" here might be said to have been Baudelaire (1821–67). The disassociation and recombination of ideas of the Cubists, the free association of ideas of the Surrealists, dreams, trance states, the poetry of preliterate people—all have been absorbed into the practice of modern poetry. This proliferation of form is not likely to end. Effort that once was applied to perfecting a single pattern in a single form may in the future be more and more directed toward the elaboration of entirely new "multimedia" forms, employing the resources of all the established arts. At the same time, writers may prefer to simplify and polish the forms of the past with a rigorous, Neoclassicist discipline. In a worldwide urban civilization, which has taken to itself the styles and discoveries of all cultures past and present, the future of literature is quite impossible to determine.

WRITINGS ON LITERATURE

Scholarly research. Research by scholars into the literary past began almost as soon as literature itself—as soon

Ancient literary scholarship

as the documents accumulated—and for many centuries it represents almost all the scholarship that has survived. The most extensive text of the Sumerian *Epic of Gilgamesh,* the first of the world's great classics, is a late Assyrian synthesis that must have required an immense amount of research into clay tablets, written in several languages going back to the beginning of Mesopotamian civilization. Many Egyptian poems and the philosophic creation myth known as the "Memphite Theology" survive in very late texts that carefully reproduce the original language of the first dynasties. Once the function of the scribe was established as essential, he invented literary scholarship, both to secure his position and to occupy his leisure. The great epoch of literary scholarship in ancient times centred on the library (and university) of Alexandria from its foundation in 324 BC to its destruction by the Arabs in AD 640. Hellenistic Greek scholars there developed such an academic and pedantic approach to literary scholarship and scholarly literature that the term Alexandrine remains pejorative to this day. To them, however, is owed the survival of the texts of most of the Greek classics. Roman literary scholarship was rhetorical rather than analytic. With the coming of Islām, there was established across the whole warm temperate zone of the Old World a far-flung community of scholars who were at home in learned circles from India to Spain. Judaism, like Islām, was a religion of the book and of written tradition, so literary scholarship played a central role in each. The same is true of India, China, and later Japan; for sheer bulk, as well as for subtlety and insight, Oriental scholarship has never been surpassed. In a sense, the Renaissance in Europe was a cultural revolution led by literary scholars who discovered, revived, and made relevant again the literary heritage of Greece and Rome. In the 19th century, literary scholarship was dominated by the exhaustive, painstaking German academician, and that Germanic tradition passed to the universities of the United States. The demand that every teacher should write a master's thesis, a doctor's dissertation, and, for the rest of his career, publish with reasonable frequency learned articles and scholarly books, has led to a mass of scholarship of widely varying standards and value. Some is trivial and absurd, but the best has perfected the texts and thoroughly illuminated the significance of nearly all the world's great literature.

Literary criticism. Literary criticism, as distinguished from scholarly research, is usually itself considered a form of literature. Some people find great critics as entertaining and stimulating as great poets, and theoretical treatises of literary aesthetics can be as exciting as novels. Aristotle, Longinus, and the Roman rhetorician and critic Quintilian are still read, although Renaissance critics like the once all-powerful Josephus Scaliger are forgotten by all but specialized scholars. Later critics, such as Poe, Sainte-Beuve, Taine, Vissarion Belinsky, Matthew Arnold, Walter Bagehot, Walter Pater, and George Saintsbury, are probably read more for themselves than for their literary judgments and for their general theorizing rather than for their applications (in the case of the first three, for instance, time has confounded almost all the evaluations they made of their contemporaries). The English critics have survived because they largely confined themselves to acknowledged masterpieces and general ideas. Perhaps literary criticism can really be read as a form of autobiography. Aestheticians of literature like I.A. Richards, Sir C.M. Bowra, Paul Valéry, Suzanne Langer, and Ernst Cassirer have had an influence beyond the narrow confines of literary scholarship and have played in our time something approaching the role of general philosophers. This has been true on the popular level as well. The Dane Georg Brandes, the Americans James Gibbons Huneker, H.L. Mencken, and Edmund Wilson—these men have been social forces in their day. Literary criticism can play its role in social change. In Japan, the overthrow of the shogunate, the restoration of the emperor, and the profound change in the Japanese social sensibility begins with the literary criticism of Moto-ori Norinaga (1730–1801). The 19th-century revolution in theology resulted from the convergence of Darwinian theories of evolution and the technical and historical criticism of the Bible that scholars had undertaken. For many modern intellectuals, the literary quarterlies and weeklies, with their tireless discussions of the spiritual significance and formal characteristics of everything from the greatest masterpiece to the most ephemeral current production, can be said to have filled the place of religion, both as rite and dogma.

BIBLIOGRAPHY. The best introduction to literature is literature itself, a systematic reading of the world's classics. Today there is a multitude of inexpensive editions of these works available, both in hardcover and paperback, many of them newly translated into English and often containing up-to-date critical notes and introductions. Writings by major critics of the English language are being issued in the "Regents Critics Series," to be followed by the writings of critics in other languages.

General works: KENNETH BURKE, *The Philosophy of Literary Form,* 2nd ed. (1967); BARRETT CLARK (ed.), *European Theories of the Drama: An Anthology of Dramatic Theory and Criticism from Aristotle to the Present Day,* rev. ed. (1965); IVOR ARMSTRONG RICHARDS, *Practical Criticism: A Study of Literary Judgment* (1929, reprinted 1968) and *Principles of Literary Criticism* (1924, reprinted 1961); GEORGE SAINTSBURY, *A History of Criticism and Literary Taste in Europe from the Earliest Texts to the Present Day,* 3 vol. (1900–04, reprinted 1961); NOWELL C. SMITH (ed.), *Literary Criticism* (1905).

Ancient to modern: *A Translation of the Latin Works of Dante Alighieri* (1904), see Letter X to Can Grande; CHARLES SEARS BALDWIN, *Ancient Rhetoric and Poetic, Interpreted from Representative Works* (1924), *Medieval Rhetoric and Poetic to 1400, Interpreted from Representative Works* (1928, reprinted 1971), and *Renaissance Literary Theory and Practice* (1939); EDWARD H. BLAKENEY (ed.), *Horace on the Art of Poetry* (1928); CECIL MAURICE BOWRA, *Primitive Song* (1962); S.H. BUTCHER, *Aristotle's Theory of Poetry and Fine Art,* with a critical text and translation of the *Poetics,* 4th ed. (1907; reprinted with corrections, 1932); INGRAM BYWATER, *Aristotle on the Art of Poetry* (1909), reprinted in Aristotle's *Poetics* and *Longinus on the Sublime,* ed. by CHARLES SEARS BALDWIN (1930); PIERRE CORNEILLE, *Oeuvres,* 3 vol. (1862), contains the "Discourse on Dramatic Poetry" and "Discourse on the Three Unities"; ALBERT S. COOK (ed.), *The Art of Poetry* (1892, reprinted 1926), contains a translation of Horace's *Art of Poetry;* J.D. DENNISTON, *Greek Literary Criticism* (1924, reprinted 1971), translations, beginning with Aristophanes; *Dryden's Essays on the Drama,* ed. by WILLIAM STRUNK (1898); ALLAN H. GILBERT (ed.), *Literary Criticism: Pluto to Dryden* (1940); EDMUND D. JONES (ed.), *English Critical Essays (Sixteenth, Seventeenth, and Eighteenth Centuries)* (1922); *Ben Jonson: Timber, Discoveries, and Conversations with Drummond of Hawthornden* in his *Works,* ed. by C.H. HERFORD and PERCY SIMPSON, 11 vol. (1925–52); LONGINUS, *On the Sublime,* Greek text with an English translation by W. HAMILTON FYFE ("Loeb Classical Library," 1927); *Plato,* trans. by LANE COOPER (1938), contains the *Phaedrus,* the *Symposium,* the *Ion,* the *Gorgias,* and parts of the *Republic* and the *Laws;* GEORGE PUTTENHAM, *The Arte of English Poesie,* ed. by GLADYS D. WILLCOCK and ALICE WALKER (1936); PAUL RADIN, *Primitive Man As Philosopher* (1927); *Sidney's Apologie for Poetrie,* ed. by J. CHURTON COLLINS (1907); G. GREGORY SMITH (ed.), *Elizabethan Critical Essays,* 2 vol. (1904); GAY WILSON ALLEN and H.H. CLARK (eds.), *Literary Criticism: Pope to Croce* (1941 and 1962); MATTHEW ARNOLD, *Essays in Criticism,* 2 vol., First and Second Series complete (1902), and *Essays in Criticism,* with an introduction by E.J. O'BRIEN, Third Series (1910); EDWIN BERRY BURGUM (ed.), *The New Criticism: An Anthology of Modern Aesthetics and Literary Criticism* (1930); SAMUEL TAYLOR COLERIDGE, *Biographia Literaria, or Biographical Sketches of My Literary Life and Opinions,* 2 vol., reprinted from the original plates (1907); BENEDETTO CROCE, *The Defence of Poetry, Variations on the Theme of Shelley,* trans. by E.F. CARRITT (1933); T.S. ELIOT, *Selected Essays, 1917–1932* (1932); *Hazlitt on English Literature: An Introduction to the Appreciation of Literature,* ed. by JACOB ZEITLIN (1913, reprinted 1970); E.R. HUGHES (trans.), *The Art of Letters: Lu Chi's "Wen Fu," A.D. 302* "Bollingen Series *XXIX* (1951); THOMAS ERNEST HULME, *Speculations: Essays on Humanism and the Philosophy of Art,* ed. by HERBERT READ (1924); JAMES GIBBONS HUNEKER, *Essays* (1929); HENRY JAMES, *Partial Portraits* (1888), see his essay, "The Art of Fiction"; and *The Art of the Novel: Critical Prefaces,* ed. with an introduction by R.P. BLACKMUR (1934); EDMUND D. JONES (ed.), *English Critical Essays of the Nineteenth Century* (1922); PHYLLIS M. JONES (ed.), *English Critical Essays: Twentieth*

Century (1933); WILLIAM PATON KER, *Collected Essays,* 2 vol. (1925, reprinted 1968); KARL MARX and FRIEDRICH ENGELS, *Sur la littérature et l'art*, ed. and trans. by JEAN FREVILLE (1936); H.L. MENCKEN, *A Mencken Chrestoinathy* (1949); PAUL ELMER MORE, *The Demon of the Absolute* (1928) and *Shelburne Essays*, 11 series (1904–21); FRIEDRICH NIETZSCHE, *Ecce Homo* and *The Birth of Tragedy,* trans. by CLIFTON FADIMAN (1927); HORATIO, *Works*, 10 vol. (1910); GEORGY V. PLEKHANOV, *Art and Society,* trans. by ALFRED GOLDSTEIN (1936), a Marxist analysis; EDGAR ALLAN POE, *Selections from the Critical Writings of Edgar Allan Poe,* ed. with an introduction by F.C. PRESCOTT (1909); EZRA POUND, *ABC of Reading* (1934) and *Literary Essays* (1954); HERBERT READ, *Reason and Romanticism* (1926); CHARLES AUGUSTIN SAINTE-BEUVE, *Causeries du lundi,* 15 vol. (1852–62; Eng. trans., 8 vol., 1909–11), contains "What Is a Classic?"; *Shelley's Literary and Philosophical Criticism,* ed. by JOHN SHAWCROSS (1909); LEO TOLSTOY, *What Is Art?,* trans. by AYLMAR MAUDE (1932); LEON TROTSKY, *Literature and Revolution,* trans. by ROSE STRUNSKY (1925, reprinted 1957); PAUL VALERY, *Littérature* (1929) and *Variété,* trans. by MALCOLM COWLEY (1927); WILLIAM CARLOS WILLIAMS, *In the American Grain* (1925); EDMOND WILSON, *Axel's Castle* (1931) and *The Triple Thinkers,* rev. ed. (1952 and 1963); EMILE ZOLA, *The Experimental Novel and Other Essays,* trans. by BELLE M. SHERMAN (1893).

Contemporary: CECIL MAURICE BOWRA, *In General and Particular* (1964); STANLEY BURNSHAW (ed.), *The Poem Itself,* rev. ed. (1967); CYRIL CONNOLLY, *The Modern Movement* (1965); PAUL GOODMAN, *The Structure of Literature* (1954); GYORGY LUKACS, *Studies in European Realism* (1964); MARSHALL MCLUHAN, *Understanding Media* (1964); HERBERT READ, *The Nature of Literature* (1956).

(K.Re.)

Literature, East Asian

The literatures of the countries of East Asia—of China, Korea, and Japan—exhibit the same strong similarities that are evident in their other arts as well as profound differences emerging from the distinctive histories and institutions of the individual lands. As in the visual and performing arts and elsewhere, China provided vital literary influences to its neighbours, in the examples of written language and of literary genres, allowing the Japanese and Korean peoples to transform the gifts through their own particular geniuses.

Generalizations about the literatures of these countries cannot be made across great periods of time—written Chinese literature is perhaps the oldest in the world—nor are they always as applicable to one country as to another. Certain tendencies, often submerged or redirected at various periods, do assert themselves in comparison to Western literatures, however, at least until the 20th century, when the full scope of Western genres was exploited by writers in the East. Songs and lyric poetry comprised some of the very earliest known examples of East Asian writings, and a vein of lyricism and of concern with the interactions of nature with man and his emotions has overweighed the development of intellectual or analytical systems of thought and exposition. For long periods in China, historical and didactic writing were considered, with poetry, to constitute literature, and fictional writings were assigned a very low place. Though the ballad emerged in some countries of East Asia, the epic form as constructed in the West was unknown. Characteristic, also, was a sense of authority from the past that, though it did not prohibit change in literary expression, served to inhibit the kind of experimentation and search for whatever is new that has impelled Western literature into relatively rapid alternations in taste and mode. Perhaps the most formative influence on the literatures of East Asia has been the religions of the region, not only Confucianism, Taoism, and Shintō that were indigenous but also Buddhism, which was imported at an early date from South Asia.

But too much can be made of such generalities. A full understanding of East Asian literatures can be achieved only through reference to the great influences on thought and philosophy which are contained in the articles BUDDHIST SACRED LITERATURE; CONFUCIAN TEXTS, CLASSICAL; HINDU SACRED LITERATURE; and TAOIST LITERATURE. See also BUDDHIST MYSTICISM; CHINESE MYTHOLOGY; HINDU MYTHOLOGY; and SHINTO. For coverage of closely related forms, see DANCE AND THEATRE, EAST ASIAN; KABUKI THEATRE; and NO THEATRE. The contemporary cultural status of the countries is incorporated in the articles CHINA; JAPAN; KOREA, NORTH; and KOREA, SOUTH.

This article is divided into the following major sections:

I. Chinese literature
  - General characteristics
  - History
    - Origins: *c.* 1400–220 BC
    - Ch'in and Han dynasties: 221 BC–AD 220
    - The Six Dynasties and Sui dynasty: AD 221–618
    - T'ang and Five Dynasties: 618–960
    - Sung dynasty: 960–1279
    - Yüan dynasty: 1279–1368
    - Ming dynasty: 1368–1644
    - Ch'ing dynasty: 1644–1912
    - Modern Chinese literature: 1916 to the present

II. Korean literature
  - Traditional forms and genres
    - Lyric poems
    - Prose
    - Oral literature
    - Literature in classical Chinese
  - Historical development
    - The earliest literature: before 57 BC
    - Literature of the Three Kingdoms: 57 BC–AD 668
    - Literature of Unified Silla: 668–935
    - Literature of Koryŏ: 935–1392
    - Literature of the early Chosŏn period: 1392–1598
    - Literature of the later Chosŏn period: 1598–1894
    - Transitional literature: 1894–1910
    - Modern literature: 1910 to the present

III. Japanese literature
  - Earliest literature
  - Classical literature: Heian period (794–1185)
  - Medieval literature: Kamakura, Muromachi, and Azuchi-Momoyama periods (1192–1600)
    - Kamakura period (1192–1333)
    - The Muromachi (1338–1573) and Azuchi-Momoyama (1574–1600) periods
  - Literature during the Tokugawa period (1603–1867)
    - Early Tokugawa period (1603–c. 1770)
    - Late Tokugawa period (*c.* 1770–1867)
  - Modern literature

(Ed.)

I. Chinese literature

GENERAL CHARACTERISTICS

Chinese literature is one of the major literary heritages of the world, with a number of distinctive phenomena and characteristics.

First of all, it has an uninterrupted history of more than 3,000 years, dating back at least to the 14th century BC. Its medium, the Chinese language, has retained its unmistakable identity in both its spoken and written aspects in spite of generally gradual changes in pronunciation, the emergence of regional and local dialects, and several stages in the structural representation of the written graphs, or "characters." Even the partial or total conquests of China for considerable periods by non-Chinese ethnic groups from outside the Great Wall failed to disrupt this continuity; for the conquerors were forced to adopt the written Chinese language as their official medium of communication because they had none of their own. Since the Chinese graphs were inherently nonphonetic, they were at best unsatisfactory tools for the transcription of a non-Chinese language; and attempts at creating a new alphabetic–phonetic written language for empire building proved unsuccessful on three separate occasions. The result was that after a period of alien domination, the conquerors were culturally assimilated (except the Mongols, who retreated en masse to their original homeland after the collapse of the Yuan [or Mongol] dynasty in 1368). Thus, there was no disruption in China's literary development.

Second, through cultural contacts, Chinese literature has profoundly influenced the literary traditions of other Asian countries, particularly Korea, Japan, and Vietnam. Not only was the Chinese script adopted for the written language in these countries but some writers adopted the Chinese language as their chief literary medium.

Third, the graphic nature of the written aspect of the Chinese language has produced a number of noteworthy effects upon Chinese literature and its diffusion: (1) Chinese literature, especially poetry, recorded in handwriting or in print purports to make an aesthetic appeal to the reader that is visual as well as aural. A poem composed of five-syllable lines, for example, could be recorded in such a way as to reveal automatically to the eye as well as the ear the uniformity of line length without making the attempt seem artificial. (2) This visual appeal of the graphs has in fact given rise to the elevated status of calligraphy in China, where it has been regarded for at least the last 16 centuries as a fine art comparable to painting. Scrolls of calligraphic renderings of poems and prose selections have continued to be hung alongside paintings in the homes of the common people as well as the elite, converting these literary gems into something to be enjoyed in everyday living. (3) On the negative side, such a writing system has been an impediment to education and the spread of literacy, thus reducing the number of readers of literature; for even a rudimentary level of reading and writing requires knowledge of more than 1,000 graphs, together with their pronunciation. (4) On the other hand, the Chinese written language, even with its obvious disadvantages, has been a potent factor in perpetuating the cultural unity of the growing millions of the Chinese people, including assimilated groups in far-flung peripheral areas. Different in function from recording words in an alphabetic–phonetic language, the graphs are not primarily indicators of sounds and can therefore be pronounced in variant ways to accommodate geographical diversities in speech and historical phonological changes without damage to the meaning of the written page. As a result, the major dialects in China never developed into separate written languages as did the Romance languages and, although the reader of a Confucian Classic in southern China might not understand the everyday speech of someone from the far north, Chinese literature has continued to be the common asset of the whole Chinese people. By the same token, the graphs of China could be utilized by speakers of other languages as their literary mediums.

Fourth, the pronunciation of the Chinese graphs has also influenced the development of Chinese literature. The fact that each graph had a monophonic pronunciation in a given context created a large number of homonyms, which would lead to misunderstanding and confusion when spoken or read aloud without the aid of the graphs. One corrective was the introduction of tones or pitches in pronunciation. As a result, metre in Chinese prosody is not concerned with the combination of syllabic stresses, as in English, but with those of syllabic tones, which produce a different but at least equally pleasing cadence. This tonal feature of the Chinese language has brought about an intimate relationship between poetry and music in China. All major types of Chinese poetry were originally sung to the accompaniment of music. Even after the musical scores were lost, the poems were, as they still are, more often chanted—in order to approximate singing—than merely read.

Relationship between poetry and music

Chinese poetry, besides depending on end rhyme and tonal metre for its cadence, is characterized by its compactness and brevity. There are no epics of either folk or literary variety and hardly any narrative or descriptive poems that are long by the standards of world literature. Stressing the lyrical, as has often been pointed out, the Chinese poet refrains from being exhaustive, marking instead the heights of his ecstasies and inspiration or the depths of sorrow and sympathy. A short poem in Chinese sometimes resembles a cablegram, wherein verbal economy is highly desirable. "Pronouns and conjunctions are almost always omitted. A word or two frequently indicates an allusion to a very complex thought or situation that may require paragraphs of elucidation ” This explains why some poems have been differently interpreted by learned commentators and competent translators.

It has also been observed that the line of demarcation between prose and poetry is much less distinctly drawn in Chinese literature than in other national literatures. This is clearly reflected in three genres. The fu, for example, "stands on the border line of poetry and prose. Strictly neither the one nor the other, it contains elements of both." It uses rhyme and metre and not infrequently also antithetic structure, but despite occasional flights into the realm of the poetic, it retains the features of prose without being necessarily prosaic. This accounts for various labels given to it in English by writers on Chinese literature—poetic prose, rhyme prose, prose poem, and prose poetry.

Another genre belonging to this category is *p'ien* wen ("parallel prose"), characterized by antithetic construction and balanced tonal patterns without the use of rhyme, which is suggestive of "a team of paired horses," as is implied in its Chinese designation. Despite the polyphonic effect thus produced, which approximates that of poetry, it has often been made the vehicle of proselike exposition and argumentation. Another genre, a peculiar mutation in this borderland, is the pa-ku wen ("eight-legged essay"). Now generally regarded as unworthy of classification as literature, for centuries (from 1487 to 1901) it dominated the field of Chinese writing as the principal yardstick in grading candidates in the official competitive civil-service literary examinations. A unique example of rhetorical gymnastics and verbal acrobatics, it exploited antithetical construction and contrasting tonal patterns to the limit by requiring pairs of columns consisting of long paragraphs, one responding to the other, word for word, phrase for phrase, sentence for sentence. A fad that stifled literary creativity and stunted intellectual advancement, despite its neat schematic architectonics and artificial euphony, it is no wonder that it was the only genre incapable of surviving the test of time.

The two kinds of prose

Chinese prose writing has been diverted into two streams, separated at least for the last 1,000 years by a gap much wider than the one between folk songs and so-called literary poems. Classical, or literary, prose (ku-wen, or wen-yen) aims at the standards and styles set by ancient writers and their distinguished followers of subsequent ages, with the Confucian Classics and the early philosophers as supreme models. While the styles may vary with individual writers, the language is always far removed from their spoken tongues. Sanctioned by official requirement for the competitive examinations and dignified by traditional respect for the cultural accomplishments of past ages, this medium became the linguistic tool of practically all Chinese prose writers. Vernacular prose (pai-hua), in contrast, comprises writings in the living tongue, the everyday language of the authors. Traditionally considered inferior, the medium was piously avoided for creative writing until it was adopted on occasion by novelists and playwrights from the 13th century on.

HISTORY

Origins: c. 1400–220 **BC.** The oldest specimens of Chinese writing extant are inscriptions on bones and tortoise shells dating back to the last three centuries of the Shang dynasty (c. 1384–c. 1122 BC) and recording divinations performed at the royal capital. These inscriptions, like those engraved on ceremonial bronze vessels toward the end of the Shang period, are usually brief and factual and cannot be considered literature. Nonetheless, they are significant in that their sizable vocabulary (about 3,400 characters, of which nearly 2,000 have been reliably deciphered) has proved to be the direct ancestor of the Chinese script of today. Moreover, the syntactical structure of the language bears a striking resemblance to later usages. From the frequent occurrences in the bone inscriptions of such characters as "dance" and "music," "drum" and "chimes" (of stone), "words" and "southern" (airs), it can safely be inferred that, by the Shang dynasty, songs were sung to the accompaniment of dance and music; but these songs are now lost.

First anthology of poetry: Shih Ching

Poetry. The first anthology of Chinese poetry, known as the Shih *Ching* ("Classic of Poetry") and consisting of temple, court, and folk songs, was given definitive form somewhere around the time of Confucius (551–479 BC).

But its 305 songs range in date from the very beginning of the Chou dynasty *(c.* 1122 BC) to around 585 BC, when the dynasty was already declining.

The ***Shih Ching*** is generally accounted the third of the five Confucian Classics, the other four of which are: (1) the **I** ***Ching*** ("Classic of Changes"), a book of divination; (2) the ***Shu Ching*** ("Classic of History"), a collection of official documents; (4) the ***Li chi*** ("Record of Rites"), a book of rituals with accompanying anecdotes; and (5) the ***Ch'un-ch'iu*** ("Spring and Autumn") annals, a chronological history of the feudal state of Lu, where Confucius was born, consisting of topical entries of major events from 722 to 481 BC. The Five Classics have been held in high esteem by Chinese scholars since the 2nd century BC. (For a discussion of the **I** ***Ching*** and ***Shu Ching,*** see below ***Prose.)***

The poems of the ***Shih Ching*** were originally sung to the accompaniment of music; and some of them, especially temple songs, were accompanied also by dancing.-(In all subsequent periods of Chinese literary history, new trends in poetry were profoundly influenced by music.) Most of the poems of the ***Shih Ching*** have a preponderantly lyrical strain whether the subject is hardship in military service or seasonal festivities, agricultural chores or rural scenes, love or sports, aspirations or disappointments of the common folk and of the declining aristocracy. Apparently, the language of the poems was relatively close to the daily speech of the common people, and even repeated attempts at refinement during the long process of transmission have not spoiled their freshness and spontaneity. In spite of this, however, when the songs are read aloud and not sung to music their prevailing four-syllable lines conduce to monotony, hardly redeemed by the occasional interspersion of shorter or longer lines; a sample (in modern pronunciation) illustrates this:

Kuan kuan tsui chiu
Tsai he chih chou
Miao t'iao shu nü
Chün tzu hao ch'iu
"Kuan kuan" *cry* the ospreys
On the islet in the river.
Lovely is the good lady,
Fit bride for our lord.

If there ever was an epic tradition in ancient China comparable to that of early India or the West, only dim traces of it persist in the written records. The ***Shih Ching*** has a few narrative poems celebrating heroic deeds of the royal ancestors, but these are rearranged in cycles and only faintly approximate the national epics of other peoples. One cycle, for example, records the major stages in the rise of the Chou kingdom, from the supernatural birth of its remote founder to its conquest of the Shang kingdom. These episodes, which, according to traditional history, cover a period of more than 1,000 years, are dealt with in only about 402 lines. Other cycles, which celebrate later military exploits of the royal Chou armies, are even briefer.

The ***Shih Ching*** exerted a profound influence on Chinese poetry that, generally speaking, has stressed the lyrical rather than the narrative element; a dependence more on end rhymes for musical effect than on other rhetorical devices; regular lines, consisting of a standard number of syllables; and utilization of intonation inherent in the language for rhythm, instead of the alternation of stressed and unstressed syllables as in Western poetry. The high regard in which this anthology has been held in China is due both to its antiquity and to the legend that Confucius himself edited it. It was elevated in 136 BC to the position of a major classic in the Confucian canon.

Ch'u tz'u

Meanwhile, another type of poetry, also originating in music and dance, had developed in the south, in the basin of the Yangtze River, an area dominated by the principality of Ch'u—hence the generic appellation ***Ch'u tz'u,*** or "songs of Ch'u." These southern songs, though adorned with end rhymes like the songs of the ***Shih Ching,*** follow a different metrical pattern: the lines are usually longer and more irregular and are commonly (though not always) marked by a strong caesura in the middle, with a syllable representing a sigh. Their effect is thus rather plaintive, and they lend themselves to chanting instead of singing. The beginning of this tradition is obscure because most of the early samples were eclipsed by the brilliant 4th/3rd-century BC compositions of the towering genius Ch'ii Yüan, China's first known poet.

Among some 25 elegies attributed to Ch'ii Yuan, the most important and longest is ***Li sao*** ("On Encountering Sorrow"), which has been described as a politico-erotic ode, relating by means of a love allegory the poet's disappointment with his royal master and describing his imaginary travels in distant regions and the realms of heaven, in an attempt to rid himself of his sorrow. Ch'ii Yiian committed suicide by drowning in the Mi-lo River; and his tragic death, no less than his beautiful elegies, helped to perpetuate the new literary genre. In contrast to the poems of the ***Shih Ching,*** which had few successful imitators, the genre created by Ch'ii Yüan was cultivated for more than five centuries, with later revivals.

Prose. Prior to the rise of the philosophers in the 6th century BC, brief prose writings were reported to be numerous; but of these only two collections have been transmitted: the ***Shu,*** or ***Shu Ching*** ("Classic of History"), consisting of diverse kinds of primitive state papers, such as declarations, portions of charges to feudal lords, and orations; and the **I,** or **I** ***Ching*** ("Classic of Changes"), a fortune-telling manual. Both grew by accretion and, according to a very doubtful tradition, were edited by Confucius himself. Neither can be considered literature, but both have exerted influence on Chinese writers for over 2,000 years as a result of their inclusion in the Confucian canon.

Earliest writings claiming individual authorship

The earliest writings claiming individual "authorship," in the loose sense of the term, are the ***Lao-tzu,*** or ***Tao-te Ching*** ("Classic of the Way of Power"), attributed to Lao-tzu, who is credited with being the founder of Taoism and who might have been an older contemporary of Confucius, and the ***Lun yü,*** or ***Analects*** (selected miscellaneous passages), of Confucius. Neither philosopher wrote extensively, and the gist of their teachings was recorded by their followers. Thus, the ***Lao-tzu*** consists of brief summaries of Lao-tzu's sayings, many in rhyme and others in polished prose to facilitate memorization. Likewise, the ***Analects*** is composed of collections of the sage's sayings, mostly as answers to questions or as a result of discussions because writing implements and materials were expensive and scarce. The circumstances of the conversations, however, were usually omitted; and as a consequence the master's words often sound cryptic and disjointed, despite the profundity of the wisdom.

By about 400 BC, writing materials had improved, and a change in prose style resulted. The records of the discourses became longer, the narrative portions more detailed; jokes, stories, anecdotes, and parables, interspersed in the conversations, were included. Thus, the ***Mencius,*** or ***Meng-tzu,*** the teachings of Mencius, not only is three times longer than the ***Analects*** of Confucius but also is topically and more coherently arranged. The same characteristic may be noticed in the authentic chapters of the ***Chuang-tzu,*** attributed to the Taoist sage Chuang-tzu, who "in paradoxical language, in bold words, and with subtle profundity, gave free play to his imagination and thought. . . . Although his writings are inimitable and unique, they seem circuitous and innocuous. Although his utterances are irregular and formless, they are unconventional and readable . . ." (quoted from the epilogue of the ***Chuang-tzu).***

Well-developed essay

The first real example of the well-developed essay, however, is found neither in the ***Mencius*** nor in the ***Chuang-tzu*** but in the ***Mo-tzu,*** attributed to Mo Ti, or Mo-tzu, a predecessor of Mencius and Chuang-tzu, whose singular attainments in logic made him a forceful preacher. His recorded sermons are characterized by simplicity of style, clarity of exposition, depth of conviction, and directness of appeal.

The prose style continued to be developed by such outstanding philosopher-essayists as Hsiin-tzu and his pupil, the Legalist Han-fei-tzu. The peak of this development, however, was not reached until the appearance of the first expertly arranged full-length book, ***Lii-shih Ch'un Ch'iu***

("The Spring and Autumn [Annals] of Mr. Lü"), completed in 240 BC under the general direction of Lü Pu-wei. The work, 60 essays in 26 sections, summarizes the teachings of the several schools of philosophy as well as the folklore of the various regions of China.

The *fu* genre

Ch'in and Han dynasties: 221 BC–AD 220. *Poetry.* Following the unification of the empire by the Ch'in dynasty (221–206 BC) and the continuation of the unified empire under the Han (206 BC–AD 220), literary activities took new directions. At the Imperial and feudal courts, the *fu* genre, a combination of rhyme and prose (see above *General characteristics),* began to flourish. Long and elaborate descriptive poetic compositions, the *fu* were in form a continuation of the Ch'u elegies, now made to serve a different purpose—the amusement of the new aristocracy and the glorification of the empire—by dwelling on such topics as the low table and the folding screen or on descriptions of the capital cities. But even the best *fu* writing, by such masters of the art as Mei Sheng and Ssu-ma Hsiang-ju, bordered on the frivolous and bombastic. Another major *fu* writer. Yang Hsiung, in the prime of his career remorsefully realized that the genre was a minor craft not worthy of a true poet. Nonetheless, the *fu* was almost universally accepted as the norm of creative writing, and nearly 1,000 pieces were produced that have survived.

A more important contribution to literature by the Han government was the reactivation in 125 BC of the Yüeh Fu, or Music Bureau, which had been established at least a century earlier to collect songs and their musical scores. Besides temple and court compositions of ceremonial verse, this office succeeded in preserving a number of songs sung or chanted by the ordinary people, including songs from the border areas, which reveal alien influences. This category—called *yiieh-fu,* after the Music Bureau—includes not only touching lyrics but also charming ballads.

Ballads

One such ballad, "The Orphan," tells of an orphan's hardships and disappointments; the form of the poem—lines of irregular length, varying from three to six syllables (or graphs)—represents the singer's attempt to simulate the choking voice of the sufferers. *Lo-fu hsing* ("The Song of Lo-fu"; also called *Me shang sang,* "Roadside Mulberry Tree"), containing 53 five-syllable lines, recounts how a pretty young lady declined a carriage ride offered her by a government commissioner. The most outstanding folk ballad of this period is *K'ung-chiieh tung-nan fei* ("Southeast the Peacock Flies"). The longest poem of medieval China (353 lines), it relates the tragedy of a young married couple who had committed suicide as the result of the cruelty of the husband's mother. The ballad was probably first sung shortly after AD 200 and grew by accretion and refinement in oral transmission until it was recorded in final form for the first time around 550. *Yiieh-fu* songs, most of which are made up mainly of five-syllable lines, became the fountainhead of a new type of poetry, *Ku shih* ("ancient style poems"); contemporary Han dynasty poets at first merely refined the originals of the folk songs without claiming credit, and later imitated their fresh and lively metre.

Prose. Prose literature was further developed during the Ch'in and Han dynasties. In addition to a prolific output of philosophers and political thinkers—a brilliant representative of whom is Liu An, prince of Huai-nan, whose work is called *Huai-nan-tzu (c.* 140 BC; "The Master of Huai-nan")—an important and monumental category of Han dynasty literature consists of historical works. Outstanding among these is the *Shilz-chi (c.* 85 BC; "Historical Records") by Ssu-ma Ch'ien. A masterpiece that took 18 years to produce, it deals with major events and personalities of about 2,000 years (down to the author's time), comprising 130 chapters and totalling more than 520,000 words. The *Shill-chi* was not only the first general history of its kind attempted in China, it also set a pattern in organization for dynastic histories of subsequent ages. An artist as well as a historian, Ssu-ma Ch'ien succeeded in making events and personalities of the past into living realities for his readers. His great successor, the poet-historian-solider Pan Ku, author of the *Han shu* (history of the Han dynasty), which totals over 800,000 words, performed a similar tour de force but did not equal Ssu-ma Ch'ien in either scope or style.

Promotion of a standard literary language in the Han dynasty

Pan Ku's prose style, though not necessarily archaic, was more consciously literary—a result of the ever-widening gap between the spoken and written aspects of the language. This anomaly was more evident in China than elsewhere. and it was to have far-reaching effects on the evolution of Chinese literary tradition. In an attempt to resolve the difficulties of communication among speakers of many dialects in the empire, a standard literary language, *ku-wen,* or *wen-yen,* was promoted from the Han dynasty on. Perpetuated for more than 2,000 years, the literary language failed to keep pace with changes in the spoken tongue, and eventually it became almost unintelligible to the illiterate masses.

The Six Dynasties and Sui dynasty: AD 221–618. After the fall of the Han dynasty, there was a long period of political division (AD 221–589), with barely four decades of precarious unification (AD 280–317). Despite the social and political confusion and military losses, however, the cultural scene was by no means dismal. Several influences upon the development of literature are noteworthy. First, Buddhism, introduced earlier, had brought with it religious chants and Indian music, which helped to attune Chinese ears to the finer distinctions of tonal qualities in their own language. Second, aggressive northern tribes, who invaded and dominated the northern half of the country from 316, were being culturally absorbed and converted. Third, the political division of the empire between the South and the North (as a result of the domination of non-Chinese in the north) led to an increase in cultural differences and to a subsequent rivalry to uphold what was regarded as cultural orthodoxy, frequently resulting in literary antiquarianism.

Poetry. Folk songs flourished in both regions. In the South, popular love songs, originating in the coastal areas, which now come increasingly under Chinese political and cultural domination, attracted the attention of poets and critics. The love songs of the North were more militant. Reflecting this spirit most fully is the "Ballad of Mu Lan," which sings of a girl who disguised herself as a warrior and won glory on the battlefield.

Ts'ao Chih and T'ao Ch'ien

Soon the number of writers of "literary" poetry greatly increased. Among them, two poets deserve special mention. Ts'ao Chih (3rd century), noted for his ethereal lyricism, gave definite artistic form to the poetry of the five-syllable line, already popularized in folk song. T'ao Ch'ien (4th–5th centuries), also known as Tao Yüan-ming, is one of China's major poets and was the greatest of this period. A recluse, he retired from a post in the bureaucracy of the Chin dynasty at the age of 33 to farm, contemplate nature, and write poetry. His verse, written in a plain style, was echoed by many poets who came after him. Using several verse forms with seemingly effortless ease—including the *fu,* for *Kuei-ch'ii-lai tz'u* ("Homeward Bound")—he was representative of the trend of the age to explore various genres for lyrical expression. One of his best loved poems is the following *ku shih* (ancient style poem), translated by Arthur Waley; it is one of 12 he wrote at different times after he had been drinking.

I built my hut in a zone of human habitation,
Yet near me there sounds no noise of horse or coach.
Would you know how this is possible?
A heart that is distant creates a wilderness round it.
I pluck chrysanthemums under the eastern hedge,
Then gaze long at the distant hills.
The mountain air is fresh at the dusk of day;
The flying birds two by two return.
In these things there lies a deep meaning;
Yet when we would express it, words suddenly fail us.

Prose. As orthodox Confucianism gradually yielded to Taoism and later to Buddhism. nearly all of the major writers began to cultivate an uninhibited individuality. Lu Chi, 3rd-century poet and critic, in particular emphasized the importance of originality in creative writing and discredited the long-established practice of imitating the great masters of the past. Still, his celebrated essay on literature (*Wen fu*), in which he enunciated this princple, was written as a *fu*, showing after all that he was a child of his own age. The 3rd/4th-century Taoist philosopher Ko Hung insisted that technique is no less essential to a writer than moral integrity. The revolt of the age against conventionality was revealed in the new vogue of *ch'ing-t'an* ("pure conversation"), intellectual discussions on lofty and nonmundane matters, recorded in a 5th-century collection of anecdotes entitled *Shih shuo* ("Sayings of the World") by Liu Yi-ch'ing. Though prose writers as a whole continued to be most concerned with lyrical expression and rhetorical devices for artistic effect. there were notable deviations from the prevailing usage in the polyphonic type of prose called *p'ien-wen* ("parallel prose"). In this form, parallel construction of pairs of sentences and counterbalancing of tonal patterns were the chief requirements. *P'ien-wen* was used especially in works concerned with philosophical disputes and in religious controversies; but it was also used in the first book-length work of literary criticism, *Wen-hsin tiao-lung* ("Carving of the Literary Dragon"), by the 6th-century writer Liu Hsieh.

Development of **poly-**phonic prose: *p'ien-wen*

Among prose masters of the 6th century, two northerners deserve special mention: Yang Hsien-chih, author of *Lo-yang Chia-lan chi* ("Record of Buddhist Temples in Lo-yang"), and Li Tao-yiian, author of *Shui Ching chu* ("Commentary on the Water Classic"). Although both of these works seem to have been planned to serve a practical, utilitarian purpose, they are magnificent records of contemporary developments and charming storehouses of accumulated folklore, written with great spontaneity and artistry.

This age also witnessed the first impact of Buddhist literature in Chinese translation, which had been growing in size and variety since the 2nd century.

T'ang and **Five** Dynasties: **618–960.** During the T'ang dynasty (618–907), Chinese literature reached its golden age.

Poetry. In poetry, the greatest glory of the period, all the verse forms of the past were freely adopted and refined, and new forms were crystallized. A new form, which had been gradually evolving for over a century, was perfected early in the dynasty and given the definitive name *lü shih* ("regulated verse"). A poem of this kind consists of eight lines of five or seven syllables—each line set down in accordance with strict tonal patterns—calling for parallel structure in the middle, or second and third, couplets. The following poem by Shen Ch'üan-ch'i. addressed to the grandfather of the great T'ang poet Tu Fu in 705, illustrates the *lü shih* form (the Cantonese used here approximates the T'ang pronunciation then in use at the capital):

T'in cheung tei fut ling tau fan
Hui kwok lei ka kin puak wan
Lok p'o fung kwong ho soh ch'i
Sung shaan cheung lai pat ham man
Naam fau cheung hoi yan hoh chue
Pak mong hang yeung ngan kei kw'an
Leung tei kong shan maan yue lei
Hoh shi ch'ung kin shing ming kwan
The sierra divides the vast heaven and the wide earth.
Far away from country and home appear the same white clouds.
The scenery on the banks of the Lo River, what does it resemble?
The pestilential vapours of the Sung Mountain are unbearable!
Floating southward on the surging seas, where is my friend?
Going northward toward Hengyang [city], how many flocks are there of swans?
Between us are mountains and rivers stretching over ten thousand miles.
How soon are we again to see our sage and enlightened emperor?

Chüeh chii verse form

Another verse form much in vogue is the *chüeh chü* ("truncated verse"). An outgrowth and a shortened version of the *lü shih*, it omits either the first four lines, the last four lines, the first two and the last two lines, or the middle four lines. Thus, the tonal quality of the *lü shih* was retained, whereas antithetic structure was made optional. These poems of four lines, each consisting of five or seven words (syllables or characters), had to depend for their artistry on suggestiveness and economy comparable to The *Rubdiydt* of Omar Khayyam and the Japanese haiku. The following example is by Li Po (romanization in Mandarin):

Pu chien Tung Shan chiu
Ch'iang-wei chi tu hua
Pai yün huan tzu san
Ming Yiieh lo shui chia
For long *l* have not seen the Eastern Hill.
How many times has the rose flowered?
Do the white clouds as then scatter themselves?
And behind whose dwelling sets the moon?

The fine distinctions of tonal variations in the spoken language had reached their height during this period, with eight tones; and rules and regulations concerning the sequence of lighter and heavier tones had been formulated. But since the observance of strict rules of prosody was not mandatory in the *ku shih* ("ancient style") form still in use, it was possible for an individual poet to enjoy conformity or freedom as he saw fit.

Of the more than 2,000 T'ang poets whose works—totalling over 48,900 pieces—have been preserved, only a few can be mentioned. Meng Hao-jan, who led the life of a recluse, excelled in poems of five-syllable lines reminiscent of the spirit of T'ao Ch'ien. Wang Wei, a musician and the traditional father of monochrome landscape painting, was also a great poet. Influenced by Buddhism, he wrote exquisite meditative verse of man's relation to nature that exemplified his own dictum that poetry should have the beauty of painting and vice versa. Li Po, one of the two major poets of the T'ang dynasty, a lover of detachment and freedom, deliberately avoided the *lü shih*, or regulation poetry, and chose the less formal verse forms to sing of friendship or wine. An example is the poem "To Tan-Ch'iu," translated by Authur Waley.

My friend is lodging high in the Eastern Range,
Dearly loving the beauty of valleys and hills.
At green Spring he lies in the empty woods,
And is still asleep when the sun shines on high.
A pine-tree wind dusts his sleeves and coat;
A pebbly stream cleans his heart and ears.
1 envy you, who far from strife and talk
Are high-propped on a pillow of blue cloud.

The works of Tu Fu

Generally considered the greatest poet of China was Tu Fu, a keen observer of the political and social scene who criticized injustice wherever he found it and who clearly understood the nature of the great upheaval following the rebellion of dissatisfied generals in 755, which was a turning point in the fortunes of the T'ang. As an artist, Tu Fu excelled in all verse forms, transcending all rules and regulations in prosody while conforming to and exploiting them. His power and passion can perhaps be suggested by a single line (translated by Robert Payne): "Blue is the smoke of war, white the bones of men."

One of the admirers of Tu Fu as a poet-historian was Po Chü-i who, like his great predecessor, was deeply concerned with the social problems of his age. Po Chü-i sought to learn from ordinary folk not only naturalness of language but also their feelings and reactions, especially at the height of his career when he wrote what he called the *Hsin yiielz-fu shih* ("New Music Bureau Poems").

At the end of the T'ang and during the Five Dynasties, another new verse form developed. Composed normally of lines of irregular length and written as lyrics to musical tunes, this form came to be known as *tz'u*, in contrast with *slrih*, which includes all the verse forms mentioned above. Since the lines in a *tz'u* might vary from one to nine or even 11 syllables, they were comparable to the natural rhythm of speech and therefore easily understood when sung.

Development of *tz'u*

First sung by ordinary folk, they were popularized by professional women singers and finally, before the end of the T'ang, attracted the attention of poets. It was not, however, until the transitional period of the Five Dynasties (907–960), a time of division and strife, that *tz'u* became the major vehicle of lyrical expression. Of *tz'u* poets in this period, the greatest was Li Yü, last monarch of the Southern T'ang, who was seized in 976 as the new Sung dynasty consolidated its power. Li Yii's *tz'u* poetry is saturated with a tragic nostalgia for better days in the south; it is suffused with sadness—a new depth of feeling notably absent from earlier *tz'u*, which had been sung at parties and banquets. The following is typical, translated by Jerome Ch'en and Michael Bullock:

Lin hua hsieh liao ch'un hung
T'ai ch'ung ch'ung
Wu nai chno lai hart yii wan lai feng
Yen chih lei
Hsiang liu tsui
Chi shih ch'ung
Tzu shih jen sheng ch'ang hen shui ch'ang tung
The red of the spring orchard has faded.
Far too soon!
The blame is often laid
on the chilling rain at dawn
and the wind at dusk.
The rouged tears
That intoxicate and hold in thrall—
When will they fall again?
As a river drifts toward the east
So painful life passes to its bitter end.

Folk literature. Besides the early *tz'u*, the end of the T'ang saw the evolution of another new folk form: *pien-wen* ("popularizations," not to be confused with *p'ien-wen*, or balanced prose), utilizing both prose and verse to retell episodes from Buddha's life and, later, non-Buddhist stories from Chinese history and folklore.

Han Yü's reform of prose

Prose. In prose writing a major reform was led by Han Yii against the peculiarly artificial prose style of *p'ien-wen*, which, cultivated by the literati for almost 1,000 years; had become so burdened with restrictive rules as to make forthright expression virtually impossible. Han Yii boldly advocated the use of Chou philosophers and early Han writers as models for prose writing. This seemingly conservative reform had, in fact, a liberalizing effect; for the sentence unit in prose writing was now given perfect freedom to seek its own length and structural pattern as logic and content might dictate, instead of slavishly conforming to the rules of *p'ien-wen*. This new freedom enabled Liu Tsung-yuan, Han Yii's chief associate in the literary reform, to write charming travel and landscape pieces. It also accelerated the development of a new genre in prose: well-made tales of love and romance, of heroic feats and adventures, of the mysterious and supernatural, and of imaginary incidents and fictionalized history. Among the 9th-century writers of such prose romances were Han Yii's pupil Shen Ya-chih and Po Hsing-chien, younger brother of the poet Po Chu-i, as well as many other literary notables. These prose romances, generally short, were written in the more natural *ku wen*, or classical prose style, for the amusement of the literati and did not reach the masses until some of the popular ones were adapted by playwrights in later ages.

Sung dynasty: 960–1279. The Sung dynasty was marked by cultural advancement and military weakness. During this period, literary output was spectacularly increased, thanks mainly to the invention and improvement of printing from the 8th century and to the establishment of public schools throughout the empire from 1044. Nearly all the literary genres in verse and prose were continued; and some trends, begun in T'ang times, were accelerated.

Prose. In prose, the reform initiated by Han Yü in the name of ancient, more straightforward style *(ku-wen)* was re-emphasized by such 11th-century writers as Ou-yang Hsiu and Su Tung-p'o. Both men held high rank in the civil service and were great painters as well as leading poets. Nevertheless, their contribution to prose writing in *ku-wen* style was as important as their poetry. The *ku-wen* movement was further supported by men whose primary interest was not belles-lettres, such as Ssu-ma Kuang, the statesman-historian, and Chu Hsi, the scholar-philosopher and principal formulator of Neo-Confucianism, both of whom advocated greater naturalness in prose writing.

Sung trends in prose fiction

In prose fiction there were two distinct trends. Short tales in *ku-wen* were written in ever greater bulk but failed to maintain the level achieved in the T'ang dynasty. The subject matter became more fragmentary and anecdotal and the style duller. In sharp contrast to the *ku-wen* school, which was still a literary language despite the movement toward naturalness of expression, there arose a school of storytelling in the vernacular. Almost purely oral in origin, these tales remained unrecorded except for the bare outlines followed by the storyteller who entertained audiences gathered in marketplaces, fairgrounds, or temple yards. These promptbooks were gradually expanded and elaborated on, so that in the 12th century they became fairly lengthy, connected stories, especially those dealing with fictionalized history. This elevation of the everyday speech of the common people as a medium of story writing of the *hua-pen* ("story roots") type was to open up new vistas in prose fiction in later periods.

Poetry. Poetry of the conventional type (shih) was cultivated by numerous rival schools, each claiming many illustrious members. On the whole, the rival literary movements were significant as steps toward greater naturalness in syntax, and a few outstanding writers approximated the spoken vernacular language. Among the many shih poets of the Sung dynasty, Lu Yu, who flourished in the 12th century, was a towering figure. A traveller and patriot, he wrote throughout his long career no fewer than 20,000 poems, of which some 9,200 have been preserved.

But it was in their utilization of the newer verse form, *tz'u*, that Sung poets achieved their greatest distinction, making *tz'u* the major genre of the dynasty. As noted above, the *tz'u* form had been popularized at first orally by women singers; and the first generation of *tz'u* writers had been inspired and guided by them in sentiment, theme, and diction; their lyrics were thus redolent with the fragrance of these women. Later in the 12th century, as men (and one great woman) of letters began to take over, the *tz'u* form reached the heights of great art. Ou-yang Hsiu and Li Ch'ing-chao, the latter generally considered the greatest woman poet of China, may be considered representatives of this trend. Li Ch'ing-chao's poems, paralleling her life, are intensely personal. They at first dealt with the joys of love, but gradually their tone darkened to one of despair, caused first by frequent and lengthy separations from her husband, who was in government service, and then by his untimely death.

Other masters of the *tz'u* were Su Tung-p'o and Hsin Ch'i-chi, the latter a soldier turned recluse. It was Hsin Ch'i-chi who imbued the writing of tz'u with new characteristics by rising above rules without breaking them, surpassing in this respect his contemporaries as well as those who came after.

Yiian dynasty: 1279–1368. Fleeing from the Chin (Ju-chen) Tatars, who captured the Sung capital in 1126, the Sung dynasty retreated southward. For almost a century and a half, China was again divided. And in spite of political reunification by Kublai Khan, founder of the Yuan, or Mongol, dynasty (beginning in 1260 in the North and comprising the whole of China in 1280), the cultural split persisted. In the South, where China's historic traditions found asylum, racial and cultural homogeneity persisted. In fact, the centre of Chinese philosophy and traditional literature never again returned north of the Yangtze Delta. But in the North new developments arose, which led to wholly new departures. First, the migration and fusion of the various ethnic groups gave birth to a common spoken language with fewer tones, which later was to become the basis of a national language; second, with the southward shift of the centre of traditional culture, the prestige of the old literature began to decline in the North, especially in the eyes of the conquerors. Thus, in contrast to the South, North China under the Yuan dynasty provided a unique milieu for unconventional literary activities.

Drama. In this period, dramatic literature came into a belated full flowering. The skits and vaudeville acts, the puppet shows and shadow plays of previous ages had laid the foundation for a full-fledged drama; but the availability of Indian and Iranian models during the Yiian dynas-

Development of a new literary type

ty may have been a more immediate cause for its accelerated growth. Many Chinese men of letters refused to cooperate with the alien government, seeking refuge in painting and writing. As the new literary type developed—the drama of four or five acts, complete with prologue and epilogue and including songs and dialogue in language fairly close to the daily speech of the people—many men of letters turned to playwriting. Between 1234 and 1368, more than 1,700 musical plays were written and staged, and 105 dramatists were recorded: moreover, there is an undetermined number of anonymous playwrights whose unsigned works have been preserved but discovered only in recent decades. This remarkable burst of literary innovation, however, failed to win the respect of the orthodox critics and official historians. No mention of it was made in the copious dynastic history, *Yiian shih*; and casual references in the collected works of contemporary writers were few. Many plays were allowed to fall into oblivion. It was not until 1615 that a bibliophile undertook to reprint, as a collection, 100 of the 200 plays he had seen. Even after ardent searches by 20th-century librarians and specialists, the number of extant Yuan dramas has been increased to only 167, hardly 10 percent of the number produced. Moreover, since the musical scores have been lost, the plays cannot be produced on the stage in the original manner.

Among the Yuan dramatists, the following at least deserve special mention. Kuan Han-ch'ing, the author of some 60 plays, was the first to achieve distinction. His *Tou-o yuan* ("Injustice Suffered by Tou-o") deals with the deprivations and injustices suffered by the heroine, Tou-o, which begin when she is widowed shortly after her marriage to a poor scholar and culminate in her execution for a crime she has not committed. Wang Shih-fu, Kuan's contemporary, wrote *Hsi hsiang chi* ("Romance of the Western Chamber"), based on a popular T'ang prose romance about the amorous exploits of the poet Yuan Chen, renamed Chang Chun-jui in the play. Besides its literary merits and its influence on later drama, it is notable for its length, two or three times that of the average Yuan play. Ma Chih-yuan, another contemporary, wrote 14 plays, of which the most celebrated is *Hart kung ch'iu* ("Sorrow of the Han Court"). It deals with the tragedy of a Han dynasty court lady, Wang Chao-chun, who, through the intrigue of a vicious portrait painter, was picked by mistake to be sent away to Central Asia as a chieftain's consort. Like the "Romance of the Western Chamber," this play has been translated into western European languages.

Characteristics of the new literary genre

This new literary genre acquired certain distinct characteristics: (1) All extant compositions may be described as operas; (2) each play normally consists of four acts following a prologue; (3) the language of both the dialogue (for the most part in prose) and the arias—which alternate throughout the play—are fairly close to the daily speech of ordinary people; (4) all of the arias are in rhymed verse, and only one end rhyme is used throughout an act; (5) all of the arias in an act are sung by only one actor; (6) nearly all of the plays have a happy ending; (7) the characters in most of the plays are people of the middle and underprivileged classes—poor scholars, bankrupt merchants, Buddhist nuns, country innkeepers, peasants, thieves, kidnappers, abductors, and women entertainers—antedating a similar trend in European drama by nearly four centuries.

At least 12 of the playwrights thus far identified were Sinicized members of originally non-Chinese ethnic groups—Mongols, Juchens, Uighur, and other Central Asians.

Poetry. Another literary innovation, preceding but later interacting with the rise of the drama, was a new verse form known as *san-ch'ü* ("non-dramatic songs"), a liberalization of the *tz'u,* which utilized the spoken language of the people as fully as possible. Although line length and tonal pattern were still governed by a given tune, extra words could be inserted to make the lyrics livelier and to clarify the relationship between phrases and clauses of the poem. The major dramatists were all masters of this genre.

Vernucular fiction. Similarly, fiction writers who wrote in a semi-vernacular style began to emerge, continuing the tradition of storytellers of the past but writing their stories rather than reciting them orally. The authors of fiction differed significantly from the prose romance writers of the T'ang and Sung. Their works not only were much longer but also were written almost entirely in the vernacular. Clearly, these fiction works were addressed to the masses and not to the authors' fellow literati. All of the early pieces of this type of book-length fiction were poorly printed and anonymously or pseudonymously published. The first known master of the art was Lo Kuan-chung, a bold social critic in his 14th-century historical and supernatural romances, the best known of which are *San Kuo chih yen-i (Romance of the Three Kingdoms*, 1925), *Shui-hu chuan (The Water Margin,* also translated as *All Men are Brothers),* and *P'ing yao chuan* ("The Subjugation of the Evil Phantoms"). The best of the three from a literary standpoint is the *Shui-hu chuan*, which gives full imaginative treatment to a long accretion of stories and anecdotes woven around a number of enlightened bandits—armed social and political dissenters—whose exploits were recorded in the official dynastic history and given fictional treatment in a Sung *hua pen* ("story root"). In the *Shui-hu chuan*, the number of characters is increased from 36 to 108; the thread of the main narrative is more complex; many subplots are introduced; and characterization and description are elaborate. The language is almost completely vernacular.

The works of Lo Kuan-chung

Ming dynasty: 1368–1644. The Yuan dynasty was succeeded by the Ming dynasty, under which cultural influences from the South—expressed in movements toward cultural orthodoxy—-again became important. Nearly all the major poets and prose writers in traditional literature were southerners, who enthusiastically launched and supported antiquarian movements based on a return to models of various ages of the past. With the restoration of competitive literary examinations, which had been virtually discontinued under the Mongols, the highly schematic *pa-ku wen* ("eight-legged essay"; see above) was adopted as the chief yardstick in measuring a candidate's literary attainments. Despite occasional protests, it continued to engage the attention of aspirants to official literary honours from 1487 to 1901.

Classical literature. Although Ming poets wrote both *shih* and *tz'u* and their output was prodigious, poetry on the whole was imitative rather than freshly creative. Tirelessly, the poets produced verses imitating past masters, with few individually outstanding attainments.

Prose writers in the classical style were also advocates of antiquarianism and conscious imitators of the great masters of past ages. Rival schools were formed, but few writers were able to rise above the ruts of conventionalism. The Ch'in-Han school tried to underrate the achievements of Han Yü and Liu Tsung-yiian, along with the Sung essayists, and proudly declared that post-Han prose was not worth reading. The Tang-Sung school, on the other hand, accused its opponents of limited vision and re-emphasized Han Yü's dictum that literature should be the vehicle of Tao, equated with the way of life taught by orthodox Confucianism. These continuous squabbles ultimately led nowhere, and the literary products were only exquisite imitations of their respective models.

The Kung-an school of prose writers

The first voice of protest against antiquarianism was not heard until the end of the 16th century; it came from the Kung-an school, named after the birthplace of three brothers, of whom the middle one was the best known. Yuan Hung-tao challenged all of the prevailing literary trends, advocating that literature should change with each age and that any attempt at erasing the special stamp of an era could result only in slavish imitation. Declaring that he could not smile and weep with the multitude, he singled out "substantiality" and "honesty with oneself" as the chief prerequisites of a good writer.

This same spirit of revolt was shared by Chung Hsing and T'an Yuan-ch'un, of a later school, who were so unconventional that they explored the possibilities of writing intelligibly without observing Chinese grammatical usages. Although their influence was not long lasting, these two schools set the first examples of a new subgenre in prose—the familiar essay.

Vernacular literature. It was in vernacular literature that the writers of this period made a real contribution. In drama, a tradition started in the Sung dynasty and maintained in southern China during the period of Mongol domination was revitalized. This southern drama, also musical and known as *ch'uan-ch'i* ("tales of marvels"), had certain special traits: (1) a *ch'uan-ch'i* play contains from 30 to 40 changes of scene; (2) the change of end rhymes in the arias is free and frequent; (3) the singing is done by many actors instead of by the hero or heroine alone; (4) many plots, instead of being extracted from history or folklore, are taken from contemporary life.

Since there were no rules regulating the structure of the *ch'uan-ch'i*, playlets approaching the one-act variety were also written. This southern theatre movement, at first largely carried on by anonymous amateurs, won support gradually from the literary men until finally, in the 16th century, a new and influential school was formed under the leadership of the poet-singer Liang Ch'en-yii and his friend the great actor Wei Liang-fu. The K'un school, initiating a style of soft singing and subtle music, was to dominate the theatre to the end of the 18th century.

Aside from drama and *ta-ch'ü* (a suite of melodies sung in narration of stories), which in the South were noticeably modified in spirit and structure, becoming more ornate and bookish—it was prose fiction that made the greatest progress in the 16th century. Three important novels took shape at that time: Wu Ch'eng-en's *Hsi-yu chi (Monkey)* is a fictionalized account of the pilgrimage of the Chinese monk Hsüan Tsang to India in the 7th century. The subject matter was not new; it had been used in early *kua-pen,* or "story root," books and Yüan drama; but it had never been presented at length in such a lively and rapid-moving narration. Of all of the 81 episodes of trial and tribulation experienced by the pilgrim, no two are alike. Among the large number of monsters introduced, each has his unique individuality. Like the *Shui-hu chuan (The Water Margin;* see above *Yuan dynasty: Vernacular fiction),* it reveals the influence of the style of the oral storytellers, for each chapter ends with the sentence "in case you are interested in what is to follow, please listen to the next instalment, which will reveal it." Unlike the *Shui-hu chuan,* which was written in a kind of semi-vernacular, the language used was the vernacular of the living tongue. For the author the choice must have been a deliberate but difficult one, for he had the novel first published anonymously to avoid disapproval. Besides eliciting numerous commentaries and "continuations" in China, it has two English translations.

Important novels of the 16th century

Lo Mao-teng's *Hsi yang chi* ("Adventure to Western Ocean") is an embellished tale of the 15th-century explorer Cheng Ho, who sailed the South China Sea and the Indian Ocean seven times and saw the northeast coast of Africa more than half a century before Columbus embarked upon his maritime venture in 1492. Utilizing the records of the journeys as he saw fit, Lo gave such free play to his imagination that his masterpiece was a combination of the traits of *Gullivers Travels* and *Robinson Crusoe.*

The title of the third novel (the author of which is unknown), *Chin P'ing Mei,* is composed of graphs from the names of three female characters. Written in an extremely charming vernacular prose style, the novel is a well-knit, long narrative of the awful debaucheries of the villain Ch'ing Hsi-men. The details of the different facets of life in 16th-century China are so faithfully portrayed that it can be read almost as a documentary social history of that age. Even the sexual perversions of the characters are so elaborately depicted that Western translators have had to resort to Latin in rendering a number of indelicate passages. The novel has been banned in China more than once, and all copies of the first edition of 1610 were destroyed.

Ch'ing dynasty: 1644–1912. The conquest of China by the Manchus, a Mongol people from the region north of China who set up the Ch'ing dynasty in 1644, did not disrupt the continuation of major trends in traditional literature. (During the literary inquisition of the 18th century, however, many books suspected of anti-Manchu sentiments were destroyed; and numerous literati were imprisoned, exiled, or executed.) Antiquarianism dominated literature as before, and excellent poetry and prose in imitation of ancient and medieval masters continued to be written, many works rivalling the originals in archaic beauty and cadence. Although the literary craftsmanship was superb, genuine creativity was rare.

Poetry and prose nonfiction. In the field of *tz'u* writing, the 17th-century Manchu poet Nara Singde (Sinicized name, Na-lan Ch'eng-te) was outstanding; but even he lapsed into conscious imitation of Southern T'ang models except when inspired by the vastness of open space and the beauties of nature. In nonfictional prose, Chin Jen-jui continued the familiar essay form.

The prose romances of P'u Sung-ling

Prose fiction. P'u Sung-ling continued the prose romance tradition by writing in *ku-wen* ("classical language") a series of 431 charming stories of the uncanny and the supernatural titled *Liao-chai chih-i (Strange Stories from a Chinese Studio).* This collection, completed in 1679, was reminiscent of the early literary tale tradition, for it contained several T'ang stories retold with embellishments and minor changes to delineate the characters more realistically and to make the plots more probable. Such traditional supernatural beings as fox spirits, assuming in these stories temporary human form in the guise of pretty women, became for the first time in Chinese fiction humanized and likable. Despite the seeming success of these tales, the author soon became aware of the limitations of the *ku-wen* style for fiction writing and proceeded to produce a vernacular novel of 1,000,000 words, the *Hsing shih yin yuan* ("Tale of a Conjugal Union to Rouse the World"). This long story of a shrew and her henpecked husband was told without any suggestion of a solution to the problems of unhappy marriages. Unsure of the reaction of his colleagues to his use of the vernacular as a literary medium, P'u Sung-ling had this longest Chinese novel of the old school published under a pseudonym.

Wu Ching-tzu satirized the 18th-century literati in a realistic masterpiece *Ju-lin wai-shih* (*c.* 1750; "Informal History of the Literati"), 55 chapters loosely strung together in the manner of a picaresque romance. Unlike P'u Sung-ling, whom he far surpassed in both narration and characterization, he adopted the vernacular as his sole medium for fiction writing.

Better known and more widely read was Ts'ao Chan's *Hung-lou meng (Dream of the Red Chamber), a* novel of a love triangle and the fall of a great family, also written in the vernacular and the first outstanding piece of Chinese fiction with a tragic ending. Because its lengthy descriptions of poetry contests, which interrupt the narrative, may seem tiresome especially to non-Chinese readers, they have been largely deleted in Western translations. Nevertheless, some Western critics have considered it one of the world's finest novels.

Drama. In drama, the Ming tradition of *ch'uatz-ch'i* ("tales of marvels") was worthily continued by several leading poets of the conventional school, though as a whole their dramatic writings failed to appeal to the masses. Toward the end of the 18th century, folk dramas of numerous localities began to gain popularity, converging finally at the theatres of Peking and giving rise to what came to be designated as Peking drama—a composite product that has continued to delight large audiences in China.

Popularity of folk dramas

19th-century translations of Western literature. By the early 19th century, China could no longer ward off the West and, after the Opium War (1839–42), China's port cities were forcibly opened to increased foreign contacts. In due course, many Western works on diverse subjects were translated into Chinese. The quality of some of these was so outstanding that they deserve a place in the history of Chinese literature. One distinguished translator was Yen Fu, who had studied in Great Britain and whose renderings of Western philosophical works into classical Chinese were acclaimed as worthy of comparison, in literary merit, with the Chou philosophers. Another great translator was Lin Shu, who, knowing no foreign language himself but depending on oral interpreters, made available to Chinese readers more than 170 Western novels, translated into the literary style of the Han historian Ssu-ma Ch'ien.

19th-century native prose and poetry. Meanwhile, writers of native fiction, especially in central and southern China, began to be seriously influenced by Western models. Using the vernacular and mostly following the picaresque romance structure of the "Informal History of the Literati," they wrote fiction usually intended for serial publication and satirizing Chinese society and culture. One of these writers was Liu E, whose *Lao Ts'an yu chi* (1903; "Travel Records of Lao Ts'an"), a fictional account of contemporary life, pointed to the problems confronting the tottering Ch'ing dynasty.

19th-century developments in poetry

Poetry, long stagnant, at last began to free itself from the shackles of traditionalism. The most prominent poet, Huang Tsun-hsien, inspired by folk songs and foreign travel, tried to write poetry in the spoken language and experimented with new themes, new diction, and new rhythm. His young friend Liang Ch'i-ch'ao not only fervently supported Huang and his associates in what they called "the revolution in Chinese poetry," but also ventured forth in new directions in prose. Liang's periodical publications, especially, exerted an extensive influence on the Chinese people in the early years of the 20th century. Fusing all the unique and attractive features of the various schools of prose writing of the past into a new compound, Liang achieved a vibrant and widely imitated style of his own, distinguished by several characteristics: flexibility in sentence structure so that new terms, transliterations of foreign words and phrases, and even colloquial expressions could be accommodated; a natural liveliness; a touch of infectious emotionalism, which the majority of his readers enjoyed. Although he was too cautious to use the vernacular, except in fiction and plays, he did attempt to approximate the living speech of the people, as Huang Tsun-hsien had done in poetry.

As part of a Westernization movement, the competitive literary examination system, which had been directly responsible for excessive conservatism and conventionality in thought as well as in literature, was abolished in 1905.

Modern Chinese literature: **1916** to the present. *1916–36.* The real literary revolution, however, did not get under way until after the establishment of the republic in 1911, when the monarchical institution of more than three millennia had ended.

Role of Hu Shih

The Chinese literary revolution, the aim of which was to bring about a thoroughgoing modernization of all aspects of Chinese literature, was touched off in 1916 by Hu Shih, who had studied in the United States. While in the U.S., he had come to advocate in private discussions with his Chinese fellow students the urgent need for a new national literature for China, to be written in the living language of the people. Hu Shih was well prepared to defend his proposal; for as early as 1906, when he was still only a preparatory school student in Shanghai, he had founded a quarterly magazine in which the vernacular was used and had published in it popular essays and an unfinished Utopian novel. Ever since his arrival in the United States in 1910, and especially since his decision in 1912 to change his field of specialization from agriculture to philosophy, he had become convinced that China's literary reform had to be consciously accelerated.

It was not until January 1917, however, that his proposals were published in Shanghai in the influential monthly *Hsin ch'ing-nien* ("New Youth") edited by Ch'en Tu-hsiu. Ch'en gave Hu Shih immediate support and declared his arguments to be irrefutable. Ironically, Hu Shih's article, "Some Tentative Suggestions for the Reform of Chinese Literature," was written in simple classical Chinese. As he himself observed later, he "was too much tempered by the moderation of a historian to be successful in revolutionary activities." His caution and willingness to listen to the opposition were effectively counterbalanced, however, by the activism of Ch'en Tu-hsiu, who also published an article, "On the Literary Revolution" (February 1917), in which he hoisted the banner of rebellion against the aristocratic, archaic classical, and regressive literary traditions of the old China.

Hu Shih's history of Chinese philosophy

In 1919 the literary revolution was given new dignity with the publication of the first volume of Hu Shih's *Chung-kuo che-hsüeh shih ta-kang (An Outline of the History of Chinese Philosophy,* 1919), written entirely in the vernacular, which had never been used before in a scholarly work. Open-minded scholars of the older generation, such as Liang Ch'i-ch'ao, soon followed suit; although Lin Shu and a few other conservatives remained resolutely opposed to the innovation. The May Fourth student movement of the same year, in which thousands of students demonstrated to protest against the weak position taken by China's delegates at the Versailles Peace Conference, which concluded World War I, spread the use of the vernacular to the whole nation. Student appeals to the unlettered masses of necessity used the language the people could understand. Many new periodicals and daily papers adopted the vernacular exclusively, and there were few creative writers of the younger generation who did not follow the new trend. Finally, in 1921, the vernacular was given official sanction by the government, being called *kuo yii,* or "national language," until the Communist regime, when the term was replaced by *p'u-t'ung-hua,* "the common language."

Meanwhile. the foundations of the literary revolution were strengthened by a whole crop of writers of varied background and training. Among the more experienced and dedicated of these were the brothers Lu Hsun (pseudonym of Chou Sbu-jen) and Chou Tso-jen, who were associated with Ch'en Tu-hsiu and Hu Shih on the monthly *Hsin ch'ing-nien.* The brothers had been brought up in the orthodox classical tradition, enriched by their acquaintance with Western literature. Lu Hsün was known especially for stories and familiar essays in which he exposed and ridiculed social foibles. In "The True Story of Ah Q" (Chinese, 1921; Eng. trans. 1926), the hero is the epitome of the ultraconservative forces in Chinese society—repugnant, obsolete, and yet so pitiable that he elicits sympathy rather than disdain and censure. "A Madman's Diary" (May, 1918) presents the jottings of a man diagnosed as insane, who, in his inspired moments, gives proof that he has not quite lost his mind; he suspects that the world is mad and he alone sober and sane. The aim of Lu Hsün's essays and epigrams is similar to that of his stories. Hurling invective at his opponents, he subjects traditional usages and customs to pitiless scrutiny and devaluation. His younger brother, Chou Tso-jen, a man of entirely different temperament, was not more conservative than Lu Hsun but a milder and gentler critic, with manners and refinements reminiscent of the best traditions of the Six Dynasties period.

The works of Lu Hsun

A new drama, modelled after Western prototypes and first introduced toward the end of the 19th century as an experiment, took firm root and flourished side by side with traditional drama of the operatic variety without displacing it. Although the achievements in poetry were not as plentiful as in fiction, new and bold experiments were continually carried on. (S.Y.C.)

1937 to the present. From 1937 to 1945 all writers, whether leftist, rightist, or nonpartisan, were in one way or another engaged in a national struggle against Japanese aggression. They all stressed the importance of literature as an effective way of boosting the morale of the people and thus helping to win the war against the Japanese. It was the leftists, however, who played the predominant role in the field of literary theory and production. In May 1942 the didactic and utilitarian function of literature was redefined by Mao Tse-tung in his "Tsai Yen-an wen-i tso-t'an-hui shang-te chiang-hua" ("Talks at the Yenan Forum on Art and Literature"). In his opinion, literature should be intended for the people—that is, workers, peasants, soldiers, and urban petty bourgeoisie. While stressing the need for popularizing literature, he attached great importance to elevating the people's level of literary appreciation. He advocated "unity of the revolutionary political content and the highest possible degree of perfection in artistic form." Most important of all, he believed that literature should eulogize "the proletariat,

Mao Tse-tung's dictum on literature

the Communist Party, New Democracy, and socialism." This kind of proletarian utilitarianism was later identified with Socialist Realism and became the guiding literary principle for Communist writers.

Under the influence of Mao's theory, countless literary works were produced in the Communist regions from 1942 to the founding of the People's Republic of China in 1949. Whatever literary form they took, they all reflected Communist policies and activities during their war against the Japanese as well as against the Chinese Nationalists. In 1951, when China's relations with the Soviet Union were still intimate and friendly, Stalin prizes for literature were awarded to three works: *T'ai-yang chao tsai Sang-kan-ho shang (The Sun Shines Over the Sang-kan River),* a novel by Ting Ling; *Pao-feng tsou-yii (The Hurricane),* a novel by Chou Li-Po; and *Pai-mao nü (The White-Haired Girl),* a musical drama by Ho Ching-chih and Ting Yi. Among the writers of the younger generation, Chao Shu-li was perhaps the most popular and distinguished. According to Communist literary critics, his works of fiction, such as *Hsiao Erh-hei chieh hun (Hsiao Erh-hei's Marriage)* and *Li Yu-ts'ai pan-hua (Rhymes of Li Yu-ts'ai),* adequately met Mao's criterion of unity of political content and artistic form.

During the 1950s, the Chinese underwent a series of ideological remolding campaigns, all of which profoundly affected literary circles. Many influential writers, such as Ting Ling, Ai Ch'ing, Hu Feng, and Feng Hsiieh-feng, were subjected to harsh criticism and denunciation because of their nonconformity with Mao's theory or the Communist Party line. During the 1956–57 movement for "letting a hundred flowers blossom and letting a hundred schools of thought contend," writers enjoyed the fullest freedom of expression, but before long that freedom, along with the movement, vanished like a dream. Socialist Realism, which remained the guiding literary principle, required that a literary work be both true to life and ideologically correct. It was not only a high standard, hard to attain, but also a principle open to various interpretations under various circumstances.

In spite of all difficulties, however, literary production continually increased from the establishment of the People's Republic until the Great Proletarian Cultural Revolution broke out in 1966. The famous authors of the older generation, however, except for a passionate few like Lao She, chose to be silent; it was the well-indoctrinated younger writers who became increasingly productive. Whatever they wrote was either directly or indirectly tied to politics, their themes changing as the major policies and objectives of the government changed. Thus, after 1949, they successively dealt with such events as the land reform, the Korean War, the suppression of counterrevolutionaries, the so-called three-anti and five-anti drives against illegal practices, the Hundred Flowers Campaign, the anti-rightist drive, and the Great Leap Forward.

Revision and re-editing of traditional works

While some authors were energetically creating works of their own, others directed their efforts toward revising and re-editing traditional masterpieces in conformity with Communist literary theory. The entire literary heritage of China was to be critically re-evaluated and reinterpreted so as to fit into the Communist ideological pattern. As a result, many traditional novels, short stories, operas, and songs appeared in newly revised versions. Some of the revised operas, such as *Liang Shan-po and Chu Ying-t'ai* and *The Story of the White Snake,* were considered great successes in terms of their readability and popular appeal. Traditional forms, such as the novel and the opera, were designated national literary forms, to be preserved and imitated. Many creative works did, in fact, appear in such forms, existing without incongruity side by side with Western types of literary works.

During the Cultural Revolution of 1966–69, practically all creative activities in literature came to a standstill, and countless writers and literary critics became victims of the Red Guards' harassment and denunciation. The literary scene in mainland China was indeed bleak. In the early 1970s, there was an indication of a swing toward relative relaxation and reinvigorated literary productivity. (T.-y.L.)

II. Korean literature

Although Korea has had its own language for several thousand years, the invention of a writing system came comparatively late. As a result, early literary activity was in Chinese characters, either in a syllabary (a set of written characters each one of which is used to represent a syllable) called *idu,* which used the characters to transcribe the sounds of Korean, or in the traditional manner of classical Chinese. Whether or not works written in classical Chinese should be considered Korean literature has been much debated, the situation being complicated by the fact that classical Chinese was extensively used in composition even into the early years of the 20th century, in spite of the creation and promulgation of a national alphabet, *han'gŭl,* in the mid-15th century.

The three categories of Korean literature

In general, then, literature written in Korea falls into three categories: works written in *han'gŭl*, those written in *idu,* and those written in classical Chinese. Although the last category may not be strictly Korean literature, its long tradition of creative development in Korean literary history makes it worthy of consideration.

TRADITIONAL FORMS AND GENRES

Since the study of Korean literature is relatively recent, a complete classification of its genres cannot be given. The most general categories, however, are lyric poems, prose, oral literature, and literature in classical Chinese.

Lyric poems. Lyric poems originally were meant to be sung rather than read, as European lyric poems originally were. The oldest are the *hyangga,* poems composed in *idu* between AD 600 and 1000, during the late middle period of the Silla dynasty and in the early period of the Koryŏ dynasty. The poems were written in four, eight, or 10 lines, the 10-line form being the most characteristic. The poets were usually of the Silla nobility––either Buddhist priests or chivalrous *hwarang* youth, leaders who were trained in civil and military virtues and who consciously cultivated sensitivity. The poems generally deal with Buddhist themes. Twenty-five *hyangga* are extant, 14 in *Samguk yusa* (1285; "Memorabilia of the Three Kingdoms") and 11 in *Kyungŏ chdn* (1075; "Life of the Poet Priest Kyunyŏ"). *Pyŏlgok*, also called *changga* or *ydyo,* was a verse form written during the middle and late Koryŏ dynasty (AD 935–1392). Characteristically, it is a group of several six-line stanzas. The poets were anonymous. So many examples of *pyŏlgok* have survived that this may be called the representative poetic form of the Koryb dynasty. The subject matter is generally love, and the style of expression is quite openly passionate.

Sijo and *kasa*

Sijo is historically the longest enduring and the most widely representative form of Korean poetry. It became an established form at the end of the Koryŏ dynasty and was written throughout the Yi dynasty. *Sijo* was still being written in the second half of the 20th century. It is usually defined as a three-line poem because the meaning of the poem falls into three distinct units. Each of the three lines generally has 14 to 16 syllables, and the total number of syllables never exceeds 45. Each line is in turn divided into smaller groups of varying lengths, the first group in the third line always having three syllables. *Sijo* frequently deals with Confucian ethical values, but there are also many poems about nature and love. The principal writers of *sijo* were members of the Confucian upper class (*yangban*) and *kisaeng* girls (like the Greek *hetairai,* highly cultivated courtesans). In the latter part of the Yi dynasty, a longer form, called *sasŏl sijo* ("narrative *sijo*"), evolved. The writers of this form, when they did not remain anonymous, were mainly common people; hence, the subject matter included more down-to-earth topics, such as trade, immorality, and corruption, as well as the traditional topic of love between the sexes. In addition, *sasŏl sijo* frequently employed slang, vulgar language, and onomatopoeia (the use of words the sound of which suggests the sense).

Another form that developed at approximately the same time as *sijo* is *kasa.* In its formative stage, *kasa* borrowed the form of the Chinese *t'zu* (lyric poetry) or fu (prose poetry). *Kasa* tends to be much longer than other forms of Korean poetry and is usually written in balanced couplets. Either line of a couplet is divided into two groups, the first having three or four syllables and the second having four syllables. The history of *kasa* is clearly divided into two periods, the division being marked by the Japanese invasion of 1592–97. During the earlier period the poem was generally about 100 lines long and dealt with such subjects as female beauty, war, and seclusion. The writers were usually *yangban*. During the later period the poem tended to be longer and to concern itself with moral instruction, travel accounts, banishment, and the writer's personal misfortunes. The later writers were usually commoners.

Akchang poems

Immediately after the founding of the Yi dynasty at the end of the 14th century and the establishment of the new capital in Seoul, a small group of poetic songs called *akchang* (literally, "words for songs") were written to celebrate the beginning of the new dynasty. In its earliest examples the form of *akchang* was comparatively free, borrowing its style from early Chinese classical poetry. Whereas the early *akchang* are generally short, the later *Yongbi ŏch'ŏn-ga* (a compilation of *akchang*) consists of *125* cantos, and *Worin ch'ŏn'-gangji-gok* seems to have originally consisted of about 500 cantos.

Prose. Korean prose literature can be divided into three groups: narratives, fiction, and essays. Narratives include myths, legends, and folktales found in the written records. The principal sources of these narratives are the two great historical records compiled during the Koryb dynasty: *Sarnguk sagi* and *Samguk yusa*. The most important myths are those concerning the sun and the moon, the founding of Korea by Tan'gun, and the lives of the ancient kings. The legends touch on place and personal names and natural phenomena. The folktales include stories about marriage, birth, ethical problems, prophecies, heroism, dreams, and animals. Because the writer of *Samguk yusa* was a Buddhist priest, some stories about the history of Buddhist temples are also recorded. The compilations made in the Koryb period preserved the stories of prehistoric times, of the Three Kingdoms, and of the Silla dynasty and have remained the basic sources for such material, but later compilations, such as *Taedong yasŭng* and *P'aerim*, were made during the Yi dynasty. Such compilations served as a major source of materials for later Yi dynasty fiction.

Korean fiction can be classified in various ways. First, there is fiction written in Chinese and that written in Korean. Second, there are the short works of one volume, "medium" works of about ten volumes, and long works of more than ten volumes. And third, there are works of *yangban* writers and those of common writers. In respect to the last classification, however, there is also a group of fictional works that combine the viewpoints of the *yangban* and the commoner. Most of this fiction was based on the narratives mentioned above, the author adding incidents and characters of his own to the original story. It is not possible to assign definite dates or authors to most of these works. The stories are generally didactic, emphasizing correct moral conduct, and almost without exception they have happy endings. Another general characteristic is that the novels written by *yangban* authors are set in China, whereas those written by commoners are set in Korea.

There has been very little study of Korean essays, but some diaries, travel accounts, letters, and other writings are of definite literary value because of the expression of emotions and experiences in them.

Oral literature. Oral literature includes all those texts that were orally transmitted from generation to generation until the invention of *han'gŭl—i.e.*, ballads, legends, mask plays, puppet-show texts, and *p'ansori* ("story singing") texts.

In spite of the highly developed literary activity from early in Korean history, many song lyrics were never recorded until the invention of *han'giil*. These orally transmitted texts are categorized as ballads and are classified according to singer (male or female), subject matter (prayer, labour, leisure), and singing style (capital area, western, and southern). The songs of many living informants have yet to be recorded, and the gradual decline of traditional singing makes the collection of these texts particularly urgent.

Legends include all those folk stories handed down orally and not recorded in any of the written records. These legends were for long the principal form of literary entertainment enjoyed by the common people. They deal, as do the legends of every major culture, with personified animals, elaborate tricks, the participation of the gods in human affairs, and the origin of the universe.

Mask plays

The mask plays are found in Hahoe, Chinju, Tan'gyŏng, Kimhae, and Tongnae in North and South Kyŏngsang provinces; Yangju in Kybnggi Province; Pongsan in Hwanghae Province; and Pukch'ong in south Hamgybng Province. The most representative plays are *Sandae Kuk* of Yangju, *Pyŏlsin kut* of Hahoe, and *Okwangdae nori* (five-actor play) of Chinju. Although the origin of these plays is uncertain, they are generally presumed to have developed from primitive communal ceremonies. Gradually, the ceremonial aspect of the plays disappeared. and their dramatic and comic possibilities were exploited. The dialogue was somewhat flexible, the actors being free to indulge in more or less satire as the occasion demanded. The plays were not performed on a stage, and there were no precise limits as to the space or time in which the performances took place. The audience also traditionally responded vocally to the play as well as passively watching it. The organization of the mask plays —through repetition and variety—achieves a remarkable effect of dramatic unity.

Only two puppet-show texts are extant, *Kkoktukaksi nori* (also known as *Pak Ch'ŏm-ji*) and *Man Sŏk-chung nori*. Both titles are derived from names of characters in the plays. No theory as to the origin and development of these plays has been formulated. The plots of the puppet plays, like those of the mask plays, are full of satiric social criticism. The characters—Pak Ch'ŏm-ji, governor of P'yŏngam, Kkoktukaksi, Buddhist monk, and Hong Tong-ji—dance and sing, enacting familiar tales that expose the malfeasance of the ruling classes.

The final type of folk literature is found in the texts of *p'ansori*, "story singing" of the Yi dynasty, which is often compared to medieval European minstrelsy. These texts were first recorded as a type of poetry, but the written forms were later expanded to be *p'ansori* novels, widely read among the common people. This transformation from poem to novel was easily accomplished, since *p'ansori* is always narrative. Originally the entire *p'ansori* performance repertoire consisted of 12 *madang* ("titles"). Although all 12 remain in novel form, only five of them are sung today. The texts evolved gradually from the legends, which provided their sources and were altered and expanded as they were passed from one performer to another.

Literature in classical Chinese. With the declining importance of Chinese characters in Korean life and the continuing promotion of *han'gŭl* by the government, Korean literature in Chinese became increasingly neglected. Many excellent works that compare very favourably with the classics of China lie unread in libraries and book collections. Fortunately, however, a considerable number of these works in Chinese were being translated into Korean in the second half of the the 20th century with the financial assistance of the government. As they are translated and more widely known, they will surely achieve the recognition due them.

HISTORICAL DEVELOPMENT

Ancient Korean songs

The earliest literature: before 57 BC. Judging from historical records and folklore, it seems that the earliest literary activity was mainly poetic. Ancient Korean songs seem to have been connected with the religious and farm life of the people. Their form was collective rather than individualistic, and they give evidence of the outstanding creativity of that age. From existing folk songs and folk dances—for example, the songs "Kuai-che-na-ch'ing-ch'ing-na-ne"—it is known that the earliest poetic songs found their theme in the praise of nature and richly expressed man's feelings concerning his environment. Though the actual function of these songs is unknown, the Chinese translations lead one to believe that they were used for collective dancing.

Three songs have been handed down in Chinese translation, "Ke-chi Ka" (or "Yongsin'gun-ga," "Song for Welcoming the Gods," found in *Sarnguk yusa*), "Huang-chou Ka" ("Song of the Yellow Birds," in *Satnguk* yusa), and "Kong-mu To-ha Ka" (or "Konghu-in," in *Haedong yoksa*). Although 'Ke-chi Ka" is related to the myth of the founding of the Karak tribe (one of the early tribes of Korea), the poetic intent of the song originally was to express sexual desire in an intense and simple way. It was first sung by women to tempt men but gradually became a kind of magic incantation, finally developing into the myth of the founding of Karak. "Huang-chou Ka" seems to be a fragment of a love song that was sung at religious ceremonies by bachelors who were trying to convince young girls to marry them and who lamented their lack of success. "Kong-mu To-ha Ka" deals with a mythological story and was probably a widely sung lyric poem. The hero, Paek-su Kwang-bu, is the equivalent of the Greek god Bacchus. and his wife is a river nymph. Both the hero and the heroine, being gods, were able to transcend life and death and to enter the water world freely. Even at her husband's death, the nymph plays the harp and sings an elegy. The story also includes human characters, such as the sailor, Kwangni chango; his wife, YB-ok; and her friend, Yo-yong.

Possible beginning of secular poetry

Another work, "Tosol-ga," is mentioned in *Samguk sagi* as having been written in the time of King Yuri of the Silla dynasty, but the poem itself has not survived. This poem is cited by historians as the beginning of secular poetry *(ka-ak,* literally, "songs and music"), and there may be some truth in this claim, as the reign of King Yuri is known to have been a time of political and social change. It was probably during this time that the roles of the tribal chiefs and the priests became clearly separated, the chiefs being divested of their. former religious powers. Hence, the artistic efforts of the people might have become more secular in outlook. It is clear, however, that the poem was free of religious elements and dealt with human emotions in a lyrical manner. It apparently was not, as the three earlier poems were, concerned with a narrative. In this respect, also, it is unusual.

Literature of the Three Kingdoms: 57 BC–AD **668.** In contrast to the literature of the earliest ages, which is characterized by collective artistic activity, that of later ages shows the effects of political, economic, and cultural changes as the peninsula increased in wealth and widened its contacts with other areas. Influences from the mainland on the traditional religious views and life patterns of the people brought about new possibilities for development from a primitive to a creative society. The introduction of Buddhism and Chinese characters to the Three Kingdoms made their literature rich and changed their world view greatly. In consequence, their artistic activity advanced far beyond collective singing and dancing to the direct expression of individual feelings of joy and anger, grief and pleasure. The heroes of this literature were human beings with individual personalities, in contrast to the more monumental tribal heroes of earlier times.

The three kingdoms of this period were Koguryŏ, in the north; Paekche, in the southwest; and Silla, in the southeast. The writers of Koguryci, the geographical location of which provided close contact with the Chinese mainland, seem to have retained something of the original pioneer spirit from the times when Koreans came from the northern regions and settled on the peninsula, and their poems tended to be heroic tales in epic form. According to *Samguk sagi,* the representative epic of Koguryŏ was the myth of the founding of Korea by King Tongmybng. The stories of Ondal, King Mich'bn, Prince Hodong, the crown prince Yuri, and others that had their origin in Koguryb are still used today as the bases for plays and movies.

In contrast to that of Koguryci, the literature of Paekche and Silla tended to be lyrical, perhaps because of the milder climate and easier life in the south. Very little literature from Paekche remains today, but the legends and songs contained in the *Samguk sagi* give a hint of the original extent and richness of Paekche's literature, which perhaps equalled that of Silla; for example, "Chŏngŭp-sa," a song passed down from Paekche through the Koryŏ and Yi dynasties, is still performed and admired in the 20th century. In addition, the considerable cultural influence of Paekche on Japan gives evidence of its high level of artistic achievement.

Silla led the other two kingdoms both politically (as proved by its subsequent unification of Korea) and artistically, in spite of the fact that it was farthest removed from contact with Chinese culture. The geographical and cultural distance from China, however, seems to have been an advantage, since the culture of Silla was able to create a true synthesis of native and foreign elements.

Absorption of Chinese culture

Literature of **Unified Silla: 668–935.** After the mid-7th century Silla absorbed Koguryŏ and Paekche and created a stable political system covering the whole nation in the 8th century. During the Silla dynasty many students were sent at government expense to study in T'ang China. The consequent absorption of Chinese culture and the flourishing of Korean Buddhism both contributed to the remarkable artistic flowering of Silla. In particular, the spiritual life of the Silla nobility—the priests and the chivalrous *hwarang* youth—was dominated by Buddhism, which thus became the driving force behind virtually all artistic activity.

The *hyangga* was the crown of Silla's literary achievement. In these poems the nobility expressed a lofty outlook on life in stories based on their own style of living. Their form was marked by a flexible, delicate, and mystical style. The influence of Buddhism can be seen in the tendency to base the poems on a philosophy of seeking a new, ideal society. The Buddhistic emphasis on future life makes it natural that the subject matter of *hyangga* should be otherworldly. Ancient historians described this flowering of Silla dynasty literature by saying that the streets were never empty of music and that all the houses were filled with song and dance throughout the night.

The work of collecting *hyangga* was either encouraged or actually done by two of the most important *hyangga* writers, the Buddhist high priest Nung-jun and Tae-guhwasang. The result of their efforts was the collection of 25 *hyangga* in *Samdaemok*. The most famous poems are "Che-mang-mae-ga" ("Religious Song to a Deceased Sister"), by the priest Wiilmybng, and "Ch'an-gi-p'a-rangga" ("Song in Praise of Ki-p'a-rang"), by the priest Ch'ungdam. Both of these men were Buddhist priests. All of the poems in the collection were either by Buddhist priests or *hwarang*.

Since the Korean language had no written form during this period of history, the Silla scholars invented a syllabary. called *idu,* that used the Chinese characters to transcribe the sounds of Korean, and it was in this form that the *hyangga* were recorded. At this time Chinese characters were also extensively used in the classical Chinese manner, and much epic literature was written in Chinese. This epic literature, most of which consisted of narratives relating to the building of Buddhist temples, was no less important than the *hyangga* in the development of Buddhist literature. The large number of narratives that appear in *Satnguk yusa* are evidence of the superior imagination of the Silla writers. In addition to these narratives, the epic literature contains outstanding stories of chivalrous deeds and sexual exploits by *hwarang* youth at the time of the founding of Unified Silla.

Influenced by Chinese novels, many narratives of the Silla dynasty became the basis of the classical novel that formed the backbone of later Yi dynasty literature.

Literature of Koryd: **935–1392.** *Hyangga* was passed from Silla to Koryŏ, but it soon became so conventionalized that writers lost interest in it. The last important poet to write *hyangga* was the 10th-century high priest Kyunyb, of whose work 11 poems remain. Even these poems are remarkable more for their refined technique than for their expressive power. By this time *hyangga* had lost its function as an expression of the poet's desire to experience the Buddhist paradise and became a formalized hymn. For the most part Koryŏ intellectuals were not interested in the too highly conventionalized *hyangga* and

A new poetic form, the *pyiilgok*

created a new poetic form called *pydlgok* (special or different song). This eventually became the typical literary form of Koryŏ. *Pyŏlgok* was intended for large-scale stage performances on festive occasions, such as P'algwan-hoe (feasts of the local gods) and Yondŭng-hoe (a festival at which people burned candles to Buddha as prayers for blessing). In contrast to the poetic forms of Silla, which had only one stanza, *pydlgok* had several stanzas. Gradually evolving during middle Koryii, *pyŏlgok* reached its definitive form in later Koryŏ.

The Koryŏ dynasty was a time of social instability. Internal and external crises abounded due to the arrogance and oppressiveness of the courtiers and the army, as well as the savage invasions of the Mongols. Under such conditions the viewpoint of the *pyiilgok* writers tended to be introspective and hedonistic. Consequently, the new intellectuals, who had adopted Chinese dualistic thought as their philosophy, were dissatisfied with *pyiilgok* and sought a different form of poetic expression. This was the genesis of *sijo,* which was to become the principal form of the Yi dynasty. At first, *sijo* was rather stilted, the result of the mixture of a classical Chinese poetic style with Korean inflectional endings. Gradually, the stiffness disappeared, and *sijo* developed its lyricism and its tender, elegant, and simple style in Korean.

The Koryii dynasty also saw the development of epic literature. The origin of this literature was in the narratives of the Silla dynasty. These epics include myths, legends, folklore, and the history of Buddhism and Buddhist temples, and they are preserved in the historical records of the time. They were modified, however, to a considerable degree by the personal desires of the scribes and the tradition of the ages. One notable class of epics is that in which the hero is a personified inanimate object, as in *Kuk-sun-jŏn* ("Tale of Pure Malt") and *Kong-bang-jŏn* ("Tale of the Square-Holed Coin"), both by Im Ch'un. The masterpieces of epic literature are *Tongmyŏng wang* ("King Tongmyŏng"), by Yi Kyu-bo, and *Chewang ŭn'gi* (Imperial records in poetic form), by Yi Sŭng-hyu. *Tongmyŏng wang* records the heroic deeds of the founder of the Koguryŏ kingdom in an extravagant style. It is written in quatrains with five Chinese characters in each line. *Che-wang-ŭngi* gives an account of Korean history in quatrains with either five or seven characters to a line.

Literature of the early **Chosŏn** period: **1392–1598.** The literature of the Yi dynasty falls naturally into two periods, with the end of the Japanese invasion (1597) serving as a convenient dividing line. The early period is notable for its poetry; the later, for its prose. Inheriting the tradition of Silla and Koryii, the writers of the early Yi dynasty raised Korean literature to new heights.

Invention of Korean alphabet

The early Yi dynasty also marks the initiation of a new era in Korean literary history with the invention of the Korean alphabet, *han'gŭl,* in 1446, soon after the founding of the Yi dynasty. This important event, which finally enabled Korean writers to record works in their native language, occurred at approximately the same time that European writers of the Renaissance were breaking away from Latin and beginning to write in their national languages. Similarly, Korean writers began to renounce their cultural dependence on Chinese culture and the Chinese language.

The extraordinary King Sejong the Great not only invented *han'gŭl,* or *hunmin-jŏngŭm* as it is also called, but he also had his scholars compile *Yongbi ŏch'ŏn-ga;* and he himself compiled *Wdrin ch'ŏn'gangji-gok*, two long poetic works, in *han'gŭl* to demonstrate its practicality as a means of literary expression. The latter poem was written in praise of the blessings of Buddha. These two poems were the prototype of a new form called *akchang.* The form was adopted by the nobility, who wrote many *akchang* with considerable stylistic variations. It was also adopted by the more conservative Confucianists, who valued the Chinese classics very highly and still sang the *Pyŏlgok-ch'e* of the Koryŏ dynasty. The popularity of the *akchang*, however, was always limited, and it was soon eclipsed by the most important forms of the Yi dynasty—*sijo* and *kasa.*

These forms owe their popularity to two factors. First, their style of expression was rich and at the same time natural, for which reason they were widely appreciated by readers. Second, they were popular with writers because the forms together provided ideal outlets for the two sides of the Confucian temper: the brief and simple *sijo* was a perfect vehicle for intense lyrical expression, whereas the longer *kasa* gave writers an opportunity to expound at greater length on the more practical aspects of Confucian thought.

The expressive content of *sijo* ranges from the idealistic union of man and nature (often coupled with the poet's pride in his poverty) to the praise of sovereigns by loyal subjects (in which poems an analogy is drawn between fidelity and romantic love) to the deeper exploration of human problems. Writers include Maeng Sa-sŏng, Yi Hyŏn-bo, Yi Hwang, and Yi I, who extolled the virtues of the life that can enjoy the Tao while being impoverished. Representative poets of *kasa* include Chdng Kug-in and Pak In-no.

Even after the invention of *han'gŭl* and of literary forms that depended on the use of *han'gŭl*, most prose continued to be written in Chinese. The important exceptions, such as *Kŭmo sinwha* by Kim Si-sup, *P'ae-gwan chap-ki* by Oh Silk-kwiin, and *Yong-jae Ch'ŏng-hwa* by Song Hyon, are of great literary value and influenced the growth and development of *han'gŭl* prose literature in the later Yi dynasty.

Shift from poetry to prose

Literature of the later **Chosŏn** period: **1598–1894.** As has been mentioned, after the Japanese invasion the emphasis in literature shifted from poetry to prose. This change represents a significant step in the evolution toward modern literature. It also reflects a basic change in the philosophical outlook of Korean society. The Yi dynasty had suffered from the rigid formalism of the Confucianists, whose ideals were based especially on the principles of the 12th-century Chinese philosopher Chu Hsi. Confucianism was gradually replaced by practical idealism. Unlike the older philosophy, which emphasized unquestioning acceptance of authoritative principles, practical idealism was based on reason and the scientific spirit of criticism. The new literary form, prose, arose in answer to the need for a vehicle by which to best express the new philosophy.

Practical idealism gave impetus to literary activity and awakened the self-consciousness of the common people. Whereas the poetry in the early period had been dominated by the nobility, in the later period it was dominated by the common people. Women also were admitted into the literary world as the principal audience for traditional novels, a further step toward modernization. The later active compilation of the anthology and the comprehensive compilation of the narratives reveal the awakening interest in rediscovering and reappraising the past.

Prose. The origins of the traditional novel and of *p'aegwan munhak* (literary tales recorded by official scribes) can be traced to the narrative literature derived from folklore. *Kŭmo sinwha*, influenced both by *p'aeg-wan munhak* and by classical Chinese novels, was the first romantic novel written in Korea. The second important novel was *Hong Kiltong-jdn* ("The Life of Hong Kiltong"), by Hŏ Kyun. Building on the style of these two works, Kim Man-jung wrote two famous novels, *Ku-un-mong* (1687–88, *Cloud Dream of the Nine*) and *Sassi namjŏng-gi* (1689–92, "Story of Lady Sa"), a satire against the institution of concubines. The most popular novels of the 18th century were all anonymous: *Ch'unhyang-jŏn* (*Fragrance of Spring*), *Simch'ŏng-jŏn* ("Tale of a Faithful Daughter of a Blindman"), *Chang-hwa-hongnyŏn-jŏn* ("Tale of Rose Flower and Pink Lotus"), and *Hŭngbu-jŏn* ("Tale of Honest Hŭngbu"). The characters in these novels were common people, and the style in which they were written was simple and natural. Together with *Hong Kiltong-jŏn*, these works form a body of literature the stories of which have become deeply rooted in the consciousness of Korea. At the end of the 18th century, this kind of literature reached its culmination, and a new style emerged that combined the outlook of the nobility and of the common people.

Palace literature

During the Yi dynasty, palace literature written by court ladies also flourished. The most significant works of this genre include *Hanjungnok* (the memoirs of Hegybng-gung Hongssi [died 1815], who was the Lady Hong of Hegybng Palace, the mother of King Chongjo); *Kech'uk Ilgi* (the diary of King Sonjo's second queen, Inmok [died 1632], who, after the assassination of her son Prince Yongch'ang, was confined in the western palace), as recorded by one of the Queen's maids in a realistic and sensitive style; and *Inhyŏn Wang-hu-jŏn*, the tragic story of Queen Inhybn and a realistic work that squarely confronted the social problems of the times.

Korean prose of *han'gŭl*, produced in smaller quantity than that in Chinese characters, includes diaries, travel accounts, women's letters, and other miscellaneous writings. Since their contents are prosaic, these works are classified as prose, but they are written in verse form because their writers, being common people, were not fully experienced in the art of adapting the prosaic language.

Poetry. The later Yi dynasty also saw a great flowering of poetry, which, remarkably, was written primarily by common people.

The most gifted poet of the period was Yun Sbn-do, who achieved what is generally considered the definitive style in *p'ydng sijo* (standard *sijo),* which he raised to new heights of expressiveness. His works are characterized by freshness of diction and structural elegance.

Gradually, *sijo* was superseded in popularity by *sasŏl sijo,* a new narrative poetic genre produced by the common people, which can be regarded as a mixture of *kasa* and *sijo.* Though both were lyric poetry, *sasdl sijo* was more realistic, included more dialogue, and was structurally more complex. It boldly allegorized facts rather than ideas and attempted to express strong human emotions freely and to expose the naked self by means of allusion, witticism, and vulgarism.

Change in the *kasa* form

Pak In-no was the most important of the traditional *kasa* writers in the later Chosbn period, although he does not reach the stature of Yun Sbn-do. He merely succeeded in expressing Confucian doctrine in a sombre spirit. The desire to re-evaluate the past and to re-create the world of literature led to changes in the *kasa,* 2s exemplified in the anonymous works of the *nae-bang kasa (kasa* written by women in their secluded quarters) and in those of the common people. The writers of the *nae-bang kasa* were mainly women from the southern regions of Korea. They depicted their joy, anger, grief, pleasure, the etiquette of entertaining guests, religious rites, principles of being a wise mother and a good wife, and the like. The *kasa* of the common people, or *p'ydng-min kasa,* was marked by the same style as the *nae-bang kasa* and played a similar role in the literary activity of the general masses.

In addition to the *nae-bang,* the nobility produced two other new categories of *kasa: kihaettg kasa,* which were accounts of their travels, and *yubae kasa,* which were accounts of their life in exile. *Iltong-changyu-ga*, written by Kim In-gyom in 1763 upon his return from an official trip to Japan, and *Ydn-haeng-ga,* written by Hong Sunhak in 1866 upon his return from an official trip to Peking, are representative of the *kihaeng kasa* style. The three most widely known *yubae kasa* are *Pukkwan-gok* ("Song from the Banished North"), written by Song Chu-sbk, after following his grandfather Song Si-ybl to his exiled place of Togwbn, in south Hamgybng Province; *Manŏnsa* ("Song of Ten Thousand Words"), by An Cho-hwan, written while he was banished to the lonely isle of Ch'uja in the south; and *Pukch'ŏn-ga* ("Song from the Banished North"), written in 1853 by Kim Chin-hyŏng, depicting the experience of life in exile in Mybngch'on, north Hamgyŏng Province.

Oral literature. Another feature of the later Yi dynasty was the formation, by the common people, of *p'ansori* ("story singing") texts. *P'ansori* seems to have originated during the reign of Sukchong (1675–1720), when *kŭnwon sŏlhwa* (original stories, thought by some to have been the basis for the *p'ansori* texts) were first sung. Their style and form was fixed by the *kwang-dae,* or professional actor-singers, and a group of amateurs in Chŏlla and Ch'ungch'bng provinces. Originally, there were 12 *p'ansori.* Later, the master *p'ansori* writer Sin Chae-hyo (1812–84) revised and readjusted them into six fixed forms during the reign of King Kojong (1863–1907).

Mask plays and puppet shows

Finally, the mask play and the puppet show should be mentioned. The representative mask play is *Sandae To-gam kŭk.* Of unknown origin, it was usually performed on a stage in 12 scenes, or acts. The masked actors followed a script that presented a story in dialogue interspersed with dances and songs. As the puppets of the *Kkoktukaksi kiik* show were made of *pak* (a gourd, rhyming with the Korean name Pak), it was also referred to as *Pak Ch'ŏm-ji kŭk* ("Old Pak's play"). Through keen satire presented in a unique and distinguished style, the contents of the masked drama and the puppet show strongly reflect the environment and the feelings of the common people of the later Yi dynasty.

Transitional literature: 1894–1910. By the time of the 1894 reforms, enough social and intellectual change had occurred to suggest the beginnings of a division between traditional and modern literature. But, just as stubborn conservative reaction did not favour sudden changes in the political and social structure, it was unavoidable that literature, too, faced a long period of transition toward its modern transformation.

Schools were established by the educational ordinance of 1895, and the organization of learned societies and "enlightenment" movements followed soon after as cultural modernization began in earnest, limited though it was to modern education and what was called the "new-life" movement. Vernacular publications, the "Independent" and the "Imperial Press," along with the establishment of the Korean Language Institute and the scientific study, consolidation, and systematization of Korean grammar, also helped open the way for the modern literary movement.

The first literary forms to appear after the 1894 reforms were the *sinsosŏl* ("new novel") and the *ch'angga* ("song"). These transitional literary forms, lying midway between traditional and modern literature, were stimulated by the adaptation of foreign literary works and the rewriting of traditional novels in the vernacular. The *ch'angga,* which evolved from hymns sung at churches and schools in the 1890s, became very popular upon the publication of the "Aeguk ka" ("National Anthem"), by Yi Yong-u, and "Tongsim ka" ("The Mind of a Boy"), by Ti Chung-wbn, in an 1896 issue of the "Independent." The early *ch'angga* had a syllabic pattern reminiscent of the *kasa,* but later, under the influence of Japanese poetry, it made wide use of a typical Japanese pattern, particularly after the publication in 1904 of "Kybngbu ch'blto ka" ("Seoul–Pusan Railroad Song"), by Ch'oe Nam-sŏn. The content of the song was chiefly an explanation and praise of the Western thought being introduced during the modernization period.

The *sinsosŏl*

The *sinsosŏl,* which, like the adapted novel, was written by professional writers, was far better literature than the *ch'angga.* As its name implies, the *sinsosŏl* ("new novel") was a transitional literary form bridging the period from the decline of the traditional novel after the 1894 reforms to Yi Kwang-su (1892–1950) and the rise of the modern novel after 1910. Three distinctly traditional elements were inherited by the *sinsosŏl.* First was the basic moral stance of the traditional novel, the reproving of vice and the rewarding of virtue. Owing to the nature of the enlightenment period, the advocates of modernization, of course, were cast as the virtuous, while the wicked were invariably conservative. Second, the development of the plot was governed by coincidence, just as in the traditional novel, and events that lacked causality were illogically connected. Finally, there was no distinction between the dialogue and the accompanying narrative, which were fused into one expository structure. But the *sinsosŏl* also had its pioneering aspects that balanced off the traditional legacy. It was written wholly in prose, whereas a considerable part of the traditional novel had been in verse. The *sinsosŏl* tried to depict a plausible human existence with backgrounds and events that more closely resembled reality than was the case in the traditional novel, which always followed certain model stories with their established plot lines and stereotyped characterizations. And, in accord with the vernacular (*ŏnmun ilch'i*) movement, initiated after the reforms of 1894, the *sinsosŏl* adopted a colloquial style. Typical *sinsosŏl* writers and their works are: Yi In-jik, *Kwi-ui sŏng* (1907, "A Demon's Voice"); Yi Hae-jo, *Chayujong* (1910, "Liberty Bell"); and Ch'oe Ch'an-sik, *Ch'uwŏlsaek* (1912, "Autumn Moon"). In their novels these writers advocated modernization, a spirit of independence, contact with advanced countries, and study abroad; they also emphasized the wonder and diffusion of science and technology, the breakdown of conventions and superstition, and the pernicious internal conflicts of a society.

Modern literature: 1910 to the present. Even while the "modern" period is said to have begun around 1910 with the poetry of Ch'oe Nam-sbn and the novels of Yi Kwang-su, it was not until the late 1930s that Korea finally achieved a modern literature that was mature in international terms.

Ch'oe Nam-sŏn and Yi Kwang-su played leading roles in the literature of enlightenment that characterized the decade after 1910. Generally, the starting point of modern Korean poetry is taken to be the poem *Hae-egesŏ sonyŏn-ege* ("From the Sea to the Youth"), published in 1908 by Ch'oe Nam-sŏn. This poem was the first attempt to break away from traditional verse forms and from the *ch'angga*. Though rather loose in its structure, the poem is a valuable example of the ideas of equality and the consciousness of modernization that are typical of early modern Korean literature. Yi Kwang-su introduced Western literature and, as a literary pioneer, wrote "Chŏlmun kkum" (1915, "A Youthful Dream") and other short stories that were aimed at fostering social enlightenment. He further contributed to the development of the Korean novel with his long work *Mujŏng* (1918, "Heartless").

The Independence Movement of 1919, an inevitable explosion of the pent-up resentment of a people who had been deprived of their sovereignty, ended with the massacre and imprisonment of many Koreans by the Japanese. Young men, utilizing the only outlet left them by the Japanese authorities with which to express their youthful ardour, pooled their efforts in publishing literary magazines. In addition to the pace-setting magazine *Paekcho* ("White Tide"), which appeared in 1922, the period also saw the emergence of *Sŏgwang* ("Aurora") and *Ch'angjo* ("The Creation") in 1919, *Kaebyŏk* ("Genesis") and *P'aehŏ* ("The Ruin") in 1920, *Kŭmsŏng* ("Venus") and "Renaissance" in 1923, and *Chosŏn mundan* ("Korean Literary Circle") in 1924. The writers experimented with Romanticism, *décadentisme* (a literary movement that sought inspiration in the morbid and decadent), Symbolism, and Neoromanticism and introduced miscellaneous *fin de siècle* (the climate of sophistication, world-weariness, and fashionable despair that characterized much Western art and literature at the end of the 19th century) literary themes in search of individual expression for their emotions. Because this sudden inrush of eclectic foreign literary thought could not be quickly absorbed and, digested, the poetic harvest of the 1920s—consisting of some memorable poems written by Yi Sang-hwa and Kim So-wŏl—was meagre. The significant short stories published in the literary magazines of the 1920s demonstrate a growing interest in outward details, the use of such devices as letters, diaries, and dialogues, human psychology and a dramatization of daily life, and the use of properties to help forward the action. Outstanding short stories of the decade include "Paettaragi" (1920, "The Sailor's Song") and "Kamja" (1928, "The Potato"), by Kim Tong-in; "P'yobonsil-ui ch'ong kaeguri" (1920, "The Green Frog in the Specimen Room"), by Ybm Sang-sbp; and "Pinch'b" (1921, "The Poor Wife"), by Hybn Chin-gŏn.

Short stories of the 1920s

If the 1920s are to be viewed as a period of upheaval in Korean literature, the 1930s may be called a period of growth in which an early, developing modern literature gave way to a mature one. Early experimental elements began to disappear from the lyric works that appeared in the magazine *Simunlzak* ("Poetic Literature," founded in 1930), from the products of the Kuinhoe (The Nine-Man Group), organized in 1934, and also from the works of the Intellectual school that employed techniques of contemporary Western literature. The stream-of-consciousness method (a first-person narrative technique that renders the apparently random thoughts and associations that go through the character's conscious mind), introduced by Yi Sang, attracted much notice around 1935, and many writers remained interested in it until the 1950s. Some writers, on the other hand, stuck to the earlier Naturalistic technique (which puts the emphasis on external action).

In this period, Chbng Chi-yang, a pioneer in late modern poetry, fully demonstrated the potential of the Korean language as a poetic medium. Yi Hyo-sŏ described his fictional reality in poetic language, as in the short story "Memilkkot p'il-murybp" (1936, "When the Buckwheat Blooms"). Kim Ki-rim also exercised great influence in poetry until the 1950s, as did Yi Sang in the novel. Toward the end of the 1930s there was a turn toward nature and the rustic life. In poetry the Ch'ongnokp'a (Blue Deer School) reflected this trend, and in fiction a similar tendency appeared in antiurban, pastoral novels. The return to the native soil was possibly a reaction to the political reality with which writers were confronted. Generally, the first half of the 1940s was a dark period in Korean literature. When World War II was at its height, the Japanese did not allow Korean writers even the escape that nature and the pastoral life offered. Korean became an outlaw language.

When the war was over, however, Korean literature made a fresh start based on the inheritance of the 1930s. In 1950, when Korean literature had just begun to take root again, the Korean War began. The literature that followed the war was so remarkably different from that which had gone before that many critics saw it as marking the beginning of a new period, which they called Post-War Literature. To many, it seemed a new generation of Korean writers had appeared, one which knew nothing of the Japanese occupation and filtered knowledge but which was making its own fresh discovery of international literary ideas after a darkness of more than ten years. The resulting new flood of Western thought reminded many of the 1920s. (B.-W.C.)

III. Japanese literature

Both in quantity and quality, Japanese literature ranks as one of the major literatures of the world, comparable in age, richness, and bulk to English literature, though the course of development has been quite dissimilar. The surviving works extend from the 7th century AD to the present; during all of this time there was never a "dark age" devoid of literary production. Not only do poetry, the novel, and the drama have long histories in Japan, but some literary genres not so highly esteemed in other countries—including diaries, travel accounts, and books of random thoughts—are also prominent. A considerable body of writing by Japanese in the Chinese classical language, of much greater bulk and importance than comparable Latin writings by Englishmen, testifies to the Japanese literary indebtedness to China. Even the writings entirely in Japanese present an extraordinary variety of styles, which cannot be explained merely in terms of the natural evolution of the language. Some styles were patently influenced by the importance of Chinese vocabulary and syntax; but others developed in response to the internal requirements of the various genres, whether the terseness of haiku (a poem in 17 syllables) or the bombast of the dramatic recitation.

Problems in reading Japanese literature

The difficulties of reading Japanese literature can hardly be exaggerated; even a specialist in one period is likely to have trouble deciphering a work from another period or genre. Japanese style has favoured ambiguity, and the particles of speech necessary for easy comprehension of a statement are often omitted as unnecessary or as fussily precise. Sometimes the only clue to the subject or object of a sentence is the level of politeness in which the words are couched; for example, the verb *mesu* (meaning "to eat," "to wear," "to ride in a carriage," etc.) designates merely an action performed by a person of quality. In many cases, ready comprehension of a simple sentence depends on a familiarity with the background of a particular period of history. The verb miru, "to see," had overtones of "to have an affair with" or even "to marry" during the 11th century, when men were generally able to see women only after they had become intimate. The long period of Japanese isolation in the 17th and 18th centuries also tended to make the literature provincial, or intelligible only to persons sharing a common background; the phrase "there was a little smoke" (*sukoshi* no *kemuri* atte) was all readers of the late 17th century needed to realize that an author was referring to the Great Fire of 1682 that ravaged the city of Edo (Tokyo).

Despite the great difficulties arising from such idiosyncrasies of style, Japanese literature of all periods is exceptionally appealing to modern readers, whether read in the original or in translation. Because it is prevailingly subjective and coloured by an emotional rather than an intellectual or moralistic tone, its themes have a universal quality almost unaffected by time. To read a diary by a court lady of the 10th century is still a moving experience because she described with such honesty and intensity her deepest feelings that the modern reader forgets the chasm of history and changed social customs separating her world from his own.

The "pure" Japanese language, untainted and unfertilized by Chinese influence, contained remarkably few words of an abstract nature. Just as English borrowed such words as morality, honesty, justice, and the like from the Continent, the Japanese borrowed these terms from China; but if the Japanese language was lacking in the vocabulary appropriate to a Confucian essay, it could express almost infinite shadings of emotional content. A Japanese poet who was dissatisfied with the limitations imposed by his native language or who wished to describe unemotional subjects—whether the quiet outing of aged gentlemen to a riverside or the poet's awareness of his insignificance as compared to the grandeur of the universe—naturally turned to writing poetry in Chinese. From the 16th century on, many words that had been excluded from poetry because of their foreign origins or their humble meanings, following the dictates of the codes of poetic diction established in the 10th century, were adopted by the practitioners of the haiku, originally an iconoclastic, popular verse form. For the most part, however, the Japanese writers, far from feeling dissatisfied with the limitations on expression imposed by their language, were convinced that virtuoso perfection in phrasing and an acute refinement of sentiment were more important to poetry than the voicing of intellectually satisfying concepts.

Effect of the language on literary forms

The Japanese language itself also shaped poetic devices and forms. Because it lacks a stress accent, meaningful rhymes (all words end in one of five simple vowels), or quantity, poetry was distinguished from prose mainly in that it consisted of alternating lines of five and seven syllables; however, if the intensity of emotional expression was low, this distinction alone could not save a poem from dropping into prose. The difficulty of maintaining a high level of poetic intensity may account for the preference for short verse forms that could be polished with perfectionist care. But however moving a tanka (verse in 31 syllables) is, it clearly cannot fulfill some of the functions of longer poetic forms; and there are no Japanese equivalents of Paradise Lost, The Rape of the Lock, or *Tintern* Abbey, Instead, the poets devoted their efforts to perfecting each syllable of their compositions, expanding the content of a tanka by suggestion and allusion and prizing shadings of tone and diction more than originality or boldness of expression.

The fluid syntax of the prose affected not only style but content. Japanese sentences are sometimes of inordinate length, responding to the subjective turnings and twistings of the author's thought; and the writers considered smooth transitions from one statement to the next, rather than structural unity, the mark of excellent prose. The longer works accordingly betray at times a lack of overall structure of the kind associated in the West with Greek concepts of literary form but consist instead of episodes linked chronologically or by other associations. The difficulty experienced by Japanese writers in organizing their impressions and perceptions into sustained works may explain the development of the diary and travel account, genres in which successive days or the successive stages of a journey provide a structure for otherwise unrelated descriptions. Japanese literature contains some of the world's longest novels and plays; but its genius is most strikingly displayed in the shorter works, whether the tanka, the haiku, the Nō plays, or the poetic diaries.

An acute literary sensibility, fostered especially by the traditions of the court, encouraged the creation of "codes" of poetic practice and of a considerable body of criticism, extending back to the 10th century, that was usually composed by the leading poets or dramatists themselves. These codes exerted an inhibiting effect on new forms of literary composition, but they also helped to preserve a distinctively aristocratic tone.

Relation to Chinese and Korean literature

Japanese literature absorbed much direct influence from China, but the characteristic literary works are strikingly dissimilar. The tradition of feminine writing, especially of such introspective works as diaries, gave a colouring to Japanese prose quite unlike the more objective, masculine Chinese writings. Although the Japanese have been criticized for their imitations of Chinese examples (even by some Japanese), the Japanese novel in fact antedates any Chinese novels by centuries; and the theatre developed quite independently. Because the Chinese and Japanese languages are unrelated, the poetry naturally took different forms, although Chinese poetic examples and literary theories were often in the minds of the Japanese poets. Japanese and Korean are probably related languages, but Korean literary influence was negligible, though Koreans served an important function in transmitting Chinese literary and philosophical works to Japan. Poetry and prose written in the Korean language were unknown to the Japanese until relatively modern times.

From the 8th to the 19th centuries Chinese literature enjoyed greater prestige among educated Japanese than their own; but a love for the Japanese classics, especially those composed at the court in the 10th and 11th centuries, gradually spread among the entire people and influenced literary expression in every form, even the songs and tales composed by humble people totally removed from the aristocratic world portrayed in classical literature.

EARLIEST LITERATURE

The first writing of literature in Japanese was occasioned by influence from China. The Japanese were still comparatively primitive and without writing when, in the first four centuries AD, knowledge of Chinese civilization gradually reached them. They rapidly assimilated much of this civilization, and the Japanese scribes adopted Chinese characters as a system of writing, although an alphabet (if one had been available to them) would have been infinitely better suited to the Japanese language. The characters, first devised to represent Chinese monosyllables, could be used only with great ingenuity to represent the agglutinative forms of the Japanese language. The ultimate results were chaotic, giving rise to one of the most complicated systems of writing ever invented. The use of Chinese characters enormously influenced modes of expression and led to an association between literary composition and calligraphy lasting many centuries.

The earliest Japanese texts were no doubt written in Chinese because no system of transcribing the sounds and grammatical forms of Japanese had as yet been invented. The oldest known inscription, on a sword that dates from about AD 440, already shows some modification of normal Chinese usage in order to transcribe Japanese names and expressions. The most accurate way of writing Japanese words was by using Chinese characters not for their meanings but for their phonetic values, giving each character a pronunciation approximating that used by the Chinese themselves. In the oldest extant works, the ***Koji-ki*** (712; "Records of Ancient Matters") and ***Nihon shoki*** (720; "Chronicles of Japan"), more than 120 songs, some perhaps dating back to the 5th century AD, are given in phonetic transcription, doubtless because the Japanese attached great importance to the sounds themselves.

In these two works, both officially commissioned "histories" of Japan (though larded with myths and legends), many sections are written entirely in Chinese; but parts of the *Koji-ki* were composed in a complicated mixture of language that made use of the Chinese characters sometimes for their meaning and sometimes for their sound.

The myths in the ***Koji-ki*** are occasionally beguiling, but the only truly literary parts of the work are the songs. The early songs lack a fixed metrical form; the lines, consisting of an indeterminate number of syllables, were strung out to irregular lengths, showing no conception of poetic form. Some songs, however, seem to have been reworked—perhaps when the manuscript was transcribed in the 8th century—into what became the classic Japanese verse form, the tanka (short poem), consisting of five lines of five, seven, five, seven, and seven syllables. Various poetic devices employed in these songs, such as the *makura-kotoba* ("pillow word"), a kind of fixed epithet, remained a feature of later poetry.

Origin of the tanka in the ***Koji-ki***

Altogether, some 500 primitive songs have been preserved in various collections. Many describe travel, and a fascination with place-names, evident in the loving enumeration of mountains, rivers, and towns with their mantic epithets, was developed to great lengths in the gazetteers (*fudoki*) compiled at the beginning of the 8th century. These works, of only intermittent literary interest, devote considerable attention to the folk origins of different place-names, as well as to other local legends.

A magnificent anthology of poetry, the *Manyō-shū* (compiled after 759), is the single great literary monument of the Nara period (710–784), although it includes poetry written in the preceding century, if not earlier. Most of the 4,500 or so poems are tankas; but the masterpieces of the ***Manyd-shii*** are the 260 ***chdka*** (long poems), ranging up to 150 lines in length and cast in the form of alternating lines in five and seven syllables followed by a concluding line in seven syllables. The amplitude of the ***chdka*** permitted the poets to treat themes impossible within the compass of the tanka—whether the death of a wife or child, the glory of the Imperial family, the discovery of a gold mine in a remote province, or the hardships of military service.

The greatest of the *Manyō-shū* poets, Kakinomoto Hitomaro, served as a kind of poet laureate in the late 7th and early 8th centuries, accompanying the sovereigns on their excursions and composing odes of lamentation for deceased members of the Imperial family. Modern scholars have suggested that the *chōka* may have originated as exorcisms of the dead, quieting the ghosts of recently deceased persons by reciting their deeds and promising that they will never be forgotten. Some of Hitomaro's masterpieces describe the glories of princes or princesses he may never have met but so convincingly as to transcend any difference between "public" expressions of grief and his private feelings. Hitomaro's ***chdka*** are unique in Japanese poetry thanks to their superb combination of imagery, syntax, and emotional strength; they are works of truly masculine expression. He showed in his tanka, however, that he was also capable of the evocative, feminine qualities typical of later Japanese poetry.

The ***chdka*** often concluded with one or more ***hanka*** ("envoys") that resumed central points of the preceding poem. The *hanka* written by the 8th-century poet Yamabe Akahito are so perfectly conceived as to make the *chōka* they follow at times seem unnecessary; the concision and evocativeness of these poems, identical in form with the tanka, are close to the ideals of later Japanese poetry. Nevertheless, the supreme works of the *Manyō-shū* are the ***chdka*** of Hitomaro, Ōtomo Tabito, Otomo Yakamochi (probably the chief compiler of the anthology), and Yamanoe Okura.

The most striking quality of the *Manyō-shū* is its powerful sincerity of expression. The poets were certainly not artless songsmiths exclaiming in wonder over the beauties of nature, a picture that is often painted of them by sentimental critics; but their emotions were stronger and more directly expressed than in later poetry. The corpse of an unknown traveller, rather than the falling of the cherry blossoms, stirred in Hitomaro an awareness of the uncertainty of human life.

The *Manyō-shū* is exceptional in the number of poems composed outside the court, whether by frontier guards or persons of humble occupation. Perhaps some of these poems were actually written by courtiers in the guise of commoners, but the use of dialect and familiar imagery contrasts with the strict poetic diction imposed in the 10th century.

The diversity of themes and poetic forms also distinguishes the *Manyō-shū* from the more polished but narrower verse of later times. In Okura's famous "Dialogue on Poverty," for example, two men---one poor and the other destitute---describetheir miserable lots, revealing a concern over social conditions that would be absent from the classical tanka. Okura's visit to China in 701, as the member of a Japanese embassy, may account for Chinese influence in his poetry. His poems are also prefaced in many instances by passages in Chinese stating the circumstances of the poems or citing Buddhist parallels.

The *Manyō-shū* was transcribed in an almost perversely complicated system that used Chinese characters arbitrarily, sometimes for meaning and sometimes for sound. The lack of a suitable script probably inhibited literary production in Japanese during the Nara period. The growing importance, however, of Chinese poetry as the mark of literary accomplishment in a courtier may also have interrupted the development of Japanese literature after its first flowering in the *Manyō-shū*.

Eighteen *Manyō-shū* poets are represented in the collection *Kaifūsō* (751), an anthology of poetry in Chinese composed by members of the court. These poems are little more than pastiches of ideas and images borrowed directly from China; the composition of such poetry reflects the enormous prestige of Chinese civilization at this time.

CLASSICAL LITERATURE: HEIAN PERIOD (794–1185)

The foundation of the city of Heian-kyō (later known as Kyōto) as the capital of Japan marked the beginning of a period of great literary brilliance. The earliest writings of the period, however, were almost all in Chinese because of the continued desire to emulate the culture of the continent. Three Imperially sponsored anthologies of Chinese poetry appeared between 814 and 827, and it seemed for a time that writing in Japanese would be relegated to an extremely minor position. The most distinguished writer of Chinese verse, the 9th-century poet Sugawara Michizane, gave a final lustre to this period of Chinese learning by his erudition and poetic gifts; but his refusal to go to China when offered the post of ambassador, on the grounds that China no longer had anything to teach Japan, marked a turning point in the response to Chinese influence.

Poetry. The invention of the kana phonetic syllabary, traditionally attributed to the 9th-century Shingon priest and Sanskrit scholar Kūkai, enormously facilitated writing in Japanese. Private collections of poetry in kana began to be compiled about 880; and in 905 the ***Kokin-shū,*** the first major work of kana literature, was compiled by the poet Ki Tsurayuki and others. This anthology contains 1,111 poems divided into 20 books arranged by topics, including six books of seasonal poems, five books of love poems, and single books devoted to such subjects as travel, mourning, and congratulations. The two prefaces are clearly indebted to the theories of poetry described by the compilers of such Chinese anthologies as the ***Shih Ching*** and ***Wen*** *hsüan*, but the preferences they express would be shared by most tanka poets for the next 1,000 years. The preface by Tsurayuki, the oldest work of sustained prose in kana, enumerated the circumstances that move men to write poetry; he believed that melan-

The first major work of kana

choly, whether aroused by a change in the seasons or by a glimpse of white hairs reflected in a mirror, provided a more congenial mood for writing poetry than the harsher emotions treated in the *Manyō-shū*. The best tanka in the *Kokin-shū* captivate the reader by their perceptivity and tonal beauty, but these flawlessly turned miniatures obviously lack the variety of the *Manyō-shū*.

Skill in composing tanka became an asset in gaining preference at court; it was also essential to a lover, whose messages to his mistress (who presumably could not read Chinese, still the language employed by men in official documents) often consisted of poems describing his own emotions or begging her favours. In this period the tanka almost completely ousted the *chdka* because the shorter poems were more suited to the lover's billet-doux or to competitions on prescribed themes, a frequent court pastime.

For the poets of the *Kokin-shū* and the later court anthologies, originality was less desirable than perfection of language and tone. The critics, far from praising novelty of effects, condemned deviations from the standard poetic diction (established by the *Kokin-shū*) of some 2,000 words and insisted on absolute adherence to the poetic conventions. Although these restrictions saved Japanese poetry from lapses into bad taste or vulgarity, they froze it for centuries in prescribed modes of expression. Only a skilled critic can distinguish a tanka of the 10th century from one of the 19th century. The *Kokin-shii* set the precedent for later court anthologies, and a knowledge of its contents was indispensable to all poets as a guide and source of literary allusions.

Love poetry occupies a prominent place in the *Kokin-shū*, but the joys of love are seldom celebrated; instead, the poets wrote in the melancholy vein prescribed in the preface, describing the uncertainties before a meeting with the beloved, the pain of parting, or the sad realization that an affair had ended. The invariable perfection of diction, unmarred by any indecorous cry from the heart, may sometimes make one doubt the poet's sincerity. This is not true of the great *Kokin-shū* poets of the 9th century —Ono Komachi, Ise, Ariwara Narihira, and Tsurayuki himself—but even Buddhist priests, who presumably had renounced carnal love, wrote love poetry at the court competitions; and it is hard to detect any difference between such poems and those of sincere lovers.

The preface of the *Kokin-shii* lists judgments on the principal poets of the collection. This criticism is unsatisfying to a modern reader because it is so terse and unanalytical; but it nevertheless marks a beginning of Japanese poetic criticism, an art that developed impressively during the course of the Heian period.

The literary diary

Prose. Ki Tsurayuki is celebrated also for his *Tosa nikki* (935; *The Tosa Diary*, 1912), the account of his homeward journey to Kyōto from the province of Tosa, where he had served as governor. Tsurayuki wrote this diary in Japanese, though men at the time normally kept their diaries in Chinese (perhaps it was in order to escape reproach for adopting this unmanly style that he pretended a woman in the governor's entourage was the author). Events of the journey are interspersed with the poems composed on various occasions. The work is affecting especially because of the repeated, though muted, references to the death of Tsurayuki's daughter in Tosa.

Tosa rzikki is the earliest example of a literary diary. Tsurayuki pretended he was a woman, but the later Heian diarists who wrote in the Japanese language were, in fact, court ladies; their writings include some of the supreme masterpieces of the literature. *Kagerd nikki (The Gossamer Years,* 1964) describes the life between 954 and 974 of the second wife of the prime minister. The first volume, related long after the events, is in the manner of an autobiographical novel; even the author confesses that her remembrances are probably tinged with fiction. The second two volumes approach a true diary, with some entries apparently made on the days indicated. The writer (known only as "the mother of Michitsuna") describes, with many touches of self-pity, her unhappy life with her husband. She evidently assumed that readers would sympathize, and often this is the case, though her self-centred complaints are not endearing. In one passage, in which she gloats over the death of a rival's child, her obsession with her own griefs shows to worst advantage; yet her journal is extraordinarily moving precisely because the author dwells exclusively on universally recognizable emotions and omits the details of court life that must have absorbed the men.

Other diaries of the period include the anecdotal *Murasaki Shikibu nikki* (Eng. trans. 1920), at once an absorbing literary work and a source of information on the court life the author described more romantically in her masterpiece *Genji inonogatari* ("The Tale of Genji," c. 1010), and *Izumi Shikibu nikki* (Eng. trans. 1969), which is less a diary than a short story liberally ornamented with poetry.

These "diaries" are closely related in content and form to the *uta monogatari* (poem tales) that emerged as a literary genre later in the 10th century. *Ise monogatari (Tales of Ise, c.* 980) consists of 143 episodes, each containing one or more poems and an explanation in prose of the circumstances of composition. The brevity and often the ambiguity of the tanka gave rise to a need for such explanations, and when these explanations became extended or (as in the case of *Ise monogatari)* were interpreted as biographical information about one poet (Ariwara Narihira), they approached the realm of fiction.

Along with the poem tales, there were works of religious or fanciful inspiration going back to *Nihon ryōi-ki*, an account of Buddhist miracles in Japan compiled by the priest Kyōkai in 822. These stories, written in Chinese, were probably used as a source of sermons by the priests with the intent of persuading ordinary Japanese, incapable of reading difficult works of theology, that they must lead virtuous lives if they were not to suffer in hell for present misdeeds. No such didactic intent is noticeable in *Taketori monogatari* (10th century; *Tale of the Bamboo Cutter*, 1956), a fairy tale about a princess who comes from the moon to dwell on earth in the house of a humble bamboo cutter; the various tests she imposes on her suitors, fantastic though they are, are described with humour and realism.

Early development of the novel

The first lengthy "novel," *Utsubo monogatari* ("The Tale of the Hollow Tree"), was apparently written between 956 and 983 by Minamoto Shitagau, a distinguished courtier and scholar. This uneven, ill-digested work is of interest chiefly as an amalgam of elements in the poem tales and fairy tales; it contains 986 tanka, and its episodes range from early realism to pure fantasy.

The contrast between this crude work and the sublime *Genji monogatari* is overwhelming. Perhaps the difference is best explained in terms of the feminine traditions of writing, exemplified especially by the diaries, which enabled Murasaki Shikibu to discover depths in her characters unsuspected by the male author of *Utsubo monogatari*. The *Genji monogatari* is not only the finest work of the Heian period but of all Japanese literature and merits being called the first important novel written anywhere in the world. *Genji monogatari* was called a work of *mono no aware* ("a sensitivity to things") by the great 18th-century literary scholar Moto-ori Norinaga; the hero, Prince Genji, is not remarkable for his martial prowess or his talents as a statesman but as an incomparable lover, sensitive to each of the many women he wins. The story is related in terms of the successive women Genji loves; each of them evokes a different response from this marvellously complex man. The last third of the novel, describing the world after Genji's death, is much darker in tone; and the principal figures, though still impressive, seem no more than fragmentations of the peerless Genji.

The success of *Genji monogatari* was immediate. The author of the touching *Sarashina nikki (As I was Crossing a Bridge of Dreams,* 1971) described how as a girl she longed to visit the capital so that she might read the entire work (which had been completed some ten years earlier). Imitations and derivative works based on *Genji monogatari*, especially on the last third of it, continued to be written for centuries, inhibiting the fiction composed by the court society.

Aware and *okashi*

Makura-no-sdshi (c. 1000; *The Pillow Book of Sei Shōnagon,* 1967) is another masterpiece of the Heian period that should be mentioned with *Genji rnonogatari.* Japanese critics have often distinguished the *aware* of *Genji monogatari* and the *okashi* of *Makura-no-sōshi. Aware* meant sensitivity to the tragic implications of a moment or gesture, *okashi,* the comic overtones of perhaps the same moment or gesture. The lover's departure at dawn evoked many wistful passages in *Genji monogatari,* but in *Makura-no-sōshi* the author noted with unsparing exactness the lover's fumbling, ineffectual leave-taking and his lady's irritation. Murasaki Shikibu's *aware* can be traced through later literature—sensitivity always marked the writings of any author in the aristocratic tradition—but Sei Shbnagon's wit belonged to the Heian court alone.

Folk tales and songs

The Heian court society passed its prime by the middle of the 11th century, but it did not collapse for another 100 years. Long after its political power had been usurped by military men, the court retained its prestige as the fountainhead of culture. But in the 12th century, literary works belonging to a quite different tradition began to appear. *Konjaku monogatari,* a massive collection of religious and folktales, drawn not only from the Japanese countryside but also from Indian and Chinese sources, described elements of society that had never been treated in the court novels. These stories, though crudely written, provide glimpses of how the common people spoke and behaved in an age marked by warfare and new religious movements. The collection of folk songs *Ryōjin hishō,* compiled in 1179 by the emperor Shirakawa II, suggests the vitality of this burgeoning popular culture even as the aristocratic society was being threatened with destruction.

MEDIEVAL LITERATURE: KAMAKURA, MUROMACHI, AND AZUCHI-MOMOYAMA PERIODS (1192–1600)

Kamakura period (1192–1333). *Poetry.* The warfare of the 12th century brought to undisputed power military men (samurai) whose new regime was based on martial discipline. Though the samurai expressed respect for the old culture, some of them even studying tanka composition with the Kybto masters, the capital of the country moved to Kamakura. The lowered position of women under this feudalistic government perhaps explains the noticeable diminution in the importance of writings by court ladies; indeed, there was hardly a woman writer of distinction between the 13th and 19th centuries. The court poets, however, remained prolific: 15 Imperially sponsored anthologies were completed between 1188 and 1439; and most of the tanka followed the stereotypes established in earlier literary periods.

The *Shin kokin-shū* anthology

The finest of the later anthologies, the *Shin kokin-shū* (1206), was compiled by Fujiwara Teika, or Sadaie, among others, and is considered by many as the supreme accomplishment in tanka composition. The title of the anthology—"the new *Kokin-shū*"—indicates the confidence of the compilers that the poets represented were worthy successors of those in the 905 collection; they included (besides the great Teika himself) Teika's father Fujiwara Toshinari (Shunzei), the priest Saigyō, and the former emperor Toba II. These poets looked beyond the visible world for symbolic meanings. The brilliant colours of landscapes filled with blossoms or reddening leaves gave way to monochrome paintings; the poet, instead of dwelling on the pleasure or grief of an experience, sought in it some deeper meaning he could sense if not fully express. The tastes of Teika especially dominated Japanese poetic sensibility, thanks not only to his poetry and essays on poetry but to his choices of the works of the past most worthy of preservation.

Prose. Teika is credited also with a novel, *Matsura-no-miya monogatari.* Though unfinished and awkwardly constructed, its dreamlike atmosphere lingers in the mind with the overtones of Teika's poetry; dreams of the past were indeed the refuge of the medieval romancers, who modelled their language on the *Genji monogatari,* though it was now archaic, and borrowed their themes and characters from the Heian masterpieces. Stories about wicked stepmothers are fairly common; perhaps the writers, contrasting their neglect with the fabled lives of the Heian courtiers, identified themselves with the maltreated stepdaughters; and the typical happy ending of such stories—the stepdaughter in *Sumiyoshi monogatari* is married to a powerful statesman and her wicked stepmother humiliated—may have been the dream fulfillment of their own hopes.

Various diaries describe travels between Kyoto and the shogun's capital in Kamakura. Courtiers often made this long journey in order to press claims in lawsuits, and they recorded their impressions along the way in the typical mixture of prose and poetry. *Izayoi nikki* (Eng. trans. 1951) tells of a journey made in 1277 by the nun Abutsu. A later autobiographical work that also contains extensive descriptions of travel is the superb *Towazu-gatari* by Lady Nijō, a work (discovered only in 1940) that provides a final moment of glory to the long tradition of introspective writing by women at court.

Although these writings in the aristocratic manner preserved much of the manner of Heian literature, works of quite different character were even more prominent in the medieval period. There are many collections of Buddhist and popular tales, of which the most enjoyable undoubtedly is the *Ujishūi monogatari* (Eng. trans. 1970)—a compilation over a period of years of some 197 brief stories. Although the incidents described in these tales are often similar to those found in *Konjaku monogatari,* they are told with considerably greater literary skill.

Gunki monogatari established as a literary genre

An even more distinctive literary genre of the period is the *gunki monogatari,* or war tale. The most famous, *Heike monogatari,* was apparently first written at the court about 1220, probably by a nobleman who drew his materials from the accounts recited by priests of the warfare between the Taira and the Minamoto families in the preceding century. The celebrated opening lines of the work, a declaration of the impermanence of all things, also states the main subject, the rise and fall of the Taira family. The text, apparently at first in three books, was expanded to 12 in the course of time, as the result of being recited with improvisations by priest-entertainers. This oral transmission may account not only for the unusually large number of textual variants but also for the exceptionally musical and dramatic style of the work. Unlike the Heian novelists, who rarely admitted words of Chinese origin into their works, the reciters of the *Heike monogatari* employed the contrasting sounds of the imported words to produce what has been acclaimed as the great classic of Japanese style. Although the work is curiously uneven, effective scenes being followed by dull passages in which the narrator seems to be stressing the factual accuracy of his materials, it is at least intermittently superb; and it provided many later novelists and dramatists with characters and incidents for their works.

Heike monogatari was by no means the earliest literary work describing warfare; and other writings, mainly historical in content, were graced by literary flourishes uncommon in similar Western works. *Ōkagami* (c. 1120?; Eng. trans. 1967), the most famous of the "mirrors" of Japanese history, undoubtedly influenced the composition of *Heike monogatari,* especially in its moralistic tone. *Hōgen monogatari* (Eng. trans. 1971) and *Heiji monogatari* (Eng. trans. 1951), chronicle warfare that antedates the events described in *Heike monogatari* but were probably written somewhat later.

War tales continued to be composed throughout the medieval period. The *Taihei-ki* (Eng. trans. 1959), for example, covers about 50 years, beginning in 1318, when the emperor Daigo II ascended the throne. Though revered as a classic by generations of Japanese, it possesses comparatively little appeal for Western readers, no doubt because so few of the figures come alive.

Characters are more vividly described in two historical romances of the late 14th century, *Soga monogarari* and *Gikei-ki* (Eng. trans. 1966), describing the life of Minamoto Yoshitsune. Though inartistically composed, these portraits of resourceful and daring heroes caught the imaginations of the Japanese; and their exploits are still prominent on the Kabuki stage.

Another important variety of medieval literature was the reflective essays of Buddhist priests. *Hōjō-ki* (1212; The Ten Foot Square Hut, 1928) by Kamo Chōmei is a hermit's description of his disenchantment with the world and his discovery of peace in a lonely retreat. The elegiac beauty of its language gives this work, brief though it is, the dignity of a classic. Chōmei was also a distinguished poet, and his essay *Mumyō-shō* (1210–12; Eng. trans. 1968) is perhaps the finest example of traditional Japanese poetic criticism.

A later priest, Yoshida Kenkō, writing during the days of warfare and unrest that brought an end to the Kamakura shogunate in 1333, the brief restoration of Imperial authority under the Emperor Daigo II from 1333 to 1335, and the institution of the Ashikaga shogunate in 1338, barely hints at the turmoil of the times in his masterpiece *Tsurezuregusa* (1330–36; Essays *in* Idleness. 1967); instead he looks back nostalgically to the past, seeking out the survivals of happier days. Kenkō's aesthetic judgments, often based on a this-worldly awareness rather surprising in a Buddhist priest, gained wide currency, especially after the 17th century, when *Tsurezuregusa* was read by every educated man.

Renga becomes preferred poetic form

The Muromachi **(1338-1573)** and **Azuchi-Momoyama (1574-1600)** periods. Poetry. In the 15th century a poetic form of plebeian origins displaced the tanka as the preferred medium of the leading poets. Renga (linked verse) had begun as the composition of a single tanka by two people and was a popular pastime even in remote rural areas. One person would compose the first three lines of a tanka, often giving obscure or even contradictory details in order to make it harder for the second person to complete the poem intelligibly. Gradually, renga spread to the court poets, who saw the artistic possibilities of this diversion and drew up "codes" intended to establish renga as an art. These codes made possible the masterpieces of the 15th century, but their insistence on formalities (*e.g.*, how often a "link" on the moon might appear in 100 links, and which links must end with a noun and which with a verb) inevitably diluted the vigour and freshness of the early renga, itself a reaction against the excessively formal tanka. Nevertheless, the renga of the great 15th-century master Iio Sōgi and his associates are unique in their shifting lyrical impulses, moving from link to link like successive moments of a landscape seen from a boat, avoiding any illusion that the whole was conceived in one person's mind.

Prose. The short stories of the 15th and 16th centuries cannot be said to have high literary value. Many still look back to the world of the Heian court, but others introduce folk materials or else elements of the miraculous in the attempt to interest readers who lacked the education to appreciate the conventional literary manner. Even though many promising stories are ruined by absurdities before their course is run, for a few moments they often give unforgettable glimpses of a society torn by disorder. The stories are anonymous, but the authors seem to have been both courtiers and Buddhist priests.

Nō plays

Drama. Unquestionably the finest literary works of the 15th century are the Nb dramas, especially those by Zeami Motokiyo. They were written in magnificent poetry (often compared to "brocade" because of the many allusions to the poetry of the past) and were provided with a structure that is at once extremely economical and free. Many are concerned with the Buddhist sin of attachment: an inability to forget his life in this world prevents a dead man from gaining release but forces him to return again and again as a ghost to relive the violence or passion of his former existence. Only prayer and renunciation can bring about deliverance. Zeami's treatises on the art of Nō display extraordinary perceptivity. His stated aims were dramatic conviction and reality, but these ideals meant ultimates to him and not superficial realism. Some Nō plays, it is true, have little symbolic or supernatural content, but the central elements of a typical program of five Nō plays were found in the highly poetic and elusive masterpieces that suggest a world invisible to the eye but evokable by the actors through the beauty of movements and speech. Unhappiness over a world torn by disorder may have led writers to suggest in their works truths that lie too deep for words. This seems to have been the meaning of yzigen ("mystery and depth), the ideal of the Nō plays. Parallel developments occurred in the tea ceremony, the landscape garden, and monochrome painting, all arts that suggest or symbolize rather than state.

LITERATURE. DURING THE TOKUGAWA PERIOD (1603-1867)

Role of printing in creating a popular literature

The restoration of peace and the unification of Japan were achieved in the early 17th century, and for approximately 250 years the Japanese enjoyed almost uninterrupted peace. During the first half of the Tokugawa period, the cities of Kyoto and Osaka dominated cultural activity; but from about 1770 Edo (the modern Tokyo) became paramount. From 1638 to 1853 Japan was closed, by government decree, to contact with the outside world. Initially, this isolation encouraged the development of indigenous forms of literature; but, eventually, in the virtual absence of fertilizing influence from abroad, it resulted in provincial writing. The adoption of printing in the early 17th century made a popular literature possible. The Japanese had known the art of printing since at least the 8th century, but they reserved it exclusively for reproducing Buddhist writings. The Japanese classics existed only in manuscript form. Perhaps the demand for copies of literary works was so small that it could be satisfied with manuscripts, costly though they were; or perhaps aesthetic considerations made the Japanese prefer manuscripts in beautiful calligraphy, sometimes embellished with illustrations. Whatever the case, not until 1591 was a nonreligious work printed. About the same time, Portuguese missionaries in Nagasaki were printing books in the Roman alphabet. In 1593, in the wake of the Japanese invasion of Korea, a printing press with movable type was sent as a present to the emperor Yōzei II. Printing soon developed into the hobby or extravagance of the rich, and many examples of Japanese literature began to appear in small editions. Commercial publication began in 1609; by the 1620s even works of slight literary value were being printed for a public eager for new books.

Early Tokugawa period **(1603–c. 1770).** Poetry. Poetry underwent many changes during the early part of the Tokugawa period. At first the court poets jealously maintained their monopoly over the tanka, but gradually other men, many of them kokugakusha ("scholars of national learning"), changed the course of tanka composition by attempting to restore to the form the simple strength of Manyd-shzi poetry. The early-18th-century poet Kamo Mabuchi was the best of the neo-*Manyō-shū* school, but his tanka rarely rise above mere competence in the ancient language.

Emergence of haiku

The chief development in poetry during the Tokugawa shogunate was the emergence of the haiku as an important genre. This exceedingly brief form (17 syllables arranged in lines of five, seven, and five syllables) had originated in the hokku, or opening verse of a renga sequence, which had to contain in its three lines mention of the season, the time of day, the dominant features of the landscape, etc., making it almost an independent poem. The hokku became known as the haiku late in the 19th century, when it was entirely divested of its original function of opening a sequence of verse; but today even the 17th-century hokku are usually called haiku.

As early as the 16th century haikai renga, or comic renga, had been composed by way of diversion after an evening of serious renga composition, reverting to the original social, rather than literary, purpose of making linked verse. As so often happened in Japan, however, a new art, born as a reaction to the stultifying practices of an older art, was "discovered," codified, and made respectable by practitioners of the older art, generally at the cost of its freshness and vitality. Matsunaga Teitoku, a conventional 17th-century poet of tanka and renga who revered the old traditions, became almost in spite of himself the mentor of the new movement in comic verse, largely as the result of pressure from his eager disciples. Teitoku brought dignity to the comic renga and made it a demanding medium, rather than the quip of a moment.

His haiku were distinguishable from serious renga not by their comic conception but by the presence of a *haigon*—a word of Chinese or recent origins that was normally not tolerated in classical verse.

Inevitably, a reaction arose against Teitoku's formalism. The poets of the Danrin school, headed by Nishiyama Sbin and Ihara Saikaku, insisted that it was pointless to waste months if not years perfecting a sequence of 100 verses. Their ideal was rapid and impromptu composition; and their verses, generally colloquial in diction, were intended to amuse for a moment rather than to last for all time. Saikaku especially excelled at one-man composition of extended sequences; in 1684 he composed the incredible total of 23,500 verses in a single day and night, too fast for the scribes to do more than tally.

The haiku was perfected into a form capable of conveying poetry of the highest quality by Matsuo Bashd. After passing through an apprenticeship in both Teitoku and Danrin schools, Bashb founded a school of his own, insisting that a haiku must contain both a perception of some eternal truth and an element of contemporaneity, combining the characteristic features of the two earlier schools. Despite their brief compass, Bashb's haiku often suggest, by means of the few essential elements he presents, the whole world from which they have been extracted; the reader must participate in the creation of the poem. Bashō's best known works are travel accounts interspersed with his verses; of these, *Oku-no-hosomichi* (1694; *The Narrow Road Through the Deep North,* 1966) is perhaps the most popular and revered work of Tokugawa literature.

Prose. The general name for the prose composed between 1600 and 1682 is *kana-zōshi,* or "kana books," the name originally having been used to distinguish popular writings in the Japanese syllabary from more learned works in Chinese. The genre embraced not only fiction but works of a near historical nature, pious tracts, books of practical information, guidebooks, evaluations of courtesans and actors, and miscellaneous essays. Only one writer of any distinction is associated with the *kana-zōshi*—Asai Rydi, a samurai who became the first popular and professional writer in Japanese history. Thanks to the development of relatively cheap methods of printing and a marked increase in the reading public, Rydi was able to make a living as a writer. Although some of his works are Buddhist, he wrote in a simple style, mainly in kana. His most famous novel, *Ukiyo monogatari* ("Tales of the Floating World," *c.* 1661), is primitive both in technique and in plot; but under his mask of frivolity Ryōi attempted to treat the hardships of a society where the officially proclaimed Confucian philosophy concealed the gross inequalities in the lots of different men.

The first important novelist of the new era was Ihara Saikaku. Some Japanese critics rank him second only to Murasaki Shikibu in all Japanese literature, and his works have been edited with the care accorded only to great classics. Such attention would surely have surprised Saikaku, whose fiction was dashed off almost as rapidly as his legendary performances of comic renga, with little concern for the judgments of posterity.

Saikaku's novels of the *ukiyo*

Saikaku's first novel, *Kōshoku ichidai otoko* (1682; *The Life of an Amorous Man,* 1964), changed the course of Japanese fiction. The title itself had strong erotic overtones, and the plot describes the amorous adventures of one man, from his precocious essays at lovemaking as a child of seven to his decision at the age of 60 to sail to an island populated exclusively by women. The licensed quarters of prostitution established in various Japanese cities by the Tokugawa government (despite its professions of Confucian morality), in order to help control unruly samurai by dissipating their energies, became a centre of the new culture. Expertise in the customs of the brothels was judged the mark of the man of the world. The old term *ukiyo,* which had formerly meant the "sad world" of Buddhist stories, now came to designate its homonym, the "floating world" of pleasure; this was the chosen world of Saikaku's hero, Yonosuke, who became the emblematic figure of the era.

Saikaku's masterpiece, *Kdshoku gonin onna* (1686; *Five Women Who Loved Love,* 1956), described the loves of women of the merchant class, rather than prostitutes; this was the first time that women of this class were given such attention. In other works he described, sometimes with humour but sometimes with bitterness, the struggles of merchants to make fortunes. His combination of a glittering style and warm sympathy for the characters lifted his tales from the borders of pornography to high art.

Saikaku was a central figure in the renaissance of literature of the late 17th century. The name Genroku (an era name designating the period 1688–1703) is often used of the characteristic artistic products: the *ukiyo-e* ("pictures of the floating world"); the *ukiyo-zōshi* ("tales of the floating world"); the Kabuki and *jōruri,* or puppet theatres; and haiku poetry. Unlike its antecedents, this culture prized modernity above conformity to the ancient traditions; to be abreast of the floating world was to be up-to-date, sharing in the latest fashions and slang, delighting in the moment rather than in the eternal truths of Nō plays or medieval poetry.

Another, darker side to Genroku culture is depicted in Saikaku's late works, with their descriptions of the desperate expedients to which men turned in order to pay their bills. Saikaku seldom showed much sympathy for the prostitutes he described; but the chief dramatist of the time, Chikamatsu Monzaemon, wrote his best plays about unhappy women, driven by poverty into their lives as prostitutes, whose only release from the sordid world in which they were condemned to dwell came when they joined their lovers in double suicides. In the world of merchants treated by Chikamatsu, a lack of money, rather than the cosmic griefs of the Nb plays, drove men to death with the prostitutes they loved but could not afford to buy.

Chikamatsu's plays for the puppet theatre

Drama. Chikamatsu wrote most of his plays for the puppet theatre, which, in the 18th century, enjoyed even greater popularity than Kabuki. His plays fell into two main categories: those based, however loosely, on historical facts or legends, and those dealing with contemporary life. The domestic plays are rated much higher critically because they avoid the bombast and fantastic displays of heroism that mark the historical dramas; but the latter, adapted for the Kabuki theatre, are superb acting vehicles.

The mainstays of the puppet theatre were written not by Chikamatsu but by his successors; his plays, despite their literary superiority, failed to satisfy the audiences' craving for displays of puppet techniques and for extreme representations of loyalty, self-sacrifice, and other virtues of the society. The most popular puppet play (later adapted also for the Kabuki actors) was *Kanadehon chūshingura* (1748, "The Treasury of Loyal Retainers") by Takeda Izumo and his collaborators; the same men were responsible for half a dozen other perennial favourites of the Japanese stage. The last great 18th-century writer of puppet plays, Chikamatsu Hanji, was a master of highly dramatic, if implausible, plots.

Later developments of tanka and haiku

Late Tokugawa period (c. 1770–1867). *Poetry.* The literature of the late Tokugawa period is generally inferior to earlier achievements, especially those of the Genroku masters. Authentic new voices, however. were heard in traditional poetic forms. Later *neo-Manyd-shii* poets such as Ryōkan, Ōkuma Kotomichi, and Tachibana Akemi proved that the tanka was not limited to descriptions of the sights of nature or disappointed love but could express joy over fish for dinner or wrath at political events. Some poets who felt that tanka did not provide ample scope for the display of such emotions turned, as in the past, to writing poetry in Chinese. The early-19th-century poet Rai Sanyd probably wrote verse in Chinese more skillfully than any previous Japanese.

Later Tokugawa poets also added distinctive notes of their own to the haiku. Yosa Buson, for example, introduced a romantic and narrative element, and Kobayashi Issa employed the accents of the common people.

Prose. A great variety of fiction was produced during the last century of the Tokugawa shogunate, but it is commonly lumped together under the somewhat derogatory heading of *gesaku* (playful composition). The word

Gesaku fiction

"playful" did not necessarily refer to the subject matter but to the professed attitude of the authors, educated men who disclaimed responsibility for their compositions. Ueda Akinari, the last master of fiction of the 18th century, won a high place in literary history mainly through his brilliant style, displayed to best advantage in *Ugetsu monogatari* (1776; "Tales of Moonlight and Rain"), a collection of supernatural tales. The ***gesaku*** writers, however, did not follow Akinari in his perfectionist attention to style and construction; instead, they produced books of almost formless gossip, substituting the raciness of daily speech for the elegance of the classical language, and relying heavily on the copious illustrations for success with the public.

The ***gesaku*** writers were professionals who made their living by sale of their books. They aimed at as wide a public as possible, and when a book was successful it was usually followed by as many sequels as the public would accept. The most popular of the comic variety of ***gesaku*** fiction was *Tōkaidō dōchū hizakurige* (1802–22; ***Hizakurige*, 1929**), by Jippensha Ikku, an account of the travels and comic misfortunes of two irrepressible men from Edo along the Tōkaidō Road to Kyōto. ***Shunshoku*** *umegoyomi* (1832; "The Plum Calendar"), by Tamenaga Shunsui is the story of Tanjirō, a peerlessly handsome but ineffectual young man for whose affections various women fight. The author at one point defended himself against charges of immorality: "Even though the women I portray may seem immoral, they are all imbued with deep sentiments of chastity and fidelity." It was the standard practice of ***gesaku*** writers, no matter how frivolous their compositions might be, to pretend that their intent was didactic.

Didactic nature of the ***yomihon***

The ***yomihon*** ("books for reading"—so called to distinguish them from works enjoyed mainly for their illustrations) were much more openly moralistic. Although they were considered to be ***gesaku,*** no less than the most trivial books of gossip, their plots were burdened with historical materials culled from Chinese and Japanese sources, and the authors frequently underlined their didactic purpose. But despite the serious intent of the ***yomihon,*** they were romances, rather than novels; and their characters, highly schematized, tended to be witches, fairy princesses, and impeccably noble gentlemen. Where they succeeded, as in a few works by Takizawa Bakin, they are absorbing as examples of storytelling rather than as embodiments of the principle of ***kanzen*** *chōaku* ("the encouragement of virtue and the chastisement of vice"), Bakin's professed aim in writing fiction.

Japanese literature in general was at one of its lowest levels at the end of the Tokugawa period. A few tanka poets and the Kabuki dramatist Kawatake Mokuami are the only writers of the period whose works are still read today. It was an exhausted literature that could be revived only by the introduction of fresh influences from abroad.

MODERN LITERATURE

Even after the arrival of Commodore Perry's fleet in 1853 and the gradual opening of the country to the West and its influence, there was little noticeable effect on Japanese literature. The long closure of the country and the general sameness of Tokugawa society for decades at a time seems to have atrophied the imaginations of the ***gesaku*** writers. Even the presence of curiously garbed foreigners, which should have provoked some sort of reaction from authors searching for new materials, at first produced little effect. The ***gesaku*** writers were oblivious to the changes in Japanese society, and they continued to grind out minor variants on the same hackneyed themes of the preceding 200 years.

It was only after the removal in 1868 of the capital to Edo (renamed Tokyo) and the declaration of the emperor Meiji that he would seek knowledge from the entire world that the ***gesaku*** writers realized their days of influence were numbered. They soon fell under attack from their old enemies, the Confucian denouncers of immoral books, and also from advocates of the new Western learning. Although the ***gesaku*** writers responded with satirical pieces and traditional Japanese fiction deriding the new learning, they were helpless to resist the changes transforming the entire society.

Early translations and their influence

Introduction of Western literature. Translations from European languages of nonliterary works began to appear soon after the Meiji Restoration. The most famous example was the translation (1870) of Samuel Smiles's ***Self-Help;*** it became a kind of bible for ambitious young Japanese eager to emulate Western examples of success. The first important translation of a European novel was *Ernest Maltravers*, by the British novelist Edward Bulwer-Lytton, which appeared in 1879 under the title *Karyū shunwa* ("A Spring Tale of Blossoms and Willows"). The early translations were inaccurate, and the translators unceremoniously deleted any passages they could not understand readily or which they feared might be unintelligible to Japanese readers. They also felt obliged to reassure readers that, despite the foreign names of the characters, the emotions they felt were exactly the same as those of a Japanese.

It did not take long. however, for the translators to discover that European literature possessed qualities unknown in the Japanese writings of the past. The literary scholar Tsubouchi Shōyō was led by his readings in European fiction and criticism to reject didacticism as a legitimate purpose of fiction; he insisted instead on its artistic values. His critical essay ***Shōsetsu shinzui*** ("The Essence of the Novel," 1885) greatly influenced the writing of subsequent fiction not only because of its emphasis on realism as opposed to didacticism but because Tsubouchi, a member of the samurai class, expressed the conviction that novels, hitherto despised by the intellectuals as mere entertainments for women and children, were worthy of even a scholar's attention.

Ukigumo (1887–89; ***The Drifting Cloud,*** 1967) by Futabatei Shimei was the first modern Japanese novel. The author was familiar with Russian literature and contemporary Western literary criticism. Futabatei wrote *Ukigumo* in the colloquial, apparently because his readings in Russian literature had convinced him that only the colloquial could suitably be used when describing the writer's own society. Despite Futabatei's success with this experiment, most Japanese writers continued to employ the literary language until the end of the century. This was due, no doubt, to their reluctance to give up the rich heritage of traditional expression in favour of the unadorned modern tongue.

Western influences on poetry

Poetry. Translations of Western poetry led to the creation of new Japanese literary forms. The pioneer collection, *Shintaishi-shō* (1882; "Selection of Poems in the New Style"), contained not only translations from English but also five original poems by the translators in the poetic genres of the foreign examples. The translators declared that although European poetry had greater variety than Japanese poetry—some poems are rhymed, others unrhymed, some are extremely long, others abrupt —it was invariably written in the language of ordinary speech:

> They never borrow words from foreign countries, nor do they pad their language with archaic words not used for a thousand years. As a consequence anyone, even a small child, can understand poetry, providing only that he is familiar with the language of the country.

The insistence on modern language and the discovery of the many forms available to the poet were not the only lessons learned from European poetry. The translators also made the Japanese public aware of how much of human experience had never been treated in the tanka or haiku forms.

Innumerable Western critics have sarcastically commented on the Japanese proclivity for imitating foreign literary models and on their alleged indifference to their own traditions. It is true that without Russian examples Futabatei could not have written *Ukigumo,* and without English examples such poets as Shimazaki Tōson could not have created modern Japanese poetry; but far from recklessly abandoning their literary heritage, most writers were at great pains to acquaint themselves with their traditional literature. The outstanding novelists of the

1890s—Ozaki Kōyō, Kōda Rohan, Higuchi Ichiyd, and Izumi Kyōka—all read Saikaku and were noticeably influenced by him. Ichiyō's short novel ***Takekurabe*** (1895; ***Growing Up,*** 1956) described the children of the Yoshiwara quarter of Edo in a realistic manner quite unlike that of the usual stories about prostitutes and their customers, but she used the language of Saikaku for her narration. Izumi Kydka, though educated partly at a Western mission school, wrote superbly in the vein of late Tokugawa fiction; something of the distant Japanese literary past prevaded even his writings of the 1930s, the final years of his life.

In poetry, too, the first products of Western influence were comically inept experiments with rhyme and with such unpromising subjects as the principles of sociology. Shimazaki Tdson's "Song of the Autumn Wind" (1896), however, is not merely a skillful echo of Shelley but a true picture of a Japanese landscape; and the irregular lines of his poem tend to fall into the traditional pattern of five and seven syllables.

A decade after the works of such English Romantic poets as Shelley and Wordsworth had influenced Japanese poetry, the translations made by Ueda Bin of the French Parnassian and Symbolist poets made an even more powerful impression. Ueda wrote, "The function of symbols is to help create in the reader an emotional state similar to that in the poet's mind; symbols do not necessarily communicate the same conception to everyone." This view was borrowed from the West, but it accorded perfectly with the qualities of tanka.

Because of the ambiguities of traditional Japanese poetic expression, it was natural for a given poem to produce different effects on different readers; the important thing, as in Symbolist poetry, was to communicate the poet's mood. If the Japanese poets of the early 1900s had been urged to avoid contamination by foreign ideas, they would have declared that this was contrary to the spirit of an enlightened age. But when informed that eminent foreign poets preferred ambiguity to clarity, the Japanese responded with double enthusiasm.

Revitalization of tanka and haiku

Even the traditional forms, tanka and haiku, though moribund in 1868, took on new life, thanks largely to the efforts of Masaoka Shiki, a distinguished late-19th-century poet in both forms but of even greater importance as a critic. Yosano Akiko, Tshikawa Takuboku, and Saitd Mokichi were probably the most successful practitioners of the new tanka. Yosano Akiko's collection *Midaregami* (1901; ***Tangled Hair,*** 1971) stirred female readers especially, not only because of its lyrical beauty but because Akiko herself seemed to be proclaiming a new age of romantic love. Takuboku emerged in the course of his short life as perhaps the most popular tanka poet of all time. His verses are filled with strikingly individual expressions of his intransigent personality. Saitō Mokichi combined an absorption with *Manyō-shū* stylistics and a professional competence in psychiatry. Despite the austere nature of his poetry, he was recognized for many years as the leading tanka poet. In haiku, Takahama Kyoshi built up a following of poets strong enough to withstand the attacks of critics who declared that the form was inadequate to deal with the problems of modern life. Kyoshi himself eventually decided that the function of haiku was the traditional one of an intuitive apprehension of the beauties of nature; but other haiku poets employed the medium to express entirely unconventional themes.

Most tanka and haiku poets continued to use the classical language, probably because its relative concision permitted them to impart greater content to their verses than modern Japanese permits. Poets of the "new style," therefore, were readier to employ the colloquial. Hagiwara Sakutarō, generally considered the finest Japanese poet of the 20th century, brilliantly exploited the musical and expressive possibilities of the modern tongue. Other poets, such as Horiguchi Daigaku, devoted themselves mainly to translations of European poetry, achieving results so compelling in Japanese that these translations are considered to form an important part of the modern poetry of Japan.

The novel between 1905 and 1941

Prose. The dominant stream in Japanese fiction since the publication of ***Hakai*** ("The Broken Commandment," 1906), by Shimazaki Tdson, and ***Futon*** ("The Quilt," 1907), by Tayama Katai, has been naturalism. Although originally inspired by the works of the 19th-century French novelist Émile Zola and other European naturalists, the movement quickly took on a distinctively Japanese colouring, rejecting (as a Confucian scholar might have rejected ***gesaku*** fiction) carefully developed plots or stylistic beauty in favour of absolute verisimilitude in the author's confessions or in his minute descriptions of the lives of unimportant people hemmed in by circumstances beyond their control.

By general consent, however, the two outstanding novelists of the early 20th century were men who stood outside the naturalist movement, Mori Ogai and Natsume Sdseki. Ōgai began as a writer of autobiographical fiction with strong overtones of German Romantic writings. Midway in his career he shifted to historical novels that are virtually devoid of fictional elements but are given literary distinction by their concise and masculine style. Sdseki gained fame with humorous novels such as ***Botchan*** (1906; ***Botchan,*** 1972), a fictionalized account of his experiences as a teacher in a provincial town. His later works have an increasingly dark and introspective character, culminating in the unfinished ***Meian*** (1916; ***Light and Darkness,*** 1971), which is a study of alienation and loneliness.

An amazing burst of creative activity occurred in the decade following the conclusion of the Russo-Japanese War in 1905. Probably never before in the history of Japanese literature were so many important writers working at once. Some of the novelists who first emerged into prominence at this time were Nagai Kafū, Tanizaki Jun-ichird, and Akutagawa Ryiinosuke. Revealing a powerful attraction to the Japanese past, tempered by a modern, essentially Western sensibility, their works were the finest of those written during the period of the 1920s and 1930s.

In some of Tanizaki's novels, especially ***Tade kuu mushi*** (1929; ***Some Prefer Nettles,*** 1955), the conflict between traditional and Western-inspired ways was a central theme. Problems of a quite different kind—of social injustice—aroused writers of the "proletarian literature" school, such as Kobayashi Takiji, to compose works that would promote revolutionary change in Japan. Still other writers, convinced that the essential function of literature was artistic rather than propagandistic, formed such opposing schools as the "Neo-sensualists," whose associates included Yokomitsu Riichi and Kawabata Yasunari. Kawabata's work, for which he was awarded the Nobel Prize for Literature in 1968, is uniquely Japanese in its lyricism and intuitive construction.

Japanese critics have divided the fiction of the period up to 1941 into many schools, each consisting of a leading writer and his disciples. Probably the most influential author was Shiga Naoya. His characteristic literary form was the closely autobiographical story, related with stylistic beauty and great intelligence but not remarkable for invention or structure. Because of Shiga's commanding presence, the ***shishōsetsu*** ("I novel") came to be more highly esteemed than outright works of fiction.

The postwar novel

The aggressive wars waged by the Japanese militarists in the 1930s inhibited literary production; and during World War II (the Pacific War of 1941–45) almost all serious writing ceased. The end of the war, however, signalled an extraordinary period of activity, both by the older generation and by new writers. Tanizaki had begun publication of his most celebrated novel, ***Sasarne-yuki*** (1943–48; ***The Makioka Sisters,*** 1957), during the war, but publication was halted by official order, and only after the war did it appear. Kawabata also completed his masterpiece ***Yukiguni*** (1935–47; ***Snow Country,*** 1956) shortly after the war. Neither writer touched on the hardships undergone during the war years or on the painful period that followed Japan's defeat. This period is vividly described, however, in the novels of Dazai Osamu; and perhaps the most powerful expression of revulsion at the war was ***Nobi*** (1951; ***Fires on the Plain***) by

Ōoka Shōhei. The dropping of the atomic bomb also inspired much poetry and prose, of which *Kuroi ame* (1967; *Black Rain)* by Ibuse Masuji was the most artistic example.

The Japan of the immediate postwar period and the prosperous Japan of the 1950s and 1960s provided the background for most of the works of Mishima Yukio, an exceptionally brilliant and versatile novelist and playwright who became the first Japanese writer generally known abroad. Mishima's best known works include *Kinkaku-ji (The Temple of the Golden Pavilion,* 1959), a psychological study of a young monk who burned a famous architectural masterpiece, and *Hōjō no Umi* (*The Sea of Fertility),* the tetralogy he completed on the day of his death. Abe Kbbb was notable among modem writers in that he managed, sometimes by resorting to avant-garde techniques, to transcend the particular condition of being a Japanese and to create myths of suffering humanity in such a work as *Suna no onna* (1960; *Woman in the Dunes,* 1964). The powerful writings of be Kenzaburō have given promise of a gifted new generation to come.

Drama. The modem Japanese theatre also began with translations and adaptations of Western plays. This new theatre originated at the end of the 19th century, when the public was still too much under the influence of Kabuki to appreciate plays without music or dance. Even in the 20th century, a distinguished dramatist such as Kishida Kunio rarely had the opportunity to see his works performed. The development of modem drama no doubt was hampered by the introduction, at about the same time, of motion pictures, which had a much greater appeal for the public. The most successful playwrights of the 1920s and 1930s, such as Mayama Seika, wrote works that, although the products of modem minds, exploited the special talents of Kabuki actors by treating historical themes and by preserving the traditional stage language. Various distinguished writers were attracted from time to time to the theatre, but they were forced to devote their major efforts to writing fiction, if only because they were so badly remunerated for their plays. It was not until after World War II that modern dramas worthy of an international audience were written and staged.

(D.Ke.)

BIBLIOGRAPHY

Chinese literature: (General works): *Ssu-pu ts'ung-k'an* (reprinted 1920–22); KARL LO, *A Guide to the Ssū pu ts'ung k'an* (1965); HU SHIH, *Pai-hua wen-hsiieh shih* ("History of Vernacular Literature," 1929); CHENG CHEN-TO, *Chung-kuo wen-hsueh shih* ("History of Chinese Literature," 1932); CH'EN SHOU-YI, *Chinese Literature: A Historical Introduction* (1961); LIU WU-CHI, *An Introduction to Chinese Literature* (1966); translations by ARTHUR WALEY too numerous to be listed; EUGEN FEIFEL (ed.), *Geschichte der chinesischen Literatur* (1945); GEORGES MARGOULIES, *Evolution de la prose artistique chinoise* (1929), *Histoire de la littérature chinoise: prose* (1949), and *Histoire de la littérature chinoise: poésie* (1951). *(Contemporary Chinese literature):* Most of the novels and plays mentioned in the article above are available in English translation. For English translations of representative modern Chinese poems and short stories, see KAI-YU HSU (ed. and trans.), *Twentieth Century Chinese Poetry: An Anthology* (1963); CHIH-TSING HSIA, with the assistance of JOSEPH S.M. LAU, *Twentieth-Century Chinese Stories* (1971); and CHI-CHEN WANG (trans.), *Contemporary Chinese Stories* (1944). TING YI, *A Short History of Modern Chinese Literature* (1959), is an informative but excessively pro-Communist survey. For concise and lucid historical accounts, see ADOLPHE C. SCOTT, *Literature and the Arts in Twentieth Century China* (1963); and the introductions to the three volumes edited by WU-CHI LIU and TIEN-YI LI, *Readings in Contemporary Chinese Literature,* rev. ed. (1964–68). C.T. HSIA, *A History of Modern Chinese Fiction,* 2nd ed. (1971), is a highly readable critical study of some representative novelists and their works. MERLE GOLDMAN, *Literary Dissent in Communist China* (1967), gives a thorough analysis of the conflict between the Chinese Communist party and the intellectuals in the 1940s and 1950s. Outstanding among monographs on individual authors are: OLGA LANG, *Pa Chin and His Writings: Chinese Youth Between the Two Revolutions* (1967); and DAVID TOD ROY, *Kuo Mo-jo: The Early Years* (1971).

Korean literature: PETER H. LEE, *Korean Literature: Topics and Themes* (1965), an introduction to Korean literature; *Anthology of Korean Poetry from the Earliest Era to the Present* (1964); and *Studies in the Saenaennorae: Old Korean Poetry* (1959), studies in the hyangga; TU-SU SO (ed. and trans.), *The Korean Literary Reader with a Short History of Korean Literature,* by DOO SOO SUH (1965), anthology of Korean literature with annotation; INSTITUTE OF ASIAN STUDIES (ed.), *Korean Studies Today: Development and State of the Field* (1970), historical and critical writings about Korean culture; INEZ KONG PAI (ed. and trans.), *The Ever White Mountain: Korean Lyrics in the Classical Sijo Form* (1965), anthology of sijo; W.E. SKILLEND, *Kodae Sosol: A Survey of Korean Traditional Style Popular Novels* (1969), catalog of Korean fiction with annotation; CHONG IN-SOP (ed. and trans.), *Folk Tales from Korea* (1952), a collection of Korean folk tales: DUK-SOON CHANG *et al.* (eds.), *The Folk Treasury of Korea: Sources in Myth, Legend, and Folktale,* trans. by TAE-SUNG KIM (1970), a collection of Korean oral literature.

Japanese literature: (General works): DONALD KEENE, *Japanese Literature: An Introduction for Western Readers* (1953) and *Landscapes and Portraits* (1971); and MAKOTO UEDA, *Literary and Art Theories in Japan* (1967), are essays on different periods of Japanese literary history. DONALD KEENE (ed.), *Anthology of Japanese Literature* (1955) and *Modern Japanese Literature* (1956), offer selections from earliest times with introductions. GEOFFREY BOWNAS and ANTHONY THWAITE, *The Penguin Book of Japanese Verse* (1964), gives examples in every form. ROBERT H. BROWER and EARL R. MINER, *Japanese Court Poetry* (1961), is an excellent study extending from the earliest period to the 14th century. EARL R. MINER, *Japanese Poetic Diaries* (1969), contains examples from the Heian period to the 19th century. *(Early and Nara periods)*: DONALD L. PHILIPPI'S translations of the *Kojiki* (1969) and *This Wine of Peace, This Wine of Laughter* (1968), incorporate recent scholarship. NIPPON GAKUJUTSU SHINKOKAI (trans.), *The Manydshzi* (1940 and 1965), is the most readable version of this great anthology. *(Heian period)*: IVAN MORRIS (trans.), *The Pillow Book of Sei Shōnagon*, 2 vol. (1967) and *As I Crossed a Bridge of Dreams* (1973); EDWARD SEIDENSTICKER (trans.), *The Gossamer Years* (1964); and ARTHUR WALEY (trans.), *The Tale of Genji* (1926–33, reprinted 1960), are free and poetic renderings of these classics. EDWIN A. CRANSTON (trans.), *The Izurni Shikibu Diary* (1969); and HELEN C. MCCULLOUGH (trans.), *Tales of Ise* (1968), are more literal versions with scholarly introductions. IVAN MORRIS, *The World of the Shining Prince* (1964), gives the social and historical background to the Heian literary masterpieces. *(Middle Ages):* DONALD KEENE (trans.), *Essays in Idleness* (1967) and *Twenty Plays of the Nō Theatre* (1970), are readable but fairly literal; ARTHUR WALEY (trans.), *The Nd Plays of Japan* (1922), gives freer and more poetic versions of the Nō plays. HELEN C. MCCULLOUGH (trans.), *Taiheiki* (1959) and *Yoshitsune* (1966), are accurate versions in somewhat old-fashioned language of these war tales. ERWIN O. REISCHAUER and JOSEPH K. YAMAGIWA (eds. and trans.), *Translations from Early Japanese Literature* (1951), gives painstakingly literal versions. D.E. MILLS (trans.), *A Collection of Tales from Uji* (1970), is readable and scholarly. *(Tokugawa period)*: WILLIAM T. DE BARY (trans.), *Five Women Who Loved Love* (1957); HOWARD HIBBETT (trans.), *The Floating World in Japanese Fiction* (1959); G.W. SARGENT (trans.), *The Japanese Family Storehouse* (1959); and IVAN MORRIS (trans.), *The Life of an Amorous Woman* (1963), are excellent translations of Saikaku's most important works with good introductions. HAROLD G. HENDERSON (ed. and trans.), *An Introduction to Haiku* (1958), is well-written and gives literal as well as artistic translations of the poems. DONALD KEENE (trans.), *Major Plays of Chikamatsu* (1961) and *Chzishingura* (1971), are representative masterpieces of the puppet drama. *(Modern period)*: IVAN MORRIS (ed.), *Modern Japanese Stories* (1961), contains examples from many different schools. YOSHIE OKAZAKI (ed.), *Japanese Literature in the Meiji Era* (1955), is overly detailed but sometimes useful for that reason. MARLEIGH G. RYAN, *Japan's First Modern Novel: Ukigumo of Futabetai Shimei* (1967), is a study and translation of an important work. EDWARD SEIDENSTICKER, *Kafū the Scribbler* (1965), is a biography of Nagai Kafū with selected translations. EDWIN MCCLELLAN, *Two Japanese Novelists* (1969), discusses the major works of Natsume Sōseki and Shimazaki Tbson. TAKAMICHI NINOMIYA and D.J. ENRIGHT (eds.), *The Poetry of Living Japan* (1957); and ICHIRO KONO and RIKUTARO FUKUDA, *An Anthology of Modern Japanese Poetry* (1957), give selections from Meiji and later poetry.

(S.Y.C./T.-y.L./B-W.C./D.Ke.)

Literature, Nonfictional Prose

Defining literature in nonfictional prose is an immensely challenging task: it may denote any writing done with an aim, however dimly perceived, of producing an impression of beauty upon the hearer or the reader. This task is much more difficult than defining poetry, which has for centuries proved a challenging enough task for many critics, even though some exterior criteria, such as prosody and music, imagery, a certain exaltation of tone, and some elevation in the themes may be invoked.

Although nonfictional prose literature differs from bald statements of fact, such as those recorded in an old chronicle or inserted in a business letter or in an impersonal message of mere information, there have been increasing attempts since the 18th century to banish all adornment, all resort to modifiers and metaphors, from writing, even from fiction or autobiography. The 19th-century French novelist Stendhal said he sought the dryness of the Civil Code in his prose, though he seldom lived up to his claims. With many other authors, the very word literature has come to have a denigrating sense, connoting a lack of sincerity and an excess of meretricious embellishments. Such a mood was not altogether absent from earlier ages; but it took on a new impetus among the modems, especially in countries of continental Europe with a long literary tradition. There, in the reaction against the bombast and the artificiality of some romantic, symbolist, and self-styled decadent authors, literature has been treated with contempt as a mendacious and idle exercise that disregards or distorts truth. With a number of late-19th-century authors such as Nikolay Gogol and even Tolstoy in Russia and Arthur Rimbaud in France, as well as several later Americans, that condemnation of literature by the very men who practiced it became a characteristic feature of a sophisticated age in revolt against sophistication.

The word literature was derived via Latin from the Greek word meaning "the knowledge of what is written." As used in a broad sense in this article, it designates writing intended to instruct (but not highly scientific and erudite writings in which no aesthetic concern is evinced), to impart wisdom or faith, and especially to please. The existence of literature, at first oral but sooner or later entrusted to writing, has proved of immense significance in the history of mankind. Some countries have left a mark in history chiefly through their folktales, epics, ballads, or prose chronicles. Others have gained a universal audience and struck the imagination of mankind chiefly thanks to their novelists, poets, moralists. The favourable interest and emotional attachment thus gained by Russia through its 19th-century authors, and by the United States and some South American nations in the 20th century, has been incalculably precious. One of the miracles of ancient Greece is that of the survival of its literary forms through 25 centuries and several continents. The emergence of some cultures to self-consciousness—their reawakening after centuries of the somnolence that often results from foreign rule—has been fostered by the romantic movement that, in the early 19th century, aroused the Scandinavians, the Magyars, the Poles, the Slovaks, the Croatians, and many others to nationalism. Through literature they recovered their lost traditions, a sense of their identity, and a will to live.

This article is one of a number dealing with the genres of literature. Comparable articles on genres of fictional prose literature, for example, are NOVEL; and SHORT STORY. A major article dealing with a particular form of nonfictional prose literature is BIOGRAPHICAL LITERATURE. Further discussion of types of nonfictional prose literature in particular cultures may be found in articles such as LITERATURE, WESTERN; and LITERATURE, EAST ASIAN; and in articles on the arts of various peoples, such as SOUTH ASIAN PEOPLES, ARTS OF. Methods of using language for literary purposes are discussed in RHETORIC; and PROSODY; and to some extent, in ORATORY.

This article is divided into the following sections:

I. Nature and elements

NATURE

The role of poetry—narrative, epic, lyrical, and in a smaller number of nations, dramatic—has been significant in the history of culture. Virtually all literatures began with poetry. In some countries, such as Ireland, or England in the age of Bede (*c.* 673–735), or France in the early Middle Ages, the development of poetry was inhibited by church-imposed religious doctrines and manuals. Even in these instances, however, the religious literature was written in the language of the church, such as Latin, or Church Slavonic in what became Russia, and therefore touched the people only indirectly. Epics composed in the vernacular, such as ***Beowulf*** in Old English, and ***The Song of Roland*** in French. contributed to the creation of a cultural and, after a fashion, a national consciousness. After 10 to 12 centuries of Christianity, when the church appeared to have established Latin in the West as the universal language, the people, having reached a more advanced degree of literacy and of refinement, caused the break-up of that community of language and of thought. Separate literatures and national moods developed their own individuality. By the 13th century a prose, often nonfictional, had arisen for entertainment, aesthetic, and moral purposes.

Literature and national consciousness

Nonfictional prose inevitably covers an almost infinite variety of themes, and it assumes many shapes. In quantitative terms, if such could ever be valid in such nonmeasurable matters, it probably includes more than half of all that has been written in countries having a literature of their own. The short story, the satire, the sonnet, the ode, and even the novel are all comparative newcomers in modern letters and encompass much more restricted areas. Some countries can hardly boast a dramatic literature of any value; others have not developed an original craft of fiction; still others for centuries lacked the individualistic spirit that appears to be required for composing lyrical poetry. Nonfictional prose, however, has flourished in nearly all countries with advanced literatures. The forms it took included political and polemical writings, biographical and autobiographical literature, and religious writings and others that popularized a philosophy or imparted a moral message. After the Renaissance, from about the 16th century onward in Europe, a personal manner of writing grew in importance. The author strove for more or less disguised self-revelation and introspective analysis, often in the form of letters, private diaries, confessions. Also of increasing importance were aphorisms after the style of the ancient Roman philosophers Seneca and Epictetus, imaginary dialogues, and historical narratives, and later, journalistic articles and extremely diverse essays. From the 19th century, writers in Romance and Slavic languages especially, and to a far lesser extent British and American writers, developed the attitude that a literature is most truly modern when it acquires a marked degree of self-awareness and obstinately reflects on its purpose and technique. Such writers were not content with imaginative creation alone: they also explained their work and defined their method in the guise of prefaces, reflections, essays, self-portraits, and critical articles. The 19th-century French poet Charles Baudelaire asserted that no great poet could ever quite resist the temptation to become also a critic: a critic of others and of himself. Indeed, most modern writers, in lands other than the United States, whether they be poets, novelists, or dramatists, have composed more nonfictional prose than poetry, fiction, or drama. In the instances of such monumental figures of 20th-century literature as the poets Ezra Pound, T.S. Eliot, and perhaps William Butler Yeats, or the novelists Thomas Mann and Andre Gide, that part of their output may well be considered by posterity to be equal in importance to their more imaginative writing.

It is virtually impossible to attempt a unitary characterization of nonfictional prose. The concern that any definition is a limitation, and perhaps an exclusion of the essential, is nowhere more apposite than to this inordinately vast and variegated literature. Ever since the ancient Greek and Roman philosophers devised literary genres, some critics have found it convenient to arrange literary production into kinds or to refer it to modes. Periodically, other critics, less prone to dogmatism and distrustful of what has been called rage for order, have ridiculed the contention that an author deliberately sets out to write a novel, a short story, a pastoral tale, a tragicomedy, or a melodrama. The French classical critics of the 17th century, and even more their followers in the 18th and 19th centuries, earnestly taught generations of Europeans that there existed distinct genres, as neatly separate from each other as mammals or vertebrates may be from oviparous or invertebrate species. The theory was driven to ridiculous lengths by Ferdinand Brunetière (1849–1906), who envisaged tragedy, comedy, bourgeois drama, fiction, and even criticism as so many genres coming to existence at a definite time, growing from infancy to childhood and maturity, and then swept away into decline. His conviction, like that of many theorists of genres, was that literature, having divorced itself from the norms of the pure genres, as he saw them, from reason, and from clarity could not but sink into utter confusion and decrepitude in the 20th century.

The hazards of classification

The classification into genres may be comforting to some commentators, aghast at the proliferation of books around them and fearful lest they be not in possession of standards guaranteeing a safe judgment of them. Confusion, however, can also be worse confounded in some doctrines. For genres may be endlessly subdivided into short or long short stories; *récits* or short novels; allegorical novels; parables and fables; didactic novels; fiction *à thèse* (*i.e.*, with a purpose), etc. And that convenient pigeon-holing presupposes that a tragedy should arouse certain emotions and not others, as Aristotle once perhaps taught; that other kinds of literature are composed with a clear and steady purpose in the author's mind; that the mimetic and the didactic do not encroach on each other; and that few writers are such naive mortals as to write merely for the purpose of giving pleasure to their audiences.

Most literary works resist such categorizing. Neither Henry Fielding in the 18th century, nor Stendhal in the 19th, nor Marcel Proust nor Alberto Moravia in the 20th century decided beforehand whether he was writing a mimetic, or a satiric, or a comic novel. Even when an author thought he knew what he was endeavouring to do, as Henry James did, he could not know whether he was not doing only that, and not much else besides. Faith in literary genres has misled many a creator into believing that laws exist for the composing of a short story, a novel, or a fable. Other creators, however, refusing to be imprisoned by those paralyzing rules, plunge into chaos, hoping ultimately to discover their own order and to set standards that will not hamper their talent. Periodically, there is a protest that existing molds have to be broken up and prevailing models ignored, so that the ascending generations may propose their own and chart new territories. Some critics, compounding the difficulty of classifying the unwieldy mass of writing accumulating around them, have even contended that the novel—the genre on which Flaubert, James, and their successors reasoned so diligently in order to endow it with autonomy—does not in truth exist as a new form, but rather as an epic, a tragedy. a comedy, a pastoral, or a form of lyricism.

ELEMENTS

Obviously, a realm as boundless and diverse as nonfictional prose literature cannot be characterized as having any unity of intent, of technique, or of style. It can be defined, very loosely, only by what it is not. Many exceptions, in such a mass of writings, can always be brought up to contradict any rule or generalization. No prescriptive treatment is acceptable for the writing of essays, of aphorisms, of literary journalism, of polemical controversy, of travel literature, of memoirs and intimate diaries. No norms are recognized to determine whether a dialogue, a confession, a piece of religious or of scientific writing, is excellent, mediocre, or outright bad, and each author has to be relished, and appraised, chiefly in his own right. "The only technique," the English critic F.R. Leavis wrote in 1957, "is that which compels words to express an intensively personal way of feeling." Intensity is probably useful as a standard; yet it is a variable, and often elusive, quality, possessed by polemicists and by ardent essayists to a greater extent than by others who are equally great. "Loving, and taking the liberties of a lover" was Virginia Woolf's characterization of the 19th-century critic William Hazlitt's style: it instilled passion into his critical essays. But other equally significant English essayists of the same century, such as Charles Lamb or Walter Pater, or the French critic Hippolyte Taine, under an impassive mask, loved too, but differently. Still other nonfictional writers have been detached, seemingly aloof, or, like the 17th-century French epigrammatist La Rochefoucauld, sarcastic. Their intensity is of another sort.

Reality and imagination. Prose that is nonfictional is generally supposed to cling to reality more closely than that which invents stories, or frames imaginary plots. Calling it "realistic," however, would be a gross distortion. The systematic attempt to convey reality faithfully in art and letters is a late development. It usually occurs in advanced fictional works, long after mankind has become wearied with the unreality of legends, dreams, and myths. Since nonfictional prose does not stress inventiveness of themes and of characters independent of the author's self, it appears in the eyes of some moderns to be inferior to works of imagination. In the middle of the 20th century an immensely high evaluation was placed on the imagination, and the adjective "imaginative" became a grossly abused cliché. Many modern novels and plays, however, were woefully deficient in imaginative force, and the word may have been bandied about so much out of a desire for what was least possessed. Many readers are engrossed by travel books, by descriptions of exotic animal life, by essays on the psychology of other nations, by Rilke's notebooks or by Samuel Pepys's diary far more than by poetry or by novels that fail to impose any suspension of disbelief. There is much truth in Oscar Wilde's remark that "the highest criticism is more creative than creation and the primary aim of the critic is to see the object as in itself it really is not." A good deal of imagination has gone not only into criticism but also into the writing of history, of essays, of travel books, and even of the biographies or the confessions that purport to be true to life as it really happened, as it was really experienced.

Imagination in nonfictional prose

The imagination at work in nonfictional prose, however, would hardly deserve the august name of "primary imagination" reserved by the 19th-century English poet Samuel Taylor Coleridge to creators who come close to possessing semidivine powers. Rather, imagination is displayed in nonfictional prose in the fanciful invention of decorative details, in digressions practiced as an art and assuming a character of pleasant nonchalance, in establishing a familiar contact with the reader through wit and humour. The variety of themes that may be touched upon in that prose is almost infinite. The treatment of issues may be ponderously didactic and still belong within the literary domain. For centuries, in many nations, in Asiatic languages, in medieval Latin, in the writings of the humanists of the Renaissance, and in those of the Enlightenment. a considerable yart of literature has been didactic. The concept of art for art's sake is a late and rather artificial development in the history of culture, and it did not reign supreme even in the few countries in which it was expounded in the 19th century. The ease with which digressions may be inserted in that type of prose affords nonfictional literature a freedom denied to writing falling within other genres. The drawback of such a nondescript literature lies in judging it against any standard of perfection, since perfection implies some conformity with implicit rules and the presence, however vague, of standards such as have been formulated for comedy, tragedy, the ode, the short story and even (in this case, more honoured in the breach than the observance) the novel. The compensating grace is that in much nonfictional literature that repudiates or ignores structure the reader is often delighted with an air of ease and of nonchalance and with that rarest of all virtues in the art of writing: naturalness.

Style. The writing of nonfictional prose should not entail the tension, the monotony, and the self-conscious craft of fiction writing. The search for *le mot juste* ("the precise word") so fanatically pursued by admirers of Flaubert and Maupassant is far less important in nonfictional prose than in the novel and the short story. The English author G.K. Chesterton (1874–1936), who was himself more successful in his rambling volumes of reflections and of religious apologetics than in his novels, defined literature as that rare, almost miraculous use of languge "by which a man really says what he means." In essays, letters, reporting, and narratives of travels, the author's aim is often not to overpower his readers by giving them the impression that he knows exactly where he is leading them, as a dramatist or a detective-story writer does. Some rambling casualness, apparently irrelevant anecdotes, and suggestions of the conclusions that the author wishes his readers to infer are often more effective than extreme terseness.

The role of eloquence

There is also another manner of writing that is more attentive to the periodic cadences and elegance of prose, in the style of the ancient Roman orator Cicero. The 19th-century English essayist William Hazlitt praised the felicities of style and the refinements of the prose of the British statesman Edmund Burke (1729–97) as "that which went the nearest to the verge of poetry and yet never fell over." A number of English writers have been fond of that harmonious, and rhetorical prose, the taste for which may well have been fostered not only by the familiarity with Cicero but also by the profound influence of the authorized version of the Bible (1611). Martin Luther's translation of the New Testament (1522) and of the Old Testament (1534) likewise molded much of German prose and German sensibility for centuries.

In the 20th century that type of prose lost favour with American and British readers, who ceased to cherish Latin orators and Biblical prose as their models. In German literature, however, in which harmonious balance and eloquence were more likely to be admired, and in other languages more directly derived from Latin, a musical style, akin to a prolonged poem in prose, was cultivated more assiduously, as exemplified in Italian in the writings of Gabriele d'Annunzio, in French in those by André Gide, and in German in ***Die Aufzeichnungen des Malte Laurids Brigge (The Notebooks*** of ***Malte Laurids Brigge)*** by the poet Rainer Maria Rilke. Such an elaborate style appears to be more easily tolerated by the readers in nonfictional writing, with its lack of cumulative continuity and, generally speaking, its more restricted size, than in novels such as Pater's ***Marius the Epicurean*** (1885) and occasionally in Thomas Mann's fiction, in which such a style tends to pall on the reader. Similarly, it is easier for the nonfictional prose writer to weave into his style faint suggestions of irony, archaisms, alliterations, and even interventions of the author that might prove catastrophic to credibility in fiction. Critics have argued that too close attention to style was harmful to the sweep necessary to fiction: they have contended that many of the greatest novelists, such as Dickens, Balzac, Dostoyevsky, and Zola at times "wrote" badly; assuredly, they treated language carelessly more than once. Essayists, historians, orators, and divines often affect a happy-go-lucky ease so as to put them on the same footing with the common reader, but they realize that language and style are vital. They must know what resources they can draw from vivid sensations, brilliant similes, balanced sentences, or sudden, epigrammatic, effects of surprise.

Author presence. The one feature common to most authors of nonfictional prose (a few staid historians and even fewer philosophers excepted) is the marked degree of the author's presence in all they write. That is to be expected in epistolary literature, and, although less inevitably, in the essay, the travel book, journalistic reporting, and polemical or hortatory prose. Although the 17th-century French religious philosopher Pascal hinted that "the ego is hateful," the author's presence is still strongly felt. This presence endows their works with a personal and haunting force that challenges, converts, or repels, but hardly ever leaves the reader indifferent. Saint Paul's epistles owe their impact—perhaps second to none in the history of the Western World—to the self that vehemently expresses itself in them, showing no concern whatever for the niceties of Attic prose. In the treatises, discourses, and philosophical argumentation of the great writers of the Enlightenment, such as Voltaire, Diderot, and Rousseau, they frequently resort to the first person singular, which results in a vivid concreteness in the treatment of ideas. To think the abstract concretely, a precept reminiscent of the 18th-century philosophers, was also the aim of the 20th-century philosophers Jean-Paul Sartre and Maurice Merleau-Ponty when they naturalized Existentialist thought in France.

The growth of personal literature in its myriad shapes is one of the striking features of modern literary evolution. In France, Italy, and Germany it invaded fiction under the banner of the personal novel. With Charles Baudelaire in poetry and with Marcel Proust in prose, memory appears to have stepped in the place of imagination as the main force in lyrical and fictional literature of the 20th century. Inevitably, such fiction is obsessed with the presence of him who is reminiscing, since "one only remembers oneself," as the philosopher Henri Bergson put it.

The very fact that the writer of nonfictional prose does not seek an imaginary projection to impart his vision, his anguish, and his delights to readers also underlines the nature of his intention. A school of critics has vigorously attacked "the intentional fallacy," which leads biographers and some literary historians to ask what an artist intended before evaluating the completed work of art. In dramatic, fictional, and lyrical works, the artist only discovers his own intentions through the process of creation; the finished work may be a product alien to him, at least to the part of him that is conscious and lucid. But in a work of apologetics or of homiletics, in a work of history or of sociology, in a critical or even in a desultory and discursive essay, and certainly in aphorisms or maxims or both, the intention of the author remains omnipresent. This intention may be disguised under the mask of a parable, under the interlocutors of a philosophical dialogue, or under the admonitions of a prophet, but the reader is never oblivious of the thinker's intent. His pleasure is not that of an addict to thrillers or of the devotee of fiction who hastens to devour incidents and metamorphoses of the characters to rush to the denouement. It is the more sophisticated enjoyment of one who shares the creator's intent and travels familiarly along with him. He respects and enjoys in those authors the exercise of an intelligence flexible enough to accept even the irrational as such.

APPROACHES

In terms of approach, that is, the attitude of the writer as it can be inferred from the writing, the distinguishing features of nonfictional prose writings are the degree of presence of the ego and of the use of a subjective, familiar tone. Such devices are also used, of course, by authors of fiction, but to a lesser extent. Similarly, the basic modes of writing—the descriptive, the narrative, the expository, and the argumentative—are found in both nonfictional literature and in fiction, but in different degrees.

The descriptive mode. Figuring prominently in the fiction of the 19th century, the descriptive mode found a much smaller role in the later novel, except perhaps in that of Russia. In nonfictional prose, however, essayists, moralists, naturalists, and others regularly evoked nature scenes. The most sumptuous masters of prose composed landscapes as elaborately as landscape painters. The French writer and statesman Chateaubriand (1768–1848), for example, who was not outstandingly successful in inventing plots or in creating characters independent from his own self, was a master of description; his writings influenced the French Romantic poets, who set the impassive splendour of outward nature in contrast to the inner anguish of mortals. The 19th-century English art critic John Ruskin had a more precise gift of observation, as revealed in his descriptions of Alpine mountains and of the humblest flowers or mosses, but his ornate and sonorous prose was the climax of a high-flown manner of writing that later read like the majestic relic of another era. American nonfictional writers of the same period such as Ralph Waldo Emerson and Henry Thoreau scrupulously described the lessons of organization, of unity, and of moral beauty to be deciphered from the vicissitudes of nature. Russian essayists vied with novelists in their minute yet rapturous descriptions of the thaw releasing the torrents of spring or the implacable force of the long Northern winters. Writers more inclined to the observation of social life, in satirical sketches of the mechanically polite and artificial habitués of salons, helped the novel of social life come into existence in several Western countries.

Narrative. The narrative element is less conspicuous in writing that does not purport to relate a story than in fictional works, but there is a role for narrative in letters, diaries, autobiographies, and historical writing. Most often, an incident is graphically related by a witness, as in letters or memoirs; an anecdote may serve to illustrate a moral advice in an essay; or an entertaining encounter may be inserted into an essay or a travel sketch. Digression here represents the utmost in art; it provides a relief from the persistent attention required when the author is pursuing his purpose more seriously. Similarly, such writing provides a pleasant contrast to the rigid structure of the majority of novels since the late 19th century. In historical writing, however, simplicity and clarity of narrative are required, though it may be interspersed with speeches, with portraits, or with moral and polemical allusions. The tampering with time and the flashbacks of modern fiction, however, would strain the credibility of a reader of history to an unbearable point. In other forms of nonfictional prose, the meandering fancy of the author may well produce an impression of freedom and of truth to life unattainable by the more carefully wrought novel. Many writers have confessed to feeling relieved when they ceased to create novels and shifted to impromptu sketches or desultory essays. The surrealist essayists of the 20th century poured their scorn on detective fiction as the most fiercely logical form of writing. In contrast, the author of essays or other nonfictional prose may blend dreams and facts, ventures into the illogical, and delightful eccentricities.

Oratory

Expository and argumentative modes. The rules of old-fashioned rhetoric apply better to expository and argumentative prose than to the other modes. These rules were first set down in ancient Greece by teachers who elicited them from the smooth eloquence of Socrates, the impassioned and balanced reasoning of Demosthenes, and others. The ancient Romans went further still in codifying figures of speech, stylistic devices, and even the gestures of the orator. Such treatises played a significant part in the education of the Renaissance humanists, of the classical and Augustan prose writers of 17th-century England and France, of the leaders of the French Revolution in the 18th century, and even in 19th-century historians and statesmen such as Guizot in France and Macaulay and Gladstone in Britain. And in the middle of the 20th century teachers of English composition, in America in particular, have attempted to reformulate and modernize those rules of rhetoric. Confronting the difficulty of setting standards of style and evaluating writing, they resorted to training students in time-honoured, and often outworn, rhetorical devices such as metonymy, anacoluthon, synechdoche, oxymoron, etc. But the sophisticated oratory of such 18th-century British orators as Richard Brinsley Sheridan, Edmund Burke, and Charles Fox or, more recently, that of Winston Churchill, hardly seems attuned to audiences in the age of television.

It has been suggested by students of German history that Adolf Hitler, in his vituperative speeches at Nuremberg in the 1930s, fascinated the Germans because they had been unaccustomed, unlike other Western nations, to eloquence in their leaders. If a large part of a population is illiterate, such as the Cubans under Fidel Castro, unending flows of eloquence may constitute a convenient means of educating the masses. Elsewhere, a more familiar and casual type of address from political leaders tends to be preferred in an era of mass media. The gift of a superior orator has been facetiously defined as that of saying as little as possible in as many words as possible. Like sermons, many types of formal address such as lectures, political speeches, and legal pleadings appear to be doomed as documents of literary value, as Burke's or Lincoln's orations and addresses were when they were learned by heart by the younger generations and helped mold the style and contribute to the moral education of men.

II. Forms of nonfictional prose

Both in the substance and in the form, nonfictional prose is much less subject to the influence of antecedents in ancient literatures, to conventions, to rules of rhetoric, and to stylistic conventions than other types of literature. Its vitality derives from its use of the language current in the author's time, flavoured with suggestions of the study of past masters. Quaintness is only bearable for a limited time, and archaisms soon detract from the impact on the readers. Virtuoso displays of verbal fireworks and of felicities of style make the writer's presence excessively obtrusive. The normal evolution of the writer's style has been from a youthful delight in the cadences and the alliterative and rhetorical effects of prose to a sinewy robustness and a deliberate stripping of adornments. Montesquieu, Rousseau, Flaubert, and Gide, to mention a few of those credited with writing of the most perspicuous and most accomplished prose, illustrate that trend.

THE ESSAY

In modern literatures, the category of nonfictional prose that probably ranks as the most important both in the quantity and in the quality of its practitioners is the essay.

Modern origins. Before the word itself was coined in the 16th century by Montaigne and Bacon, it was called a "treatise," and its attempt to treat a serious theme with consistency deprived it of the seductive charm relished in the later examples of that form of literature. In this sense, the word "essay" would hardly fit the didactic tone of Aristotle's ***Rhetoric*** or his ***Metaphysics.*** There were, however, ancient masters of an early form of the essay, such as Ciceso discoursing on the pleasantness of old age or on the art of "divination"; Seneca, on anger or clemency; and Plutarch, more superficially and casually, on the passing of oracles. The relentless desire to analyze one's own contradictions, especially among Christians, who, like Saint Paul, were aware of their duality and of "doing the evil which they would not," also contributed to the emergence of the essay. But Christian writing tended to be highly didactic, as may be seen in the work of Saint Augustine of the 5th century, or of the 12th-century theologian Abélard, or even in the Latin writings on "the solitary life" or on "the scorn of the world" by the 14th-century Italian poet Petrarch. Not until the Renaissance, with its increasing assertion of the self, was the flexible and deliberately nonchalant and versatile form of the essay perfected by Montaigne.

Montaigne's pre-eminence as an essayist

Montaigne left his mark on almost every essayist who came after him in continental Europe, and perhaps even more in English-speaking countries. Emerson made him one of his six ***Representative Men*** along with others of the stature of Plato, Shakespeare, and Goethe. Hazlitt lauded Montaigne's qualities as precisely those that "we consider in great measure English," and another English romantic writer, Leigh Hunt, saw him as "the first man who had the courage to say as an author what he felt as a man." And the 20th-century poet T.S. Eliot declared him to be the most important writer to study for an insight into the literature of France. With Montaigne, the essay achieved for the first time what it can achieve better than any other form of writing, except perhaps the epistolary one: a means of self-discovery. It gave the writer a way of reaching the secret springs of his behaviour, of seizing the man and the author at once in his contradictions, in his profound disunity, and in his mobility. The essay was symbolic of man's new attitude toward himself, revelling in change, and hence in growth, and forsaking his age-old dream of achieving an underlying steadfastness that might make him invulnerable and similar to the gods. Now he set out to accept himself whole, with his body and his physical and behavioural peculiarities, and thereby repudiate medieval asceticism. He would portray his foibles and unworthiness, hoping to rise above his own mediocrity, or, at the other extreme, he would exalt himself in the hope that he might become the man he depicted. Montaigne in his essays pursued an ethical purpose, but with no pompousness or rhetoric. He offered an ideal that was adopted by his successors for 200 years: perfecting man as a tolerant, undogmatic, urbane social being. But, unlike medieval Christian writers, he would not sacrifice to others the most dearly cherished part of himself. To others he would lend himself, but his personality and his freedom were his own, and his primary duty was to become a wiser human being.

No essayist after Montaigne touched on so many varied aspects of life with such an informal, felicitous, and brilliant style. The later writers who most nearly recall the charm of Montaigne include, in England, Robert Burton, though his whimsicality is more erudite, Sir Thomas Browne, and Laurence Sterne, and in France, with more self-consciousness and pose, Andrd Gide and Jean Cocteau.

The uses of the essay. In the age that followed Montaigne's, at the beginning of the 17th century, social manners, the cultivation of politeness, and the training of an accomplished gentleman became the theme of many essayists. This theme was first exploited by the Italian Baldassare Castiglione in his ***Il cortegiano*** (1528; ***The Courtier).*** The influence of the essay and of genres allied to it, such as maxims, portraits, and sketches, proved second to none in molding the behaviour of the cultured classes, first in Italy, then in France, and, through French influence, in most of Europe in the 17th century. Among those who pursued this theme was the 17th-century Spanish Jesuit Baltasar Gracián in his essays on the art of worldly wisdom.

With the advent of a keener political awareness with the age of Enlightenment, in the 18th century, the essay became all-important as the vehicle for a criticism of society and of religion. Because of its flexibility, its brevity, and its potential both for ambiguity and for allusions to current events and conditions, it was an ideal tool for philosophical reformers. ***The Federalist Papers*** in America and the tracts of the French Revolutionaries, are among the countless examples of attempts during this period to improve the condition of man through the essay.

The advantage of this form of writing was that it was not required to conform to any unity of tone or to similar strictures assigned to other genres since it was for a long time not even considered a genre. After ponderous apologies for traditional faith failed to repulse the onslaught of deism and atheism, traditionalists of the 19th and 20th centuries, such as Burke and Coleridge, abandoned unwieldy dogmatic demonstrations in favour of the short, provocative essay. In the defense of the past, it sewed as the most potent means of educating the masses. French Catholics, German pietists, and a number of individual English and American authors confided to the essay their dismay at what they saw as modern vulgarity and a breakdown of the coherence of the Western tradition. Essays such as Paul Elmer More's long series of *Shelburne* ***Essays*** (published between 1904 and 1935), T.S. Eliot's ***After Strange Gods*** (1934) and ***Notes Towards*** *the* ***Definition*** of ***Culture*** (1948), and others that attempted to reinterpret and redefine culture, established the genre as the most fitting to express the genteel tradition at odds with the democracy of the new world.

Journalistic essays

The proliferation of magazines in the United States, and the public's impatience with painstaking demonstrations and polemics, helped establish the essay just as firmly as a receptacle for robust, humorous common sense, unpretentiously expressed, as in the writings of Oliver Wendell Holmes (1809–94). Creative writers resorted to it to admonish their compatriots when they seemed too selfishly unconcerned by the tragedies of the world. Archibald MacLeish, for instance, did so in ***A Time to Speak*** (1941). Lewis Mumford, Allen Tate, and other literary and social critics became crusaders for moral and spiritual reform; others seized upon the essay for scathingly ironical and destructive criticism of their culture: for example, James Gibbons Huneker (1860–1921), an admirer of iconoclasts and of egoists, as he called them, proposed European examples to Americans he deemed to be too complacent and lethargic; and, more vociferously still, H.L. Mencken (1880–1956), a self-appointed foe of prejudices, substituted his own for those he trounced in his contemporaries.

In other new countries. or in cultures acquiring an awareness of their own ambitious identity, the essay became semipolitical, earnestly nationalistic, and often polemical, playful, or bitter. Such essays sometimes succeeded in shaking the elite out of its passivity. In Uruguay, for example, Josd Enrique Rod6 (1872–1917), in an analogy to the characters in Shakespeare's ***Tempest,*** compared what should be the authentic South American to the spirit Ariel, in a work thus entitled, in contrast to the bestial Caliban, representing the materialism of North America. In Canada Olivar Asselin (1874–1937) used the essay to advocate the development of a genuine French-Canadian literature. Among the older cultures of Europe, Salvatore Quasimodo (1901–1968), the Italian poet and Nobel laureate, appended critical and hortatory essays to some of his volumes of verse, such as *Il falso* ***e vero verde*** (1956; "The True and False Green"). Other European heirs to this tradition of the essay include Stefan Zweig and Hugo von Hofmannsthal in Austria and Thomas Mann and Bertolt Brecht in Germany; their sprightly and incisive essays on the arts can be traced to the 19th century German philosopher Arthur Schopenhauer.

Essays for entertainment

One of the functions of literature is to please and to entertain; and the essay, as it grew into the biggest literary domain of all, did not lose the art of providing escape. Essayists have written with grace on children, on women, on love, on sports, as in Robert Louis Stevenson's collection ***Virginibus Puerisque*** (1881), or Willa Cather's pleasant reflections in ***Not under Forty*** (1936). Ernest Renan (1823–92), one of the most accomplished French masters of the essay, found relief from his philosophical and historical studies in his half-ironical considerations on love, and Anatole France (1844–1924), his disciple, and hosts of others have alternated playful essays with others of high seriousness. Sports, games, and other forms of relaxation have not been so often or so felicitously treated. Izaak Walton's ***The Compleat*** ***Angler*** (1653), however, enjoys the status of a minor classic, and the best of the modern Dutch essayists, Johan Huizinga (1872–1945), has reflected with acuteness on ***Homo*** *ludens,* or man at play. A Frenchman, Jean Prdvost (1901–44), who was to die as a hero of the Resistance to the German occupation of France during World War II, opened his career as an essayist with precise and arresting analyses of the Plaisirs des sports (1925). But there are surprisingly few very significant works, except in chapters of novels or in short stories, on the joys of hunting, bullfighting, swimming, or even, since Anthelme Brillat-Savarin's overpraised essay, Physiologie du *goût* (1825; "The Physiology of Taste") on gourmet enjoyment of the table.

Serious speculations, on the other hand, have tended to overburden the modern essay, especially in German and in French, and to weigh it with philosophy almost as pedantic as that of academic treatises, though not as rigorous. The several volumes of Jean-Paul Sartre's Situations, published from 1947 on, constitute the most weighty and, in the first two volumes in particular, the most original body of essay writing of the middle of the 20th century. Albert Camus' *Mythe* de Sisyphe (1942; Myth of Sisyphus) and his subsequent *Homme révolté* (1951; The Rebel) consist of grave, but inconsistent and often unconvincing, essays loosely linked together. Émile Chartier (1868–1951), under the pseudonym Alain, exercised a lasting influence over the young through the disjointed, urbane, and occasionally provoking reflections scattered through volume after volume of his essays, entitled Propos.

Political essays

Apart from philosophical speculation, which most readers prefer in limited quantities, the favourite theme of many modern essays has been speculation on the character of nations. It is indeed difficult to generalize on the national temper of a nation or on the characteristics of a given culture. The authors who have done it—Emerson in his essay on English Traits (1856), Hippolyte Taine in his studies of the English people, Alexis de Tocqueville in his Democracy in America (1835, 1840)—blended undeniable conclusions with controversial assertions. Rather than systematic studies, desultory essays that weave anecdotes, intuitions, and personal remarks, ever open to challenge, have proved more effective in attempting to delineate cultures. In the 20th century, the masters of this form of writing have been among the most able in the art of essay writing: Salvador de Madariaga in Spanish, Hermann Keyserling in German, and Elie Faure in French. Some nations are much more prone than others to self-scrutiny. Several of the finest Spanish essayists were vexed by questions of what it meant to be a Spaniard, especially after the end of the 19th century when Spain was compelled to put an end to its empire. Angel Ganivet in his essay on *Idearium español* (1897; Spain: An Interpretation), Ortega y Gasset in *España* invertebrada (1922; Invertebrate Spain), and Miguel de Unamuno in almost every one of his prose essays dealt with this subject. A Spanish-born essayist, George Santayana (1863–1952), was one of the most accomplished masters of written English prose; because of his cosmopolitan culture and the subtlety of his insights, he was one of the most percipient analysts of the English and of the American character.

Laments on the decline of the essay in the 20th century have been numerous since the 1940s, when articles in most journals tended to become shorter and to strive for more immediate effect. As a result, the general reader grew accustomed to being attacked rather than seduced. Still, the 20th century could boast of the critical essays of Virginia Woolf in England, of Edmund Wilson in America, and of Albert Thibaudet and Charles du Bos in France, all of whom maintained the high standards of excellence set by their predecessors of the previous century. It is regrettable that, in the language in which the best modern essays have been written, English, the term "essay" should also have acquired the connotation of a schoolboy's attempts at elementary composition. For the essay requires vast and varied information, yet without pedantry or excessive specialization. It must give the impression of having been composed spontaneously, with relish and zest. It should communicate an experience or depict a personality with an air of dilettantism, and of love of composition, and it should make accessible to the reader knowledge and reflection and the delight of watching a fine mind at work. The essayist should possess the virtues that one of the most influential English essayists, Matthew Arnold, praised in Culture and Anarchy (1869): "a passion . . . to divest knowledge of all that was harsh, uncouth, difficult, abstract, professional, exclusive; to humanize it."

HISTORICAL AND CRITICAL FORMS

History as literature. Among the ancient Greeks and Romans, history was the branch of literature in which the most expert and the most enduring prose was written. It only recovered its supreme rank in nonfictional prose in the 18th century. Earlier, however, at the beginning of the 16th century, in Florence, Italy, Niccolb Machiavelli and Francesco Guicciardini prepared the way for history to become great literature by marrying it to the nascent science of politics and by enlarging its scope to include elements of the philosophy of history. In the 18th century Voltaire, tersely and corrosively, and Edward Gibbon, with more dignity, established history again as one of the great literary arts. In the 19th century, their lessons were taken to heart, as writers and readers realized that, in Thomas Carlyle's words, "every nation's true Bible is its history." In some nations: historians, together with epic and political poets, instilled into the people a will to recover the national consciousness that had been stifled or obliterated. Macaulay's ambition, to see history replace the latest novel on a lady's dressing table, was endorsed as an eminently reasonable and beneficent ambition by scholars throughout the 19th century. After an eclipse during the first half of the 20th century, when erudition and distrust of elaborate style prevailed, the poetry of history has again been praised by the most scrupulous practitioners of that discipline. Poetry, in that context, does not mean fiction or unfaithfulness to facts, or a mere prettification, which would be tantamount to falsification; rather, it is the recognition that, as G.M. Trevelyan, Regius professor at Cambridge University, proclaimed "The appeal of history to us all is in the last analysis poetic." Few historians today would wholly agree with the once sacrosanct formula of Leopold von Ranke (1795–1886) that their task is to record the past as it really took place. They well know that, for modern history, facts are so plentiful and so very diverse that they are only meaningful insofar as the historian selects from them, places them in a certain order, and interprets them. Since World War II, as history drew increasingly on sociology, anthropology, political and philosophical speculation, and psychoanalysis, the conviction that objectivity could be maintained by a scholar dealing with the past came to be questioned and in large measure renounced. The Italian philosopher Benedetto Croce's laconic warning that all history is contemporary history (*i.e.*, bound to the historian's time and place, hence likely to be replaced by another one after a generation) has come to be generally accepted. Nietzsche, who had sharply questioned the historical methods of his German countrymen in the 1870s, stressed the need to relate history to the present and to present it in a living and beautiful form, if it is to serve the forces of life. "You can only explain the past," he said, "by what is highest in the present."

Interpretive history

In Germany, Italy, Spain, England, America, and most of all in France, where the vogue of sheer, and often indigestible, erudition was never wholeheartedly adopted, more literary talent may have gone into historical writing than into the novel or the short story. Many reasons account for the brilliance, and the impact, of this branch of nonfictional prose. Modern man has a powerful interest in origins—of civilization, of Christianity, of the world initiated by the Renaissance, or the French Revolution, or the rise of the masses. History invites an explanation of what is in terms of its genesis, not statically but in the process of becoming. The breadth of men's curiosity has expanded significantly since the 18th century, when belief in the absolutes of religious faith tended to be supplanted by greater concern for the relative world in which men live, move, and exist. A primary factor in the increasing importance of history is the bewilderment concerning the revolutions that occurred in or threatened so many countries in the latter part of the 20th century. As fiction, philosophy, and the exact sciences failed to provide a plausible explanation, many anguished readers turned to the record of brutal change in earlier periods. The historians who addressed themselves to those immense subjects, with their myriad ramifications, often composed monumental works of a synthetical character, such as those of Arnold Toynbee or Henri Pirenne, but they also cultivated the essay. Sometimes these essays appeared as short and pregnant volumes of reflections, such as Isaiah Berlin's ***Historical Inevitability*** (1954), sometimes in collections of articles that first appeared in magazines.

20th-century academic critics

Criticism. If the 19th century was the golden age of history, the 20th deserves the title of the age of criticism. The spread of higher education in advanced countries resulted in vast numbers of young scholars learning to write critical and analytical essays and theses. The training of imagination, however, or even of sensitivity to works of art, failed to keep pace with the development of the critical faculties, probably because it poses much more difficult problems to educators. The mass of writing produced in book form combined with the limited leisure time for reading created a need for a growing number of advisers to the public. Many gifted persons were drawn to literary criticism and, to a lesser degree, to criticism of music, painting, architecture, and the cinema. When they apply themselves to current productions, these critics inevitably proffer value judgments, informing their readers of the merits and demerits of a book or of the success or failure of a show. Thus, a significant function of these critics is to convert their readers--or to prevent their conversion—into purchasers, devotees, supporters, and patrons. Critics who seek to express themselves in less hurried fashion in regard to creative work prefer to write on the past, to re-evaluate, for example, a Shakespearean play or a 19th-century Russian novel. Criticism of the second sort has tempted most academic writers on literature and the arts in the 20th century. They feel, by instinct and by profession, closer to the past. Thus, they would be less likely to repeat the disastrous mistakes of their predecessors, who had failed to acknowledge the greatness of Beethoven, Holderlin, Shelley, Baudelaire, Cézanne, and most other great talents between 1800 and 1920. These academic critics have perfected techniques for the analysis of texts and for the discovery of structure hidden behind disorder. They can discern symbolic or esoteric meanings in the simplest of poems and plays. One of their faults is that, having acquired more subtlety than the creator himself, and being eager to descry layers of meaning or of allusions unperceived by others, they tend to display chiefly their own finer subtlety. The often unsophisticated, or uncouth, imaginative author, or the instinctive craftsman, such as the poet Rimbaud, the sculptor Brancusi, or the painter Soutine, becomes a mere pretext for the abstruse lucubrations of the critic.

Authors of imaginative works have bitterly inveighed against critics who allegedly failed to understand them ever since the Romantic era. Each artist must believe in the integrity and in the uniqueness of what he is accomplishing. Artists who are original are of necessity men who refuse to be the followers of an earlier tradition. Since they are different from their predecessors, they realize that they must overstress that difference combatively, that they must themselves create the taste according to which they will eventually be appreciated. The proliferation of critics around them thus leads them to become more provocative iconoclasts than they might naturally have been.

The 20th century has seen a burgeoning of critics and reviewers, many of them highly sophisticated, invading more and more journals with their prose, in New York, London, and Paris. Many literary talents who in earlier times might have been drawn to fiction or to drama have been attracted to criticism instead. Criticism offers a readier outlet for their writing, and, owing to the support of academic presses and foundations and to teaching and lecturing positions, it enjoys generous subsidies. Criticism has been called the substitute Muse of the universities.

Still, some of the huge mass of criticism that has grown up around literature and the arts since the Renaissance is nonfictional prose of merit. Most likely to endure is the occasional criticism by creators: the pronouncements of Molière, Wordsworth, or Goethe, and their judgments of others, are enlightening and irreplaceable. No statue was ever raised to a critic, remarked the irate and long-neglected Finnish composer Jean Sibelius. Nonetheless, critical works that belong to the literary heritage of the Western world include such work as the defenses of poetry of Joachim du Bellay (1522–60) and John Dryden (1631–1700), the partial but unforgettable verdicts of Dr. Samuel Johnson (1709–1784), and the critical works of Denis Diderot (1713–84) and Gotthold Lessing (1729–81).

The leisurely critical essay of the 19th century

A leisurely, Epicurean, urbane, and gently ironical type of critical essay was practiced in the 19th century. It is read to this day by students of letters and by those nostalgic for a bygone age, which they envision as less subject to the worship of vulgarity than their own. The French critic Charles Sainte-Beuve (1804–69), notwithstanding his gross unfairness to almost all his contemporaries, remains the master of a type of biographical and psychological criticism unmatched by his sucessors. Hippolyte Taine (1828–93), when he was not carried away by his rigid formulas, composed remarkable and very influential essays on Balzac, Goethe, and on a number of English writers. He attempted to explain the appearance of a work of genius and to analyze it in its inner structure. Less professional and more intimate and humorous in tone were several British critics of the late Victorian era, such as Leslie Stephen (1832–1904) and George Saintsbury (1845–1933). The most sedulously artistic master of prose among them, and long the most prestigious, was Walter Pater (1839–94).

That art has not completely died in the 20th century. Remy de Gourmont and Paul Valéry in France, Virginia Woolf in England, Lionel Trilling in America, and Dámaso Alonso in Spain have sustained the tradition of writing penetrating and original critical essays that stand as minor literary masterpieces in their own right. They never forgot that the critic must necessarily live at second hand and that his craft remains the handmaid of creation. The American critic Edmund Wilson, especially in his early volumes of essays on French and Russian authors and on Dickens, was, if not the most creative, certainly the most influential man of letters of his generation. Most of these critics were appraisers of already recognized works rather than pioneers venturing into the new and establishing young reputations. These critics, and probably few others, rank among the masters of nonfictional prose, along with the creators themselves, when they condescend to write as critics either of their own work or of that of others, such as Goethe, Stefan George, Victor Hugo, Baudelaire, Zola, Gide, T.S. Eliot, or D.H. Lawrence.

DOCTRINAL. PHILOSOPHICAL, AND RELIGIOUS PROSE

The question of how much of doctrinal writing, dealing with faith, ethics, and philosophy, can be called literature can only be answered subjectively by each reader, judging each case on its own merits. There have been philosophers who felt in no way flattered to be included among what they considered unthinking men of letters. The prejudice lingers in some quarters that profundity and clarity are mutually exclusive and that philosophy and social sciences therefore are beyond the reach of the layman. On the other hand, many writers, while often profound and fastidiously rigorous in their thought, such as Paul Valéry, have vehemently objected to being called philosophers. Nonetheless, a vast number of philosophical works owe their influence and perhaps their greatness to their literary merits.

In periods when philosophical speculation became very abstruse, as in Germany in the 19th century, men of letters often acted as intermediaries between the highly esoteric thinkers and the public. Much of the impact of the erudite 19th-century German philosopher Georg Wilhelm Friederich Hegel was due to the more easily approachable writings of those who took issue with him, such as the existentialist thinker Søren Kierkegaard, or to those who reinterpreted him, such as Karl Marx. Similarly, the thoughts of 20th-century German phenomenologist Edmund Husserl achieved wider circulation by receiving more literary expression in the writings of Jean-Paul Sartre. In modern Europe, the men of letters of Germany were long the most deeply imbued with abstract philosophy. After World War II, however, French writers appeared to take on a zest for abstract speculation, for turgid prose, and for the coining of abstruse terms. Much of French literature in the years after the war has been characterized as "literature as philosophy."

A very few philosophers have reached greatness by evolving a coherent, comprehensive system, ambitiously claiming to account for the world and man. Such harmonious constructions by the greatest philosophers, such as Descartes and Spinoza, might be compared to epic poems in sometimes embracing more than there actually appears to be between heaven and earth. These philosophical systems were conceived by powerful imaginative thinkers whose creative abilities were not primarily of an aesthetic order. The ability and the ambition to produce such systems has appeared in very few countries or cultures. The Slavic, the Spanish, and Spanish-American cultures have been richer in thinkers than in philosophers; that is, in men who reflected on the problems of their own country, who attempted to evolve a philosophy from history, or who applied a broad view to moral or political questions, rather than in men who constructed abstract philosophical systems.

More and more in the 20th century, the sciences that are called in some countries "social" and in others "humane" have replaced the all-encompassing philosophical systems of past ages. In Spain, Miguel de Unamuno (1864–1936) and José Ortega y Gasset (1883–1955) marked the thought and the sensibility of Spanish-speaking peoples far more than systematic philosophers might have done. Their writing, which disdains impeccable logic, is no less thought-provoking for being instinct with passion and with arresting literary effects.

Russian philosophical essays

In Russia, the doctrinal writers whose thought was most influential and often most profound were also those whose prose was most brilliant. They generally centred their speculations on two Russian preoccupations: the revival of Christian thought and charity in the Orthodox faith; and the relationship of Russia to Western Europe, branded by the Slavophiles as alien and degenerate. The consistency of ancient Greek and later Western thinkers, from Aristotle through Descartes, was of scant concern to them, but in the vitality of their style, some of these Russian theorists were masters, whose turbulent, paradoxical ideas were taken to heart by novelists, poets, and statesmen. Among these masters, Aleksandr Herzen (1812–70) combined romantic ardour and positivism, formulating a passionately Russian type of socialism; he left his mark in autobiography, political letters, fiction, and chiefly philosophy of history in ***From the Other Shore*** (1851). Nikolay Danilevsky (1822–85), a scientist who turned to philosophy, attempted to convince his compatriots that the manifest destiny of their country was to offer a purer and fresher ideology in lieu of that of the decadent West. V.V. Rozanov (1856–1919) was an apocalyptic prophet preaching an unusual interpretation of Christian religion; a number of his intuitions and passionate assertions are found in the novel ***The Possessed*** (1871–72), by Fyodor Dostoyevsky, whose own nonfictional prose is of considerable quality and conviction. The strangest and most contradictory, but also the most brilliant prose writer, among those thinkers who were torn between East and West, between a jealous Orthodox faith and the attraction of Catholic Rome, was Vladimir Solovyov (1853–1900). He blended the most personal type of visionary mysticism with an incisive humour in a manner reminiscent of Kierkegaard. His philosophical essay-dialogue-treatise, ***Three Conversations on War, Progress and the End of Human History*** (1900), is representative of the nonfictional Russian prose that, while not widely known outside Russia, is as revealing as the Russian novel of the permanent contradictions and aspirations of the Slavic character.

The American novel was never as close to the religious and political ideas of the country as imaginative literature was elsewhere. Yet, the role of nonfictional prose in the American literature of ideas is significant, as can be seen in several of Emerson's philosophical essays and addresses; in Walt Whitman's *Democratic* ***Vistas*** (1871); in William James's pleasantly written essays on religious experience and on sundry psychological and ethical topics; in George Santayana's dexterous and seductive developments on beauty, on nature, on poets, on the genteel tradition, all envisaged with ironical sympathy. Irving Babbitt (1865–1933), Thorstein Veblen (1857–1929), and Lewis Mumford (1895–) are among the many American writers who, in the 20th century, maintained the tradition of writing on abstract or moral themes with clarity and elegant simplicity. Earlier, Thomas Jefferson and Benjamin Franklin had expressed their lay philosophy in a manner they wished to be widely accessible.

In France the tradition ***haute vulgarisation — "higher*** vulgarization" or popularization—never died and was seldom slighted by the specialists. There, and to a slightly lesser extent in Britain, much of the most valuable writing in prose was an elucidation of the view of life underlying the creations of eminent men in many fields. Such doctrinal writing, expounding innermost convictions and sometimes representing a diversion from more intensive pursuits, constitutes a by no means negligible portion of the writings of the philosopher Bertrand Russell, of the poet William Butler Yeats, and others. The novelist or the poet may well use nonfictional prose to purge his own anger, to give vent to his vituperation against his confrères, and to relieve his imagination of all the ideological burden that might otherwise encumber it. D.H. Lawrence preserved the purity of his storyteller's art by expressing elsewhere his animadversions against Thomas Hardy or Sigmund Freud. Albert Camus stripped his fiction and short stories of the ideological musings found in his philosophical volumes. Marcel Proust succeeded in incorporating many abstract discussions of the value of art, love, and friendship in his very original and loose type of fiction; but his great work, A ***la recherche du temps perdu*** (1913–27; *Remembrance* ***of Things Past)*** might well have gained even more from the excision of those dissertations and the writing of more volumes like his ***Ckroniques*** (1927) or *Contre Sainte-Beuve* (1954). The masters of nonfictional prose in French in the 20th century have been those thinkers who were also superb stylists and who deemed it a function of philosophy to understand the aesthetic phenomenon: Henri Bergson (1859–1941), Paul Valéry (1871–1945), and Gaston Bachelard (1884–1962). No more poetical advocate of reverie has arisen in the 20th century than ***La Poétique de la*** *rêverie* (1960; ***The Poetics of Reverie)*** and the posthumous collection of essays, ***Le Droit de*** *rêver* (1970; ***"The Right to Dream"),*** by Bachelard, who was also a philosopher of science. A major influence on him, as on several earlier poets endowed with profound intellect, such as Baudelaire and Valéry, was Edgar Allan Poe, the impact of whose essays on poetics, on cosmology, and on dreams and reveries has been immense and beneficent. More than a century after his death, many of Poe's American compatriots have conceded that the storyteller and the poet in Poe counted for less, as his European admirers had divined, than the writer of critical and doctrinal prose rich in dazzling intuitions.

Religious literature

Although lectures, articles, and other prosaic admonitions have tended to take their place, sermons, funeral orations, allegories, and the visions of eternal punishment brandished by theologians constitute some of the most unforgettable prose. This form of nonfictional prose literature dates from before the Christian Era; Jewish thought and style were molded by commentaries on the Old Testament and compilations of the wisdom of the sages. Later, and more nearly literary, works of this nature include Sebastian Brant's didactic, poetical, and satirical ***Narrenschiff*** (1494; ***Ship of Fools),*** and the mystic writings of Jakob Bohme (1575–1624) in Germany, the moving sermons of Jón Vidalin (1666–1720) in Iceland. In England, Richard Baxter (1615–91) and John Bunyan

(1628–88) were among the most eloquent of the 17th-century Puritans who composed doctrinal works of literary merit; along with the epic poet John Milton (1608–74), whose prose works hardly count for less than his poetry, they exercised a powerful influence on the English language through their doctrinal prose. Their contemporary, the Anglican Jeremy Taylor (1613–67), wrote the most sustained and dignified prose of an age that, on the continent, would be called Baroque. A little later, in northern Europe, the Norwegian Ludvig Holberg (1684–1754), who spent most of his life in Denmark and became best known as a comic writer, also advised his contemporaries how to live morally in his *Ethical Thoughts* and other didactic treatises. The Swede Emanuel Swedenborg (1688–1772), less gifted as a writer but far more original in his blend of mysticism and science, outshone all previous Scandinavians in impressing the imagination of other Europeans. No less influential, Søren Kierkegaard (1813–55), because of his stimulating ambiguities, his bold treatment of traditional theology and philosophy, and his extraordinary ability to write vivid, biting, and provoking prose, was, a century after his death, one of the most potent forces in the literature and thought of Western civilization.

Many 20th-century readers experience a feeling of remoteness in this kind of doctrinal writing, which stems in part from a lack of vital interest in the beliefs it embodies and from a coolness toward religious dogmatism or fanaticism. Intolerance has shifted from religion to the domain of politics. But the contemporary estrangement from that rich literary heritage is due also to a distrust of high-flown eloquence. Cotton Mather's *Essays to do good* (1710) has few readers in present-day New England, despite that region's Puritan tradition, and Jonathan Edwards (1703–58), a writer of great spiritual warmth and imaginative style who was the first of the great prose writers of America, is admired today chiefly by specialists.

A less sonorous style, one that does not ring so monotonously ornate to the reader's ears, is now preferred. In Spain, Antonio de Guevara (*c.* 1481–1545), a preacher who was at his best in his familiar and satirical moments, and St. Teresa of Avila (1515–82), in her records of her mystical ecstasies, have withstood the changing tides of taste. The French also succeeded in maintaining their appreciation of their two greatest religious writers, Pascal and Bossuet, at the very top of the nonfictional prose writers; both are still revered and occasionally imitated. Pascal took over traditional theology and treated it as literature; his unfinished *Pensées* have exercised far more influence than the rationalism of the greatest French philosophers on the sensibilities of the French. Bossuet's orations reveal the magnificent but refrigerating decorum that seems inseparable from eulogies of the dead—a genre that precludes full sincerity and cultivates tremulous emotion to a dangerous degree. Bossuet's sermons and treatises, however, include masterpieces of simple, terse, direct oratory, which show him as the majestic defender of the unity of faith, of absolutism, and of tradition. His was the last significant endeavour in the 17th century to arrest the flow of relativism and of rebellious individualism, which had engulfed Western civilization with the Renaissance, the Reformation, and Humanism. The two most brilliant writers of religious prose in France in the 20th century were Pierre Teilhard de Chardin (1881–1955), a poetical writer with a luxury of images, and Simone Weil (1909–43), more terse and restrained; they steered a middle course between dogmatism and humility in luring the lay reader to their ardent expressions of conviction.

POLITICAL, POLEMICAL, AND SCIENTIFIC PROSE

In the 20th century, political, economic, and social thought has attempted to reach scientific precision through the use of quantitative data, processing machines, and mathematical formulas. Through such means, other disciplines eventually were elevated to the status of sciences. Literature lost a great deal as a result of this scientific urge, and political and economic thought may have lost even more; for example, the ability to be understood, and perhaps applied, by men of affairs and leaders of nations. The result has been that momentous decisions may be made independent of political theory, which is more often called upon to explain them afterward. Albert Einstein remarked that politics is much more baffling and difficult than physics, and that consequences of errors in politics are likely to make far more difference to the world than the miscalculations of science. Politics is often defined as the art of the possible; it is also an art of improvisation, since the fleeting occasions must be grasped when they appear, and risks must be taken without a full array of scientific data. Like military action, however, political action can be studied in historical writings and in the literary testimonials of men who ran the affairs of their country. Thucydides, Cicero, Caesar, Milton, Burke, Napoleon, and Jefferson were such men of action who were also endowed with uncommon literary gifts. In varying degrees, Benjamin Disraeli, Winston Churchill, Woodrow Wilson, Clemenceau, Lenin, and de Gaulle owed some of their insight and effectiveness to their literary efforts.

Authors, however, are by no means infallible in dealing with the unpredictable course of political life. Interpreting and channelling public opinion proved insuperably difficult, for example, to Alphonse de Lamartine in the revolutionary period of 1848–49 in France, to the bookish Aleksandr Kerensky during the 1917 revolution in Russia, and to a number of brilliant writers who attempted to guide the Spanish Republic in the 1930s. Crowds often can be moved more readily by vapid, repetitious, or inflammatory speeches than by profound or wise counsel. Abraham Lincoln's Gettysburg Address, Churchill's speeches during Britain's finest hour early in World War II, and de Gaulle's lofty eloquence regarding the crises of three decades in France were admired less when they were delivered than afterward. As they are collected, studied, and engraved in the mental make-up of millions of future citizens, such speeches have an effectiveness second to no other form of nonfictional prose. Novels may exercise immense influence through the acute social criticism they embody, but their impact upon the sensibility and the behaviour of their readers is probably less than that of political prose.

Spanish political essays

Although the Spanish language cannot boast of any political thinker comparable to Plato, Machiavelli, or Rousseau, it may boast a large number of fine writers on political topics. Generally, these writers reveal a restrained and terse style, like the poets of Spain, the Latin country least addicted to inflation of language. Garcilaso de la Vega (1539–1616), the son of an Inca mother, wrote with courage and talent of the Peruvians and other cultures of the New World cruelly wrecked by their Catholic conquerors. The Argentinian Domingo Faustino Sarmiento (1811–88) fought in battle and with his pen against his country's dictator and left a masterpiece of social insight, written with rare effectiveness, *Facundo* (1845). Miguel Ángel Asturias (1899–1974), from Guatemala, scathingly depicted the evils of dictatorship in Central America. Like many others in South America, where versatility is not uncommon, Francisco de Miranda (1750–1816) of Venezuela was both a political writer and a statesman.

Italy, after Machiavelli, failed to produce political writers of very great eminence, even during its liberation and unification in the 19th century. The universal thinker Benedetto Croce (1866–1952), however, had the courage to publish, during the Fascist era, the most impassioned defense of liberty in volumes such as *La Storia come pensiero e come azione* (1938; *History as the Story of Liberty*). Another Italian, but from another political direction, Antonio Gramsci (1891–1937), one of the most intelligent exponents of Communism in western Europe, was aware of the vital significance of literary form to spread political ideas. He bitterly deplored the lack, in his country, of a popular literature that reflected the morality and sentiment of the people.

In France political speculation was more comprehensive: few political theoreticians have proved as influential as the philosophers of the Enlightenment, especially Montesquieu and Rousseau. It was the good fortune of the French that during their Revolution at the end of the 18th century and throughout the 19th century, its keenest political minds were also writers of admirable prose. Tocqueville's observations became a sacred text for many a student of America and of pre-Revolutionary France. Since the French seldom give ideas serious consideration unless they are well expressed, however, it was a misfortune that most political speculation after the Napoleonic age was written by gifted, often brilliant, conservatives, such as Joseph de Maistre, Auguste Comte, Frédéric Le Play, Renan, Taine, and Charles Maurras. Those advocating a socialistic view, such as Jean Jaurès and the more elegant and genteel Léon Blum, failed to express their theories in classic prose. The level of political comment in the magazines and newspapers in France is consistently high, but the writers tend to be either too clear-sighted or too arrogant to grant their statesmen a chance to act. "Fair play" is an untranslatable phrase in French, and politics in France, unlike some other countries, is never regarded as a game or sport. Rather, it is a passionate affair of the heart and intellect, conducted in a mood of intransigence. The English essayist Walter Bagehot (1826–77), observing the French at the time of the 1851 Coup d'Etat, commented wryly that "the most essential quality for a free people, whose liberty is to be progressive, permanent and on a large scale, is much stupidity. . . . Stupidity is nature's favorite resource for preserving steadiness of conduct and consistency of opinion."

English and American political works, from the 17th century on, excel all others; they constitute the richest form of nonfictional prose in the English language. John Milton's *Areopagitica* (1644) and his other political pamphlets are monuments of political prose that survive to this day as classics. Edmund Burke's *Letter to a Noble Lord* (1796) was praised a century and a half after its composition as the greatest piece of invective in the English language. William Godwin's *Political Justice* (1793) does not compare in the majesty of its prose to those supreme models, but it did inflame Shelley and other men of letters of the time. Walter Bagehot wrote equally well on literature, politics, and economics, and *The Economist*, which he founded, was the best written weekly of its kind in any language. John Stuart Mill and Thomas Carlyle also helped to maintain the tradition of political and social thought expressed as literature through the 19th century.

The decline of polemics

Polemical prose significantly declined in the modern era. Few moderns express the rage for invective seen in the verse of satirists such as the ancient Roman Juvenal, or Alexander Pope in 17th-century England, or even in the writings of Christian disputants such as Martin Luther. Voltaire rejoiced in flaying not only his enemies but also some, such as Montesquieu and Rousseau, who were fundamentally in agreement with him in the fight against the religion of his age. Literary polemics of a high order were employed against the cultural imperialism of the French in Gotthold Lessing's *Hamburgische Dramaturgie* (1767–69; *Hamburg Dramaturgy).* Besides these examples, the polemics of more recent periods seems tame, or else gross and venomous. Later practitioners of the literature of insult include Émile Zola, particularly in his celebrated article on the Dreyfus affair, *J' Accuse* (1898). Later writers, however, often overreach themselves; their rhetoric sounds vapid and their epigrams strained.

The rift between the two cultures, scientific and humanistic, is probably not as pronounced or final as it has been alleged to be. Around the time the division was enunciated, in mid 20th-century, it was possible to point to a number of eminent scientists who were also masters of prose writing—Jules-Henri Poincaré, Jean Rostand, and Gaston Bachelard in France; Bertrand Russell and Alfred North Whitehead in England; and Rent Dubos and Robert Oppenheimer in the U.S. The peril for scientists who undertake to write for laymen appears to lie in a temptation to resort to florid language and to multiply pretentious metaphors and elaborate cadences in their prose. Some scientists who wrote on astronomy, on anthropology, and on geology have not altogether escaped that pitfall: Sir James Jeans, Loren Eiseley, Sir James Frazer, Teilhard de Chardin. The marriage of the "two cultures" in one mind, which was no less concerned with scientific truth than with beauty of form, was found frequently in older times; Aristotle, Hippocrates, Galileo, Newton, and Goethe all showed strong interest in both. The popularization of science reached a level of a lucid and elegant art with the writings of Bernard de Fontenelle (1657–1757) in French, Francesco Algarotti (1712–64) in Italian, and later, with a masterpiece of scientific rigour expressed in flexible and precise prose, *Introduction à l'étude de la médecine expérimentale*, by the physiologist Claude Bernard (1813–78).

OTHER FORMS

Reportage. Journalism often takes on a polemical cast in countries in which libel laws are not stringent. Polemical journalism flourished in continental Europe when a journalist's insults could be avenged only in a duel; one of the great journalists of this heroic era of the press in France, Armand Carrel, died in such a duel with another journalist in 1836. Most journalistic literature, however, deserves none of the ill-repute that is associated with its more polemical expressions. Rather, it is a remarkably elastic form, as adaptable to sarcasm and the puncturing of illusions as to reflection, subtle persuasion, and infectious geniality. Among the eminent writers who explored its possibilities in the 18th century, Joseph Addison and Richard Steele offered models of polished English prose in the journals *The Tatler* and *The Spectator*, and Jonathan Swift and Oliver Goldsmith also used it effectively in England. In France Voltaire, the novelist Abbé Prévost, and the dramatist Pierre-Carlet de Marivaux all found effective use for the form. By the 19th century, most eminent men of letters attempted to broaden their audiences by means of articles and essays in the press, and in the 20th century, the influence of journalism pervaded the most important works of some authors. Some of the works of G.B. Shaw and H.G. Wells, for example, were reminiscent of journalism in the manner in which they sought topical controversy and challenged social and political prejudices. Many of the finest essays of Virginia Woolf, John Middleton Murry, and Aldous Huxley represented British literary journalism at its most intelligent level. In America, the more heterogeneous public to which authors must address themselves and, later, the competition of the audiovisual media, were not propitious to the flowering of literary journalism of that type. In a more ephemeral genre, that of political reflections couched in clear, pungent style, Walter Lippmann composed models of commentaries on politics and ethics.

Ego in journalistic style

The more self-centred and passionate writers seldom succeeded in journalistic prose as well as those who could forget their ego and adapt their style to a public that wanted to be entertained, moved, or convinced, perhaps, but whose attention span extended no further than the 15 minutes of a train ride or of a hurried breakfast. In France, Proust dreamt for years of appearing as a journalist on the first column of the journal *Le Figaro*. But he and his contemporaries Gide, Claudel, and Valéry, and, later, the imperious and nervous André Malraux, did not conform to the limitations of the newspaper article. On the other hand, Colette, Paul Morand, and François Mauriac proved conspicuously successful in writing the brief, gripping, taut article dear to readers of many of the better continental dailies and weeklies.

The insidious appeal of journalistic writing to thinkers, novelists, and poets is similar to the siren charm of conversation for the author who enjoys talking brilliantly at dinner parties. As Oscar Wilde ruefully remarked, conversationalists and journalists, intent on reporting on the ephemeral, pour whatever genius is theirs into their lives, and only their talent into their works.

Aphorisms and sketches. Authors of maxims and aphorisms, on the contrary, strive for the brevity of inscriptions on medals and public buildings and for a diamond-like resistance to the devastation of time upon diffuse and padded writing. This form is periodically revived. In modern letters, in the latter half of the 20th century, a condensed and enigmatic sort of prose was preferred to poetry by several poets, who invested their sensations, their illuminations, or their reflections with the mystery and éclat of aphorisms. Among the French. who have always favoured the maxim for philosophical, psychological, and ethical advice, a great poet, René Char (1907–), came to be more and more fascinated by that epigrammatic form, harking back to the ancient Greek philosopher whom he admired most, Heracleitus. Char found in the aphorism a means of "pulverizing language" and of allowing the isolated words or groups of words, freed from rhetoric and from the exigencies of clarity, to emerge like rocks from a sunken archipelago. Other French prose writers, including Camus, Char's warmest admirer, and Malraux, likewise scattered through their prose works striking aphorisms that summed up the sense of a situation or the experience of a lifetime. French novels, from the 18th century through the 20th, reflect the influence of the unforgettable maxims coined by the 17th-century moralists Pascal, La Rochefoucauld, and La Bruyère. The novelist could never long resist the seduction of brevity, the challenge of condensing wisdom into a neat, usually bitter, formula, which usually suggested to the reader not to expect overmuch from life and to take revenge upon its little ironies by denouncing it in advance.

Maxims and other pointed and epigrammatic phrases of the sort the ancient Romans called *sententiae* can become too sophisticated or can too obviously strive for effect. This form of expression reached its point of perfection, balancing profundity and solidity of content with pointedness of form, with the moralists of the 17th and 18th centuries in France, whom Nietzsche ranked above all other writers. They included Pascal and La Rochefoucauld and, later, Stbastien Chamfort (1740/41–94), a satirical pessimist often quoted by Schopenhauer and Joseph Joubert (1754–1824). This form, even more than poetry, represents the most economical means of communicating long experience and for imparting moral advice. In a very few words, or at most a few lines, an aphorism may enclose enough matter for the plot of a novel. It may trounce the prejudices of snobbery more vigorously than a long, meandering novel of manners. The greatest of the 19th-century poets, Goethe, Novalis, Leopardi, Vigny, and Baudelaire, as well as painters such as Delacroix, Cézanne, Degas, and later Braque, cherished the epigrammatic, incisive form of expression. One of the advantages of the aphorism or *pensée* is that it can easily produce an impression of depth when it may be only a commonplace pungently expressed. Another is that it allows several approaches to a subject by the skilled prose writer. If he is of a fiery temperament, prone to enthusiasms and lashing out in wrath against what he deems to be false, he can, like Nietzsche, embrace contradictions and sponsor opposed attitudes. If Epictetus, Pascal, and Nietzsche had expressed their reflections consistently and systematically, their works would probably be forgotten. Nonetheless, as Pascal shrewdly remarked. the aphoristic prose style is, of all the manners of writing, the one that engraves itself most lastingly in the memories of men.

That form, in verse and in prose, probably constitutes the most widespread form of literature. It is found in many nations that long lived without fiction, epics, or even popular poetry. It is found in ancient sayings that interlard the speeches of the 20th-century leaders both of the U.S.S.R. and of China; in the book of Proverbs of the Bible; in the Qur'ān; in the Afrikaans language of South Africa in the 20th-century writings of J. Langenhoven. Proverbs, maxims, riddles, and even conundrums make up a large part of African Negro folklore. The animal tales of these people also provide lessons in the form of aphorisms that are neither as platitudinous nor as didactic as Aesop's fables.

Portraits and character sketches

Portraits and sketches are a form of literature that thrives in cultures in which the court, the salon, or the café plays an important role. The few examples left by the ancient Greeks, such as by Theophrastes, pale beside the vivid portraits of real individuals drawn by the ancient Roman historian Tacitus and by the impassioned orator Cicero. In the Classical age of 17th-century France, the character sketch was cultivated in the salons and reached its summit with La Bruyère. That form of writing, however, suffered from an air of artificiality and of virtuosity. It lacked the ebullience and the imagination in suggesting telltale traits that characterize the portraits of the duc de Saint-Simon (1675–1755). Collections of sketches and characters, however, tend to strike the reader as condescending and ungenerous insofar as the writer exempts himself of the foibles he ridicules in others.

Humorous essays

The humorous article or essay, on the other hand, is a blend of sympathy and gentle pity with irony, a form of criticism that gently mocks not only others but the mocker himself. Humour strikes deep roots in the sensibility of a people, and each nation tends to feel that its own brand of humour is the only authentic one. Its varieties of humorous writing are endless, and few rules can ever be formulated on them. Humorous literature on the highest literary level includes that of Cervantes in Spain, of Sterne, Lamb, and Thackeray in England. of Jean Paul in Germany, and of Rabelais, Montaigne, and Voltaire in France. Romantic authors have, as a rule, been too self-centred and too passionate to acquire the distance from their own selves that is essential to humour. In the 20th century, some of the most original examples of what has been called the "inner-directed smile" are in the works of the Argentine Jorge Luis Borges (1899–) and by one of the writers he admires most, the English essayist G.K. Chesterton (1874–1936). In both writers, and in other virtuosos of the intellectual fantasy, there is a persistent refusal to regard themselves as being great, though greatness seems to be within their reach. The humorist will not take himself seriously. Chesterton hides the depth of his religious convictions, while Borges facetiously presents his prodigious erudition and indulges in overelaborate and flowery prose. Borges likes to put on and take off masks, to play with labyrinths and mirrors, but always with a smile. By sketching what appear to be fanciful portraits rather than overtly fictional stories, he creates a half-imaginary character whose presence haunts us in all his writings—that of the author himself.

Dialogues. The dialogue form has long been used as a vehicle for the expression of ideas. It is especially cherished by authors eager to eschew the forbidding tone of formality that often accompanies the expression of serious thought. The writer of a dialogue does not directly address his public, but instead revels in the multiple facets of ideas. By playing this dialectical game he can appear to present contrary views as their respective proponents might and then expose the errors of those he opposes, leading the readers to accept his own conclusions. The advantages of the dialogue are clear: ideas that might have remained abstruse and abstract become concrete and alive. They assume dramatic force. A constant element in the dialogue is irony; etymologically, the term derives from a form of interrogation in which the answer is known beforehand by the questioner. The earliest models of the genre, by the ancient Greeks Plato and Lucian, have never been excelled. Sophistry is another element of the dialogue. In Plato and in the dialogues of Pascal's *Provinciales* (1656–57; "Provincial Letters"), the protagonist plays with the naiveté of his opponents, who always end by surrendering. The writer of a dialogue cannot affect the same casual and self-indulgent attitude as the author of a personal essay since the characters and their statements must be plausible. Nor can he pursue an argument consistently, as he might in a critical, historical, or philosophical essay. Something must persist in the dialogue of the spontaneity and the versatility of an actual conversation among witty and thoughtful people.

There was much seriousness and occasionally some pedantry in early dialogues in several literatures. The dialogues of Bardesanes (154–222) in Syriac, rendered into English as *On Fate*, are on the subject of the laws of the country. A hundred years earlier, Lucian, who was also Syrian, proved himself a master of flowing and ironical Greek prose in his satirical dialogues. The Italian Renaissance writer Pietro Aretino (1492–1556) proved himself the equal of Lucian in verve in his ***Dialogues*** using the same mold and the same title as Lucian. Others who used the dialogue form included Castiglione and Pietro Bembo (1470–1547) in Italy; and in Spain Juan Luis Vives (1492–1540), León Hebreo (1460–*c.* 1521), and Juan de Valdés (*c.* 1498–1541), who treated questions of faith and of languages in dialogues. The genre flourished in the 18th century: Lessing, Diderot, and the Irish philosopher George Berkeley. Diderot's works largely consist of sprightly, iambling, and provocative discussions between the various aspects of his own remarkable mentality. Bold conjectures, determined onslaughts on prejudices, insights into physiology and biology, and erotic fantasies all enter into his dialogues. In the 19th century a number of complex literary personalities, who were capable of accepting the most diverse, and even conflicting points of view, such as Renan and Valéry, had a predilection for the dialogue. Among the devices used by authors of dialogue — many of whom lacked the sustained inventiveness required by fiction— was to attribute their words to the illustrious dead. The French prelate Fénelon, for example, composed *Dialogues des morts* (1700–18), and so did many others, including the most felicitous master of that prose form, the English poet Walter Savage Landor, in his *Imaginary Conversations* (1824) and *Pentameron* (1837).

Travel and epistolary literature. The literature of travel has declined in quality in the age when travel has become most common — the present. In this nonfictional prose form, the traveller himself has always counted for more than the places he visited, and in the past, he tended to be an adventurer or a connoisseur of art, of landscapes, or of strange customs who was also, occasionally, a writer of merit. The few travel books by ancient Greek geographers, such as Strabo and Pausanias of the 1st and 2nd centuries AD, are valuable as a storehouse of remarks on ancient people, places, and creeds. Travel writing of some literary significance appears in the late-13th-century writings of Marco Polo. Works of a similar vein appeared in the 17th century in the observations of Persia two French Huguenots, Jean-Baptiste Tavernier and Jean Chardin, whose writings were lauded by Goethe. Many books of documentary value were later written by English gentlemen on their grand tour of the Continent. The 18th-century Italian egotist Casanova and his more reliable and sharper compatriot Giuseppe Baretti (1719–89) also produced significant travel writings.

The form comprises many of the finest writings in prose during the Romantic age. Not only were the Romantics more alive to picturesqueness and quaintness but also they were in love with nature. They were eager to study local colours and climates and to depict them in the settings for their imaginative stories. Also, travel gave the Romantic writer the illusion of flight from his wearied self. The leisurely record of Goethe's journey to Italy in 1786–88 counts more readers than most of his novels. The *Letters* of *a Russian Traveller* (1795) by Nikolay Karamzin is one of the earliest documents in the development of Russian Romanticism. Ivan Goncharov (1812–91), the Russian novelist who stubbornly limited his fiction to his own geographical province, recorded in *Frigate Pallas* his experience of a tour around the world. Nowhere else in the whole range of literature is there anything comparable to *Saint-Petersburg* (1913), by a virtuoso of poetic style, Andrey Bely; it is a travel fantasy within a city that is both real and transfigured into a myth. Neither James Joyce's Dublin nor Balzac's Paris is as vividly recreated as the former Russian capital in Bely's book. Other travel writers of note include the multinational Lafcadio Hearn (1850–1904), who interpreted Japan with sensitivity and insight. Earlier, two other Westerners wrote on Asia, the English historian Alexander W. Kinglake (1809–91), in *Eothen* (1844), and, more incisively, the French diplomat Joseph-Arthur, comte de Gobineau (1816–82); both blended a sense of the picturesqueness of the East with shrewdness in the interpretation of the people. One of the most thoughtful and, in spite of the author's excessive self-assurance, most profound books on Asia is *Das Reisetagebuch eines Philosophen* (1919; *Travel Diary of a Philosopher),* by the German thinker Hermann Keyserling (1880–1946). With an insatiable interest in countries, Keyserling also interpreted the soul of South America and, less perceptively, analyzed the whole spectrum of European nations. Among the thousands of travel books on Italy, there are a few masterpieces of rapturous or humorous prose: in English, the writings of D.H. Lawrence on Sardinia, on Etruscan Italy, and on the Italian character are more lucid and less strained than other of his prose cogitations. Venice, "man's most beautiful artifact," as Bernard Berenson called it, inspired Rousseau, Chateaubriand, Maurice Barrès, Anatole France, and hundreds of other Frenchmen to write some of their finest pages of prose. After World War I, there was a distinct yearning for new possibilities of salvation among war-ridden Europeans, dimly descried in Asia, in Russia, or in America, and travel literature assumed a metaphysical and semireligious significance. The mood of the writers who expressed this urge was somewhat Byronic; they were expert at poetizing the flight from their own selves. Blaise Cendrars (1887–1961) in his novel *Emmkne-moi au bout du monde* (1956; "Take me away to the end of the world"), epitomizes the urge to seek adventures and a rediscovery of oneself through strange travels. The very theme of travel, of the protagonist being but a traveller on this earth, has been, from Homer's *Odyssey* onward, one of the most laden with magical, and symbolical, associations in literature. Countless authors have played moving and delicate variations on it.

The letter as a genre

Of all the branches of nonfictional prose, none is less amenable to critical definition and categorization than letter writing. The instructions of the ancient grammarians, which were repeated a thousand times afterward in manuals purporting to teach how to write a letter, can be reduced to a few very general platitudes: be natural and appear spontaneous but not garrulous and verbose; avoid dryness and declamatory pomp; appear neither unconcerned nor effusive; express emotion without lapsing into sentimentality; avoid pedantry on the one hand and banter and levity on the other. Letters vary too much in content, however, for generalizations to be valid to all types. What is moving in a love letter might sound indiscreet in a letter of friendship; an analysis of the self may fascinate some readers, while others prefer anecdotes and scandal. La Bruyère, at the end of the 17th century, remarked that women succeed better than men in the epistolary form. It has also been claimed that a feminine sensibility can be seen in the letters of the most highly acclaimed male masters of this form, such as Voltaire, Mirabeau, Keats, and Baudelaire. Advice to practitioners of the art of letter writing usually can be expressed in the often-quoted line in Shakespeare's *Hamlet:* "To thine own self be true." The English biographer Lytton Strachey (1880–1932), a copious and versatile letter writer himself, wrote: "No good letter was ever written to convey information, or to please its recipient: it may achieve both those results incidentally; but its fundamental purpose is to express the personality of the writer." There are, however, numerous and even contradictory ways of expressing that personality.

Although critics have issued endless disquisitions on the craft of fiction and other genres, they have generally remained silent on the epistolary genre, though it has sometimes been the form of prose that outlives all others. Ever since the expression of the writer's personality became one of the implicit purposes of writing in the 18th century, the letters of such eminent authors as Diderot, Rousseau, Byron, and Flaubert have probably offered at least as much delight as any of their other writings. Impressive monuments of scholarship have been erected on the presentation of the complete letters of Thackeray, George Eliot, Swinburne, and Henry James. The literatures of France and England are notably richer in letter writing of the highest order than are the literatures of the United States and Germany. Contrary to many pessimistic predictions regarding the effect on letter writing of modern means of communication, such as the telephone, together with an apparently increasing penchant for haste, some of the richest, most revealing, and most thoughtful letters of all times have been written in the 20th century; those of the English writers Katherine Mansfield and D.H. Lawrence are paramount among them.

Personal literature. The cult of the ego (that is, a preoccupation with self-analysis) is a late development in the history of literature. There were, to be sure, men in ancient times who were absorbed in their own selves, but there is almost no autobiographical literature from ancient Greece and, in spite of Cicero and Pliny the Younger, there is little from ancient Rome. The confession, made as humble as possible and often declamatory in the exposition of the convert's repented sins, was an outgrowth of Christianity; masters of confessional literature were Saint Augustine, Petrarch, and the English Puritans. Autobiographical writing took a different form in the 18th century in the work of men who would have agreed with Goethe that personality is the most precious possession. After the publication of Rousseau's *Confessions* in France in 1781, the passion for looking into one's heart (and other organs as well) spread to other literatures of western Europe. Many a novelist thereafter kept a precise record of his cogitations, anxieties, and harrowing moments of inability to create. Poets and painters, including Delacroix, Constable, and Braque, have often done the same. There is only a very tenuous separation between fiction of this sort from nonfiction; the introspective novel in the first person singular has much in common with a diary, or a volume of personal reminiscences. In his long novel *A la recherche du temps perdu,* Proust revealed himself in three ways—as the author, as the narrator, and as the characters who are projections of his own self. An autobiography once was ordinarily written toward the end of a life, as a fond recollection or an impassioned justification of a lifetime's deeds. More and more, it has come to be written also by men and women in their prime. The names of writers whose autobiographical writings have become classics is legion. Henry Adams (1838–1918) owes his place in American letters chiefly to his book on his education; in 20th-century English letters, Osbert and Sacheverell Sitwell, Leonard Woolf, and Stephen Spender may similarly survive in literature through autobiographical works. André Gide, always uncertain of his novelist's vocation, felt more at ease laying bare the secret of his life in autobiographies and journals.

Although imaginative fiction has probably suffered from excesses of introspection and of analyses of the author's own artistic pangs, knowledge of man's inner life has been enriched by such confessions. The most profound truths on human nature, however, have been expressed not in the form of autobiography but in its transposition into fiction. Readers generally have found more truth in literature created from the possibilities of life than from the personal record of the one life that the author has lived.

In conclusion, the variety of nonfictional prose is prodigious. It can be written on almost any conceivable subject. Almost any style may be used, from casual digressions or sumptuous and sonorous sentences to sharp maxims and elliptical statements. But nonfictional, prose seldom gives the reader a sense of its being inevitable, as does the best poetry or fiction. Nonfiction prose seldom can answer positively the question that Rilke and D.H. Lawrence suggest that any potential writer should ask: Would I die if I were prevented from writing?

BIBLIOGRAPHY. The following comprehensive works, touching on the literatures of a great many nations, are useful to consult: H.M. and N.K. CHADWICK, ***The Growth of Literature,*** 3 vol. (1932–40); RAYMOND QUENEAU (ed.), ***Encyclopédie de la Pléiade***, vol. 1, *Histoire des Litte'ratures,* 3 vol. (1956–58); and J.T. SHIPLEY (ed.), ***Encyclopedia of Literature, 2*** vol. (1946). The last two works contain articles by many authors on all the literatures of the world, treated historically. On the essay, see R.D. O'LEARY, ***The Essay*** (1928), which analyzes the essay theoretically and examines several categories of essayists. DAVID DAICHES gives a brief and lively introduction to his anthologies of essays: ***A Century of the Essay: British and American*** (1951), and ***More Literary Essays*** (1968). On letter writing in general, the best piece is by GUSTAVE LANSON, in French: Introduction to ***Choix de lettres du XVIIe siècle***, 5th rev. ed. (1898), reprinted in Lanson's, ***Essais de méthode, de critique et d'histoire littéraire,*** pp. 243–258 (1965). On personal literature, the diary, autobiography, and the questions raised by those forms of prose, see HENRI PEYRE, ***Literature and Sincerity*** (1963).

(H.M.P.)

Literature, Western

Diverse as they are, European literatures, like European languages, are parts of a common heritage. Greek, Latin, Germanic, Baltic and Slavic, Celtic, and Romance languages are all members of the Indo-European family. (Finnish and Hungarian and Semitic languages of the eastern Mediterranean, such as Hebrew, are not Indo-European. Literatures in these languages are, however, closely associated with major Western literatures and are treated here.) The common literary heritage is essentially that originating in ancient Greece and Rome. It was preserved, transformed, and spread by Christianity and thus transmitted to the vernacular languages of the European Continent, the New World, and other regions settled by Europeans. To the present day, this body of writing displays a unity in its main features that sets it apart from the literatures of the rest of the world. The subject is treated under seven major period divisions: ancient, medieval, Renaissance, 17th century, 18th century, 19th century, and 20th century. The section on each period is further subdivided into subsections on each of the literatures that flourished then.

The first, ancient, period ended with the fall of the Roman Empire in the West (AD 476). Between 800 BC and that date, ancient civilization had created the major types, or genres—epic, tragedy, comedy, lyric, satire, history, biography, and prose narrative—that were to determine much of the future formal development of Western literature. It also bequeathed to its successors a vast collection of mythological lore and a complex philosophical literature that decided the course of Western thought.

The second period, the Middle Ages, extending to about the end of the 15th century, is defined by the establishment of Christianity and the prevalence of a Christian world view in the territories of the former Roman Empire. Though it did not retain its classically refined form, Latin remained an international language in the West, while the vernacular languages gradually came into their own as vehicles of written literature. By the end of the period there existed major works and writers in most of the chief European vernaculars.

By about the middle of the 15th century a new spirit was discernible in some European writing, but the end of the Middle Ages and the beginning of the Renaissance was not an occurrence that can be assigned a single definite date. Gradually, the medieval Catholic world view changed, giving way to a more secular conception of men and the universe. This third period was characterized by a renewed interest in learning, especially a revival of classical learning (Humanism). Renaissance literature also showed the effects of the Protestant Reformation and a sense of cultural expansion that resulted from explorations and discoveries of the time.

The remaining periods, divided according to century from the 17th to the 20th, are somewhat arbitrary divisions; they are justified by a few fundamental developments that affected the literature of each period.

The 17th century in most European countries was characterized by political and religious turbulence. Literature was responsive to these disturbances as well as to increasingly skeptical and rationalist tendencies that accompanied the rise of science as the prevailing intellectual mode in the West.

The next period, the 18th century, is accurately described as the Age of Reason or the Enlightenment. The rational aspect of the period was reflected in a literature of wit, argument, and satire. At the same time, however, an interest in the passions, the sensibility, and other irrational qualities in man inspired another kind of writing: Gothic prose, with its emphasis on terror and the supernatural; poetry concerned with the cultivation of the emotions and the sublime; and the psychological novel.

The effects of the French Revolution (1789) and the

Industrial Revolution were felt in the writing of the 19th century, during which time literary interests developed earlier were explored further in the works of Romantic poets and great 19th-century novelists.

The typical features of 20th-century literature first began to emerge after World War I, and essential forms and themes of modern writing were to be seen in ferment and growth from the 1920s onward.

These sections, with their major headings, appear as follows:

I. Ancient literature

CHARACTERISTICS OF ANCIENT LITERATURE

The stark fact about ancient Western literature is that the greater part of it has perished. Some of it had been forgotten before it was possible to commit it to writing; fire, war, and the ravages of time have robbed posterity of most of the rest; and the restitutions that archaeologists and paleographers achieve from time to time are small. Yet surviving writings in Greek and far more in Latin have included those that on ancient testimony marked the heights reached by the creative imagination and intellect of the ancient world.

Four ancient civilizations in the Near East and in the West, Babylon and Assyria, Egypt, Greece, and Rome, and a fifth, the culture of the Israelites in Palestine, each came into contact with one or more of the others. The two most ancient, Assyro-Babylonia, with its broken clay tablets, and Egypt, with its rotted papyrus rolls, make no direct signal to the modern age; yet Babylon produced the first full code of laws and two epics of archetypal myth which came to be echoed and re-echoed in distant lands; and Egypt's mystical intuition of a supernatural world caught the imagination of the Greeks and Romans. Hebrew culture exerted its greatest literary influence on the West because of the place held by its early writings as the Old Testament of the Christian Bible; and this literature profoundly influenced Western consciousness through translation from about the time of St. Augustine of Hippo (AD 354–430) onward into every vernacular language as well as into Latin. Until then, Judaism's concentrated spirituality set it apart from the Greek and Roman world.

Though influenced by the religious myths of Mesopotamia, Asia Minor, and Egypt, Greek literature has no direct literary ancestry and appears as self-originated. Roman writers looked to Greek precept for themes, treatment, and choice of verse and metre. Rome eventually passed the torch on to the early Middle Ages, by which time Greek had been subsumed under a wholly Latin tradition and was only rediscovered in its own right at the Renaissance—the "classical" tradition afterward becoming a threat to natural literary development, particularly when certain critics of the 17th century began to insist that the subjects and style of contemporary writing should conform with those employed by Greece and Rome.

All of the chief kinds of literature were established by the Greeks and Romans, and later developments have for the most part been secondary extensions. The Greek epic of Homer was the model for the Latin of Virgil; the lyric fragments of Alcaeus and Sappho were echoed in the work of Catullus and Ovid; the history of Thucydides was succeeded by that of Livy and Tacitus; but the tragedy of the great Athenians of the 5th century BC had no worthy counterpart in Roman Seneca nor had the philosophical writings of Plato and Aristotle in those of any ancient Roman, for the practical Romans were not philosophers. Whereas Greek writers excelled in abstraction, the Romans had an unusually concrete vision, and, as their art of portraiture shows, were intensely interested in human individuality.

In sum, the work of these writers and others and perhaps especially that of Greek authors expresses the imaginative and moral temper of Western man. It has helped to create his values and to hand on a tradition to distant generations. Homer's epics extend their concern from the right treatment of strangers to behaviour situations of deep involvement among hero rivals, their foes, and the overseeing gods; the tragedies of Aeschylus and Sophocles are a sublime expression of man's breakthrough into moral awareness of his situation. Among Roman authors an elevated Stoicism stressing the sense of duty is common to many, from Naevius, Ennius, and Cato to Virgil, Horace, and Seneca. A human ideal is to be seen in the savage satire of Juvenal and in Anacreon's songs of love and wine, as it is in the philosophical thought of Plato and Aristotle. It is given voice by a chorus of Sophocles, "Wonders are many, but none is more wonderful than man, the power that crosses the white sea. . . ." The human ideal held up in Greek and Latin literature, formed after civilization had emerged from earlier centuries of barbarism, was to be transformed, before the ancient world came to its close, into the spiritual ideal of Judeo-Christianity, whose writers foreshadowed medieval literature.

ASSYRO-BABYLONIAN LITERATURE

Akkadian, the language of ancient Babylonia and Assyria, is the earliest recorded Semitic language and hence of considerable importance for comparative linguistics. Its writing system was borrowed from Sumerian, the world's oldest written language. In the Assyro-Babylonian culture that flourished on the banks of the Tigris and Euphrates rivers, Akkadian replaced Sumerian as a spoken language in the 3rd millennium BC, though Sumerian continued in written usage almost to the end of the active use of Akkadian in the 1st century AD.

Many of the records of this civilization are of linguistic and historical rather than literary interest. There are numerous word lists of names of gods and of common words, the purpose of which was encyclopaedic: the systematic collection of all possible words and concepts. These lists may also have served pupils as spelling guides. Lists giving a word sign or signs with the Sumerian pronunciation and the Akkadian equivalent arranged in columns served as vocabularies, various types of which remained in use until the end of Babylonian scholarly activity. Of historical as well as linguistic interest are the earliest royal inscriptions, simple dedications of objects or buildings of the type "To god X has A, son of B, king of the city Y, given this vase" or "built this temple!' The Assyrians further developed this pattern into what are known as "annals," detailed accounts of a king's military exploits for each year. Although popular and sometimes even mythological stories appear in some of these entries, their importance is chiefly historical.

The *Epic of Creation* and the *Epic of Gilgamesh*

Assyro-Babylonian literary texts have been found on tablets that are dated as early as *c.* 2500 BC and were about myths, epics about the deeds of early kings, hymns, and wisdom literature (proverbs). Of the greatest literary significance are two epics. ***Enuma elish*** (the ***Epic of Creation***) told how Marduk, the city god of Babylon, slew the monster Tiamat and created the world out of her body, a deed for which the other gods rewarded him by granting him universal rule. The ***Epic of Gilgamesh,*** probably the most important work in Akkadian, related the story of the Sumerian hero Gilgamesh and his companion Enkidu.

The fullest text extant is that on tablets found at Nineveh in the library of Ashurbanipal, king of Assyria (668–627 BC). The epic was a creation of the Babylonians. It incorporated Sumerian stories about Gilgamesh and his friend and introduced a new motif, Gilgamesh's quest for immortality after the death of Enkidu. Of particular interest in the epic is the Babylonian version of the Flood, analogous to Noah's Flood in the Bible, which is recounted to Gilgamesh by a survivor, Utnapishtim. The Gilgamesh story, which was current in Asia Minor in languages other than Akkadian, may possibly have influenced the Greek ***Odyssey;*** and themes from it may occur also in the folklore of the Pacific.

ANCIENT EGYPTIAN LITERATURE

Egyptian literature has survived largely in the form of hieratic texts on papyrus, or school copies written on tablets and ostraca (potsherds and flakes of limestone). Some 70 works have been identified, but many of them are fragmentary and can represent only a fraction of the literary achievement of the ancient Egyptians, most of which is lost. The texts may be classified as historically and mythologically based popular romances, secular and religious poems, model letters, and collections of moral precepts or "instructions."

Among the popular tales that have survived are the story of "King [Cheops] Khufu and the Magicians," a lively series of episodes involving famous ancient kings; "The Shipwrecked Sailor," about a castaway on an island who is entertained by a monster serpent; "The Tale of the Two Brothers," which seems to be involved with the myth of Osiris, god of the underworld, and the theme of the struggle between good and evil; and "The Contendings of Horus and Seth," in which the gods play semi-comic roles. Other popular tales contain no supernatural elements: "The Tale of the Eloquent Peasant" tells of a peasant whose eloquent pleading secures him justice; "The Story of Sinuhe" relates how an exiled fugitive receives the pharaoh's pardon and returns home; and "The Report of Wenamun" describes the adventures of an envoy sent to Phoenicia to buy timber.

Didactic literature of Egypt

The Maxims and Instructions belong to the genre of didactic or wisdom literature, of which the sayings of Ptahhotep, of Kagemni, of Ani, and of Amenemope are examples. They were generally expressed as advice given by an old man to his son, or a king to his heir; some of them contain parallels with Hebrew wisdom literature.

Similarly moral, but of a different, pessimistic flavour are the "Songs of the Harpers" engraved on the walls of tombs. The most remarkable text of this sort is "The Man Who Was Tired of Life," in which a man seems to be meditating on suicide. Such works are in contrast to poems in praise of the sun god and other deities. Although few complete myths have survived in literary form, temple inscriptions contain long ritual texts, often with a mythological content, and the Book of the Dead, a hodgepodge of magic and mythology, contains many references to the world of the gods. Full treatment of Egyptian legends is to be found mainly in the works of later Greek or Latin authors who were responsive to the richness and antiquity of Egyptian mythology.

ANCIENT HEBREW LITERATURE

Literature in Hebrew was produced uninterruptedly from the early 12th century BC, and certain excavated tablets may indicate a literature of even greater antiquity. From 1200 BC to AD 200, Hebrew was a spoken language in Palestine, first as biblical Hebrew, then as Mishnaic Hebrew, a later dialect that does not derive directly from the biblical dialect and one that gained literary status as the Pharisees began to employ it in their teaching in the 2nd century BC. The Hebrew language, whether spoken at the time or not, always adapted itself to the needs of changing literary tastes. It was at various times the vehicle for sophisticated religious thought, scientific precision, and fervid mysticism. It is through the enormous literary influence of the Bible that Hebrew, directly or in translation, has chiefly affected Western literature.

Pre-exilic period, c. 1200–587 BC. All that is preserved of the literature of this period is slightly more than 20

The principle of parallelism in biblical poetry

of the 39 books included in the Old Testament (the remainder being from the next period). In all probability poetry preceded prose. Biblical poetry was based on the principle of parallelism; *i.e.*, the two halves of a verse express the same idea, either by repeating it in different words or by stressing different aspects of it. Examples are found in the book of Psalms: "But they flattered him with their mouths; they lied to him with their tongues" (Ps. 78:36); "He turned their rivers to blood, so that they could not drink of their streams" (Ps. 78:44). To this "rhyme of meaning" was added a simple rhythm, consisting mainly in having each half of a line divided into an equal number of stressed words. There were also folk songs, to which belonged perhaps large parts of the Song of Solomon, dirges, epic chants, and psalms. The use of various forms of poetry in the work of the prophets appears to be a later literary development.

The earlier prose texts were still very close to poetry in structure and language. The first real prose may well have been some of the laws recorded in the Pentateuch. In Jeremiah and Deuteronomy a high standard of prose rhetoric was achieved: some of the conversations in the historical books were attempts to reproduce in writing the style of ordinary talk. (See also BIBLICAL LITERATURE.)

Period of the Second Temple, 538 BC–AD 70. The literary output of this period was large, only part of it belonging to the biblical canon. The biblical Hebrew of the writings was artificial because it had ceased to be spoken and had been replaced by Aramaic, a related Semitic language, and Mishnaic Hebrew. Works that are included among the Dead Sea Scrolls belonged to this period. Some of these works provide evidence of a new kind of writing, the homiletic, or sermonizing, commentary to the Bible called Midrash. The only work of real literary merit among the scrolls was the fervent personal poetry of the Hymns of Thanksgiving.

Parts of the biblical books of Ezra and Daniel and certain works among the Dead Sea Scrolls are in an early form of Aramaic. This period also began to provide translations, called Targums, into a slightly later Aramaic of most of the Hebrew Bible.

Talmudic literature. In contrast to the works of the Bible and the Second Temple were the collections of writings concerned with Jewish civil and religious law. Whereas the former were lengthy writings bearing the imprint of their authors or editors, early rabbinic literature consisted entirely of collections of individual statements loosely strung together. The individual paragraphs exhibit the influence of Hellenistic rhetoric. Collections that follow the arrangement of biblical books are called Midrash, as opposed to works such as the Mishna, where the material is arranged according to subject. The Mishna was the main work of the period c. 100 BC–AD 200. The following period, AD 200–500, was notable for two main innovations: the appearance of an additional literary centre in Babylonia (southern Iraq), where Jewry flourished in contrast to the oppressive rule of Rome and later Byzantium in Palestine; and the literary use of the spoken local dialects of Aramaic alongside Hebrew. The Talmuds produced by Palestine and Babylonia in this period contained a large proportion of Haggada, statements dealing with theological and ethical matters and using stories, anecdotes, and parables to illustrate certain points. This material was later influential on Hebrew fiction of the Middle Ages and of the modern period. (See also TALMUD AND MIDRASH.)

The Mishna and Midrash

ANCIENT GREEK LITERATURE

The ancient Greeks comprised several racial groups that settled, in successive waves of migration, in Greece and along the coast of Asia Minor, on numerous Aegean islands, including Crete, and in Sicily and southern Italy (together known as Magna Graecia). By about 1000 BC there existed three groups, each speaking its own dialect of Greek: the Dorians in southern Greece, where Sparta was their chief city; the Aeolians in central and northern Greece, whence they colonized Lesbos and the neighbouring coast of Asia Minor; and the Ionians in Attica, the area around Athens, and the islands of the central Aegean, and in their settlements on the shore of Asia Minor opposite. The Ionian dialect in its Attic form later became the literary language of all Greece.

Greek literature has a continuous history extending from the 1st millennium BC to the present day, and in it are to be found most of the acknowledged literary genres—epic, drama, lyric poetry, and various types of prose works. From the beginning its authors were Greeks living not only in Greece proper but also in the peripheral territories. Later, after the conquests of Alexander the Great, as Greek became the common language of the eastern Mediterranean lands and then of the Byzantine Empire, Greek literature was produced over a much wider area and also by persons whose mother tongue was not Greek. The cultural centre of the Greek-speaking world shifted gradually during the centuries in which its literature was produced. Until the 5th century BC, the coastal cities of Asia Minor were dominant. Athens became the centre in the 5th and 4th centuries BC. Thereafter Alexandria was the seat of Greek literary activity until the entire area was taken into the Roman Empire. Of the literature of ancient Greece only a relatively small proportion survives, yet its influence on other literatures throughout the ages has been very great.

Stylistic periods. The history of ancient Greek literature may be divided into three periods: pre-Classical (to the end of the 6th century BC); Classical (5th and 4th centuries BC); Hellenistic and Grew-Roman (3rd century BC onward). These periods are summarized below, and the article then treats of the literary genres that arose in them and of the writers associated with the genres.

Pre-Classical period, to the end of the 6th century BC. The Greeks created poetry before they made use of writing for literary purposes, and from the beginning their poetry was intended to be sung or recited. (The art of writing was little known before the 7th century BC. The script used in Crete and Mycenae [Linear B] is not known to have been employed for other than administrative purposes, and after the destruction of the Mycenaean cities it was forgotten.)

Its subject was myth—part legend, based sometimes on the dim memory of historical events; part folktale; and part primitive religious speculation. But since the myths were not closely associated with religious ritual, even though they often treated of gods and heroic mortals, they were not authoritative and could be varied by a poet to express new concepts.

Thus, at an early stage Greek thought was advanced as poets refashioned their materials; and to this stage of pre-Classical literature belonged the epics of Homer, the ***Iliad*** and the ***Odyssey***, retelling intermingled history and myth of the Mycenaean Age. These two great poems, standing at the beginning of Greek literature, established most of the literary conventions of the epic poem. The didactic poetry of Hesiod (c. 700 BC) was probably later in composition than Homer's epics and, though different in theme and treatment, continued the epic tradition.

Types of early Greek poetry

The several types of Greek lyric poetry originated in the pre-Classical period among the poets of the Aegean islands and of Ionia on the coast of Asia Minor. Archilochus of Paros, of the 7th century BC, was the earliest Greek poet to employ the forms of elegy (in which the epic verse line alternated with a shorter line) and of personal lyric poetry. His work was very highly rated by the ancient Greeks but survives only in fragments; its forms and metrical patterns—the elegiac couplet and a variety of lyric metres—were taken up by a succession of Ionian poets. At the beginning of the 6th century, two poets composing in the Aeolic dialect of Lesbos, Alcaeus and Sappho, produced lyric poetry mostly in the metres named after them that Horace was later to adapt to Latin poetry. No other poets of ancient Greece entered into so close a personal relationship with the reader as Alcaeus, Sappho, and Archilochus do. They were succeeded by Anacreon of Teos, in Ionia, composing his lyrics in the Ionian dialect, like Archilochus. Choral lyric, with musical accompaniment, belonged to the Dorian tradition and its dialect, and its representative poets in the period were Alcman at Sparta and Stesichorus of Sicily.

Both tragedy and comedy had their origins in Greece. "Tragic" choruses are said to have existed in Dorian Greece around 600 BC, and in a rudimentary dramatic form tragedy became part of the City Dionysia at Athens about 534. Comedy, too, originated partly in Dorian Greece and developed in Attica, where it was officially recognized rather later than tragedy. Both were connected with the worship of Dionysus, god of fertility and demonic possession.

Written codes of law were the earliest form of prose and were appearing by the end of the 7th century, when knowledge of reading and writing was becoming more widespread. No prose writer is known earlier than Pherecydes (*c.* 550 BC) of Syros, who wrote about the beginnings of the world; but the earliest considerable author was Hecataeus of Miletus, who wrote about both the mythical past and the geography of the Mediterranean and surrounding lands. To Aesop, a semi-historical, semi-mythological character of the mid 6th century, have been attributed the moralizing beast fables copied by later writers.

Classical period, 5th and 4th centuries FC. True tragedy was created by Aeschylus and reached its culmination with Sophocles and Euripides in the second half of the 5th century. Aristophanes, the greatest of the comic poets, lived on into the 4th century, but the Old Comedy did not survive the fall of Athens in 404.

The sublime themes of Aeschylean tragedy, in which man stands answerable to the gods and receives awe-inspiring insight into their purposes, are exemplified in the *Oresteia,* his only trilogy that is complete in all its three plays. The tragedy of Sophocles made a progress toward both dramatic complexity and naturalness while remaining orthodox in its treatment of religious and moral issues. The youngest of the three great tragic playwrights, Euripides, handled his themes on the plane of skeptical enlightenment and doubted the traditional picture of the gods. Corresponding development of dramatic realization accompanied the shift of vision: the number of individual actors was raised to three, each capable of taking several parts.

The three stages of dramatic comedy

The Old Comedy of Aristophanes was established a little later than tragedy but preserved more obvious traces of its origin in ritual; for the vigour, wit, and indecency with which it keenly satirized public issues and prominent persons clearly derived from the primitive ribaldry of the Dionysian festival. Aristophanes' last comedies show a transition, indicated by the dwindling importance of the chorus, toward the Middle Comedy, of which no plays are extant. This phase in its turn was followed toward the beginning of the 3rd century by the New Comedy, introduced by Menander, which turned for its subjects to the private fictional world of ordinary people. Later adaptations of New Comedy in Latin by Plautus and Terence carried the influence of his work on to medieval and modern times.

In the 5th century, Pindar, the greatest of the Greek choral lyrists, stood outside the main Ionian-Attic stream and embodied in his splendid odes a vision of the world seen in terms of aristocratic values that were already growing obsolete. Greek prose came to maturity in this period. Earlier writers as Anaxagoras the philosopher and Protagoras the sophist used the traditional Ionic dialect, as did Herodotus the historian. His successors in history, Thucydides and Xenophon, wrote in Attic. Thucydides was the most intellectually brilliant of the three, and his style and authority make him the foremost Greek historian. The oldest extant piece of Attic prose was an anti-democratic essay on the constitution of Athens, falsely attributed to Xenophon but dated *c.* 430.

The works of Plato and Aristotle, of the 4th century, are the most important of all the products of Greek culture in the intellectual history of the West. They were preoccupied with ethics, metaphysics, and politics as man's highest study and, in the case of Aristotle, extended the range to include physics, natural history, psychology, and literary criticism. They have formed the basis of Western philosophy and, indeed, they determined, for centuries to come, the development of European thought.

This was also a golden age for rhetoric and oratory, first taught by Corax of Syracuse in the 5th century. A study also raising questions of truth and morality in argument, it was of concern to the philosopher as well as to the advocate and the politician and was expounded by teachers among whom Isocrates was outstanding. The orations of Demosthenes, a statesman of 4th-century Athens and the most famous of Greek orators, are pre-eminent for force and power.

Hellenistic and Greco-Roman periods. In the huge empire of Alexander the Great, Macedonians and Greeks composed the new governing class; and Greek became the language of administration, a new composite dialect based to some extent on Attic and called the Koine, or common language. Everywhere the traditional city state was in decline, and the individual was becoming aware of his isolation and seeking consolidation and satisfaction outside corporate society. Artistic creation now came under private patronage, and, except for Athenian comedy, compositions were intended for a small, select audience that admired polish, erudition, and subtlety.

Importance of the Museum at Alexandria

An event of great importance for the development of new tendencies was the founding of the Museum, the shrine of the Muses with its enormous library, at Alexandria. The chief librarian was sometimes a poet as well as tutor of the heir apparent. The task of accumulating and preserving knowledge begun by the Sophists and continued by Aristotle and his adherents was for the first time properly endowed. Through the researches of the Alexandrine scholars, texts of ancient authors were preserved. Frequently the scholars were the poets of the Hellenistic age, and their scholarship influenced their poetry.

The Hellenistic period lasted from the end of the 4th to the end of the 1st century BC. For the next three centuries, until Constantinople became the capital of the Byzantine Empire, Greek writers were conscious of belonging to a world of which Rome was the centre.

The *Iliad* and the *Odyssey*

The genres. *The epic tradition.* At the beginning of Greek literature stood the two great epics, the *Iliad* and the *Odyssey,* poems with roots reaching far into the Mycenaean age, perhaps to 1500 BC, but traditionally ascribed to Homer; in something like their present form they probably date back to the 8th century. The *Iliad* was the tragic story of the wrath of Achilles, son of a goddess and richly endowed with all the qualities that make men admirable. Being slighted by Agamemnon, the Greek leader, he refuses to take any further part in the war against Troy. The Greeks become hard pressed, Agamemnon in vain offers amends; and eventually the plight of the Greeks worsens. Achilles is so far moved by pity that he sends his friend Patroclus in his own armour to their aid; Patroclus saves the Greeks but is killed himself; Achilles, in a frenzy of grief, returns to the war and kills Hector, the Trojan hero who had slain Patroclus, and outrages his corpse. Finally he brings himself to restore the body to Priam, Hector's father, and the poem ends in reconciliation. Achilles' end is outside the poem, but from the beginning the brilliant hero is shadowed by the knowledge of early death. With his readiness to sacrifice all to honour, Achilles embodies the Greek heroic ideal; and the contrast between his superb qualities and his short and troubled life reflects the sense of tragedy always prevalent in Greek thought. Whereas the *Iliad* is tragedy, the *Odyssey* is tragicomedy. It is an enriched version of the old folk tale of the wanderer's return and of his triumph over those who were usurping his rights and persecuting his wife at home. Odysseus too represents a Greek ideal. Though by no means inadequate in battle, he works mainly by craft and guile; only so can he defeat the superior strength of the Cyclops and the superior numbers of the suitors who have taken possession of his palace. During his ten-year journey home from Troy he loses all his comrades and ships, and back in Ithaca he moves disguised as a beggar in his own palace. Through all these hazards it is by mental superiority that he survives and prevails.

Both poems were based on plots that grip the reader, and the story is told in language that is simple and direct, yet eloquent. The *Iliad* and the *Odyssey,* though they are the oldest European poetry, are by no means primitive. They marked the fulfillment rather than the beginning of the literary form to which they belong. They were essentially oral poems, handed down, developed, and added to over a vast period of time, a theme upon which successive nameless poets freely improvised. The world they reflect is full of inconsistencies; weapons belong to both the Bronze and Iron Ages, and objects of the Mycenaean period jostle others from a time five centuries later. Certain mysteries remain: the date of the great poet or poets who gave structure and shape to the two epics; the social function of poems that take several days to recite; and the manner in which these poems came to be recorded in writing probably in the course of the 6th century BC.

In the ancient world the ***Iliad*** and ***Odyssey*** stood in a class apart among epic poems. Of these, there were a large number known later as the epic cycle. They covered the whole story of the wars of Thebes and Troy as well as other famous myths. A number of shorter poems in epic style, the Homeric Hymns, are of considerable beauty.

Didactic poetry

Didactic poetry was not regarded by the Greeks as a form distinct from epic. Yet the poet Hesiod belongs to an altogether different world from Homer. He lived in Boeotia in central Greece about 800 BC. In his ***Works and Days*** he describes the ways of peasant life and incidentally describes the dreary Boeotian plain afflicted by heat, cold, and the oppression of a "gift-devouring" aristocracy. He believed passionately that Zeus cares about right and wrong and that Justice is his daughter. His other surviving poem, the ***Theogony,*** attempts a systematic genealogy of the gods and recounts many myths associated with their part in the creation of the universe. Near Eastern influence is clearly to be seen, especially in some of the cruder speculation about the origin of the universe.

By the end of the 6th century the epic tradition was a spent force until its revival in the Hellenistic period, and the few composers of epic narrative left little but their names.

Lyric poetry. Hesiod, unlike Homer, tells something of himself, and the same is true of the lyric poets. Except for Pindar and Bacchylides at the end of the Classical period, only fragments of the works of these poets survive. There had always been lyric poetry in Greece. All the great events of life as well as many occupations had their proper songs, and here too the way was open to advance from the anonymous to the individual poet.

Kinds of lyric poetry

The word "lyric" covers many sorts of poem. On the one hand, poems sung by individuals or chorus to the lyre, or sometimes to the flute, were called melic; elegiacs, in which the epic hexameter, or verse line of six metrical feet, alternated with a shorter line, were traditionally associated with lamentation and a flute accompaniment; but they were also used for personal poetry, spoken as well as sung. Iambics (verse of iambs, or metrical units, basically of four alternately short and long syllables) were the verse form of the lampoon. Usually of an abusive or satirical—burlesque and parodying-character, they were not normally sung.

If Archilochus of Paros was really writing as early as 700 BC, he was the first of the post-epic poets. The not very numerous fragments suggest a trenchant personality to which he gave uninhibited expression. They reflect the turbulent life of an embittered adventurer. Scorn both of men and of convention is the emotion that seems uppermost, and Archilochus was possessed of tremendous powers of invective. Though much of his poetry was unedifying, he was regarded throughout antiquity as a major poet.

Of lesser stature than Archilochus were his successors, Semonides (often mistakenly identified with Simonides) of Amorgos and Hipponax of Ephesus.

Like the iambic writers, the elegiac poets came mostly from the islands and the Ionian regions of Asia Minor. Chief among them were Callinus of Ephesus and Mimnermus of the neighbouring city of Colophon, a contemporary of the Athenian statesman Solon. On the mainland of Greece, Tyrtaeus roused the spirit of the Spartans in their desperate struggle with the Messenian rebels in the years after 650. His martial poems are perhaps valued more for historical than for literary interest. The same is to some extent true of the poems in elegiac, iambic, and trochaic (the latter a metre basically of four alternately long and short syllables) metres by Solon, who used his poetry as a vehicle for propaganda. Xenophanes (born about 560 BC) rather in the same way used his poems to propagate his revolutionary religious and ethical ideas. The elegiacs attributed to Theognis seem to be poems of various dates suitable for use at drinking parties. Many of them were actually by Theognis himself (about 540 BC). Some give uninhibited expression to his hatred of the lower class rulers who had ousted the aristocracy of Megara; others are love poems to the boy Cyrnus; still others are gnomic, commonplaces of Greek wisdom and morality.

About the beginning of the 6th century a new kind of poetry made its appearance in the island of Lesbos. It was composed in the local Aeolic dialect by members of the turbulent and factious aristocracy. Alcaeus (born about 620 BC), absorbed in political feuds and in civil war, expressed with striking directness searing hate and blind exultation. With the same directness and infinite grace, Sappho, a younger contemporary who seems to have enjoyed a freedom unknown to the women of mainland Greece, told of her loves and hates, though little is known of her relations with the girls named in her poems. The surviving works by their successor in personal lyric, Anacreon of Teos, suggest a more convivial amorousness.

Choral lyric was associated with the Dorian parts of the Greek mainland and the settlements in Sicily and south Italy, whereas poetry for solo performance was a product of the Ionian coast and the Aegean islands. Thus choral song came to be conventionally written in a Doric dialect.

Choral lyric, which had lyre and flute accompaniments, was highly complicated in structure. It did not use traditional lines or stanzas; but the metre was formed afresh for each poem and never used again in exactly the same form, though the metrical units from which the stanzas, or strophes, were built up were drawn from a common stock and the form of the strophe was usually related to the accompanying dance. This elaborate art form was connected mainly with the cult of the gods or, as in the case of Pindar, the celebration of the victors in the great Hellenic games.

The earliest poet of choral lyrics of whose work anything has survived was Alcman of Sparta (about 620 BC). Somewhat later Stesichorus worked in Sicily, and his lyric versions of the great myths marked an important stage in the development of these stories. Simonides of Ceos, in Ionia, was among the most versatile of Greek poets. He was famed for his pathos, but today he is best known by his elegiac epitaphs, especially those on the Greek soldiers who fell in the struggle against Persia.

Pindar's odes

The supreme poet of choral lyric was Pindar from Thebes in Boeotia (born 518 or possibly 522–died after 446 BC), who is known mainly by his odes in honour of the victors at the great games held at Olympia, Delphi, the Isthmus of Corinth, and Nemea. In honouring athletes in the hour of glory, when they were exalted almost to the level of the heroes of old from whom they claimed descent, Pindar felt he was making no unworthy use of his gift. The last of the lyric poets was Bacchylides (flourished 5th century BC), whose work, though often exquisite, is empty, reflecting the declining significance of myth.

Tragedy. Tragedy is generally believed to have developed from the dithyramb, the choral cult song of the god Dionysus. Arion of Lesbos, who worked at Corinth in about 600, was the first to write serious poetry in this medium. Thespis (6th century BC), possibly combining with dithyrambs something of the Attic ritual of Dionysus of Eleutherae, invented tragedy by introducing an actor who conversed with the leader of the chorus. These performances became a regular feature of the great festival of Dionysus at Athens about 534 BC. Aeschylus (525/524–456), created true tragedy by introducing a second actor, though his drama was still centred in the chorus, to whom, rather than to each other, his actors directed themselves.

At the tragic contests at the Dionysia each of three competing poets produced three tragedies and a satyr play, or burlesque, in which there was a chorus of satyrs. Aeschylus, unlike later poets. usually made of his three tragedies a dramatic whole, treating a single story, as in the *Oresteia,* the only complete trilogy that has survived. His main concern was not dramatic excitement and the portrayal of character but rather the presentation of human action in relation to the overriding purpose of the gods.

The tragedies of Aeschylus, Sophocles, and Euripides

His successor was Sophocles (*c.* 496–406 BC), who abandoned for the most part the practice of writing in trilogies, reduced the importance of the chorus, and introduced a third actor. His work too was based on myth, but whereas Aeschylus tried to make more intelligible the working of the divine purpose in its effects on man's life, Sophocles was readier to accept the gods as given and to reveal the values of life as it can be lived within the traditional framework of moral standards. Sophocles' skill in control of dramatic movement and his mastery of speech were devoted to the presentation of the decisive, usually tragic, hours in the lives of men and women at once "heroic" and human, such as Oedipus.

Euripides (*c.* 480–406), last of the three great tragic poets, belonged to a different world. When he came to manhood, traditional beliefs were scrutinized in the light of what was claimed by Sophist philosophers, not always unjustifiably, to be reason; and this was a test to which much of Greek religion was highly vulnerable. The whole structure of society and its values was called into question. This movement of largely destructive criticism was clearly not uncongenial to Euripides. But as a dramatic poet he was bound to draw his material from myths, which, for him, had to a great extent lost their meaning. He adapted them to make room for contemporary problems, which were his real interest. Many of his plays suffer from a certain internal disharmony, yet his sensibilities and his moments of psychological insight bring him far closer than most Greek writers to modern taste. There are studies, wonderfully sympathetic, of wholly unsympathetic actions in the *Medea* and *Hippolytus;* a vivid presentation of the beauty and horror of religious ecstasy in the *Bacchae*; in the *Electra*, a reduction to absurdity of the values of a myth that justifies matricide; in *Helen* and *Iphigeneia in Tauris*, melodrama with a faint flavour of romance.

Comedy. Like tragedy, comedy arose from a ritual in honour of Dionysus, in this case full of abuse and obscenity connected with averting evil and encouraging fertility. The parabasis, the part of the play in which the chorus broke off the action and commented on topical events and characters, was probably a direct descendant of such revels. The dramatic element may have been derived from the secular Dorian comedy without chorus, said to have arisen at Megara, which was developed at Syracuse by Epicharmus (*c.* 530–c. 440). Akin to this kind of comedy seems to have been the mime, a short realistic sketch of scenes from everyday life. These were written rather later by Sophron of Syracuse; only fragments have survived but they were important for their influence on Plato's dialogue form and on Hellenistic mime. At Athens, comedy became an official part of the celebrations of Dionysus in 486 BC. The first great comic poet was Cratinus. About 50 years later Aristophanes and Eupolis refined somewhat the wild robustness of the older poet. But even so, for boldness of fantasy, for merciless invective, for unabashed indecency, and for freedom of political criticism, there is nothing like the Old Comedy of Aristophanes, whose work alone has survived. Cleon the politician, Socrates the philosopher, Euripides the poet were alike the victims of his masterly unfairness, the first in *The Knights;* the second in *The Clouds;* and the third in the *Thesmophoriazusae* and *The Frogs;* whereas in *The Birds* the Athenian democracy itself was held up to a kindlier ridicule. Aristophanes survived the fall of Athens in 404, but the Old Comedy had no place in the revived democracy.

The gradual change from Old to Middle Comedy took place in the early years of the 4th century. Of Middle Comedy, no fully developed specimen has survived. It seems to have been distinguished by the disappearance of the chorus and of outspoken political criticism and by the growth of social satire and of parody; Antiphanes and Alexis were the two most distinguished writers. The complicated plots in some of their plays led to the development of the New Comedy at the end of the century, which is best represented by Menander. One complete play, the *Dyscolos,* and appreciably fragments of four others out of Menander's 105 plays are extant on papyrus. New Comedy was derived in part from Euripidean tragedy; its characteristic plot was a translation into terms of city life of the story of the maiden—wronged by a god—who bears her child in secret, exposes it, and recognizes it years after by means of the trinkets she had put into its cradle. The god becomes a young man about town, the child a courtesan with whom a young man about town of the next generation is involved and whom he joyfully marries when she is revealed as the daughter of free parents.

History. The first great writer of history was Herodotus of Halicarnassus, who was also a geographer and anthropologist. The theme of his history, written in large part for Athenian readers, is the clash between Europe and Asia culminating in the Persian War. The account of the war itself, which occupies roughly the second half of the work, must have been composed by means of laborious inquiry from those whose memories were long enough to recall events that happened when Herodotus was a child or earlier. The whole history, though in places badly put together, is magnificent in its compass and unified by the consciousness of an overriding power keeping the universe and mankind in check. The fluent simplicity of his prose gives Herodotus a peculiar charm.

The histories of Thucydides and Xenophon

Thucydides (*c.* 460–c. 400) was perhaps the first person to apply a first class mind to a prolonged examination of the nature of political power and the factors by which policies of states are determined. As a member of the board of generals he acquired inside knowledge of the way policy is shaped. After his failure to save Amphipolis in 424, he spent 20 years in exile, which he used as an opportunity for getting at the truth from both sides. The result was a history of the war narrowly military and political but of the most penetrating quality. Thucydides investigated the effect on individuals and nations both of psychological characteristics and of chance. His findings were interpreted through the many speeches given to the characters of his history. There is little apparent bias in Thucydides, but his admiration of the Athenian statesman Pericles and detestation of his opponent Cleon are easy to see.

Just as Thucydides had linked his work to the point at which Herodotus had stopped, so Xenophon (431–died before 350) began his *Hellenica* where Thucydides' unfinished history breaks off in 411. He carried his history down to 362. His work was superficial by comparison with that of Thucydides, but he wrote with authority of military affairs and accordingly appears at his best in the *Anabasis,* an account of his participation in the enterprise of the Greek mercenary army, with which the Persian prince Cyrus tried to expel his brother from the throne, and of the adventurous march of the Greeks, after the murder of their leaders by the Persians, from near Babylon to the Black Sea coast. Xenophon also wrote works in praise of Socrates, of whom his appreciation was very incomplete.

No other historical writing of the 4th century has survived except for a substantial papyrus fragment containing a record of events of the years 396–395. Later, history declined to being merely a province of rhetoric.

Rhetoric and oratory. In few societies has the power of fluent and persuasive speech been more highly valued than it was in Greece, and even in Homer there are speeches that are pieces of finished rhetoric. But it was the rise of democratic forms of government that provided a great incentive to study and instruction in the arts of persuasion, which were equally necessary for political debate in the assembly and for attack and defense in the law courts.

Origin of the study of rhetoric

The formal study of rhetoric seems to have originated in Syracuse *c.* 460 BC with Corax and his pupils Tisias and Gorgias (died *c.* 376); Gorgias was influential also in Athens. Corax is reputed to have been the first to write a handbook on the art of rhetoric, dealing with such topics as arguments from probability and the parts into which speeches should be divided. Most of the Sophists had pretensions as teachers of the art of speaking, especially Protagoras, who postulated that the weaker of two arguments could by skill be made to prevail over the stronger, and Prodicus of Ceos, who distinguished five shades of meaning.

Antiphon (c. 480–411), the first professional speech writer, was an influential opponent of democracy. Three speeches of his, all dealing with homicide cases, have been preserved, as have three "tetralogies," sets of two pairs of speeches containing the arguments to be used on both sides in imaginary cases of homicide. In them primitive ideas are expressed concerning bloodguilt and the duty of vengeance. Antiphon's style is bare and rather crudely antithetical. Gorgias from Sicily, who visited Athens in 427, introduced an elaborate balance and symmetry emphasized by rhyme and assonance. Thrasymachus of Chalcedon made a more solid contribution to the evolution of a periodic and rhythmical style.

Andocides (c. 440–died after 391), an orator who spent much of his life in exile from Athens, wrote three speeches containing vivid narrative; but as an orator he was admittedly amateurish. Lysias (c. 455–died after 380) lived at Athens for many years as a resident alien and supported himself by writing speeches when he lost his wealth. His simple, lucid style has obvious charm. His speeches, some of them written for litigants of humble station, show dexterous adaptation to the character of the speaker; though the most interesting of all is his own attack on Eratosthenes, one of the Thirty Tyrants imposed on Athens by the Spartans in 404 BC; this speech contains a brilliant account of the Thirty's reign of terror.

Orators of the 4th century BC

The 12 extant speeches of Isaeus, who was active in the first half of the 4th century BC, throw light on aspects of Athenian law. Isocrates, who was influential in Athens for half a century before his death in 338, perfected a periodic prose style that, through the medium of Latin, was widely accepted as a pattern; and he helped give rhetoric its predominance in the educational system of the ancient world. In his writings, which took the form of speeches but were more like pamphlets, Isocrates shows some insight into the political troubles besetting Greece, with its endless bickering between cities incapable of cooperation. Yet he lacks incisiveness and is usually monotonous in style.

The greatest of the orators was Demosthenes (384–322), supreme in vehemence and power, though lacking in some of the more delicate shadings of rhetorical skill. His speeches were mainly political, and he is best remembered for his energetic opposition to the rise of Macedon under its king Philip II, embodied in the three "Philippics." After Demosthenes, oratory faded, together with the political setting to which it owed its pre-eminence. Two more. 4th-century-BC writers need only be mentioned, Hyperides (c. 390–322) and Lycurgus (c. 390–324).

Plato and Aristotle

Philosophy. Philosophic prose, the greatest literary achievement of the 4th century, derived from Socrates (who himself wrote nothing) and his characteristic method of teaching by question and answer, which led naturally to the dialogue. Alexamenus of Teos and Antisthenes, both disciples of Socrates, were the first to use it; but the greatest exponent of Socratic dialogue was the Athenian Plato (c. 428–348/347). Shortly after Socrates' death in 399 Plato wrote some dialogues, mostly short; to this group of work belong ***The Apology, Protagoras,*** and ***Gorgias.*** In the decade after 385 he wrote a series of brilliant works, ***Phaedo, Phaedrus, Symposium,*** and ***The Republic.*** His Socrates is the most carefully drawn character in Greek literature. Subsequent dialogues became more austerely philosophical; Socrates tended increasingly to be a mere spokesman for Plato's thought; and in the last of his works, the *Laws*, he was replaced by a colourless "stranger." Plato's style is a thing of matchless beauty, though ancient critics, who were likely to entangle themselves in the rules they had invented, found it too poetical. All human experience is within its range; it fits itself to every nuance of a developing argument, rises to the heights of earnest eloquence in discourse on man's destiny, and reflects with equal faithfulness the wit and gaiety of a drinking party and the grandeur of Socrates in the condemned cell.

Plato's pupil Aristotle (384–322) was admired in antiquity for his style; but those of his works that have survived are all of the "esoteric" sort, intended for use in connection with his philosophical and scientific school, the Lyceum. They are without literary grace, and at times they approximate lecture notes. His works on literary subjects, the ***Rhetoric,*** and above all, the ***Poetics,*** had an immense effect on literary theory in the centuries after the Renaissance. In the ancient world, Aristotelian doctrine was known mainly through the works of his successor Theophrastus (*c.* 372–288/287), now lost except for two books on plants and a famous collection of 30 ***Characters,*** sketches of human types much imitated by English writers of the 17th century.

With Theophrastus, Attic prose died out for a time. Technical prose in this period was produced in abundance; and rhetoric, a mere literary exercise divorced from political influence, became ornate and flowery in manner.

Late forms of poetry. The creative period of the Hellenistic age was practically contained within the span of the 3rd century BC. To this period belonged three outstanding poets: Theocritus, Callimachus, and Apollonius of Rhodes. Theocritus (*c.* 310–250), born at Syracuse, is best known as the inventor of bucolic mime, or pastoral poetry, in which he presented scenes from the lives of shepherds and goatherds in Sicily and southern Italy. He also dramatized scenes from middle class life; and in his second idyll the character Simaetha, who tries by incantations to recover the love of the man who has deserted her, touches the fringe of tragedy. He also used another Hellenistic form, the epyllion, a short scene of heroic narrative poetry in which heroic stature is often reduced by playful realism and delicate psychology. In his hands the hexameter attained a lyric purity and sweetness unrivalled elsewhere. He was the first of the nature poets, succeeded by Moschus and Bion.

The ***Aitia*** of Callimachus

Callimachus (flourished about 260) was a scholar as well as a poet. His most famous work, of which substantial fragments survive, was the ***Aitia,*** an elegiac poem describing the origins of various rites and customs. It was heavy with learning but diversified by passages of entertaining narrative. His six hymns show immense poetic expertise but no religious feeling, for the gods of Olympus had long since become obsolete. Callimachus also wrote epigrams, and fragments survive of *iambi* ("iambs"). The form was widely used throughout the 3rd century to denounce the vanities of the world. Sometimes, in a mixture of prose and verse, these pieces had links with satire; and their chief exponents were Bion the Borysthenite, Menippus of Gadara, Cercidas and Phoenix of Colophon.

Callimachus avoided epic in favour of the greater intensity possible in shorter works. The last surviving classical Greek epic was written by his successor at Alexandria, Apollonius of Rhodes (born about 295). Apollonius' account of the voyage of the Argonauts is so full of local legend that the coherence of the poem is lost; but the story of Medea's wild passion for Jason, the leader of the Argonauts, is marked by a new sort of romantic awareness that is fully realized in the episode of Dido's passion for Aeneas in Virgil's *Aeneid.*

The desire to combine learning with poetry led to the revival of didactic verse. The ***Phaenomerza*** of Aratus of Soli (*c.* 315–c. 240) is a versification of a treatise on the stars by Eudoxus of Cnidus (c. 390–c. 340). Chance has preserved the poems of Nicander (probably 2nd century) on the unlikely subjects of cures for bites and antidotes to poisons.

The mimes of Herodas (3rd century), short realistic sketches of low life in iambic verse, have affinities with the non-pastoral mimes of Theocritus. They perhaps give a hint as to the character of the literature of popular entertainment, now largely lost. Mime, especially pantomime, was the main entertainment throughout the early Roman Empire.

After the middle of the 3rd century, poetic activity largely died away, though the great period of scholarship at Alexandria and at Pergamum was still to come. The names of a few poets are known: Euphorion (born about 275) of Chalcis and Parthenius (flourished 1st century BC), the teacher of Virgil. Thereafter Greek poetry practically ceased, apart from a sporadic revival in the 4th century AD. An exception exists in the case of epigrammatic poetry in elegiac couplets, surviving mainly in two compilations, the Planudean and Palatine anthologies.

Late forms of prose. Almost all of the great mass of Hellenistic prose—and later prose, historical, scholarly, and scientific—has perished. Among historians Polybius (c. 200–c. 118 BC), the most outstanding, has survived in a fragmentary condition. Present at Rome when it was succumbing to the first influences of Greek literature, he wrote mainly of events of which he had direct experience, often with great insight; his work covered the period from 264 to 146. Diodorus Siculus' universal history (1st century BC) is important for the sources quoted there. The most considerable of lost historians was Timaeus (c. 350–c. 250), whose history of the Greeks in the west down to 264 provided Polybius with his starting point. Later historians were Dionysius of Halicarnassus (flourished about 20 BC); Appian of Alexandria (2nd century AD), who wrote on Rome and its conquests; and Arrian (c. AD 96–c. 180) from Bithynia, who is the most valuable source on Alexander the Great.

The criticism of Longinus and Dionysius of Halicarnassus

The most important works of criticism, of which little has survived, were by Dionysius of Halicarnassus and the obscure Longinus. Longinus' treatise *On the Sublime* (c. AD 40) is exceptional in its penetrating analysis of creative literature. The *Bibliotheca* attributed to Apollodorus (c. 180 BC) is a handy compendium of mythology.

Scientific work such as the astronomy and geography of Eratosthenes (c. 276–c. 194) of Alexandria is known mainly from later summaries; but much that was written by the mathematicians, especially Euclid (flourished c. 300 BC) and Archimedes (c. 287–212), has been preserved.

Much survives of the writings of the physician Galen (AD 129–199). His contemporary Sextus Empiricus is an important source for the history of Greek philosophy. The survey of the Mediterranean by Strabo in the time of Augustus preserved much valuable information; and so, in a more limited field, did the description of Greece by Pausanias (2nd century AD). Greek achievement in astronomy and geography was summed up in the work of Ptolemy of Alexandria in the 2nd century AD.

Greek became the language of the large settlement of Jews at Alexandria, and the Septuagint, the Greek version of the Old Testament, was completed by about the end of the 2nd century BC. Much of the Apocrypha was composed in Greek, and the New Testament was written in popular Greek (Koine). Of the early Christian writers in Greek the most notable were Clement of Alexandria (c. AD 150–c. 215) and Origen (c. AD 185–c. 254), together with Clement of Rome and Ignatius of Antioch.

Plutarch's *Lives*

The *Parallel Lives* of famous Greeks and Romans by Plutarch (c. AD 46–c. 119) of Chaeronea in Boeotia was for centuries one of the formative books for educated Europeans. Great figures from an idealized past are presented for the edification of the lesser men of his own day; and the anecdotes with which the *Lives* abound are of various degrees of credibility. They belong to biography rather than to history, though they are an important source for historians. A number of shorter works on a wide variety of subjects have come down under the title *Moralia*, which show the intellectual tide of Greece on the ebb.

There was much concern over a question that had been argued ever since the days when Athens had ceased to be a free city: to what extent was Attic prose a norm that writers and especially orators were bound to follow? Many had shunned it in favour of a more ornamental Asiatic style. But at the end of the 1st century AD there was a revival of the Attic dialect. Speeches and essays were written for wide circulation. This revival is known as the Second Sophistic movement, and chief among its writers were Dio Chrysostom (1st century AD), Aelius Aristides (2nd century), and Philostratus (early 3rd century). The only writer of consequence, however, was Lucian (c. 120/125–c. 190). His works are mainly slight and satirical; but his gift of humour, even though repetitive, cannot be denied. *Lives and Opinions of Famous Philosophers* was a valuable work of the 3rd century by Diogenes Laertius, a writer otherwise unknown.

Philosophical activity in the early empire was mainly confined to moralizings based on Stoicism, a philosophy advocating a life in harmony with nature and indifference to pleasure and pain. Epictetus (born about AD 55) influenced especially the philosophic Roman emperor Marcus Aurelius (121–180), whose *Meditations* have taken their place beside works of Christian devotion. Many of Plutarch's *Moralia* were Platonic, with vaguely mystical tendencies; but Plotinus (c. 205–260/270) was the last major thinker in the classical world, giving new direction to Platonic and Pythagorean mysticism.

Erotic romances

The latest creation of the Greek genius was the novel, or erotic romance. It may have originated as early as the 1st century BC; but its roots reach back to such plays of triumphant love as the lost *Andromeda* of Euripides, to the New Comedy, to Xenophon's daydreams about the education of Cyrus, and to the largely fictitious narratives that were one extreme of what passed for history from the 3rd century BC onward. Of these last, the best known examples are the Alexander romances, a wildly distorted and embroidered version of the exploits of Alexander the Great, which supplied some of the favourite reading of the Middle Ages. Erotic elegy and epigram may have contributed something and so may the lost *Milesian Tales* of Aristides of Miletus (c. 100 BC), though these last appear to have depended on a pornographic interest that is almost completely absent from the Greek romances. Only fragments survive of the Ninos romance (dealing with the love of Ninos, legendary founder of Nineveh), which was probably of the 1st century BC; but full-length works survive by Chariton (2nd century AD), Heliodorus (3rd century AD), Xenophon (2nd or 3rd century AD) of Ephesus, and Achilles Tatius (2nd century AD). All deal with true lovers separated by innumerable obstacles of human wickedness and natural catastrophe and then finally united. *Daphnis and Chloe* by Longus (between 2nd and 3rd century AD) stands apart from the others because of its pastoral rather than quasi-historical setting. The works of Dictys Cretensis and Dares Phrygius belong to the same period. They claim to give a pre-Homeric account of the Trojan War. The Greek originals are almost wholly lost, but the Latin version was for the Middle Ages the main source for the story of Troy.

ANCIENT LATIN LITERATURE

Latin literature was a product of the Roman Republic and the Roman Empire. After Rome fell, Latin remained the literary language of the Western medieval world until it was superseded by the Romance languages it had generated and by other modem languages.

Literature in Latin began as translation from the Greek, a fact that largely conditioned its development. The Greeks themselves valued originality of treatment above invention. Latin literature was even more traditional in its use of former writers as sources of stock motifs or arguments. Whereas the Greeks excelled in abstraction, the Romans had concrete vision, illuminating Greek generalizations with vivid illustration. They were more distinguished as verbal artists than as thinkers. Intensely interested in human individuality, their genius was moral and political. Their noblest ideal was *humanitas*, a blend of culture and kindliness, approximating the quality of being "civilized" as the word is used today of individuals.

Little need be said of the preliterary period. Hellenistic influence came from the south, Etrusco-Hellenic from the north. Improvised farce, with stock characters in masks, may have been a native invention from Campania (the countryside of modern Naples). The historian Livy traced quasi-dramatic satura (medley) to the Etruscans. The statesman-writer Cato and the scholar Varro said that in former times the praises of heroes were sung after feasts, sometimes to the flute, which was perhaps an Etruscan custom. If they existed, these carmina *convivalia*, or festal songs, would be behind some of the legends that came down to Livy. There were also the rude verses improvised at harvest festivals and weddings and liturgical formulas, whose scanty remains show alliteration and assonance. The nearest approach to literature must have been in public and private records and in recorded speeches.

Stylistic periods. Latin literature may be conveniently divided into four periods: early writers, to 70 BC; Golden Age, 70 BC–AD 18; Silver Age, AD 18–133; later writers. These periods are outlined below, and their literary genres and the writers associated with them are described afterward.

Greek influence of Livius Andronicus

Early writers. The ground for Roman literature was prepared by the capture in 272 BC of the Spartan colony of Tarentum and the subsequent influx of Greek slaves, some of whom were put to tutoring young Roman nobles. Among them was Livius Andronicus, who was later freed and who is considered to be the first Latin writer. In 240 BC, to celebrate the victory over Carthage, he composed a genuine drama adapted from the Greek. His success established a tradition of performing such plays alongside the cruder native entertainments. He also made a translation of the Odyssey. For his plays Livius adapted the Greek metres to suit the Latin tongue; but for his Odyssey he retained a traditional Italian measure, as did Gnaeus Naevius for his epic on the First Punic War (the first of three waged by Rome against Carthage). Scholars are uncertain as to how much this metre depended on quantity or stress. Ennius (239–169) from Calabria, who was half Greek, adopted and latinized the Greek hexameter for his epic *Annales.* His range of subjects was remarkable: he acquainted Rome with the Hellenistic world, but unfortunately his work survives only in fragments.

The Greek character thus imposed on literature made it more a preserve of the educated elite. Now emerged coteries in Rome, such as the circle formed around the Roman consul and general Scipio Aemilianus (185/184–129), who had led Rome to final victory against Carthage in the Third Punic War. This circle contained the statesman-orator Gaius Laelius; the Greek Stoic philosopher Panaetius; the Greek historian Polybius; the satirist Lucilius; and an African-born slave of genius, the comic playwright Terence. Greek influence increased after the acquisition of Greece as a Roman province and the ensuing influx of Greek scholars. Greek soon became a second language to educated Romans. Early in the 1st century BC, however, Latin declamation established itself, and, borrowing from Greek, it attained polish and artistry.

Plautus, the leading poet of comedy, is one of the chief sources for colloquial Latin. But Ennius, following Livius, sought to heighten epic and tragic diction by using archaisms; and from his time onward, with a few exceptions such as the short poems (nugae) of Catullus, Lucilius' and Horace's satires, and Petronius' Satyricon, the language of literature and its rhythms became ever more divorced from that of the people, until reaction came in the 2nd century AD.

The Ciceronian period

Golden Age, 70 BC–AD ***18.*** The Golden Age of Latin literature spanned the last years of the republic and the virtual establishment of the Roman Empire under the reign of Augustus (27 BC–AD 14). The first part of this period, from 70 to 42 BC, is justly called the Ciceronian. It produced writers of distinction, most of them also men of action, among whom Julius Caesar stands out. The most prolific was Varro, "most learned of the Romans," but it was Cicero, statesman, orator, poet, critic, philosopher, who developed the Latin language to express abstract and complicated thought with clarity. Subsequently, prose style was either a reaction against, or a return to, Cicero's. As a poet, although uninspired, he was technically skillful. He edited the *De rerum natura* of the philosophic poet Lucretius. Like Lucretius, he admired Ennius and the old Roman poetry and, though apparently interested in Hellenistic work, spoke ironically of its extreme champions, the nedteroi ("newer poets").

After the destruction of Carthage and Corinth in 146 BC, prosperity and external security had allowed the cultivation of otiunz ("leisure"), living for oneself, and hence literature of self-expression and literature for entertainment. In this climate flourished the *neōteroi*, a new school of poets, largely non-Roman Italians from the north, who introduced the mentality of "art for art's sake." None is known at first hand except Catullus of Verona. These poets reacted against the grandiose, the Ennian tradition of "gravity," and their complicated allusive poetry consciously emulated the Callimacheans of 3rd-century Alexandria. The Neoteric influence persisted into the next generation through Cornelius Gallus to Virgil.

Virgil, born near Mantua and schooled at Cremona and Milan, chose Theocritus as his first model. His Eclogues aimed at recapturing Theocritus' cadences and at creating beauties of sound to match the beauties of the idyllic life of shepherds, depicted in a fanciful landscape. Epicurean, escapist poems, they are shot through with topical allusions, and in the fourth he already appears as a national prophet. Virgil was drawn into the circle being formed by Maecenas, Augustus' chief minister. In 38 BC he and Varius introduced the promising young poet Horace to Maecenas; and by the final victory of Augustus in 30 BC, the circle was consolidated.

The Augustan Age

With the reign of Augustus began the second phase of the Golden Age, known as the Augustan Age. It gave encouragement to the classical notion that a writer should not so much try to say new things as to say old things better. The rhetorical figures of thought and speech were mastered until they became instinctive. Alliteration and onomatopoeia (accommodation of sound and rhythm to sense), overdone by the Ennians and therefore eschewed by the nedteroi, were now used effectively with due discretion. This studied perfection of form characterizes the odes of Horace; elegy too became more polished.

The decade of the first impetus of Augustanism, 29–19 BC, saw the publication of Virgil's Georgics and the composition of the whole Aeneid by his death in 19 BC; Horace's Odes, books I–III and Epistles, book I; in elegy, books I–III of Propertius (also of Maecenas' circle) and books I–II of Tibullus, with others from the circle of Marcus Valerius Messalla Corvinus, and doubtless the first recitations by a still younger member of his circle, Ovid. About 28 or 27 BC Livy began his monumental historical work.

Maecenas' circle was not a propaganda bureau. True, he seems to have suggested occasionally that some public theme was ripe for celebration; witness the ingenious poems of graceful refusal. But the tone of these indicates that refusing was easy and that he was as much friend as patron. Poetic promise was the qualification for entry; Propertius, when admitted, was simply a youth with an anti-Caesarian background who had gained favour with passionate love elegies. He and Horace quarrelled, and after Virgil's death the circle faded out. Would-be poets now abounded, such as Horace's protégés, who occur in the Epistles; Ovid's friends, whom he remembers wistfully in exile; and Manilius, whom no one mentions at all. Poems were recited in literary circles and in public, hence the importance attached to euphony, smoothness, and artistic structure. They thus became known piecemeal and might be improved by friendly suggestions. When finally they were assembled in books, great care was taken over arrangement, which was artistic or significant (but not chronological).

Meanwhile, in prose the Ciceronian climax had been followed by a reaction led by Sallust. In 43 BC he began to publish a series of historical works in a terse, epigrammatic style studded with archaisms that eschewed the rhythmic copiousness of Cicero. Later, eloquence, deprived of political influence, migrated from the forum to the schools, where cleverness and point counted rather than rolling periods. Thus developed the epigrammatic style of the younger Seneca and, ultimately, in an extreme form, of Tacitus. Spreading to verse, it conditioned the witty couplets of Ovid, the tragedies of Seneca, and the satire of Juvenal. Though Livy stood out, Ciceronianism only found a real champion again in the rhetorician Quintilian.

Silver Age, AD 18–133. After the first flush of enthusiasm for Augustan ideals of national regeneration, literature under the empire paid the price of political paternalism. Without freedom, it was subtly sterilized; and Ovid was but the first of many writers actually suppressed or inhibited by fear. Only Tacitus and Juvenal, writing under tolerant emperors, turned emotions pent up under Domitian's reign of terror into the driving force of great literature. Late Augustans such as Livy already sensed that Rome had passed its summit. Yet the title of Silver Age is not undeserved by a literary century that produced, in addition to Tacitus and Juvenal, the elder Seneca, his son, his nephew Lucan, the two Plinys, Quintilian, Petronius, Statius, Martial, and, of lesser stature, Manilius, Persius, Valerius Flaccus, Silius Italicus, and Suetonius.

Later writers. The decentralization of the empire under Hadrian and the Antonines weakened the Roman pride and passion for liberty. Romans began again to write in Greek as well as in Latin. The "new sophistic" movement in Greece affected the "novel poets" such as Florus. An effete culture devoted itself to grammar, metric, and philology, to archaism and preciosity. After Juvenal, 250 years elapsed before Ausonius of Bordeaux (4th century AD) and the last of the true classics, Claudian (flourished about 400), appeared. The interim produced only the anonymous ***Pervigilium Veneris*** ("Vigil of Venus"), presaging the Middle Ages in its vitality and touch of stressed metre. Ausonius, though in the pagan literary tradition, was a Christian and was contemporary with a truly original Christian poet, the Spaniard Prudentius. Henceforward, Christian literature overlaps pagan and generally surpasses it.

The ***Meditations*** of Marcus Aurelius

In prose these centuries have somewhat more to boast, though the greatest work by a Roman was written in Greek, the ***Meditations*** of the emperor Marcus Aurelius. ***Elocutio novella,*** a fashionable blend of archaisms and colloquial speech, is seen to best advantage in Apuleius (born about 125). Other writers of note were Aulus Gellius and Macrobius. The 4th century AD was the age of the grammarians and commentators, but in prose some of the most interesting work is again Christian, by Tertullian, Ambrose, Jerome, and Augustine; Boethius (*c.* 480–524), the last Roman philosopher and probably a Christian, lived after imperial Rome had fallen. His ***Consolation of Philosophy*** was enormously popular throughout medieval Europe.

The genres. ***Comedy.*** Roman comedy was based on the New Comedy fashionable in Greece, whose classic representative was Menander. But whereas this was imitation of life to the Greeks, to the Romans it was escape to fantasy and literary convention. Livius' successor, Naevius, who developed this "drama in Greek cloak" ***(fabula palliata),*** may have been the first to introduce recitative and song, thereby increasing its unreality. But he slipped in details of Roman life and outspoken criticisms of powerful men. His imprisonment warned comedy off topical references, but the Roman audience became remarkably alert in applying ancient lines to modern situations and in demonstrating their feelings by appropriate clamour.

Plautus and Terence

Unlike his predecessors, Plautus specialized, writing only comedy involving high spirits, oaths, linguistic play, slapstick humour, music, and skillful adaptation of rhythm to subject matter. His plays must be thought of almost as comic opera. Part of the fun consisted in the sudden intrusion of Roman things into this conventional Greek world. "The Plautine in Plautus" consists in pervasive qualities rather than supposed innovations of plot or technique. Nevertheless, he is not a favourite Latin author because of his stereotyped characters and situations.

As Greek influence on Roman culture increased, Roman drama became more dependent on Greek models. Terence's comedy was very different from Plautus'. Singing almost disappeared from his plays, and recitative was less prominent. From the Greek dramatist Menander he learned to appreciate and exhibit refinements of psychology and to construct ingenious plots; but he lacked comic force, being uninterested in raising laughs from the crowd. His pride was refined language—the avoidance of vulgarity, obscurity, oaths, or slang. His characters were less differentiated in speech than those of Plautus, but their discourse has an elegant charm. The society Terence portrayed was more sensitive than that of Plautine comedy; lovers tended to be loyal and sons obedient. His historical significance has doubtless been enhanced by the loss of nearly all of the comedies of Menander.

Though often revived, plays modelled on Greek drama were rarely written after Terence. The Ciceronian was the great age of acting, and in 55 BC Pompey gave Rome a permanent theatre. Plays having an Italian setting and Italian dress now came into vogue, their framework being Greek New Comedy but their subject Roman society. A native form of farce was also revived. Under Julius Caesar, this yielded in popularity to verse mime of Greek origin that was realistic, often obscene, and also full of quotable and often edifying apothegms. Finally, when mime gave rise to the dumb show of the ***pantomimus*** with choral accompaniment and when exotic spectacles had become the rage, Roman comedy faded out.

Tragedy. Livius introduced both Greek tragedy ***(fabula*** *crepidata*, "buskined") and comedy to Latin. He was followed by Naevius and Ennius, who, like the Hellenistic Greeks, loved Euripides. Pacuvius, probably a greater tragedian, liked Sophocles, and heightened tragic diction even more than Ennius. His successor, Accius, was more rhetorical in style and impetuous in spirit. The fragments of these poets betoken grandeur in "the high Roman fashion," but they also have a certain ruggedness. They did not always deal in Greek mythology: occasionally they exploited Roman legend or even recent history. The Roman chorus, unlike the Greek, performed on stage and was inextricably involved in the action.

Classical tragedy ceased being composed after Accius, though its plays were constantly revived. Writing plays, once a function of slaves and freedmen, became a pastime of aristocratic dilettantes. Such writers had probably no thought of production: post-Augustan drama was for reading. The nine extant tragedies of the younger Seneca were clearly not written for public performance. They are melodramas of horror and violence. They exhibit faults latent or incipient in Euripides: sensational realism and rhetorical cleverness. Seneca strains after each point. Characterization is crude, and Stoic moralizing obtrusive. The iambic verse is competent but facile. Yet he was the great model for 16th- and early 17th-century tragedy, especially in France. Shakespeare's ***Richard III*** and the ghost scenes in ***Julius Caesar, Hamlet,*** and ***Macbeth*** are reminiscent of him.

Epic and *epyllion.* Livius' pioneering ***Odyssey*** was, to judge from the fragments, primitive, as was the ***Bellum Punicum*** of Naevius, important for Virgil because it began with the legendary origins of Carthage in Phoenicia and Rome in Troy. But Ennius' ***Annales*** soon followed. This compound of legendary origins and history was in Latin, in a transplanted metre, and by a poet who had imagination and a realization of the emergent greatness of Rome. In form his work must have been ill-balanced; he almost ignored the First Punic War in consideration of Naevius and became more detailed as he added books about his own times. But his great merit shines out from the fragments—nobility of ethos matched with nobility of language. On receptive spirits, such as Cicero, Lucretius, and Virgil, his influence was profound. For a century and a half he was unapproached in epic.

Virgil's ***Aeneid***

Little is known of the "strong epic" for which Virgil's friend Varius is renowned, but Virgil's ***Aeneid*** was certainly something new. Recent history would have been too particularized a theme. Instead, Virgil developed Naevius' version of Aeneas' pilgrimage from Troy to found Rome. The poem is only superficially an Odyssey of travel (with an interlude of love) followed by an Iliad of conquest, and is more a symbolic epic of deep contemporary relevance. The hero is Rome, Aeneas is the archetypal Roman, and the subject is his conversion from an individualistic leader to the dedicated servant of Rome's destiny. He learns what he must do from a succession of ever clearer prophetic promises. His fault was to have lingered at Carthage. The command to leave the Carthaginian queen Dido shakes him ruthlessly out of the last great temptation to seek individual happiness. But it is only the vision of Rome's future greatness, seen when he visits Hades, that kindles obedient acceptance into imaginative enthusiasm. It was just such a sacrifice of the individual that the Augustan ideal demanded. The second half of the poem represents the fusing in the crucible of war of the civilized graces of the Trojans with the manly virtues of the Italians. The tempering of Roman culture by Italian hardiness was another part of the Augustan ideal. So was a revival of interest in ancient customs and religious observances, which Virgil could appropriately indulge. The verse throughout is superbly varied, musical, apt, and rhetorical in the best sense,

With his ***Hecale,*** Callimachus had inaugurated the short, carefully composed hexameter narrative that is called epyllion to replace grand epic. The ***Hecale*** had started a convention of insetting an independent story. Catullus inset the story of Ariadne on Naxos into that of the marriage of Peleus and Thetis, and the poem has a monotonous beauty. But the story of Aristaeus at the end of Virgil's ***Georgics,*** with that of Orpheus and Eurydice inset, shows what heights epyllion could attain.

Ovid's *Metamorphoses*

Ovid's ***Metamorphoses*** is a nexus of some 50 epyllia with shorter episodes. He created a convincing imaginative world with a magical logic of its own. His continuous poem, meandering from the creation of the world to the apotheosis of Julius Caesar, is a great Baroque conception, executed in swift, clear hexameters. Its frequent irony and humour are striking. Thereafter epics proliferated. Statius' ***Thebaid*** and inchoate ***Achilleid*** and Valerius' ***Argonautica*** are justly less read now than they were. Lucan's unfinished ***Pharsalia*** has a more interesting subject, the struggle between Caesar and Pompey, whom he favours. He left out the gods. His brilliant rhetorical style almost makes the poem a success, but it is too strained, and the treatment too episodical.

Didactic poetry. Ennius essayed didactic poetry in his ***Epicharmus,*** a work on the nature of the physical universe. Lucretius' masterpiece, ***De rerum natura,*** was its first noteworthy successor, driven as it is by Lucretius' anxiety to rid himself of fear of death and the supernatural. He is not an original thinker but attained an imaginative conception of the effect of Epicurus' atomism on everything in and around him. His poem is truly didactic; some of the finest poetry is in digressions, but the grandeur of the total vision is missed if the detailed and impassioned argument is shirked.

The appearance of this poem must have profoundly affected Virgil, but his poetic reaction was delayed for some 17 years; and the ***Georgics,*** though deeply influenced by Lucretius, were not truly didactic. Countrybred though he was, Virgil wrote for literary readers like himself, selecting whatever would contribute picturesque detail to his impressionistic picture of rural life. The ***Georgics*** portrayed the recently united land of Italy and taught that the idle Golden Age of the ***Eclogues*** was a mirage: relentless work, introduced by a paternal Jupiter to sharpen men's wits, creates "the glory of the divine countryside." The compensation is the infinite variety of civilized life. Insofar as it had a political intention, it encouraged revival of an agriculture devastated in wars, of the old Italian virtues, and of the idea of Rome's extending its works over Italy and civilizing the world.

Ovid's ***Ars amatoria*** was comedy or satire in the burlesque guise of didactic, an amusing commentary on the psychology of love. The ***Fasti*** was didactic in popularizing the new calendar; but its main object was clearly to entertain, and true didactic poetry was diminishing.

Satire. Satura meant a medley. The word was applied to variety performances introduced, according to Livy, by the Etruscans. Literary satire begins with Ennius, but it was Lucilius who established the genre. After experimenting, he settled on hexameters, thus making them its recognized vehicle. A tendency to break into dialogue may be a vestige of a dramatic element in nonliterary *satura.* Lucilius used this medium for self-expression, fearlessly criticizing public as well as private conduct. He owed much to the Cynic-Stoic "diatribes" (racy sermons in prose or verse) of Greeks such as Bion; but in extant Hellenistic literature he is most clearly presaged by the fragments of Callimachus' iambs. "Menippean" satire, which descended from the Greek prototype of Menippus of Gadara and mingled prose and verse, was introduced to Rome by Varro.

The satire of Horace and Juvenal

Horace saw that satire was still awaiting improvement: Lucilius had been an uncouth versifier. ***Satires*** I, 1–3 are essays in the Lucilian manner. But Horace's nature was to laugh, not to flay. and his incidental butts were either insignificant or dead. He came to appreciate that the real point about Lucilius was not his denunciations but his self-revelation. This encouraged him to talk about himself. In ***Satires*** II he developed in parts the satire of moral diatribe presaging Juvenal. But he was still composing ***sermones*** ("talks"), whereas his successor Persius blended Lucilius, Horace, diatribe, and mime into sermons in verse full of contemporary rhetoric. The great declaimer was Juvenal, who fixed the idea of satire for posterity. Gone was the personal approach of Lucilius and Horace. Anger, bottled up under Domitian's reign, may have become a self-indulgent habit of mind or have been cultivated for the purpose. Essentially, Juvenal was writing for effect; but his epigrammatic power makes him stimulating and at his best he is brilliant.

The younger Seneca's ***Apocolocyntosis*** was a medley of prose and verse, but its pitiless skit on the deification of the emperor Claudius was Lucilian satire. The ***Satyricon*** of Petronius is also Menippean inasmuch as it contains varied digressions and occasional verse; essentially however, it comes under fiction.

With Lucilian satire may be classed the fables of Augustus' freedman Phaedrus, the Roman Aesop, who in his beast fables could not resist contemporary social and political allusions.

Iambic, lyric, and epigram. The short poems of Catullus were called by himself ***nugae*** ("trifles"). They vary remarkably in mood and intention, and he uses iambic metre normally associated with invective not only for his abuse of Caesar and Pompey but also for his tender homecoming to Sirmio. Some of these poems are fine lyrics. Catullus and the Neoterics markedly influenced later poets in this medium.

Horace was a pioneer. In his ***Epodes*** he used iambic verse to express devotion to Maecenas, though in general he adopted with the metre the offensive posture of the Greek poet Archilochus. But his primary aim was to create literature, whereas his models had been venting their feelings. There are several poems in this collection with no Archilochian element. In the ***Odes*** he adapted other Greek metres to the Latin language. The resultant verse forms compensated for loss of fluent spontaneity with artistic elegance and dignity. Horace claimed immortality as the introducer of early Greek lyric to Latin. The ***Odes*** rarely show the passion nowadays associated with lyric; but the qualities of their originals are transfigured in them, forming something peculiarly Horatian and well-nigh perfect as meditated works of art.

Martial went back to Catullus for his metres and for his often obscene wit. He fixed the notion of epigram for posterity by making it characteristically pointed. But Martial is certainly no Catullus; he was a paltry-spirited flatterer. His social information is nevertheless valuable.

Catullus' love poems

Elegy. The elegiac couplet of hexameter and pentameter (verse line of five feet) was taken over by Catullus, whose temperament broke in tradition filling elegy with personal emotion. One of his most intense poems in this metre, about Lesbia, extends to ***26*** lines; another is a long poem of involved design in which the fabled love of Laodameia for Protesilaus is incidentally used as a paradigm. These two poems make him the inventor of the "subjective" love elegy dealing with the poet's own passion. Gallus, whose work is lost, established the genre; and Tibullus and Propertius developed it and smoothed out the metre.

Propertius' first book is still Catullan in that it seems genuinely inspired by his passion for Cynthia: the involvement of Tibullus is less certain. Already in book II Propertius also is beginning to write love elegy as a form having its own commonplaces and conventions, which were often borrowed from other forms. Tibullus' elegy is constructed of sections of placid couplets with subtle transitions. These two poets established the convention of the "soft poet," valiant only in the campaigns of love, immortalized through them and the Muses. Propertius was at first impervious to the Augustan ideals then taking shape, glorying in his abject slavery to love and his naughtiness *(nequitia),* though later he became acclimatized to Maecenas' circle.

Tibullus, a lover of peace, country life, and old religious customs, had charm and grace, and a quiet humour. Propertius, too, could be charming, but he was far more. He often wrote impetuously, straining language and associative sequence with passion or irony or sombre imagination. He could betray bad taste in overdoing mythological parallels, but he was a true and exciting poet.

Ovid had a different aim: not to unburden his soul, nor to move the reader's, but to entertain. In the *Amores* he is outrageous and amusing in the role adopted from Propertius, his Corinna being probably a fiction. Elegy became his characteristic medium. He carried the couplet of his predecessors to its logical extreme, characterized by parallelism, regular flow and ebb, and a neatness that pointed his wit. He had no successor worthy of the name.

Cicero's influence on oratory

Rhetoric** and **education. Speaking in the forum and law courts was the essence of a public career at Rome and hence of educational practice. After the 2nd century BC, Greek art affected Latin oratory. The dominant style in Cicero's time was the "Asiatic"—emotional, rhythmical, and ornate. Cicero, Asiatic at first, early learned to tone down his style. Criticized later by the revivers of plain style, he insisted that style should vary with subject. But in public speaking he held that crowds were swayed less by argument than emotion. He was the acknowledged master speaker from 70 BC until his death (43 BC). He expounded the history of Roman oratory in the *Brutus* and his own principles and practice in the *De oratore.*

The establishment of monarchy robbed eloquence of its public importance, but rhetoric remained the crown of education. Insofar as this taught boys to marshal material clearly and logically and to express themselves cogently, it performed the function of the modern essay; but insofar as the temptations of applause made it strained and affected, it did harm.

In the *De oratore*, Cicero had pleaded (through Crassus) that an orator's training should be in all liberal arts. Education without rhetoric was inconceivable; but what Cicero was proposing was to graft onto it a complete system of higher education. Quintilian, in his *Institutio oratoria,* went back to Cicero for inspiration as well as style. Much of that work is conventional, but the first and last books in particular show admirable commonsense and humanity; and his work greatly influenced Renaissance education.

History. Quintus Fabius Pictor wrote his pioneering history of Rome during the Second Punic War, using pontifical and family annals (*i.e.*, dating events from records kept officially by the priestly college and privately by great families) and writing in Greek. His immediate successors followed suit. Latin historical writing began with Cato's *Origines.* After him there were as many historiasters, or worthless historians, as the poetasters disdained by Cicero. The first great exception is Caesar's *Commentaries,* a political apologia in the guise of unvarnished narrative. The style is dignified, terse, clear, and unrhetorical.

Sallust took Thucydides as his model. He interpreted, using speeches, and ascribed motives. In his extant monographs *Bellum Catilinae* and *Bellum Jugurthinum*, he transcended the annalistic method, using history to emphasize the decadence of the dominant caste. The revolution in style he inaugurated gives him importance in an account of Latin literature. Though a Caesarian, he at least affected a Thucydidean detachment. He rebuked the extravagance of the dying republic.

Livy began his 40 years' task as Augustus came to power. His work consummated the annalistic tradition. If in historical method he fell short of modern standards, he had the literary virtues of a historian. He could visualize and vividly describe past events and interpret the participants' views in eloquent speeches. He inherited from Cicero his literary conception of history, his copiousness, and his principle of accommodating style to subject. Indeed, he was perhaps the greatest of Latin stylists. His earlier books, where his imagination has freer play, are the most readable. In the later books, the more historical the times become, the more disturbing are his uncritical methods and his patriotic bias. Livy's work now is judged mainly as literature.

The histories of Tacitus

Tacitus, on the other hand, stands higher now than in antiquity. Though his anti-imperial bias in attributing motives is plain, his facts can rarely be impugned; and his evocation of the terrors of tyranny is unforgettable. He is read for his penetrating characterizations, his drama, his ironical epigrams, and his unpredictability. His is an extreme development of the Sallustian style, coloured with archaic and poetic words, with a careful avoidance of the commonplace.

Suetonian biography apart, historiography thereafter degenerated into handbooks and epitomes until Ammianus Marcellinus appeared. He was refreshingly detached, rather ornate in style, but capable of vivid narrative and description. He continued Tacitus' account from Domitian's death to AD 378, more than half his work dealing with his own times.

Biography and letters. The idea of comparing Romans with foreigners was taken up by Cornelius Nepos, a friend of Cicero and Catullus. Of his ***De viris illustribus*** all that survive are 24 hack pieces about worthies long dead and one of real merit about another of his friends, Atticus. The very fact that Atticus and Tiro decided to publish nearly 1,000 of Cicero's letters is evidence of public interest in people. Admiration of these fascinating letters gave rise to letter writing as a literary genre. The younger Pliny's letters, anticipating publication, convey a possibly rose-tinted picture of civilized life. They are nothing to his spontaneous correspondence with Trajan, where one learns of routine problems, for instance with Christians' confronting a provincial governor in Bithynia. The letter as a verse form, beginning with striking examples by Catullus, was established by Horace. Some of his ***Epistles*** I look like real letters, though others approximate to "talks."

Suetonius' biographies

Suetonius' lives of the Caesars and of poets contain much valuable information, especially since he had access to the imperial archives. His method was to cite in categories whatever he found, favourable or hostile, and to leave this raw material to the judgment of the reader. The *Historia Augusta,* covering the emperors from 117 to 284, is a collection of lives in the Suetonian tradition. Tacitus' ***Agricola*** was an admiring, but not necessarily overcoloured, biographical study.

Some of the most valuable autobiography was incidental, such as Cicero's account of his oratorical career in the *Brutus.* Horace's largely autobiographical ***Epistles*** I was significantly sealed with a miniature self-portrait. Ovid, in exile and afraid of fading from Rome's memory, gave an invaluable account of his life in *Tristia* IV.

Philosophy and learning. The practical Roman mind produced no original philosopher. The only name that demands consideration is Cicero's. He was trained at Athens in the eclectic New Academy, and eclectic he apparently remained, seeking a philosophy to fit his own constitution rather than a logical system valid for all. He used the dialogue form, avowedly in order to make people think for themselves instead of following authority. Essentially, he was a philosophic journalist, composing works that became one of the means by which Greek thought was absorbed into early Christian thinking. The ***De officiis*** is a treatise on ethics. The dialogues do not follow the Platonic or dialectic pattern but the Aristotelian, in which speakers expounded already formed opinions at greater length.

Nor were the Romans any more original in science. Instead, they produced encyclopaedists such as Varro and Celsus. Pliny's ***Natural History*** is a fascinating ragbag, especially valuable for art history, though it shows to what extent Hellenistic achievement in science had become confused or lost.

Horace's *Ars poetica*

Literary criticism. Cicero's *Brutus* and the tenth book of Quintilian's *Institutio* ***oratoria*** provide examples of general criticism. Cicero stressed the importance of a well-stocked mind and of native wit against mere handbook technique. By Horace's day, however, it had become more timely to insist on the equal importance of art. Some of Horace's best criticism is in the ***Satires*** (I, 4 and 10; II, 1), in the epistle to Florus (II, 2), and in the epistle to Augustus (II, 1), an outspoken vindication of the Augustans against prejudiced archaists. But it was his epistle to a Piso and his sons (later called ***Ars*** *poetica*) that was so influential throughout Europe in the 18th century. It was a cunningly devised treatise based on a hellenistic critic, Neoptolemus of Parium. It supported, among acceptable if trite theses, the dubious one that poetry is necessarily best when it mingles the useful (particularly moral) with the pleasing. Much of the work concerned itself with drama. The Romans were better at discussing literary trends than fundamental principles—there is much good sense about this in Quintilian, and Tacitus' *Dialogus* is an acute discussion of the decline of oratory.

Fiction. Republican and early imperial Rome knew no Latin fiction beyond such things as Sisenna's translation of Aristides' ***Milesian Tales.*** But two considerable works have survived from imperial times. Of the ***Satyricon*** attributed to Petronius there are extracts from books XV and XVI. It was apparently a rambling picaresque novel. The disreputable characters have varied adventures and talk lively colloquial Latin. The satirical description of the vulgar parvenu Trimalchio's banquet is justly famous. Apuleius' ***Metamorphoses (The Golden Ass)*** has a hero, Lucius, who has accidentally been changed into an ass. After strange adventures he is restored to human shape by the great goddess Isis. Many passages, notably the story of Cupid and Psyche, have a beauty that culminates in the apparition of Isis and in the initiation of the hero, now assimilated to the author, into her mysteries.

II. The Middle Ages

CHARACTERISTICS OF THE PERIOD

The term medieval, "belonging to the Middle Ages," is used here to refer to the literature of Europe and the eastern Mediterranean from as early as the establishment of the Eastern Roman, or Byzantine, Empire about AD 300 for medieval Greek, from the period following upon the fall of Rome in 476 for Latin, and from about the time of Charlemagne and the Carolingian Renaissance he fostered in France (*c.* 800) to the end of the 15th century for most written vernacular literatures. The term is thus given a wider application here than a use that restricts it to the years between 1000 and *c.* 1450.

In the centuries following the fall of the Western Roman Empire one major event occurred that permanently altered the course of European development. The establishment of Christianity throughout the territories that formed the Roman Empire meant that Europe was exposed to and tutored in the systematic approach to life, literature, and religion developed by the early Church Fathers. In the West, the fusion of Christian and classical philosophy formed the basis of the medieval habit of interpreting life symbolically. Through St. Augustine, Platonic and Christian thought were reconciled: the permanent and uniform order of the Greek universe was given Christian form; nature became sacramental, a symbolic revelation of spiritual truth. Classical literature was invested with this same symbolism; exegetical, or interpretative, methods first applied to the Scriptures were extended as a general principle to classical and secular writings. The allegorical or symbolic approach that found in Virgil a pre-Christian prophet and in the *Aeneid* a narrative of the soul's journey through life to paradise (Rome) belonged to the same tradition as Dante's allegorical conception of himself and his journey in ***The Divine Comedy.***

The role of the church in preserving ancient literatures

The church not only established the purpose of literature but preserved it. For various reasons, access to Greek learning was blocked in western Europe until the 12th and 13th centuries; St. Benedict's monastery at Monte Cassino in Italy was established in 529, and other monastic centres of scholarship followed, particularly after the 6th- and 7th-century Irish missions to the Rhine and Great Britain and the Gothic missions up the Danube. These monasteries were able to preserve the only classical literature available in the West through times when Europe was being raided by Goths, Vandals, Franks, and, later, Norsemen in succession. The classical Latin authors so preserved and the Latin works that continued to be written predominated over vernacular works throughout most of the period. St. Augustine's ***City of God,*** the Venerable Bede's ***Ecclesiastical History,*** the Danish chronicle of Saxo Grammaticus, for example, were all written in Latin, as were most major works in the fields of philosophy, theology, history, and science.

The main literary values of the period are found in vernacular works. The pre-Christian literature of Europe belonged to an oral tradition that was reflected in the ***Poetic Edda*** and the sagas, or heroic epics, of Iceland, the Anglo-Saxon ***Beowulf,*** and the German ***Song of Hildebrand.*** These belonged to a common Germanic alliterative tradition, but all were first recorded by Christian scribes at dates later than the historical events they relate, and the pagan elements they contain were fused with Christian thought and feeling. The mythology of Icelandic literature was echoed in every Germanic language and clearly stemmed from a common European source. Only the Scandinavian texts, however, give a coherent account of the stories and personalities involved. Numerous ballads in different countries also reflect an earlier native tradition of oral recitation. Among the best known of the many genres that arose in medieval vernacular literatures were the romance and the courtly love lyric, both of which combined elements from popular oral traditions with those of more scholarly or refined literature and both derived largely from France. The romance used classical or Arthurian sources in a poetic narrative that replaced the heroic epics of feudal society, such as ***The Song of Roland,*** with a chivalrous tale of knightly valour. In the romance, complex themes of love, loyalty, and personal integrity were united with a quest for spiritual truth, an amalgam which was represented in every major western European literature of the time. The love lyric has had a similarly heterogeneous background. The precise origins of courtly love are disputed, as is the influence of a popular love poetry tradition; it is clear, however, that the idealized lady and languishing suitor of the poets of southern and northern France were imitated or re-interpreted throughout Europe—in the Sicilian school of Italy, the minnesingers (love poets) of Germany, and in a Latin verse collection, ***Carmina Burana.***

Medieval drama began in the religious ceremonies that took place in church on important dates in the Christian calendar. The dramatic quality of the religious service lent itself to elaboration that perhaps first took the form of gestures and mime and later developed into dramatic interpolations on events or figures in the religious service, and this elaboration increased until drama became a secular affair performed on stages or carts in town streets or open spaces. The players were guild craftsmen or professional actors and were hired by towns to perform at local or religious festivals. Three types of play developed: the mystery, the miracle, and the morality. The titles and themes of medieval drama remained religious but their pieces' titles can belie their humorous or farcical and sometimes bawdy nature. One of the best known morality plays was translated from Dutch to be known in English as ***Everyman.*** A large majority of medieval literature was anonymous and not easily dated. Some of the greatest figures—Dante, Chaucer, Petrarch, and Boccaccio—came late in the period, and their work convincingly demonstrates the transitional nature of the best of medieval literature, for, in being master commentators of the medieval scene, they simultaneously announced the great themes and forms of Renaissance literature.

MEDIEVAL LATIN LITERATURE

The history of Latin literature in the Middle Ages reflected directly the history of the society that gave it birth. The vast corpus of written Latin produced in the West during the 1,000 years that separated the end of the Western Empire from the Renaissance could never have come into being but for the widespread dominion of Rome during the previous 500 years. So long as Rome was a cultural metropolis, a unitary classical literature remained a possibility, while the maintenance of Latin as the language of the empire's governing class prevented any growth of rival vernacular literatures. When the power of Rome was broken by barbarian invasions and truncated by the rise of Byzantium as the separate eastern half of the empire and when its religion gave way to Christianity, persistence of anything that could be called a Latin literature was made possible only by the church. Rome became the centre of a new ecumenical culture, and learning and literature were preserved and produced in monasteries and monastic and cathedral schools and later in nascent universities.

Latin prose of the Middle Ages consisted—apart from official documents, chronicles, laws, treaties, diplomatic correspondence, and the like—of theological, philosophical, and scientific treatises, devotional exercises, works of biblical exegesis and ecclesiastical controversy, history, biography, hagiography, and personal correspondence: writers had practical rather than literary ends in view. The poetry was not so utilitarian or so predominantly religious. It comprised secular as well as sacred epics, didactic poems, and "occasional" verse of every description; but, except for its hymns and romantic lyrics, it is of negligible literary merit.

The contrast between classical and post-classical literature, both in quantity and in character, is striking; and in both respects it is easily accounted for: only a few—and on the whole, no doubt, the best--of the classical writers have survived, while the bulk of the medieval texts, good, bad, and indifferent, have been preserved; and there was nothing in classical times to match the influence that Christianity exercised over literate persons in the Middle Ages: the church dominated their lives, and religion coloured their minds and saturated their writings. Religion operated both positively, by supplying subject matter and inspiration and providing new, liturgical verse forms, and negatively, by treating the classics as pagan productions to be read principally as an accessory to biblical studies. Nevertheless, classical texts were studied in the monasteries and transcribed; certain of them, such as those of the Roman poet Virgil, were "baptized" for Christian use by being interpreted allegorically, while the classical liberal arts still served as the basis of medieval education in the trivium (grammar, rhetoric, and logic) and the quadrivium (geometry, arithmetic, astronomy, and music), with medicine and architecture also counted among the liberal arts. Concentration on these *artes* (arts) at the expense of the *auctores* (authors) and a debased conception of rhetoric were fatal to literary quality and led poets to indulge in acrostics and similar involved and often trivial forms that delighted the medieval mind.

The classical liberal arts

Christian Latin literature, written in a new and varied vocabulary that was conditioned by specifically Christian needs, began in the 3rd and 4th centuries with a group of African writers who had been trained in the schools of grammar and rhetoric. The most important of these were Tertullian, Minucius Felix, Cyprian, Arnobius, and Lactantius. The problem of the attitude to be taken by educated Christians toward secular literary culture was discussed by St. Augustine, bishop of Hippo, in his *De doctrina Christiana* (On *Christian Doctrine*). *A* former teacher of rhetoric himself, Augustine gave a qualified sanction to an acceptance by Christians of classical pagan learning and methods. The influence of secular classics was also a problem for St. Jerome, but in the Vulgate (his translation of the Bible), which was to exercise a profound influence throughout the Middle Ages, he produced a masterpiece of Christian Latin. Most prose of the 4th century was concerned with doctrinal matters; only the *Confessiones (Confessions)* of St. Augustine could be said to survive as literature.

The close of the 4th century saw the end of poetry that was both pagan in feeling and classical in form; Ausonius and Claudian were its last representatives; Paulinus, bishop of Nola, stood at the parting of the ways; Prudentius, a Spaniard who created the Christian ode, delivered a Christian message in classical accents. The most original literary creation of the 4th century was the Latin Christian hymn. In it, classical quantitative metres were displaced by rhymed or unrhymed measures depending on syllabic symmetry, accent, and rhythm. The hymns of St. Ambrose were the first exploration of this, the richest vein of Medieval Latin poetry.

Between 500 and 700 the chief contribution made to literature came about through the foundation of the monasteries, which provided centres for a transmission of classical texts. Among original works calling for mention were those by two pagan writers of the 5th century: a commentary on Cicero's *Somnium Scipionis (Dream of Scipio*) by Macrobius and *De nuptiis Philologiae et Mercurii* ("On the Marriage of Philology and Mercury"), an introduction to the seven liberal arts by Martianus Capella—writings that were to have influence later in medieval schools. The *Etymologiae* ("Etymologies") of Isidore, bishop of Seville, served as an encyclopaedia for many generations. In the hymns of Venantius Fortunatus a fervent Christian feeling is expressed in classical form. The *De consolatione philosophiae* (*On the Consolation of Philosophy)* of Boethius, a lofty philosophic work in prose and verse, retained its hold on the literate world throughout the Middle Ages and beyond.

The 7th and 8th centuries were a sterile period: learning sought refuge from political turmoil on the Continent in the monastic establishments of Britain and Ireland, whence missionary scholars reconveyed it across the English Channel. Most notable among British scholars were Aldhelm, remarkable more for learning than for literary grace, and the Venerable Bede, whose history of the English Church is a monument of scholarship and historical genius.

The close of the 8th century saw a revival of letters under the strong and enlightened government of Charlemagne, after whom the period is named Carolingian. He summoned Alcuin, himself a poet as well as a scholar, to come from England in order to help in the organization within the Frankish domains of schools designed to educate a literate clergy. A flourishing band of poets frequented Charlemagne's court—at their head Theodulf of Orléans, an accomplished classicist who succeeded Alcuin in charge of the palace school at Aachen—and a plentiful flow of verses, sacred and profane, in classical and in rhythmic metres, issued from the Frankish monasteries. Two poets showed especial originality in lyric measures: Walafrid Strabo and Gottschalk, both pupils of Rabanus Maurus, a learned abbot of Fulda, another of whose pupils, Lupus of Ferrières, displayed in his letters a remarkable passion for literature and the classics.

The Carolingian Renaissance

If the Carolingian Renaissance deserves its name, however, it is by reason of its regard for literature, its zeal in scholarship, and its expertise in copying manuscripts, rather than its original literary productions. The most memorable prose work of the period was the *De divisione naturae* ("On the Division of Nature") by John Scotus Erigena, an Irish scholar, in which is expounded a philosophical theory of the universe that draws on Greek sources. Otherwise, there was nothing in the Latin prose of the 9th century that deserved to survive by reason of literary merit.

Among the numerous theologians, chroniclers, authors of saints' lives, and poets of the 10th and 11th centuries, there were few that produced literature of outstanding importance, religious or secular. The most notable development in the 11th century was in the secular lyric poem, which revealed for the first time, in metre and stanzaic form, the influence of the religious "sequence." An original poetical creation of the Latin Middle Ages, the sequence was written to be sung at the mass, between the reading of the epistle and the gospel. A number of anonymous poems illustrating the use of the sequence in secular pieces occur in the Cambridge Songs, an 11th-century anthology composed in Germany.

At the turn of the 11th and 12th centuries a good deal of verse written for specific occasions and formed on classical models came out of the French cathedral schools. Examples are the poets Marbod of Rennes, Hildebert of Lavardin, and Baudry of Bourgueil, all of whom also composed hymns in accentual measures.

In the 12th century these cathedral schools, precursors of the universities, replaced monasteries as centres of literature. A wider range of studies was established, with a more intelligent devotion to auctores (authors). At the cathedral school at Chartres in France appeared a group of writers who reflected the influence of the philosophy of Plato. Most notable of these were Bernard Silvestris and Alain de Lille (Alanus de Insulis), both of whom wrote works of which later poets writing in the vernacular made use. In England, John of Salisbury, a famous medieval Humanist, owed much to his masters at Chartres. Two other English authors who wrote in Latin deserve mention: the *Historia regum Britanniae* (The History *of* the Kings of Britain) by Geoffrey of Monmouth had an unprecedented success in a century that took a great interest in romantic tales; and Walter Map, a member of the same literary circle at Canterbury to which John of Salisbury belonged, produced a mixture of gossip, folklore, and satire in his *De nugis curialium* ("On the Jests of the Courtiers").

Other great names of the 12th and 13th centuries—Peter Abelard and Thomas Aquinas, Bernard of Clairvaux, Duns Scotus, and Roger Bacon—are memorable for their contributions to philosophy or religion. In verse, the triumphs of this period were its hymns: Abelard displayed originality in the creation of new forms of stanza, while Adam of St. Victor brought the "sequence" form to perfection. In the 13th century a new note of personal passion, infused by the influence of St. Francis of Assisi, pervaded both religious life and religious literature, finding triumphant expression in the "Laude" ("Praise") of Jacopone da Todi and in such Latin hymns as the Stabat Mater.

A wave of new feeling had for a century or more been making itself felt also in the secular verse of the time, both in Latin and in vernacular writings. Its products were preserved in such collections as the *Carmina Burana*—love lyrics, drinking songs, and poems whose subject matter is as various as their form and invariably spontaneous, unliterary, fresh. Most of them were the work of "wandering scholars" and other unknown hands, but some authors can be identified: Walter of Châtillon, a mysterious Hugh Primas, and an anonymous "Archpoet," some of whose poems are good examples of a rollicking tavern verse.

Italian precursors of the Renaissance

Long before 1453, when the fall of Constantinople brought into the West a wave of Greek texts and scholars able to expound them, there had been writers of Latin in Italy who rejected medieval barbarities of form and vocabulary and, in defiance of ecclesiastic disapproval, turned to classical writers as models for their own compositions. Notable among these precursors of the Renaissance were Albertino Mussato (1261–1329), writer of tragedies modelled on those of Seneca; Petrarch, of epistles modelled on Horace and of the epic Africa; and Giovanni Boccaccio, author of *De genealogia deorum gentilium* ("On the Genealogy of the Gods of the Gentiles"), an encyclopaedia of mythology. Teachers like Guarino Veronese and Vittorino da Feltre and librarians and collectors of manuscripts like Coluccio Salutati and Poggio Bracciolini ministered to a passion for classical learning and humane literature that by the middle of the 15th century pervaded the schools and universities of Italy. This passion was fostered by powerful patrons—the Medici of Florence, the Aragonese court of Naples, and the papal court at Rome under Pius II (pope from 1458 to 1464)—and by associations of literary men. Recently established presses helped to disseminate works of current as well as of classical writers in Latin. By 1500 there existed in Italy a flourishing Latin literature written by men who cared (sometimes over much) about form, modelled their style (sometimes too mechanically) upon classical prototypes, and used Latin prose and verse as a vehicle for an expression of their personalities and for conveying an adult and humane view of life.

FRENCH LITERATURE IN THE MIDDLE AGES

Human settlement was ancient in Gaul, or France, especially in the south and centre. During the 1st millennium BC, much of the country was conquered by Celts. Their tribal names survive in the names of French cities, but knowledge of Celtic myth and legend comes almost entirely from insular sources, not France. It was the Romans who introduced written literature, and until the 12th century all documents and almost all written literature were in Latin.

The political bases of modern France were laid by the Franks, who ruled in what is now southern Belgium and in the Rhineland c. AD 500–900 and eventually controlled much of central Europe. The subjugation of Saxony by Charlemagne in 772–804 undermined the stability of the heathen north, and consequent attacks by Vikings and the political immaturity of the Frankish ruling classes led to the collapse of Charlemagne's empire (the Carolingian Empire). A western kingdom slowly emerged and took the old name Francia, "France." In this kingdom two forms of "rustic" or vernacular Latin arose: the langue d'oïl in the north and the langue d'oc in the south (*oïl* and *oc* being the two different pronunciations of the word for "yes"), each dialect becoming the vehicle of an important literature. Lyric poetry developed in the south, while a more varied, mainly narrative literature arose in the north.

Many beside the Franks contributed to the cultural heritage of northern France. After the Roman withdrawal, Celts reoccupied what is modern Brittany. Significantly, the best known themes of Old French literature, those of the Arthurian cycle, were Celtic in origin, King Arthur being a hero to Cornishmen and Bretons. Norse settlements in 911 formed the basis of the powerful duchy of Normandy, which dominated the political development of France to 1210, and the conquest of England and south Wales by Norman dukes strengthened ties between France and the Nordic and Celtic worlds: some of the oldest and finest French narrative poems, dating from the 12th century—St. Brendan, Tristan, *Haveloc*—described the adventures of Irish, Cornish, and Anglo-Scandinavian heroes. In 1154, when Henry Plantagenet, duke of Normandy, count of Anjou, and husband of Eleanor of Aquitaine, became Henry II of England, this Anglo-Norman state was incorporated into a vast Angevin empire, stretching from southeastern Ireland to the foothills of the Pyrenees.

The international status of Old French literature. The emergence of Old French, the langue d'oïl, as a vehicle of song, story, law, and administration in countries far beyond France was attributable to demographic change, foreign conquests of feudal barons, and the revival of literary studies in the cathedral schools of northern France. During the formative period of French literature (1050–1210) French barons were pre-eminent in a new kind of knightly warfare. The First Crusade, dominated by French commanders, established a kingdom in Palestine and feudal principalities in Syria. In 1204 a French ruler was placed on the throne of Byzantium. While Anglo-Norman and Anglo-Angevin poets composed couplet verse on Celtic and Danish legends, settlers in the Middle East used prose to describe their battles. Old French literature between c. 1100 and c. 1400 was widely dispersed, but everywhere it reflected the tastes, opinions, and interests of the military caste responsible for its spread.

Oral poetry and the chansons de geste. The brief written texts of the 9th and 10th centuries were little more than philological curiosities, and the 11th century was devoid of any texts in Old French. A manuscript of the ***Chanson de Roland (Song of Roland),*** greatest of the Old French epics, contains the only important literary text that is certainly datable before 1150; the manuscript is Anglo-Norman, but the poem itself, in assonanced stanzas of 10-syllabled lines, is continental and derives from an ancient tradition of oral poetry.

The contribution of the minstrels

After the collapse of the Carolingian dynasty, first attempts to unite the warrior caste were made in the 11th century, and, supported by reformed Benedictine monasteries under Cluny, military aid was organized for the Christian principalities of northern Spain. Warriors, merchants, and ecclesiastics travelled along the routes that converged on the Pyrenean passes of Roncevaux and the Somport, and thence to a pilgrimage centre, Santiago de Compostela, and for their entertainment a body of epic poetry deriving its themes from all parts of France came into being. Oral narratives on traditional themes were developed by minstrels (jongleurs), servants of great barons and ecclesiastics, but occasionally men of knightly rank or of clerical training. The main subject matter of these epic cycles—the decline of the Carolingian dynasty—depended on deep-rooted memories of the 8th-century struggles between Franks, Gascons, and Mozarabic (Arabicized Christian) emirs for control of southern France. These memories coalesced in the poetic imagination with later rebellions and dynastic conflicts.

The title chansons de geste ("dynastic histories") was occasionally given by their authors themselves to about 80 epic poems in written form, largely anonymous, at the core of which were legends concerning Charlemagne as emperor and champion of Christendom, though many chansons de geste were about individual barons, such as Girart de Roussillon, Doon de Mayence, Ogier the Dane, the Loherains (a group from Lorraine), and Raoul de Cambrai. The poems all have in common their essentially heroic themes and treatment, with a particular predilection for the description of great battles—whether of Christians against "Saracens," of the army of one great baron against another, father against son, brother against brother—and of single combats between individual champions. The ***Chanson de Roland*** was set in the context of Charlemagne's wars against the Saracens. Roland, hero of the Battle at Roncesvalles (or Roncevaux), was portrayed as a paragon of an unyielding warrior, victorious in defeat. The theme of individual honour and revenge was subordinated to that of crusading Europe united against the infidel. The 12th-century ***Pèlerinage de Charlemagne*** ("The Pilgrimage of Charlemagne") related various adventures of Charlemagne and his peers—Roland, Olivier, Guillaume d'Orange, Ogier the Dane, and others—in the context of Charlemagne's pilgrimage to Constantinople and Jerusalem. Guillaume d'Orange was the central hero of about 24 epics of the 12th and 13th centuries. There is little doubt that Guillaume was the same person as Wilhelmus, a Frankish nobleman and cousin of Charlemagne. ***Huon de Bordeaux*** involved fantastic materials, the hero being aided in his tasks by Oberon, king of faerie; that epic was known, in an English translation, to Edmund Spenser, Shakespeare, and John Keats.

Romance. Another form of narrative, difficult to define in formal terms, originated in 12th-century France. Formally, romance is distinguished from epic by being written in rhyming octosyllabic couplets instead of assonanced stanzas, or laisses. Romances were traditionally classified according to subject matter. Those based on the "matter of Britain" were concerned with stories about the legendary exploits of King Arthur's knights. Those based on the "matter of Rome," more accurately called ***romans d'antiquite'*** ("romances of antiquity"), dealt with Alexander the Great, Aeneas, and the legends of Thebes and Troy. A third group, known as *romans d'aventure* ("romances of adventure"), which contains a variety of themes, is best illustrated by such narrative poems as ***Floire et*** *Blanchefleur* and *L'Escoufle* ***(c.*** 1200) by Jean Renart.

Distinction between romance and epic and religious narrative

Romance as a separate genre is distinguished from epic and religious narrative by the fact that it recounted fictions for the delight and wonder they evoked, and not for any religious, political, didactic, or moral purpose. The romances of love, chivalry, and adventure that arose in 12th-century France were, at the outset, a creation of professional writers who had been educated in the study of the medieval trivium—grammar, rhetoric, and logic—and wrote for the amusement of aristocratic patrons. The term ***romanz,*** originally meaning the "vernacular" as opposed to Latin, became used about 1140 for vernacular translations or adaptations of Latin works the content of which was historical or pseudo-historical. Thus Wace's translation of Geoffrey of Monmouth's ***Historia*** *regum Britanniae* was known as ***Li Romanz de Brut*** (after Brut or Brutus, the mythical founder of Britain), and an anonymous adaptation of Virgil's ***Aeneid,*** as ***Li Romanz*** *d'Enéas.* It was with these and similar works—the story of Alexander the Great and the tales of Thebes and Troy—that the romance, both as a name and as a literary genre, began its evolution.

The Alexander romances, in style and form an epic rather than a romance, related the fictional biography of Alexander the Great as transmitted to the Middle Ages through Latin intermediaries. The ***Roman d'Alexandre,*** to which several authors contributed, was begun by Alberic de Briançon, grew up piecemeal over more than half a century, and was given final shape by Alexandre de Bernai. Like the chansons de geste, much of it dealt with wars and battles, but it is noteworthy for its fantastic elements, and especially for a section on the marvels of India by Lambert le Tort, and it was mainly to these exotic and marvellous features that it owed its great popularity.

The ***Roman de Thebes*** ("Romance of Thebes"), ***Roman d'Enéas*** ("Romance of Aeneas"), and ***Roman de Troie*** were all composed in the period 1150–65. They medievalized the classical world, in that they showed the warriors of antiquity behaving very much like medieval knights. More noteworthy is their technique, which reflected the influence of contemporary rhetoric in elaborate set descriptions and long passages in which the characters indulge in internal debate or self-examination, generally in connection with affairs of the heart. ***Troie*** and ***Enéas*** had a strong love interest, based on Ovid's conception of love as a restless malady. ***Troie*** also gave the story of faithless Brisei's, who abandoned Troilus, which. as expanded by Giovanni Boccaccio, formed the basis for Chaucer's ***Troilus and Criseyde*** and for Shakespeare's ***Troilus and Cressida.***

With Geoffrey of Monmouth's ***History of the Kings of Britain*** (1137), adapted into French verse in Wace's ***Brut*** (1155), a new world of imagination, that of Britain and the court of King Arthur, was opened up to writers of romance. Chrétien de Troyes (flourished 1165–80) was by far the most outstanding writer in this kind, and the impact of the Arthurian legend and courtly romance on later literature would have been less decisive had it not been handled by so accomplished a writer. Chrétien's romances, ***Erec et Enide, Cligès, Le Chevalier de la*** *charette* ("The Knight of the Cart"), ***Yvain,*** and ***Perceval,*** or ***Li conte del graal,*** were not limited to one type of subject. Chrktien told with equal zeal a story of a marriage between a lord and a poor girl ***(Erec);*** a love story of a young knight and the wife of the Emperor of Byzantium (*Cligès*); a tale of how Lancelot, the lover of Guinevere, the wife of his overlord, King Arthur, rescues his mistress from captivity ***(Le Chevalier de la charette);*** the adventures of a man who marries the widow of a man he has killed in single combat and who undertakes a series of adventures in which he is helped by a grateful lion ***(Yvain);*** and the story of a young and inexperienced knight's progress toward maturity, combined with the first known attempt at introducing the Grail theme into the Arthurian world ***(Perceval).*** Among the other themes that occur in Old French romances was that of the persecuted child regaining, after many adventures, his lost inheritance; ***Huon de Bordeaux,*** technically an epic, was a variant on this theme. The exceptionally popular ***Floire et Blanchefleur,*** which was translated into

English, Norse, Middle High German, and Middle Dutch, was a romance built around another popular theme: that of the separation and reunion of young lovers. But by far the most important medieval love romance was the story of Tristan and Iseult, originally composed in the form in which it is known in the second or third quarter of the 12th century, but preserved in its adaptations only, both in French and in other European languages—German, English, Old Norse, and Italian. The narrative poems of Béroul and Thomas (late 12th century) and Gottfried von Strassburg's *Tristan (c.* 1210) rank among the great masterpieces of the Middle Ages.

What Chrétien de Troyes's intention was in introducing the Grail theme into the last of his Arthurian romances it is hard to say, but the work had important consequences for the development of romance. About 1200, Robert de Borron, a Burgundian knight, produced a poem dealing with the origins of the Grail. Robert's apparent aim was to connect the Grail (interpreted as the vessel of the Last Supper) with the coming of a redeemer in the days of King Arthur. This notion of a predestined Grail hero appearing during Arthur's reign became the subject of two prose romances, *Perceval* and *Perlesvaus.* These works marked the beginnings of Arthurian prose romance, and prose thereafter became the leading romance form. The Arthurian romance in its classical stage was most fully developed in an anonoymous "vulgate" cycle of romances with its complex harmony of interwoven tales about the maimed Fisher King, the bleeding lance, the Waste Land, the love of Lancelot for Guinevere, the wizardry of Merlin, and the treachery of Mordred. Most of these tales are readily available to English readers in Sir Thomas Malory's *Morte Darthur*, the great example of creative adaptation, which secured the survival of Arthurian romance in the modern age.

Early lyric poetry. It was the poetic imagination of the trouvkres that shaped early lyric poetry in northern France. Poet-musicians of the 12th and 13th centuries writing in the langue d'oïl, they were largely influenced by the southern troubadours (see below). The word trouvère probably meant in the first place "maker of tropes" (that is, interpolations into the ecclesiastical chant). Whereas the troubadour school had reached its zenith by 1150, lyric poetry in the north is rarely attested before the Third Crusade (1189),

As in Provence, many of these lyrics were an expression of courtly love. The essential feature of this "love service" was the worship of the beloved by the lover, who conceived of her as a being altogether more excellent than himself and who was ennobled by the servitude of love and the suffering it entailed. This courtly attitude received unusual emphasis in all the lyric poetry of this period.

Northern verse forms

The verse forms of the north were remarkably similar to those of the troubadours: the solemn chanson, the medieval equivalent of the classical ode, usually addressed by the poet to his lady, but sometimes dealing with crusading themes; a variety of dance songs, characterized by a refrain, such as the rondeau, virelai, and ballade; and pieces in quasi-dramatic form, such as the *pastourelle* (pastoral) and the *aube* (dawn song). The *chanson de toile* was a song supposedly sung by a lady at her needlework. The term *sirventois,* which derives from the Provençal *sirventes,* meant a ribald satire suitable to a hired trooper *(sirvent),* but in the north it was soon applied to the pious hymns of the "servant" of the Virgin Mary. All these types of poetry were written to be sung.

Many of the lyrics were anonymous, but the names of some authors are known. Guy de Coucy and Blondel de Nesle acquired a legendary status: the former died on the crusade of 1203, while Blondel's name is remembered in association with that of Richard I of England. The most eminent trouvkre was Thibaut IV of Champagne, king of Navarre. Nearly half the extant trouvkre lyrics were the work, not of feudal lords, but of the citizen-poets of Arras, who formed themselves into a guild. The ablest among them was Adam de la Halle, who with Jean Bretal and others carried on the tradition of the *jeu-parti,* or debate in verse. A still more original note was struck by Colin Muset, who moved about the châteaus of the upper Marne Valley. He distinguished sharply between grasping and generous patrons and combined the conventional themes of young love with a realistic portrayal of his personal tastes (a good fire in winter, fat capons, and strong wine). The autobiographic and gastronomic theme was taken up by the versatile Rutebeuf, much of whose work involved controversy and satire aimed at the friars. Rutebeuf was the last of the trouvkres and the first modern poet and all-around man of letters.

Drama. The earliest drama in France, as elsewhere in Europe, consisted of liturgical plays, of which the best known were *mystères* (meaning originally "ceremonial acts," but by the 14th century "dramatic performances," and not to be taken as the exact French equivalent of the English "mystery" plays). These used biblical material, whereas the *miracles* were based on non-biblical legends about the saints. The offices of the church, themselves essentially dramatic, required little more than development by writers of *mystères.* The occasional revels of the lower clergy, such as the Feast of Fools, celebrated at Christmas and New Year, helped toward a development of drama distinct from that embedded in the liturgy.

The *mystères* and the *miracles*

The earliest dramatic works to use the vernacular were all anonymous 12th-century pieces: the *Sponsus*, partly in Provençalized French, partly in Latin; the Beauvais *Daniel* (in Latin, but with some liturgical refrains in French); and the *Mystère* (or *Jeu) d'Adam*, in French (from an apparently Anglo-Norman original) except for its Latin stage directions and readings. The earliest surviving *miracles* in French, Jehan Bodel's *Jeu de Saint h'icolas* (1200) and Rutebeuf's *Miracle de Thhophile (c.* 1261), were outstanding examples of this genre. Typical of the earliest known secular pieces were the short *Courtois d'Arras* (before 1228), a lively dramatization with a contemporary setting of the parable of the prodigal son; and *Le Garcon et I'aveugle* ("The Boy and the Blind Man"). The first wholly secular comic work of any worth was Adam de la Halle's *Jeu de la feuillée* (written after 1275), which poked fun at Adam's family, friends, and fellow citizens in a mixture of satire and fantasy and used characters from folklore (*e.g.*, the "three fairies" of many traditional tales). His *Jeu de Robin et Marion* (*c.* 1280) was a dramatized *pastourelle,* with songs and dances.

From the end of the 13th century onward, plays dealing with the Passion of Christ were produced with success. The *Passion de Semur* was characteristic of the type: consisting of 9,500 lines, it took two days to perform, began with the creation, and ended with the Ascension. Arnoul Greban's *Mystère de la Passion* (1450), which required four days to perform, is the most famed, poetic, and tragic of French passion plays.

Such plays, popular until the mid-16th century, began to give way in the 15th century to *moralifks, farces,* and *soties,* three kinds of comedy difficult to distinguish. The *moralité* could be serious or comic but was always didactic in intention; as it was often tedious, it was given variety by the introduction of farce. The *sotie,* or comedy played by *sots* ("fools," who dressed in costumes to suit their roles), was openly satirical. The *sotie* eventually perished as a result of its use in political attack; for example, Pierre Gringoire's *Jeu du prince des sots et mère sotte* ("Play of the Prince of Fools and Mother Fool," 1512), aimed at Pope Julius II, was later forbidden, being replaced as a weapon by prose pamphlets and by more general forms of satire. The *farce* had neither moral nor political intention and therefore was in no danger of extinction. Farcical interludes were interpolated in the *mystères* and were used to liven up the *moralitks.* In subject matter and style, medieval *farces* resembled the fabliaux (see below), their probable source. Treatment is narrative rather than dramatic, and liveliness of dialogue makes up for their lack of structurally planned plots.

The *sotie*

Allegory. Allegory, the capacity to find recurrent patterns in nature and to see analogies in different realms of human experience, was acquired by medieval man in school days (for instance, by reading Ovid with moral interpretations of the myths). While allegory is a kind of parallelism (as in treating man's life as a journey beset with dangers and distractions), the poetic symbol is somewhat different: it has been defined as "a clear visual image" that is "given much more intensity by having a meaning." The Grail is a symbol in this sense, and so is the cluster of images associated with the rose in the famous allegory of Guillaume de Lorris, the ***Roman de la rose*** (*c.* 1230; "The Romance of the Rose").

The *Roman de la rose*

The ***Roman de la rose*** began with a dreamer who enters a walled garden and looks into a transparent pool or wellspring, wherein he sees the rose reflected; the pool may represent the beloved's eyes, but, says the poet, it is also that well in which Narcissus drowned when he gazed too long at his own image. Upon this world of ambivalent symbols a formal allegory was superimposed: the dreamer seeks to obtain the rose and is aided or rebuffed by actors who personify the beloved's feelings, Fair Welcome, Disdain, and so forth. Before the outcome of the lover's quest had been revealed, Guillaume abandoned the work, unfinished; it was continued (*c.* 1275) by Jean de Meung, who used it as a vehicle for a series of sophistic debates. Nature and Genius expound the purpose of sex and procreation; Hypocrisy gives an inside account of the machinations of the mendicant friars in the University of Paris; and various political doctrines are canvassed. This most popular but most chaotic of medieval poems, which began as a delicate poetic fancy and ended as an "outline of modern knowledge," showed the medieval laity at the crossroads, politically illiterate yet eager to share in the heady speculations of the medieval university clergy.

Old French literature in its classical phase. A new start had been made in national history with the ***Grandes Chroniques de Saint-Denis*** (1274 onward) and the life of Louis IX by his companion, Jean de Joinville. The latter work is an attractive account of the saintly king, of which the core relates to day-to-day experiences during an ill-fated attack on the Nile Delta in the crusade of 1248–49, but it was retouched by the writer (1304–09) after his patron had been canonized.

Didactic literature of many kinds flourished at this period. Philippe de Rémi, sire de Beaumanoir (*c.* 1250–96), a romance writer, compiled the customary laws of his local region, the Beauvais (the ***Coutume de Beauvaisis,*** "Custom of the Beauvais"), first of a long series of such attempts. French was becoming the official language of the law in England and in the surviving coastal towns of the Kingdom of Jerusalem. The earliest French grammatical treatises were written in England in the 15th and 16th centuries and complete scientific encyclopaedias had been composed in French verse by Gautier (or Gossouin) de Metz and by the Italian Brunetto Latini.

The popularity of fabliaux

The 14th century brought an enormous extension in the scope and popularity of existing genres. Fabliaux (from the Old French noun ***fable),*** mainly produced in the 13th century, were preserved in manuscripts dating from the 13th and 14th centuries. These short narrative poems had a fair amount of realistic detail and were written to entertain and to arouse laughter. Sometimes the basis was a simple situation that could have been mirrored in real life, and the merriment provoked by many fabliaux depended on either indecorous or frankly obscene situations and observations. The most celebrated authors of fabliaux were the professional poets Jehan Bodel, Rutebeuf, and Jean de Condé, though there were also some talented amateur authors such as Henri d'Andeli and Philippe de Rémi, sire de Beaumanoir. Fables, especially the animal tales belonging to the cycle of Reynard the Fox (the ***Roman de Renart),*** also grew in popularity.

Though, before the 15th century, verse was the favourite medium, moral and educational works were also early written in prose; and household books containing, with accounts, family histories and suchlike material have been preserved from as early as the 13th century. Travellers' tales, especially by pilgrims to the Holy Land, culminated in the famous ***Voyages*** of "Sir John Mandeville" (flourished 1356), written in Anglo-Norman French. To the 14th century also belonged the ***Livre du Chevalier de la Tour Landry,*** written by the Chevalier to educate his daughters.

Eustache Deschamps, a journalist in verse who used the ballade to express the ephemeral interests of the knightly classes, is also notable as the first writer on French versification, in his ***Art de dictier,*** forerunner of a long series of ***Artes poeticae,*** while the greatest poet and musician of the age was Guillaume de Machaut. Deschamps's counterpart in prose was the last and greatest of the medieval chroniclers, Jean Froissart; for all his merits as a diarist in a brillant, disturbed epoch, Froissart is less interesting politically than are the historians of the preceding age (from Guernes de Pont-Sainte-Maxence to Joinville and Philip of Novara), who were both more critical and closer to the sources of power.

The true position of French in the literary Europe of the 14th century can be gauged most accurately from its reflection in writers from countries that had already moved some way toward literary maturity. Geoffrey Chaucer in England, Dafydd ap Gwilym in Wales, the Arcipreste de Hita in Castile, Giovanni Boccaccio in Italy were all in various ways steeped in French—its stories and its poetic metres, its manners and its rhetoric—yet each preserved and defined his national ethos and his personal originality within a European framework that, at the level of lay culture, had become distinctively and inescapably French.

The Middle French period (1380–1530). The later stages of the Hundred Years' War, and the papal exile at Avignon, form a new era in western Europe, in which the Valois rulers of France at first enjoyed immense influence. This was challenged by England and Flanders, and later by the ramifying Burgundian empire, which collapsed with the death of Charles the Bold in 1477. French ambitions revived in the 1490s with the first of a long series of invasions of Italy and intermittent occupations of Piedmont and Milan; but after the disastrous Battle of Pavia (1525), at which Francis I was taken prisoner, the Habsburg Empire, while retaining its hold on Flanders and Hainaut, annexed by treaty the whole of Artois as far as Arras and kept it for over a century. In the course of these wars, the centres of French culture shifted southward to the Loire Valley and Lyons, which at the end of this period had become the great mercantile and cultural centre of France.

In this age, France, no longer the main centre of European learning and chivalry, became a nation-state vying with others in a Europe whose military techniques and political geography had utterly changed. Under Charles V and in the early reign of Charles VI, a deeper devotion to the monarchy can be perceived; a small group of Humanists under court patronage—first the great translators Pierre Bersuire and Nicole d'Oresme, then Jean de Montreuil and Gontier Col and the first deeply learned woman writer in Europe, Christine de Pisan—mark the real origins of the Renaissance in France. The madness of Charles VI in 1392 led to a series of disasters for France, accompanied by clan feuds; at Agincourt (1415) the French chivalry was massacred by English crossbowmen, and the London government took over. Alain Chartier, secretary to the proscribed dauphin and the most versatile writer of his age, composed a dramatic poem depicting the plight of four ladies whose lovers had been killed, taken prisoner, or reported missing, or had simply fled with ignominy from the fatal field. In a terror-ridden postwar society, in which the bourgeoisie held great wealth, women in some circles greatly outnumbered men; new weapons had destroyed an ancient social caste, and a curiously modern, experimental attitude prevailed. There were "new men" proclaiming by costly tournaments and grotesque grave monuments a chivalry and piety that were confirmed by nothing in their ancestry or mores. Outrageous fashions in dress and bizarre views of sexual relations prevailed, leading to resounding literary controversies such as those initiated by Christine de Pisan, with her bitter, feminist critique of the ***Roman de la rose,*** and by Alain Chartier, with his sparkling dialogue between an arrogantly celibate lady and her despairing lover ***(La Belle Dame sans merci).***

To contemporaries, Alain Chartier was the great poet of the age; most modern critics, however, regard Charles d'Orléans as a superior writer of graceful poems in which an artificially contrived and imposed pattern of versification helps, by its recurrent lines and modulated rhythms, to provide something of a musical accompaniment even to poems not written for music. Though his ballades are inferior to those of François Villon, Charles's rondels (or rondeaux) are unequalled. Villon (died after 1463) was an incomparable figure in later medieval French literature. His main works, Le Testament and Le *Lais* (or Le Petit Testament), were written in eight-line octosyllabic stanzas, interspersed with lyrical poems in ballade or rondeau form. There was little in Old French literature to compare with the best of these, while the composition of the works as a whole was marked by both vigour and pathos.

Humanism, translation from the Greco-Latin classics, and Italian influence had already modified the language; a new and decisive influence was the use of French for administrative purposes throughout the kingdom. The final departure of the English from the provinces, some of which they had controlled for three centuries, led to the establishment by the monarchy of local high courts, or *parlements*, staffed by a legal caste, or noblesse de robe, deeply attached to the royal cause, who used French in their public deliberations but dialect, often Provençal or Gascon, in daily life. The pellucid grace, the phonetic purity, and the delicate sense of semantic values, characteristic of Old French in its classical phase, soon disappeared when French became the vernacular of an embryo nation-state.

Individualist aspects of the Renaissance, obsession with public success, and the rise and fall of military leaders found expression in the great school of contemporary historians that flourished at the court of Burgundy: Georges Chastellain (died 1474), Jean Molinet (died 1507), and Oliver de la Marche (died 1502). The greatest was Philippe de Commynes (died 1511), who left the service of Charles the Bold for that of Louis XI in 1472 and after a complex diplomatic career at court ended as ambassador to Venice during the French invasions of Italy in 1494–95. His Memoires, which span four decades, have won him the title of the first modern French historian.

An age of self-conscious modernity produces on the whole a literature of ideas—and nothing fades more quickly. Hence 15th-century literature is little read and studied, even by scholars. But the old medieval literature, with its deep humour and humanity, survived in the *nouvelle*, successor to the fabliau, and in dramatic literature, farces, and mystery plays, which are the living literature of the age. Lyric verse, wedded to music, preserved the ancient trouvère tradition and and even developed it to unprecedented achievements. Polyphonic music after Machault reached dizzy heights of mathematical complexity, and it was quite commonplace to sing simultaneous texts in different languages and on different themes, much as in certain scenes of modern opera. France, England, Flanders, Burgundy, and northern Italy entered into a kind of cultural symbiosis, which was not definitely modified until the Reformation era, when, with the northern breach with Rome and the decisive shift of French culture to a new Lyonnese centre, another literary period began.

ANGLO-NORMAN LITERATURE

The literature in the French dialect of medieval England is also known as Norman-French or Anglo-French. Though it had been introduced to English court circles in Edward the Confessor's time, its history really began with the Norman Conquest in 1066, when it became the vernacular of the court, the law, the church, schools, universities, parliament, and later of municipalities and of trade. For the aristocracy it became an acquired tongue and its use a test of gentility, and the more widely it was known the more it was corrupted. It was introduced into Wales and Ireland and used to a limited extent in Scotland before and during the wars of independence. The earliest extant literary texts belonged to the reign of Henry I, the latest to that of Henry IV. The diminution of friendly relations with France during the Hundred Years' War started an increasing use of English, the last strongholds of a French dialect being parliament and the law, in both of which it still survives in a few formulas.

Most types of literary works were represented in Anglo-Norman as in French, with a slight difference of emphasis. The chanson de geste was an exception. The type was not unknown in England (*e.g.*, the only surviving manuscript of the assonanced version of the *Chanson de* Roland is Anglo-Norman), but there seem to be no original works of the kind. Conversely, Anglo-Norman works were known, copied, or imitated on the Continent. One important difference between continental and Anglo-Norman literature is that the Fourth Lateran Council of 1215 led to an outpouring of doctrinal and devotional works for the laity in England not paralleled in France, which perhaps explains the fact that in the early periods England was often in advance of the Continent in the development of new literary forms. History was popular both in Normandy and on the Continent; and although, after the Norman Conquest, Latin replaced English in documents and chronicles, examples of both are found in Anglo-Norman. Religious houses caused lives of native saints to be written, and the nobility had a taste for romances about imaginary English ancestors. Thus social and political differences between the two countries prevented Anglo-Norman literature from being a mere imitation of French.

Influence of the Fourth Lateran Council

Religious and didactic writings. In the 12th century the oldest substantial Anglo-Norman prose work, "The Book of Kings," was written in England, as were many versions of the Psalter. Sanson de Nanteuil translated into verse the proverbs of Solomon, with commentary; and in the 13th century Robert of Greatham wrote the Sunday Gospels for a noble lady. The same century saw the beginning of the magnificent series of Anglo-Norman apocalypses, best known for their superb illustrations, which served as a model for a series of tapestries at Angers, France. In the 14th century an Anglo-Norman Bible was begun, though never completed. Anglo-Norman was rich in literature of legends of saints, of which Benedeit's "Voyage of St. Brendan" was perhaps the oldest purely narrative French poem in the octosyllabic couplet. Wace led the way in writing a saint's life in standard form but was followed by Anglo-Norman writers in the 12th century who wrote numerous lives, many connecting religious houses with their patron saints.

The oldest play in French, the Mystdre *d'Adam*, is Anglo-Norman. The resurrection play La Seinte *Resureccion* was probably 12th century but was rewritten more than once in the 13th century; it is unique in having stage setting and action described in passages of verse that seem to have been designed to be read during the performance. There were a few religious allegories, the most important, the "Castle of Love," being the oldest in French.

The Fourth Lateran Council of 1215 led to the compilation of instructive works, the oldest and most attractive being the Merure de seinte *église* ("Mirror of Holy Church") by St. Edmund of Abingdon. In the 13th–14th century countless treatises appeared on technical subjects—manuals for confession, agriculture, law, medicine, grammar, and science, together with works dealing with manners, hunting, hawking, and chess. Spelling treatises produced in the late 13th, 14th, and 15th centuries are valuable for the light they shed on continental French as well as Anglo-Norman.

Romances. Anglo-Norman literature was well provided with romances. In the 12th century one Thomas wrote a courtly version of the Tristan story, which survived in scattered fragments and was used by Gottfried von Strassburg in *Tristan* und *Isolde* as well as being the source of the Old Norse, Italian, and Middle English versions of the story. Béroul's Tristan, also 12th century, was probably written in England, but by a Norman; *Waldef*, a long, confused story of an imaginary king of East Anglia and his sons, has passages of remarkable originality. In the 12th century some romances were composed in the form of the chanson de geste; for example, ***Horn,*** by Master Thomas, which is connected with the Middle English ***Horn Childe and Maiden Rimnild.*** Yet another Thomas wrote the ***Roman de toute chevalerie*** ("Romance of All Chivalry"), an independent version of the Alexander romance and the source of the Middle English romance ***King Alisaunder.*** In the 13th century the more courtly type of romance reappeared in *Adamas et Idoine* and in *Amis et Amiloun,* perhaps derived from the same source as the Middle English poem ***Amis and Amiloun.***

The *Lais* of Marie de France

Lais and fabliaux. Marie de France, earliest named French woman poet, about whom little is actually known, wrote fables based on an English source and 12 narrative lays (dedicated, probably, to Henry II of England) in octosyllabic rhymed couplets. She claimed that they had Breton lays as their originals. The ***Lais*** combined realistic and fairy-tale elements, and their author was skillful in the analysis of love problems and often showed a keen interest in contemporary life. A few fabliaux have been found copied in manuscripts from religious houses, probably for exemplary purposes.

Political and historical writings. Anglo-Norman lyrics were few in number and unoriginal in form, but numerous political satires and songs were written. Fragments of political songs are found in Peter Langtoft's ***Chronicle,*** which begins as a ***Brut—a*** complete chronicle of British history—but became a source for the times of Edward I; while the Dominican Nicholas Trevet wrote a prose chronicle of European history from which Chaucer derived his "Man of Law's Tale." Earlier than these was an Anglo-Norman verse, ***Estoire des Engleis,*** by Geffrei Gaimar (*c.* 1140), the earliest chronicle in French. Two magnificent biographies of the 1st earl of Pembroke (William Marshal) and of Edward, the Black Prince, were written for English patrons by foreigners. Official documents were often in Anglo-Norman, and the ***Yearbooks,*** unofficial reports of cases in the common pleas, ran from the reign of Edward I to that of Henry VIII. English began to be used in Parliament alongside French in the late 14th century.

Natural history and science. One of the earliest writers in Anglo-Norman, Philippe de Thaon or Thaün, wrote ***Li Cumpoz (The Computus),*** the first French bestiary, and a work on precious stones. Simund de Freine based his ***Roman de philosophie*** on Boethius, to whom the 13th-century ***Petite Philosophie*** also owes much. All these are allegorical or didactic.

MEDIEVAL PROVENCAL

The literature written in the oc-speaking southern France and, for a limited period, in Italian and Spanish cultural dependencies can be divided into three periods, two of which fall outside the scope of this section: the first, a period of unity during which a literary language called "classic Provenqal" was universally written, if not spoken (11th–15th century); the second, a period of dispersion in which a variety of provincial dialects were developed (16th–18th century); and the third, a period of partial recovery (19th–20th century).

Poetry. The oldest piece of Provenqal verse extant, a refrain in a Vatican manuscript, probably belonged to the 10th century. More important are a fragment of didactic verse preserved in an Orléans manuscript and the beginning of an anonymous adaptation of Boethius' treatise ***On the Consolation of Philosophy.*** At the close of the 11th century, there were the poems of William IX, duke of Aquitaine (Guilhem VII of Poitiers), 11 strophic pieces (in stanza form with repeat lines) meant to be sung. Several were love songs, and the most important expressed the writer's regret for the frivolity of his past and apprehension at bidding farewell to his country and his son. The contrast between Boethius' poem and the stanzas of William IX indicates that by the 11th century Provenqal poetry was developing in various directions. Apparently Romance poetry sprang out of a popular poetry from Roman times, which was gradually refined, in keeping with its aristocratic audience. The origins of the poets were indicated by Latin chroniclers, who mention ***ioculares,*** men of a class not highly regarded, whose profession consisted in amusing their audience by juggle tricks, exhibiting performing animals or by recitation and song. They were called ***joglars*** in Provençal, ***jongleurs*** in French. From among them rose the troubadours, who originally may have been *joglars* skilled in poetry.

Provenqal love songs

Provenqal love songs reflected social conditions in the Midi (southern France) under feudalism. Daughters of territorial lords were married for political reasons and welcomed attentions of courtiers, who addressed songs of love to them. As the poets were usually far beneath the ladies in social status, they wrote in a most guarded style. This profession of "courtly," or "chivalrous," love became a matter of convention, but it would be a mistake to assume that the real experiences were as platonic as the expression of them.

The troubadours. By the end of the 11th century a clear distinction had been drawn between the lower sort of *joglars* and refined poets, troubadours of both lower and noble ranks of society. William IX had been the first of the troubadours. In the first part of the 12th century, Cercamon, a Gascon, composed pastorals, and his pupil Marcabrun wrote about 40 pieces, several of which were concerned with contemporary history. Jaufré Rudel of Blaye, a nostalgic singer of the ***amor de lonh*** (distant love), is scarcely less famous. Slightly later in the same century Bernart de Ventadour composed songs of elegant simplicity, some of which may be taken as perfect specimens of Provenqal poetry. His contemporary Bertran de Born is famous for the part he is said to have played both with his sword and his *sirventes* (a form of Provençal lay) in the struggle between Henry II of England and his rebel sons. Other troubadours include Arnaut Daniel, a master of complicated versification and difficult form; Guiraut de Borneil, an acknowledged master of ***trobar clus*** or "close" style, though he also composed songs of charming simplicity; Arnaut de Mareuil, noteworthy for his exquisite delicacy of sentiment; the somewhat eccentric Peire Vidal of Toulouse; the chivalrous Raimbaut de Vaqueyras; Folquet de Marseille (a monk who became bishop of Toulouse); the truculent monk of Montaudon; and the satirical Peire Cardenal.

Effects of the Albigensian Crusade

When in the first decades of the 13th century the Albigensian Crusade had ruined a large number of nobles of the Midi, the profession of troubadour ceased to be lucrative, and many went to spend their last days in the north of Spain and Italy, where Provenqal poetry was esteemed. Following their example, other poets began to compose in Provenqal, but from the middle of the 13th century they began to abandon the foreign tongue and took to the local dialects. About the same time in the Midi itself poetry had died out save in a few places, and in the 14th century works were mainly for instruction and edification: the poetry of the troubadours was dead.

Chansons de geste and historical poems. Provenqal literature has some highly important specimens of the chansons de geste (poems in stanzas in indefinite length, with a single rhyme). The first place belongs to ***Girart de Roussillon,*** a poem of 10,000 lines, which related the struggles of Charles Martel with his vassal Gerard of Roussillon; it was a recast of an older poem no longer extant, probably of French or Burgundian origin. Midway between legend and history was the Provenqal chanson on the siege of Antioch. To history proper belonged a chanson of the crusade against the Albigensians, which, in its present state, is composed of two poems. The first was the work of Guilhem de Tudela; the second, different in language, style, and opinions, was a masterpiece of Provenqal literature.

Narrative poems. Only three Provenqal romances of adventure have survived: ***Jaufré, Blandin de Cornoalha,*** and ***Guillem de la Barra.*** Connected with the romance of adventure was the ***novel*** (in Provenqal, plural, ***novas),*** which was originally an account of an event recently happened. Some of them could be ranked with the most graceful works in Provenqal literature. Two were by the Catalan author Ramon Vidal de Besalú: the ***Castia-gilos*** was an elegant treatment of a story of the husband who disguises himself as his wife's lover, and the other was a recital of a question of the law of love. Mention may also be made of ***Novas del Papagai*** by Arnaut de Carcassès, in which the principal character is an eloquent parrot, who assists his master's amorous enterprises. Novels came to be extended to the proportions of a long romance, and ***Flamenca*** was a poem of over 8,000 lines, in which a lady by ingenious devices eludes the vigilance of her jealous husband: no book in medieval literature had more quickness of intellect or was more instructive about the manners and usages of polite society in the 13th century.

Didactic and religious poetry. The more important works were a very early Boethius poem already mentioned, an early metrical translation of the ***Disticha de moribus*** of Dionysius Cato, and Daude de Prades's ***Auzels cassadors***, a source for the study of falconry. More original were some compositions whose general object was the education of ladies of rank. Of metrical lives of saints, there remain about a dozen, including a life of St. Fides, a life of St. Enimia by Bertran of Marseilles, and one of St. Honoratus of Lérins, distinguished for variety and elegance of versification but almost entirely a translation from Latin, by Raimon Feraut.

Drama. The dramatic literature of the Midi consisted of mysteries and miracle plays seldom exceeding 2,000 or 3,000 lines and generally belonging to the 15th or 16th century. The oldest appears to be the "Mystery of St. Agnes," written in Arles. Later came a mystery on the Passion of Christ and another on the marriage of the Virgin, the latter partly adapted from a French poem of the 13th century.

Prose. In the 12th century, in Languedoc, sermons whose importance is more linguistic than literary were written, and, about the same time, in Limousin, St. John's Gospel was being translated. Various translations of the New Testament and of some parts of the Old were made in Languedoc and Provence during the 13th and 14th centuries. To the 13th century belonged certain lives of troubadours intended to be prefixed to and to explain their poems, many of them being written before 1250, when the first anthologies of troubadour poetry were compiled. To the same period belonged ***Las Razos de trobar*** of the troubadour Ramon Vidal de Besalu (treatise on various points of grammar and the poetic art). A leading prose work of this period was a treatise on grammar, poetry, and rhetoric known as ***Leys d'amor,*** composed in Toulouse. The decay of Provençal literature, caused by political circumstances, arrived too soon to allow a full development of prose.

MEDIEVAL ENGLISH LITERATURE

The Old English period. Since the effects of the Norman Conquest on English literature began to be apparent in the early 12th century, pre-Chaucerian writings are conveniently divided between an Old English (Anglo-Saxon) and an Early Middle English period, with the division at about 1100.

Poetry. The Angles, Saxons, and Jutes that settled in Britain in the 5th and 6th centuries AD had an oral poetic tradition, but no poetry from the pre-Christian period has survived except as fragments in gnomic verses and heroic poetry composed in Christian times. The metre of Old English poetry, with a rich poetic diction suited for heroic themes, had its roots in the pagan past. It was sometimes sung to the harp, not only by minstrels but also by humbler performers, the peasants in the Venerable Bede's story of Caedmon. According to Bede, whose ***Historia ecclesiastica gentis Anglorum (Ecclesiastical History of the English Nation)*** is a main source for Anglo-Saxon history, it was a Northumbrian peasant, Caedmon, who first used the traditional metre and diction for Christian religious poetry. He entered the monastery of Whitby, where he composed many poems, but only the first nine lines of a hymn on the creation have survived. Biblical poems are also contained in the Junius (Caedmon) manuscript and three poems known collectively as ***Christ and Satan.*** This manuscript was one of the four manuscript collections of Old English poetry, all written *c.* 1000. The others were the Beowulf manuscript, The Vercelli Book, and The Exeter Book. Most of the poetry in these manuscripts dated from the 8th or 9th century. Outside these collections, there remain a few short pieces in 8th-century form from various sources.

Manuscript sources of poetry in Old English

Since Old English literature was preserved almost entirely by religious houses, it is not surprising that the extant poetry is largely religious. Besides didactic poetry and some close handlings of biblical material, such as Genesis and Daniel (from the Junius manuscript), there is in Exodus (from the same source) a heroic picture of Moses by an original and imaginative poet and in "Judith" (from the Beowulf manuscript) a colourful apocryphal story. Heroic tradition also influenced "Elene" and "Juliana" by Cynewulf, and two other poems of his are extant, "The Fates of the Apostles" and "The Ascension" ("Crist" II). All his work was marked by clarity and metrical skill, using rhyme as an additional ornament. By far the most remarkable religious poem was "The Dream of the Rood," with its moving account of the crucifixion placed in a dream framework in which the dreamer sees the cross dominating the universe; in depth of feeling and expression this is one of the greatest religious poems.

The terms lyric and elegiac are applied to The Exeter Book poems: "The Wanderer," "The Seafarer," "The Ruin," "The Wife's Complaint," "The Husband's Message," and "Deor." All were poems written with great power and told of the sentiments of individuals in typical situations. "Deor" made use of heroic legends, for in it a legendary minstrel consoles himself for loss of his post by remembering how various heroes have passed through affliction to better times. Had The Exeter Book been lost, the existence of this body of sensitive and poignant lyric would never have been suspected.

Only the fragments "Waldere" and "The Fight at Finnesburg" remain to give a glimpse of the dramatic quality of heroic narrative poetry. Allusions to such stories were used for various purposes in "Deor" and in ***Beowulf.*** In this poem, the most important of all that survive, the narrative of the hero's exploits against monsters was placed in a context of the heroic past. ***Beowulf*** is more than a heroic poem; it has deep religious significance, the monsters being depicted as the enemies of God and the human acts of violence, disloyalty, and ambition being shown to bring retribution. In diction, vividness, and narrative power, it takes first place among surviving poetry from this age.

The heroic and religious in ***Beowulf***

Finally, there was some poetry on historical themes, notably, "The Battle of Brunanburh" and the fragmentary "Battle of Maldon" (written soon after 991), a dramatic and vigorous narrative showing that heroic ideals of loyalty and courage were still fully alive.

Prose. The writing of Old English prose began soon after the conversion of the English to Christianity following the mission led by St. Augustine of Canterbury to southeast England in 596–597. The laws of Aethelberht I were written down c. 602–603, though they survived only in a 12th-century manuscript. The turning of the Lord's Prayer and the creed into English were equally early. Seventh-century kings of Kent issued legal codes, and in the 9th century it was common for wills and other deeds to be written in English; before the century closed King Alfred claimed that many persons could read English. From his reign (871–899) survived the first literary prose, and, though much was based on Latin originals, it is a mistake to dismiss it as uninteresting. Even the translation of St. Gregory's ***Dialogues*** by Werferth and Alfred's rendering of ***Cura Pastoralis (Pastoral Care)*** contained passages not in the original works; the version of Bede's ***History*** translated from the Latin included sections in vivid, natural prose. Of far greater literary interest are Alfred's adaptations of Boethius' ***Consolation of Philosophy*** and of St. Augustine's ***Soliloquies,*** which used other sources and also included original passages of high quality. The version of Orosius' ***Historia adversus paganos*** ("History Against the Pagans"), attributed to Alfred, was freely rendered: it was greatly compressed but retained a general theme of the working out of God's will in history and revealed an individual point of view influenced by heroic tradition.

It was in Alfred's reign that ***The Anglo-Saxon Chronicle***

The Anglo-Saxon Chronicle

was first compiled. It recorded briefly events prior to the 9th century, except that an untypically vivid account of a feud was inserted into the annal for 755. The events of the 9th century were related in plain narrative styles, which continued to the death of Edward the Elder in 924. Chronicle writing was in abeyance for some 50 years; and when it was resumed, its writing shared the great advance that characterized the writing of the period after the 10th century monastic revival, which was one of the great periods of English prose. Its most famous exponents were Abbot Aelfric and Archbishop Wulfstan of York, both in Aethelred's reign. Aelfric wrote in a restrained style of great beauty and elegance, while Wulfstan ranged from impassioned appeal to measured statement of religious truths. Much of the content of the works of this time came from Latin sources but was so handled as to make the results literature in their own right. Bishop Aethelwold of Winchester translated the Benedictine Rule, and the Gospels and parts of the Old Testament were also translated at this time. As a result of all this activity, a remarkable vocabulary developed. Interest in science was shown by Byrhtferth's ***Manual*** and by Aelfric's ***De temporibus,*** both based on Bede's scientific works. Finally, some literature of entertainment has come down. Three pieces on marvels in the Beowulf manuscript, a fragmentary ***Life of St. Christopher, Alexander's Letter to Aristotle,*** and ***Wonders of the East*** were written in indifferent prose, but the Greek romance ***Apollonius of Tyre*** was lucidly and cogently rendered.

Throughout the period, from the late 7th century on, Latin works were produced; yet it was the creation of a large body of vernacular literature, in verse and prose, that was one of the most remarkable achievements of the Anglo-Saxons, without parallel among other European races of early times. The excellent writing of the ***Chronicle*** continued, and great activity in the production of manuscripts, some of them beautifully illuminated, was maintained throughout the 11th century. In literature (as in several other respects) the civilization of the Anglo-Saxons was superior to that of the Normans, who conquered them in 1066.

Early Middle English period. The extant texts suggest that the Norman Conquest had little immediate effect on English writing: during William the Conqueror's reign even royal writs and charters continued sometimes to be issued in English. Changes came about gradually, through influence from continental culture.

Prose. ***The Anglo-Saxon Chronicle*** continued until 1154, maintaining and at the same time modifying old traditions. The last two sections contain a famous description of the anarchy in Stephen's reign and the black comedy of Henry of Angély's abbacy at Peterborough. The abandonment of English historical writing was mainly due to a growing use of Latin for all work intended as more than ephemeral entertainment: a consequence not so much of the Conquest as of the 12th-century revival of learning throughout western Europe. A brilliant school of Anglo-Latin chroniclers arose, the earliest being Edmer of Canterbury and the best known William of Malmesbury. By the 1130s Latin chronicling merged into prose fiction, with Geoffrey of Monmouth's ***History of the Kings of Britain,*** one of the earliest surviving treatments of the Arthurian legend, here blended with the legend of how Britain had been colonized by Brutus of Troy. Latin was used even for works of local interest and for lives of such contemporary English saints as Wulfric of Haselbury, William of Norwich, and Aelred of Rievaulx. In this field French verse also soon came to be used. Before 1140 Geffrei Gaimar had compiled his ***Estoire des Engleis,*** and other Anglo-Norman metrical chronicles included Wace's ***Roman de Rou.*** By about 1300, metrical chronicles and saints' lives were appearing in English, the best known of such chronicles being those of Robert Mannyng and Robert of Gloucester.

Early Middle English devotional writing

Yet despite competition from Latin and French, English prose was still used for devotional writing, presumably because in this type of work neither intellectual pretension nor aristocratic patronage counted for as much as general intelligibility. The most famous English prose works of this period were the collection known as the ***Katherine Group*** and a long treatise on the religious life known as the ***Ancren Riwle*** or ***Ancrene Wisse*** ("Guide for Anchoresses"). Though contrasting in style (the ***Katherine Group*** being old-fashioned in diction and heavily alliterative, the ***Ancrene Wisse*** supple in style and language, including numerous French loanwords), they had in common the influence of continental thought.

Verse. For verse the range of genres was wider, and competition and influence from Latin and French traditions were more complex. Latin hymns had been current wherever the Roman Church had reached; and, with the patronage of secular literature in the hands of the Anglo-Norman aristocracy, French verse flourished in post-Conquest England. The oldest and best manuscript of France's great early national epic, the ***Chanson de Roland,*** was Anglo-Norman. With many in England bilingual and some trilingual, Latin and French verse could not but affect English modes. By the late 12th century these influences had transformed English prosody. While Layamon's ***Brut*** used an irregular metre, a compromise between the Anglo-Saxon alliterative line and the French octosyllabic couplet, the ***Ormulum,*** a series of metrical homilies composed about the same time by Orm, used a 14-syllable line of Latin origin, with alliteration only for ornament and emphasis. ***The Owl and the Nightingale*** (late 12th century) used an adaptation of the commonest French metre, the octosyllabic couplet, soon to be established as the standard English metre for narrative and discursive writing, and the poem exhibited a poised acceptance of continental themes as well as forms: staging a debate between the two birds, the poet ranged over many topics—witchcraft, the church, and marriage—although whether the birds had any consistent symbolism is disputed. Much of the poem's merit lies in its diction, natural and lively, yet elegant.

Layamon's ***Brut***

Layamon's ***Brut*** contrasts with the ***Owl*** in every way. Translated from Wace's ***Roman de Brut*** (his translation of Geoffrey of Monmouth's Latin ***History of the Kings of Britain),*** it rendered its material in Anglo-Saxon heroic terms, suppressing the chivalric elements. Archaic even in its own time, the diction assimilated King Arthur to early Germanic heroes. The first Arthurian work in English, Layamon's ***Brut*** may have remained the only one until ***Arthour and Merlin*** appeared toward the end of the 13th century.

As a bourgeois counterpart to the aristocratic Arthurian cycle stood the ***Roman de Renart,*** irreverent and satirical. How well this was known and appreciated in England is implied by widespread use of motifs from it in carvings; but only one pre-Chaucerian tale, ***The Vox and the Wolf,*** has survived. In the same manuscript was ***Dame Sirith,*** the only pre-Chaucerian English representative of the realistic, usually ribald, verse tale known as the fabliau.

Other narrative genres are more amply represented. The verse romance was widely cultivated in English by the mid-13th century, the earliest example extant being probably ***King Horn (c.*** 1225). Most early romances in English seemed meant for popular rather than aristocratic audiences. ***Havelok the Dane,*** which told how a Danish prince in exile as a Grimsby fisherman's son regained his inheritance, gave a scullion's view of castle life and endowed its hero with working class virtues, while ***Floris and Blauncheflur***, translated from a French tale of Oriental origin, excised sentimental embellishment so as to allow concentration on the plot. Within the romance genre was the "Breton lay," a tale of love and magic, often with a Celtic setting, as in the ***Lais*** of Marie de France. The most attractive English "Breton lay," ***Sir Orfeo,*** which retold the Orpheus legend in Celtic terms, is notable for the simple elegance of its structure and diction.

Early Middle English lyric works

The most novel genre in early Middle English was the lyric. In Old English lyric metres were unknown, "Deor" and some of the ***Charms*** representing the nearest approach; nor did the dominant themes of love and springtime much concern Old English poets. In western Europe generally, the vernacular lyric began only in about 1100, with the work of the Provençal troubadours, but spread rapidly appearing in English by the early 13th century. Before 1170 simple hymns were composed by St. Godric of Finchale, and, according to Giraldus Cambrensis, whose contemporary accounts are a valuable source for this period, dance songs with refrains were current among the peasantry in the late 12th century. The earliest English secular lyric extant was *c.* 1200; the earliest noted with music were "Mirie it is while sumer ilast," from about 1225, and the famous "Sumer is icumen in." From the mid-13th century on, lyric anthologies survived in fair number, the most famous being the "Harleian Miscellany," probably compiled 1330–40. Its contents are mixed, the same hand copying secular and religious pieces, verse and prose, French, Latin, and English.

Thus the early Middle English period showed English literature being transformed by influences from the Continent, but gradually and without losing its own identity.

From Chaucer to the Renaissance. ***Poetry: the alliterative tradition.*** The middle of the 14th century saw the beginning of a new fertility in the composition of long poems in an alliterative metre. Among the early group were very varied themes—love romances such as ***William of Palerne*** (before 1361); quasi-historical fragments on the life of Alexander; political satire such as ***Winner and Waster;*** and religious legend such as ***Joseph of Arimathie*** (*c.* 13-50), treating of the Holy Grail. Surviving records give an impression that alliterative verse had suddenly been taken up again after a long interval, and the movement is generally called the alliterative revival. The differences from Old English technique were due largely to developments in the language, and the "revival" should therefore be seen as a striking expansion of an existing way of writing. It was a western and northern movement, contrasting with the rhyming technique widespread in London and the east. This kind of verse continued to be written in Scotland up to the early 16th century. There is a considerable volume of it, and the best work, all apparently 14th-century, is equal to the finest that Chaucer wrote. The "heroic romance" ***Morte Arthure*** is virile and noble and served as Sir Thomas Malory's model for part of his ***Morte Darthur. Piers Plowman*** (earliest version *c.* 1370; attributed to William Langland) expressed in a complex allegory, with unique power and richness, the author's faith in man's need for Christian truth and charity. The poem is long and irregularly constructed; its language is often unpolished and its metre loose. But its intensity of feeling and range of thought have no parallel. It is a magnificent statement of the mind and faith of the late 14th century and one of the greatest of English poems.

Piers Plowman and ***Sir Gawayne and the Grene Knight***

In a very different tone was the chivalric romance ***Sir Gawayne and the Grene Knight,*** in which the author told a tale of enchantment in an Arthurian setting; but he depicted his characters and his hero's moral dilemma with such subtle understanding that the poem becomes a serious reflection upon human conduct: Gawayne is represented as a devout Christian, humanly imperfect. The alliterative lines are tautly constructed, and the vocabulary is extraordinarily rich—strongly influenced by French but coloured also by dialect words of northwest England. The blend of sophisticated atmosphere, psychological depth, and regional language produces an effect unlike that of any other work of the time. In the same manuscript as ***Sir Gawayne*** and with so many stylistic similarities that they are generally attributed to the same author were two alliterative poems of moral teaching, "Patience" and "Purity" (Cleanness), and an ostensibly elegiac poem called "Pearl," in which the poet sees a vision of his daughter who died in infancy and is instructed by her in submission to God's will. He conducts a theological debate with intricate technique, showing in descriptions an enjoyment of colour and light as in ***Sir Gawayne;*** yet for all this complexity he conveys a poignant sense of personal grief.

Poetry: the Chaucerian tradition. The alliterative revival reached great heights, yet it was the end of a line. The future lay mainly with the tradition of Middle English verse modelled on French in the early Middle English period. Soon after the alliterative revival began this other tradition acquired new life from the early works of Geoffrey Chaucer, which, however, for all their skill, still owed an enormous debt to France—the influence of the *Roman* ***de la rose,*** which Chaucer says he translated, was very great. The ***Book of the Duchess*** and the ***House of Fame*** take up the humorous detachment foreshadowed in ***The Owl and the Nightingale*** and employ the same verse form. It was with the *Parliament of Fowles* that Chaucer established himself as a powerful new genius. It has some philosophical depth as well as humour, and the dialogue looks forward to the superb conversation scenes of ***Troilus and Criseyde.*** In the ***Parliament*** Chaucer used, for the first time in a major English poem, the ten-syllable line, which has been important ever since. He adopted it from French, with the additional example of the Italian poets, whom he had begun to read and imitate. The *Parliament* was arranged in stanzas of seven lines, the "rime royal" that Chaucer used again in ***Troilus and*** *Criseyde.*

Continental influences on Chaucer

It was in the "Prologue" to the ***Legend of Good Women*** that he first used the ten-syllable line in couplets, which he continued to use in most of ***The Canterbury Tales.*** Chaucer's absorption of continental influences is most strongly seen in ***Troilus and Criseyde;*** his plot was derived from Boccaccio, much philosophical background from Boethius and Dante, much of his attitude to his leading characters from conventions of courtly love—but he transformed Boccaccio's direct narrative, especially by deepening the treatment of character. Yet though ***Troilus and Criseyde*** may be considered Chaucer's masterpiece, it cannot finally stand against ***The Canterbury Tales.*** Though the tales vary in interest, the finest of them are immensely rich in the subtle play of incident and character, and the pilgrims who tell them are revealed, in the linking passages and in the incomparable "Prologue," with the greatest shrewdness and penetration.

Chaucer was early recognized as a great master of English poetry but seemed less pre-eminent than he does today. John Gower, an older contemporary, was much praised up to Elizabethan times. Like Chaucer, he was a court poet, and his principal English work consisted also of a series of stories, ***Confessio*** *amantis* ***(Confession of a Lover).*** He also wrote two long didactic allegories, the ***Mirour de*** *l'omme* ("Mirror of Mankind") in French and ***Vox clamantis*** in Latin, which merit the address "O moral Gower" with which Chaucer dedicated to him ***Troilus and Criseyde.*** Gower's work helped to consolidate the literary use of the cultivated language of London and the rhymed verse characteristic of the south and east. After Chaucer the practice of rhyme and all that accompanies it in diction and manner became dominant.

Chaucer's successors acknowledged their debt to him. Thomas Hoccleve, in ***The Regement of Princes,*** called him "the first fynder of our faire langage." Hoccleve's works were for the most part naïvely didactic, and one of his main faults was the affecting of Latinate vocabulary, which became a very common habit. It was much practiced, especially in his shorter poems, by John Lydgate, who also felt intense admiration for Chaucer and imitated him as far as he could, writing ***The Falle of Princis*** in the rime royal that Chaucer had used in ***Troilus and Criseyde.*** Lydgate's long poems were verbose and laboured, but some shorter ones, such as the satire of ***Chichevache and Bycorne,*** were more successful. He was greatly admired in his own day and his influence on other writers was greater than his merits.

After Lydgate the Chaucerian, or "courtly," tradition persisted mainly in minor works. Many short love lyrics have been preserved, some by known authors such as William de la Pole and Charles d'Orléans, the best of them being highly accomplished, though generally skillful variations upon familiar themes. Of longer poems perhaps the finest was the anonymous ***The Flower and the Leaf,*** a graceful "debate" poem in rime royal, purporting to be written by a woman. The tradition continued into the 16th century.

Non-courtly poetry. The courtly tradition was not the only poetic field cultivated in the 15th century: there was

Carols and ballads

a great body of religious poetry. The prayers and songs were often rather formal; the best were the simplest, perhaps the finest of all the famous "I syng of a mayden that is makeles." This song was close in spirit to Christmas carols, and it was in the 15th century that the carol first appeared in English in great numbers, most of those that survive being anonymous. An important group was formed by the "lullaby" carols, which descended from the type of a 14th-century song, "Lullay lullay, litel child, whi wepest thou so sore?"

Beside the richness of the devotional verse secular non-courtly poetry was limited in interest, but there were highlights. The outstanding poem was "The Nut-Brown Maid," a delightfully managed dialogue on women's constancy. This, with its theme of the outlaw, had affinities with the Robin Hood ballads, and these too flourished in the 15th century, though they must have originated much earlier. How far late forms of ballads such as "Chevy Chase" and "Sir Patrick Spens" date from this period is debatable; but there was a wide range of ballad themes: dramatic scenes of battle, romantic elopements and pursuits, supernatural visitations. Nearly all were northern or Scottish in language and setting, and many achieved, with their austere technique, great power of feeling. Political themes appeared in a good deal of 15th-century verse, whose topics extended from a carol of thanksgiving for the victory of Agincourt to coarse triumph over a defeated opponent. The most important political poem was the anonymous *Libelle of Englyshe Polycye* (c. 1436–37). There was also still a public for chivalric romances, several of which were translated or adapted from French, whereas others were more or less independent compositions, though heavily derivative in incident and language. Some show a fair control of plot and some capacity for lively description, but in the main they are shapeless and unoriginal. Much of this later romance fully merited the ridicule that Chaucer poured upon the genre in *Sir Thopas.*

The works of John Skelton

Toward the end of the century came the unique figure of John Skelton, who fit into no tradition yet who was rather a reactionary than a man of the Renaissance. His most characteristic verse, a succession of short rhyming couplets, was all his own and was most effective in passages of violent action or invective but could also be strangely successful in tender moods—the lyrics in the "Garlande of Laurell" (1523) have a freshness that looks forward to the great flowering of Elizabethan song. In the keenness of his mind, the depth of his human sympathy, and his mastery of words Skelton was the greatest English poet of his time.

The Wakefield cycle

Drama. The beginnings of English religious drama belonged to the 13th century, but only two fragments have survived from c. 1300. The most important texts were the cycles from Chester, York, Wakefield, and the so-called *Ludus Coventriae* or *Hegge Plays,* which probably were from Norfolk. The Chester cycle covered in 25 plays Bible stories of the history of mankind from the Fall of Lucifer to the Day of Judgment. The work of the "Wakefield Master," who contributed five plays and parts of others to the Wakefield cycle, was distinguished by a stanza form combining rhyme and alliteration, and a humorously realistic treatment of character and original command of colloquial language. The Wakefield plays were the finest in variety of interest and in literary quality. All these cycles were plays for performance on the Feast of Corpus Christi and were acted by craft guilds. There were also a number of individual plays, which were mainly narrative treatments of their subjects, not associated with cycles, some on the lives of saints—called miracle plays in distinction from the mysteries, which were based on Scripture. Beside them arose a different type of play in which the characters were abstract qualities or types and the plot was allegory; these are called moralities, the most important of the earlier ones being *The Castell of Perseverance* and of the later *Everyman.* They rise to simple solemnity in treating the theme of death and the fate of man's soul.

Prose. In the later medieval period English prose took on a new importance. The devotional tradition of Richard Rolle and Walter Hilton was carried on in *The Cloud of Unknowing* (probably late 14th century) and other anonymous treatises: in *The Sixteen Revelations of Divine Love* by Dame Julian of Norwich; and in such work as that of Nicholas Love, especially his translation (printed 1486) of the *Meditationes Vitae Christi* as *The Mirror of the Blessed Life of Jesus Christ.* Much of this prose is beautifully lucid, simple and easy in movement (some has a rhythmical cadence which removes it from the field of normal prose), but well before the 14th century ended plain prose began to be used for a much greater range of subject. John Wycliffe was the author or source of controversial works in which clarity of style matched vigour of thought, and, in addition, he inspired a great translation of the Bible that is known by his name. The later version, attributed to John Purvey, won great popularity, maintained until William Tyndale brought out his New Testament in 1526. Chaucer also wrote rather stiff prose in "The Tale of Melibee," "The Parson's Tale," and his rendering of Boethius, while more important as a symptom of the trend of writing was his use of prose for a scientific purpose in *A Treatise of the Astrolabe.*

From the later 14th century, English prose came to be used for all kinds of secular purposes. John of Trevisa translated Ranulph Higden's universal history *Polychronicon* and a great 13th-century encyclopaedia, *De proprietatibus rerum.* There were several English versions of Mandeville's *Travels,* traveller's tales of the East, originally in French. Many chronicles of national and local affairs were written, including over 120 manuscripts of the *Brut.* Three translations of the 13th-century *Gesta Romanorum,* a collection of anecdotes and tales, met a popular taste for stories. The style of such works was close to the language of conversation, and this is found again in private letters, of which the collections of the Paston family of Norfolk is the best known. These were the first documents of their kind in English, and, although not literature in the ordinary sense, they give an intimate picture of the life of the time, the best showing a remarkable command of language. A comparable plainness and vigour characterized the remarkable autobiography of Margery Kempe, a mystic who, apparently illiterate, dictated an account of her religious experiences and pilgrimages.

The greatest writer of the century was Sir Thomas Malory, whose collection of Arthurian stories was printed by William Caxton in 1485 under the title *Morte Darthur* (finished 1469–70). From sources in English verse and French prose Malory composed a masterpiece of dignified romance and nostalgic tragedy, in a style conspicuous for its sonority of rhythm that yet used the simple language of his day.

The first printing in England

The introduction of printing into England in 1476 by Caxton revolutionized the dissemination of literary works. He made accessible to a wider public than ever before most of the important English writings of the preceding century and also published many translations both by others and by himself. Caxton's prologues and epilogues were livelier than some of the books they accompany, for he was keenly concerned to write readily intelligible English, and, by circulating large numbers of books in a fairly homogeneous type of English, he contributed much to the stabilization of the language of London as the general literary medium.

The new learning. For most of the 15th century, English literature was closely linked to what had gone before and was little affected by the new Humanism that had arisen in Italy in the 14th century. Cultivated Englishmen were aware of the movement but as patrons rather than practitioners, and it was not until the beginning of the 16th century that the full tide of Renaissance Humanism reached England. Even then the effects on literature were slow. An epoch-making work of English Humanism, the *Utopia* (printed 1516) of Sir Thomas More, was written in Latin, still the language of scholarship. More did write English prose clearly, but it continued the informality of earlier prose rather than imitating classical models. Good English prose was written, too, by men concerned with education—Sir Thomas Elyot, Sir John Cheke, Sir Thomas Wilson, Roger Ascham—and by the Bible translators Tyndale and Miles Coverdale. The educationists, too, became involved in the controversy over “inkhorn terms”—words borrowed wholesale from Latin by scholars and stylists writing in English, often to excess and with ludicrous results. Nevertheless, practice in the classics did much to give form to the English sentence; yet its strongest rhythms derived from the simplicities of English speech. Edmund Spenser inherited his fable from Malory and much of his language from Chaucer. Though great masterpieces at the end of the medieval period were few, its importance for the understanding of the wealth that followed was immense.

Scottish literature in Scots and English. Apart from poems of doubtful authorship the earliest extant literature appeared in the second half of the 14th century. No doubt much earlier literature has been lost, as neither John Barbour's ***Bruce*** (completed 1376) nor ***Susannah, or Pistill of Susan*** (*c.* 1360, ascribed to Huchown of the Awle Ryale) can be called primitive in technique. These poems represent two traditions, the former being in octosyllabic rhyming couplets, the latter having the alliterative rhyming stanza that reached Scotland through Middle English alliterative writers of the north and became a characteristic Scottish poetic form. To the first tradition belonged Andrew of Wyntoun's ***Orygynale Cronykil*** (*c.* 1420), a history of Scotland; and ***The Buik of Alexander*** and ***Legends of the Saints,*** which belonged to Barbour's time and region. The second tradition developed romance and fantasy and included Richard Holland's ***Buke of the Howlat***, an anonymous ***Golagrus and Gawain, Rauf Coilyear,*** and shorter pieces such as ***The Gyre-Carling***, William Dunbar's ***Kynd Kittok,*** and Robert Henryson's ***Sum Practysis of Medecyne.*** The rather dry style of John Barbour suited his own patriotic narrative but had little influence on later poetry. His ***Bruce,*** national epic in kind if not in value, was an isolated poem. The more flamboyant alliterative tradition had a lasting effect well beyond the medieval period and was valuable in forging a link between popular and sophisticated verse, which in Scotland was not broken to the same extent as in England.

The ***makaris,*** or Scottish Chaucerians

The great period of the ***makaris*** ("makers"; *i.e.*, poets), or Scottish Chaucerians (c. 1425–1550), included the four chief pre-union poets: Robert Henryson, William Dunbar, Gavin Douglas, and Sir David Lyndsay. To them must be added the author of ***The Kingis Quair*** (probably James I of Scotland, and the author of ***Schir William Wallace*** (traditionally Harry the Minstrel, Blind Harry). Scottish poetry has never been so confident, dexterous, or varied as in the century that produced Henryson's ***Testament of Cresseid,*** Dunbar's ***Tua Mariit Wemen and the Wedo***, Douglas' ***Eneados,*** and Lyndsay's drama ***Ane Satyre of the Thrie Estaits.*** The rough, precarious culture that nourished this literature has often the appearance of a rejuvenated medievalism rather than of a premature Renaissance. What was not medieval was a keen linguistic consciousness and a desire to expand the vernacular, which followed the political self-determination of Scotland in the 14th century. The elaborate style of the ***makaris*** has been criticized as artificial and excessive; but such excesses, which had their parallel in Elizabethan England, were a necessary stage in the development of a literary medium. Gavin Douglas, justifying his borrowings from other tongues in his translation of Virgil, spoke of his desire for "fouth (copiousness) of language." His powerful descriptions burst with words and have a rich etymological and idiomatic texture.

The ***makaris*** gained from Chaucer an ideal of poetic utterance, a rhetoric or a diction, and metrical forms like the decasyllabic couplet. Nevertheless, this was only one influence among many; others included the elements of satire and fantasy (from romance to farce), which had always been strong in Scottish poetry. Both Chaucer and Dunbar were connected with the court circles of their day and addressed a courtly and educated audience; yet Dunbar as well as Gavin Douglas and Lyndsay had a good deal of social give-and-take, perhaps not unexpected in a small country that had been only partly and haphazardly feudalized. "A man's a man for a' that" is a recurring statement in Scottish literature.

MEDIEVAL CELTIC LITERATURE

Irish Gaelic. The Celtic language spoken in Ireland from the earliest historic period, often described as Gaelic or Goidelic, probably reached Ireland not long before the beginning of the Christian Era. The earliest evidence consists of sepulchral inscriptions in the ogham alphabet not unlike Scandinavian runes. These inscriptions, usually ascribed to the 4th and 5th centuries, show an archaic state of language. Writings in the Roman alphabet date from 8th-century glosses in Old Irish, but works of 7th- or even 6th-century composition are preserved in manuscripts of a much later date.

The functions of the ***filid***

Irish literature was from the earliest times aristocratic, kept in the hands of a professional class, the ***filid.*** These poets, together with other men of learning, such as druids, judges, and historians, probably belonged to the same privileged ruling class as the landed aristocracy, and the early laws defined their status. Their duties were to know the sagas and genealogies and to compose encomiastic poems; that is, to serve the past and present glory of the ruling class. The ***filid*** were a proud, even arrogant, class, whose demands were extravagant and could be enforced by the threat of a lampoon (***a'er***), which could not only take away a man's reputation byt also damage him physically and even kill him. The laws set out penalties for abuse of the ***a'er,*** and a belief in its powers continued up to modern times. The official work of the ***filid*** has been preserved in fragments of annals and treatises.

Early verse. The earliest verse has been preserved mainly in passages incorporated into later documents, both literary and legal; most of these passages have suffered in transmission and are difficult to elucidate. There remains also the ***Eulogy*** on St. Columba (*c.* 521–597) in rhetorical short sentences linked by alliteration, ascribed to Dallán Forgaill, chief poet of Ireland. This device of alliterative rhythmical prose was used again in the sagas. Probably the oldest actual metre was that in which two half lines were linked by alliteration—a system reminiscent of early Germanic verse. Rhyme was used from the 7th century; the requirement was only that there should be identity of vowel and syllabic length and that consonants should belong to the same phonetic class —a system found also in early Welsh. The quatrain (seven or eight syllables to a line and rhyme between second and fourth lines) was derived from Latin hymn metres. A metre later to become the most popular of all was the ***debide*** (literally "cut in two"), in which the quatrains consisted of two couplets with the two lines of each couplet rhyming.

Much early verse was of an official nature, but that of the church was hardly more lively than that of the ***filid,*** who often affected a deliberately obscure style. More interesting was the 10th-century ***Psalter,*** a biblical history in 150 poems, which contained much strange learning. But it was not ponderous works that constituted the real glory of Irish verse but rather the personal ones by anonymous poets who composed poems such as the famous address to Pangur, a white cat. They avoided complicated metres and used a language that had been cultivated for centuries, with a freshness of insight denied to the ***filid.*** That the ***filid*** could, however, adapt their technique was shown by an 11th-century poem on the sea, where the preface, choice of theme, and free use of metaphorical expressions all suggest Scandinavian influence. This and other nature poetry carried on a tradition of native lyrics, sagas, and seasonal songs that showed remarkable sensitivity to natural beauty. The monastic movement in the Irish Church also provided a strong impetus to nature poetry. Exile from men brought hermits very close to their environment. This almost Franciscan poetry had an especial appeal to monastic scribes, so that a considerable amount of it has been preserved.

Historical verse arose partly because remembrance of the past was an important part of the work of the ***filid;***

some of the earliest poems were metrical genealogy. As time went on the necessity for compendiums of information grew, and these were again often in metrical form. They were especially numerous toward the end of the early period. In a long poem, Fianna bhtar in Emain ("The Warriors Who Were in Emain"), Cináed ua Artacáin summed up the saga material, while Fland Mainistrech collected the work of generations of filid who had laboured to synchronize the history of Ireland with that of the outside world. Equally important is a great collection, in prose and verse forms, called the Dindshenchus, which gave appropriate legends to famous sites of Ireland. Indeed, the development of a loose debide form, making rhyming easy, produced mnemonic verse on all sorts of subjects.

The early Irish epic

Early prose. The early Irish epic was a prose narrative that usually contained non-narrative passages in a heightened style, often in dialogue form. The resemblance between this and the type of epic found in early Sanskrit suggests that the tradition went back to Indo-European times. The oldest sagas were composed in a language that suggests that they were first written down in the 7th and 8th centuries, from an oral tradition. These were imperfectly preserved, since there is no extant manuscript earlier than the 12th century; Scandinavian invasions in the 9th and 10th centuries disrupted literary studies and everything else. Not until the 11th century did life become sufficiently settled for works to be collected in monasteries, but this revival was uneven. The great codex The Book of the Dun Cow, written early in the 12th century, showed older treatments of saga material than are found in the Book of Leinster, written years later. Only the shortest stories were preserved in forms approximating to the original; the rest have suffered from uncritical editorial handling, and the rhythmic prose used in heightened passages of older stories became unintelligible. Although they tended to be grouped in cycles, they ranged over a period from the earliest fables to the 11th century, and in the catalogs they were arranged by type rather than by period, under headings such as stormings, cattle raids, courtships, battles, voyages, and so on. Thus they have preserved a picture of primitive society—fighting from chariots, the taking of heads as trophies, the position of the druids, the force of taboo—for which there is little or no evidence from strictly historical sources.

The Ulaid cycle

The most important cycle was that of the Ulaid, a people who gave their name to Ulster. Conchobar, king of the Ulaid; Cú Chulainn, a boy warrior; Medb, queen of Connaught; and Noisi and Déirdre, doomed lovers, were outstanding in early Irish literature, and it was on elements from the sagas of the Ulaid that the nearest approach to an Irish epic was built—Thin *Bó Cúailnge* (The Cattle Raid of Cooley). It is probable that even when first put together, in the 7th or 8th century, the *Táin* was not very coherent; it is striking chiefly for the terse vividness of its narrative and the wit of its dialogue. The longer stories are no more than loosely connected incidents. Undoubtedly the finest section is that in which Fergus, an exile from Ulster, recalls the deeds of Cú Chulainn's youth; this story of the boyhood of a hero is obviously the work of an accomplished artist. But the value of the *Táin* generally lies in that, as it was being continually worked over, it provides a record of the degeneration of Irish style. The Thin collected around itself a number of ancillary stories, including that of its revelation to the filid in the 7th century by the ghost of Fergus and the tragedy of Noisi and Déirdre. (The latter has come down to modern times, notably through the work of J.M. Synge and William Butler Yeats.)

In this period stories with an origin outside a recognized tradition began to appear. An exotic element was represented by The Taking of Troy and The Story of Alexander, which appeared in the oldest saga lists, but classical learning had comparatively small effect until the next period. The same was true of the stories of Finn; the traditions went back to an early period, but their real development was at the time when the filid were no longer in control. The "wild man of the woods" cycle associated with Suibne Geilt had its origins in Strathclyde, where Irish and Brythonic literature must have been in contact at an early date; this mixture of hagiographical, saga, and nature material was one of the most attractive stories of the later period.

Although Irish was used for didactic purposes from the earliest period, the material was mostly derived from Latin, and that belonging to the earliest period consisted mainly of monastic rules, exempla to illustrate a moral, and anecdotes about saints. The lives of the saints were mainly works of fantasy, increasingly incorporating elements from folklore and saga material. The emphasis was always on the miraculous, but they contain a great deal of information about everyday life in early Ireland and are valuable as social documents. Another important type of religious work was the vision, the best known example of which was The Vision of Adamnan, whose soul is represented as leaving his body for a time to visit heaven and hell under the guidance of an angel. Both the saints' lives and the visions tended to degenerate into extravagance, so that parodies were composed, notably Aislinge Meic Con *Glinn* (The Vision of *MacConglinne*).

Theory has always been important in Irish cultural life, and the filid built up a considerable body of academic speculation. Very little has survived, but there are fragments on the nature of inspiration and the origin of language, combined with practical instructions on matters of metrics and style. At an early stage this learning became interpenetrated with that of Latin, and a technical vocabulary was built up to deal with Latin, as well as Irish, grammar. In several ancient texts discussion of the art of the poet was mixed with questions relating to his legal status.

The rise of the bards

In the 12th century the oldest surviving codices were being written in monasteries, preserving literature many centuries earlier with remarkable fidelity, while a new literary order arose that elaborated verse forms to a far greater extent than before and used a language that must have been very near the vernacular of the day. After the 12th century the hereditary bardic families became custodians of Irish literature and continued in that function until the collapse of the Gaelic polity. They were at first completely divorced from ecclesiastical influence, but soon they were sending their sons into the recently introduced orders, above all to the Franciscans, who were to become by the end of this period the greatest custodians of the tradition.

Later verse. The bardic reform of verse was sweeping. The language was brought up-to-date, the large number of metres used by the filid was greatly reduced, and a greater strictness in rhyme and lavishness in ornament was insisted on. The scope of the verse narrowed; the bulk was praise poetry, addressed to the poet's patron or to God. No longer associated with monastic foundations, the bards were dependent upon patronage for the high standard of living that they claimed as their right, for the bards (who trained for six or seven years) were no less arrogant than the filid whom they replaced. Modern taste finds little of interest in even the best of official bardic poetry; it is the occasional verse that shows real worth. One of the earliest poets of the great bardic family of Ó Dálaigh, Muireadhach Albanach, left, as well as a stirring defense of his action in killing a tax collector, an extremely fine elegy on the death of his wife. A new note was brought into Irish literature by the Norman invaders, that of courtly love, and the combination of this exotic element with the wit and felicity of the bardic style left a number of enchanting poems. A different kind of departure from praise poetry was the croshnacht, a poem in which verse was frequently interspersed with humorous or satirical prose passages.

The Fenian cycle

Later *prose*. Most native prose of this period was concerned with the hero Finn and his war band (*fian*). Stories about Finn, Oisin, Caoilte, and the rest must have existed among the people for many centuries. The outstanding work was Agallamh *na Seanórach* ("The Interrogation of the Old Men"), in which Oisin and Caoilte are represented as surviving the Battle of Gabhra and living on until the time of Patrick, whom Caoilte accompanies through Ireland, explaining to him the origin of the names of the places that they visit. But whereas in the Ulster cycle the introduction of the hero of former times was a way of authenticating the story, in the Fenian stories it was a nostalgia for past glory. The Fenian stories never received such careful literary treatment as did those of the older tradition, and the old form was soon abandoned for prose tales and ballads, which were easier reading and may be regarded as the beginnings of popular, as opposed to professional, literature in Irish. The metres represented a drastic simplification of the bardic technique, and a distinct change in theme occurred as this literature passed into the hands of the people.

Retellings of stories popular with the *filid* continue to be found, but these steadily dropped out of favour. Sometimes they were combined with folktale elements, as was the case with the very old saga of Fergus mac Léti, which was rewritten, perhaps in the 14th century, to include a story of a people of tiny stature—the leprechauns. Most important of all, a flood of translations from Latin and English began. The stories of Marco Polo, Sir John Mandeville, Prester John, and Guy of Warwick, as well as classical and Arthurian stories in their medieval adaptations, became well-known in Ireland. The new religious orders were responsible for the translation of many spiritual and devotional works, and the churchmen made the experiment, remarkable for the time, of handling philosophical material in the vernacular. There was also a considerable amount of technical writing; most important was the grammatical and metrical teaching, which was set out in elaborate tracts, giving examples of correct and incorrect usage. Continental teaching seems to have superseded the native tradition during this period.

By the 15th century the printing press was beginning to make literature available to larger numbers in most European countries. In Ireland, in spite of the fact that Irish was the language of most of the country and that the literature was flourishing at that period, the Irish did not control towns and had no access to printing. Irish literature remained an aristocratic one, the preserve of those who could afford to maintain the writers and supply them with costly vellum; inevitably this literature had little effect on the masses of the people.

The Lismore compilation

Scottish Gaelic. The earliest extant document containing Gaelic matter is ***The Book of Deer,*** in which portions of the Gospels in Latin, written in an Irish hand, with illuminations, were accompanied by marginalia in Gaelic and Latin. Those in Gaelic were of the 12th century but are without literary interest. Many other manuscripts of later date belong to the common Scots-Irish tradition and contained versions of heroic tales and sagas, poetry, ecclesiastical writings, and miscellaneous matter. The most important of these was the ***Book of the Dean of Lismore,*** an anthology of verse compiled between 1512 and 1526 by Sir James MacGregor, dean of the diocese of Lismore in Argyllshire, and his brother Duncan. The poems fall into three main groups: those by Scottish authors, those by Irish authors, and ballads concerned with Ossian, the mythical warrior and bard. This was the earliest extensive anthology of heroic Gaelic ballads in either Scotland or Ireland. The poems by Scottish authors range from about 1310 to 1520. The bard best represented was Fionnlagh Ruadh, bard to John, chief of clan Gregor (died 1519). There were three poems by Giolla Coluim mac an Ollaimh, a professional poet at the court of the Lord of the Isles and almost certainly a member of the MacMhuirich bardic family, the famous line of hereditary bards whose work spans nearly 500 years from the 13th century to the 18th. He lamented the downfall of the Isles family in the late 15th century. Perhaps the most notable of the other poets were Giolla Críost Brúilingeach and two women, Aithbhreac Inghean Coirceadail and Isabella, countess of Argyll.

Taliesin and Aneirin

Welsh. ***Poetry.*** Welsh literature has extended in an unbroken tradition from about the middle of the 6th century to the present day, but, except for two or three short pieces, all pre-Norman poetry has survived only in manuscripts written from the 12th century onward to the 15th, and it is clear that many of the poems are very much older than the manuscripts in which they have survived. Welsh had developed from the older Brythonic by the middle of the 6th century. In the ***Historia Brittonum (c.*** 800) attributed to the Welsh author or compiler Nennius, references are made to Welsh poets of the 6th century. Works by two of them, Taliesin and Aneirin, have survived. Taliesin wrote odes, or ***awdlau,*** in praise of the warlike deeds of his lord, Urien of Rheged, a kingdom in present-day southwest Scotland. To Aneirin is attributed a long poem, Y ***Gododdin,*** commemorating an ill-starred expedition sent from the region where Edinburgh stands today to take Catraeth (Catterick, Yorkshire) from the invading Saxons. The background, inspiration, and social conventions of the poems of Taliesin and Aneirin are typically heroic, the language is direct and simple, and the expression terse and vigorous. These poems, and others that have not been preserved, set standards for later ages. The alliterative verse and internal rhyme found here were developed by the 13th century into the intricate system of consonant and vowel correspondence, or consonant correspondence and internal rhyme, called ***cynghanedd.***

The heroic tradition of poetry existed in Wales proper and was continued after the break with North Britain in the mid-7th century; the earliest surviving example is a poem, perhaps by Taliesin, in praise of Cynan Garwyn of Powys, whose son Selyf was slain in the Battle of Chester. This poem struck a note that remained constant in all Welsh eulogies and elegies down to the fall of the Welsh bardic system: Cynan is the bravest in the field, the most generous in his home, all others are thrall to him and sing his praises.

The period between the 7th and 10th centuries is represented by a few scattered poems, most of them in the heroic tradition. The few authentic poems still extant include "The Eulogy of Cadwallon" (by Afan Ferddig?), the elegy on Cynddylan ap Cyndrwyn of Powys in the first half of the 7th century, and, more peaceful in atmosphere, "The Eulogy of Tenby," by an unknown South Wales poet toward the end of the 9th century. Early vaticinatory poetry is represented by "The Great Prophecy of Britain," a stirring appeal to the Welsh to unite with other Britons (those of the north, of Cornwall, and of Brittany), with the Irish, and with the Norse of Dublin to oppose the Saxons and to refuse the unjust demands of their "great king," probably Athelstan of Wessex. Poetry outside the main bardic tradition is preserved in ***englynion*** (stanzas of three of four lines), a dialogue between Myrddin and Taliesin, and in "The Song of the Wind," a riddle poem that contains the germ of the later convention known as ***dyfaliad,*** or comparison.

The poems of Llywarch Hen

The poems associated with the name Llywarch Hen are the verse remains of at least two sagas composed toward the middle of the 9th century by an unknown storyteller of Powys, whose basic material was the traditions that had gathered around the historical Llywarch and Heledd, sister to Cynddylan ap Cyndrwyn. In these, prose (now lost) was used for narrative and description and verse for dialogue and soliloquy. The metrical form was embellished by alliteration, internal rhyme, and incipient ***cynghanedd.*** The theme of both sagas was lamentation for the glory that once had been. The background was the heroic struggle of the Welsh of Powys against the Saxons of Mercia. Many other such sagas have failed to survive, though there do exist fragments of poetry preserved in the ***Black Book of Carmarthen*** (12th century), which were parts of soliloquies or dialogues from those lost sagas. Examples are a conversation between Arthur and the porter Glewlwyd Mightygrasp, a monologue of Ysgolan the Cleric, verses in praise of Gereint, son of Erbin, and a fragment of what may be an early native version of the Drystan and Esyllt (Tristan and Iseult) story. Other groups of verses in the same manuscript show that there once existed a legend of Myrddin Wyllt, a wild man of the woods who went mad at the sight of a battle, a legend associated with Suibne Geilt in Ireland and with Lailoken in Scotland. This Myrddin (later better known as Merlin) had the gift of prophecy. The historical poet Taliesin also became the central figure in a folk tale that was given literary form in the 9th or 10th century, but that has survived only in certain monologues preserved in ***The Book of Taliesin*** and in a 16th-century text, ***Hanes Taliesin*** ("Story of Taliesin"). As portrayed in the monologues Taliesin was the young bardic prodigy who surpassed everyone in his knowledge of the past, present, and future. Hence he too, like Myrddin, came to be regarded as a prophet.

Nature, a source of similes in the heroic poetry and of symbolism in verse remains of the sagas, was sometimes a subject of song in its own right. Generally, treatment of the subject was remarkable for its sensitive objectivity, its awareness of form, colour, and sound, and its concise, often epigrammatic, expression. In mood, matter, and form (that of the ***englyn)*** it often overlapped with gnomic poetry, which consisted of sententious sayings about man and nature. Most gnomic and nature poetry was probably produced in the 10th and 11th centuries by poets other than professional bards. Toward the end of the pre-Norman period a few poems on religious, biblical, and other subjects showed acquaintance with non-native legends. Saga poetry gradually died out, for prose became the recognized medium for storytelling.

The poets of the princes

With the consolidation of the principality of Gwynedd under Gruffudd ap Cynan (1054–1137) and his descendants, court poetry arose in the country, composed by the ***gogynefeirdd,*** or poets of the princes, who continued and developed the tradition of their predecessors, the *cynfeirdd.* In the bardic system, bards were divided into grades, the upper grade, or ***pencerdd*** ("chief of song"), being a high officer of the court whose duty it was to sing the praises of his lord and his family and of God and the saints. He was forbidden to sing of love and nature, and his choice of song was prescribed. Under him came the ***bardd teulu*** ("the bard of the king's war band), who did for the king's household what the ***pencerdd*** did for the king himself. He also was restricted but might sing of love and nature and such songs as would please ladies but be distasteful to warriors. Last of all came many kinds of ***cerddorion*** ("minstrels"), who were restricted in subject and might be permitted ribaldry and satire and who told the ***cyfarwyddyd,*** or oral tale.

Across this classification cuts another, based on an entirely different principle—the grading of the bards according to degrees of proficiency. This classification was educational, and its essential feature, the relation of disciple and teacher, remained into the modern period. It led, in the time of Henry IV, to the holding of an eisteddfod, or a session of bards to confer certificates of proficiency and to prevent the lower orders from proliferating and drifting into mendicancy. One of the results of a bardic system of this type was a remarkable conservatism in literature. Most of the 13th-century bards used a conventional diction that was consciously archaic in its vocabulary, grammar, and idiom and incomprehensible to anyone uneducated in poetry.

Bardism often went by families, and the first names of the new period were those of Meilyr, his son Gwalchmai, and his grandson Meilyr ap Gwalchmai, who were attached to the court of Gwynedd at Aberffraw. Gwalchmai in his "Praise of Owain," displayed one characteristic of all the ***gogynfeirdd,*** description of water, whether of river or sea. Official bards of this period all used the same material in the same way. Song and its modes were prescribed for them; the poetry of the bardic tradition was not measured by depth of thought but by exquisiteness. It sought not to interpret life but to adorn it. Two poet-princes, Owain Cyfeiliog of Powys and Hywel ab Owain of Gwynedd, stood out from contemporary bards. Cyfeiliog's most famous work was the ***Hirlas*** ("The Long Gray Drinking Horn"), in which he described his warriors making merry after a victorious raid; Hywel ab Owain's departure from convention was even more striking; for the first time in Welsh literature love of country and of beauty in the modern sense appeared: land and sea and women and the Welsh language spoken in cultured accent by his lady love awoke in him the feeling of awe and wonder. The ***gogynfeirdd*** alternated throughout this period between ***marwnad*** ("elegy") and ***moliant*** ("eulogy"), and the period closed with the most famous of all the elegies, sung by Gruffudd ab yr Ynad Coch after the death in 1282 of Llywelyn ap Gruffudd, last native prince of Wales.

The religious verse of the ***gogynfeirdd*** was generally simpler in style than the eulogies and elegies. A set type was the ***marwysgajn*** ("deathbed song"), in which the poet, sensing the approach of death, confessed his sins and prayed for forgiveness. Other religious poems were in praise of God and the Trinity, in honour of saints, on the torments of hell, and on the birth of Christ. They illustrate the gradual widening of the bardic horizon.

With the princes and their pageantry, there passed away the older modes and conventions of Welsh poetry. An audience that could once understand the intricate and involved ***awdl*** (a long stanza with a single rhyme throughout) could no longer find the means to educate themselves for the understanding of it. The old metres still remained. but the language became simpler. Poets in the years between the English conquest (1282) and Dafydd ap Gwilym (flourished ***c.*** 1340–70) seem to have returned to an earlier poetic fashion or to have been influenced by new ideas from other lands. Whereas in the early period exquisiteness was sought in archaic precision and in the suggestion of older modes, the new poets employed colour and form to an extent hitherto unknown in Welsh poetry and unparalleled later. Famous names in the 14th century were those of Gruffudd ap Maredudd, Gruffudd ap Dafydd, and Casnodyn.

The golden age of the ***cywydd***

The conquest of Wales by Edward I transferred the patronage of court poetry at Gwynedd and Powys from prince to landed aristocracy. The *pencerdd,* or chief poet, lost his superiority over the lower bardic ranks, who were no longer restricted in choice of content and style, and who, especially in South Wales (where the Norman Conquest had been established for a whole century before the conquest of Gwynedd), became more vocal as the older bardic song began to decline. The new poets of the south were well established before their works began to be preserved. The most important of them was Dafydd ap Gwilym, a contemporary of Chaucer, who in his first period wrote according to two entirely distinct traditions: on the one hand he wrote *awdlau,* poems in strict alliterative verse, in which the same rhyme was used throughout and which followed the traditions of the later *gogynfeirdd;* on the other, he wrote ***cywyddau,*** freer verse more in the style of the ***beirdd teulu,*** the lower class of poets, with couplets of seven-syllable lines rhyming in alternately stressed and unstressed final syllables. Dafydd may have invented the *cywydd*, a verse form consisting of couplets of seven-syllable lines rhyming in alternately stressed and unstressed final syllables, but his main advance was a simplicity that he introduced to his poetic language. His successors followed his lead; the old diction became obsolete, and he thus established the standards of modern Welsh. The substance of his poetry was also new, for he seems to have borrowed many of his themes from the wandering minstrels and trouvères of France. He wrote some love poetry but is best known for the descriptive nature poetry this included.

Dafydd's influence was twofold: the ***cywydd*** was established as the leading form, and the new subjects were recognized as fit themes for poetry. One of his contemporaries was Gruffudd ab Adda, who went much farther than Dafydd in the direction of a modern conception of nature; another was Iolo Goch, whose poem to the husbandman shows traces of English ideas, as seen in ***Piers Plowman.*** Llywelyn Goch Amheurug Hen wrote some early poems in the ***gogynfeirdd*** tradition, but his "Elegy to Lleucu Llwyd" successfully combined the Welsh elegy tradition with the imported serenade form. Other poets almost contemporary with Dafydd were Gruffudd Llwyd ap Dafydd, who wrote two superb poems to Owen Glendower, the rebel hero, and Rhys Goch Eryri, who is chiefly remembered for his quarrel on the nature of true poetry with Llywelyn ab y Moel and Siôn Cent, the last an elusive figure to whom are attributed semi-political poems in the form of prophecies and a number of songs lamenting the decadence of the world.

Dafydd ap Gwilym's successors

In the 15th century the cywydd was purified and refined. Although Dafydd Nanmor was inferior to most of Dafydd ap Gwilym's contemporaries in treatment of his subject and in imagination, in his mastery of the cywydd form he had no equal. Further advances in the cywydd metre were made by Lewis Glyn Cothi and Guto'r Glyn, in whose work a real consciousness of nationhood among the Welsh is detected, perhaps for the first time.

Prose. The earliest examples of Welsh prose were utilitarian: notes and glosses on Latin texts dealing with weights and measures, a record of an agreement, a list of church dues, and a commentary about a problem in astronomy. Shortly before the middle of the 10th century Howel (Hywel) Dda, according to tradition, had the Welsh laws codified. The oldest extant version is in Latin, but the Welsh version in The Black Book of Chirk (c. 1200) is derived from an earlier exemplar.

The Mabinogion collection of tales

The stylistic merits of the legal texts were reflected in a more conscious literary use of prose by storytellers (*cy*farwyddiaid), who recited oral tales made up of a medley of mythology, folklore, and heroic elements. Some of these were recorded in writing; the most famous collection is the Mabinogion, preserved in The White Book of Rhydderch (c. 1300–25) and *The* Red Book of *Hergest* (c. 1375–1425). The Mabinogion was composed of 11 tales, all anonymous, that were conscious literary compositions, based on older oral material. The greatest were the four related stories "The Four Branches of the *Mabi*nogi," composed in the second half of the 11th century by a writer from Dyfed. The author of Culhwch and *Olwen* (c. 1100), using the same basic material that underlies the "Four Branches," appears to have kept closer to the oral tale, but his inferior stylistics presaged the later decadent areithiau ("rhetorics"), which were in part parodies of the Mabinogion. Three of the Mabinogion tales, "Owein and Luned" (or "The Lady of the Fountain"), "Geraint and Enid," and "Peredur Son of Efrawg," represented a transition from purely native tales to those composed under Norman influence. The basic material was largely native, but there was French influence in certain names, social conventions, and atmosphere. These tales correspond to the Yvain, Erec, and Perceval of Chrétien de Troyes, and the exact relationship between the Welsh and French texts has long been a subject of debate. These romances, with their extreme sophistication, showed a decline from the native tales with their directness, restraint, and disciplined selection of material.

Many translations from Latin and French played a part in the evolution of a prose that could express aspects of human thought and activities not often touched upon in the tales. Most of them dated from the 13th and 14th centuries and were probably made by monks. Notable among them were translations from the Latin of Geoffrey of Monmouth's *Historia regum Britanniae* and the French of the Queste *del* Saint Graal. Their prose was largely experimental and influenced in varying degrees by the language and style of the originals, but at its best it was a not unworthy development of the prose of the law tracts and native tales.

In summary, the development of medieval prose was hampered by a lack of alliance between the heirs of Welsh literary tradition and those who cultivated the wider learning of western Europe. The inspiration for the fashioning of a prose, able to express all facets of thought and activity, arose toward the middle of the 16th century from the upheavals of the Renaissance and the Reformation; the foundations were laid by the Humanists, both Protestant and Catholic.

Cornish. The oldest remains of Cornish are proper names in the Bodmin Gospels and in the Domesday Book, 10th-century glosses on Latin texts, and a 12th-century vocabulary based on Aelfric's Latin–Anglo-Saxon glossary. The first literary text was a Middle Cornish fragment of a dramatic nature (c. 1400) in which a girl is offered as wife, praised for her virtues, and counselled on her behaviour. The 15th-century long poem *Pascon* Agan *Arluth* ("The Passion of Our Lord") has only slight literary merit.

The Ordinalia trilogy

The most important Cornish literary remains are a series of long dramatic compositions related to the miracle and morality plays that included a 15th-century trilogy called the *Ordinalia*. The first part, Origo Mundi, related the main events of biblical history; the second, Passio *Domini*, tells the story of Christ from the temptation to the crucifixion. The third, *Ressurectio Domini*, ends with the Resurrection and the Ascension. The dialogue is generally dull and there is little dramatic invention except in the scenes of ribaldry, but it is greater in bulk than all other remains of Cornish. Similar in form and technique but of greater variety was a play written in 1504 by "Dominus Hadton"—*Beunans* Meriasek, on the life of Meriasek, patron saint of Camborne. Also highly derivative was the play Gwreans an Bys ("The Creation of the World"). Later literary remains include a folktale, a few songs, translations of passages of Scripture, and several versions of the Apostles' Creed.

MEDIEVAL GERMAN LITERATURE

Old High German. Although Gothic was the earliest recorded Germanic language and the only east Germanic language on which there is trustworthy information, only a translation of the Bible by Ulfilas (c. 311–c. 383), a Gothic bishop, has survived in fragments. It is of linguistic rather than literary interest.

The first written records of the western German tribes dated from the second half of the 8th century. There is, however, evidence of an earlier, orally transmitted literature, consisting of short Heldenlieder (songs celebrating the exploits of famous heroes), hymns connected with pagan religious rites, battle songs, and laments for the dead. Although none of these was recorded, their substance formed the basis of later popular heroic epics. As in other areas, the first significant texts were a product of the efforts to spread Christianity. Many were translations from Latin, for example, the Alemannic version (c. 800) of De fide *catholica* ("Concerning the Catholic Faith") by Isidore of Seville (c. 570–636), Spanish churchman and scholar, and the translation of Boethius' *Consolation* of Philosophy made about 1000 by Notker Lateo, of the Swiss monastery of St. Gall.

The few extant verse works were more original, although religious and didactic in subject. The largest of these, the Evangelienbuch (c. 870; "Book of the Gospels") by the monk Otfrid of Weissenburg, presented the life of Christ in a form rivalling that of the Heldenlieder. Although handicapped by didactic passages, the poem is noteworthy as the first German work to replace the alliteration of Germanic tradition with the end rhyme of Medieval Latin verse.

Pagan works

Despite the church's opposition to interest in anything pagan, there survived a few pagan works, such as the Zauberspriiche — magic spells to protect domestic animals or for use against the common cold and other everyday concerns. The fragmentary *Hildebrandslied* (c. 800; "Song of Hildebrand")—a grim Germanic account of a duel between father and son—is more interesting and is important as the sole relic of heroic verse in Old High German. Throughout the period the vernacular was sporadically used as a literary medium, but scholars continued to prefer Latin. The movement for ascetic reform, which spread from the monastery of Cluny in France all over western Europe in the first half of the 10th century, discouraged churchmen from writing for the laity, and hence there was even less recorded literature in the vernacular for about a century. By the time the vernacular was written again, important changes had occurred in the language and the old ecclesiastical didacticism had been replaced by courtly feudalism.

Middle High German. The changes concomitant with these events marked the transition from Old to Middle High German, one important point being that the knight replaced the cleric as a poet. Early works still reflected the clerical tradition, but already in Heinrich von Melk's Von des *tôdes* gehugede (c. 1160; "Remembrance of Death"). the feudal knight's love of combat was used to illustrate the transient nature of life. The main literary forms at this time were *Minnesang* (love lyric) and epic. Poets, mainly noblemen, expressed their love according to courtly convention or told traditional tales of combat and romance.

The epic of local tradition had been created by the *Spielmann* (wandering minstrel), who collected and recited heroic songs on popular legends and began to link together the *Heldenlieder* to form longer epic narratives. Probably the earliest such composition was King *Rother* (c. 1160), a repetitive tale of violent bride abduction. The poets often provided interesting, sometimes amusing, commentaries on the life of their times. *Salman und Morolf* treated the conflict between Christian and pagan communities, reflecting an interest aroused by the Crusades also evident in *Orendel* and Sankt Oswald. With the *Alexanderlied* (c. 1130), a free rendering of a French work of the same name, German poets began drawing on successful French epics.

The Nibelung*enlied*

The anonymous Austrian *Spielmann* who wrote the *Nibelungenlied* ("Song of the Nibelungs") displayed great dramatic skill and poetic ability in recounting a well-integrated story of the hero Siegfried that combined songs of heroic legends about historical events and an epic concerning the destruction of the Burgundians, or Nibelungs, in 437. Gudrun, composed c. 1210, centred on the heroine Gudrun's steadfastness in the face of abduction, while a collection, Das *Heldenbuch* ("The Book of Heroes"). includes romances about Theodoric the Great, a popular hero in south Germany.

Court epics. The nationalistic epics, popular principally in the south, were counterbalanced in the west by the development of court epics based on French models. The first notable work, by Eilhart von Oberg, dealt with the Arthurian subject of Tristant und *Isolde* (c. 1170). Although Dutch, Heinrich von Veldeke (Henric van Veldeke) established himself as the father of the German court epic with his Aeneid (c. 1175–86), which was based on a French source. The three most famous poets of the court epic were Hartmann von Aue, Wolfram von Eschenbach, and Gottfried von Strassburg. Their work, dated largely between 1190 and 1210, was concerned with the knightly virtues of moderation, constancy, loyalty, and the duty of service to superiors and to God. Hartmann was an artistic and lucid storyteller. In Erec and *Iwein*, based on the French romances by Chrétien de Troyes, he handled the conflict between private inclination and public responsibility that existed for the medieval knight, and in Gregorius and Der *Arme Heinrich* ("Poor Heinrich"), illustrated man's relationship with God. Wolfram von Eschenbach's famous Parzival employed highly original imagery in a subtle and ambitious treatment of man's search for truth in his relationship with God, its hero Parzival progressing in maturity to become keeper of the Holy Grail. Gottfried von Strassburg took Thomas of Brittany as his source for his Tristan epic; a careful master of form and artistry, he was less concerned with otherworldly considerations and concentrated on an ideal primacy of love in this life.

Later epic poets cannot equal these three, though emulators abounded. Heinrich von Türlin's epic Din Krone ("The Crown") followed Hartmann but was excessively long and badly written. Despite the prolific writing of Rudolf von Ems and Konrad von Wiirzburg, the epic tradition gradually weakened. Although Meier *Helm*brecht (c. 1250), by Wernher the Gardener, movingly described the lawlessness and violence of decadent feudal society, both the social level of its subject and its realistic approach announced the end of courtly literature. By the mid-15th century a new satirical and didactic tone of middle class realism had taken over.

Courtly lyric. The other important literary form, *Minnesang*, the courtly love lyric, like the court epic, borrowed its content and stanza form from French or Provençal models. Its poets were minor noblemen who often became court poets to the higher nobility or travelled as minstrels throughout Germany, Austria, and Switzerland. Many of the poems representative of *Minnesangsfrühling* ("springtime of *Minnesang*") were recorded in an early-14th-century manuscript collection, the Codex Manesse, attributed to Swiss-born Rudiger Manesse (1224–1304). About 1190 Reinmar von Hagenau became court poet to the dukes of Austria, shortly after he had tutored Walther von der Vogelweide in Vienna.

The love lyrics of Walther von der Vogelweide

Walther's love lyrics, which united the courtly style for which Reinmar became famous with popular, natural love poetry, constituted the finest collection of medieval German lyrics, and his influence was reflected in later poets throughout Austria and the Tirol. His equally famous Spriiche—moral and political poems—dealt with his relationship with various patrons, the struggle between empire and papacy, and the spiritual value of crusade and pilgrimage. This interest in satire and worldly wisdom was continued by the Bavarian Neidhart von Reuenthal and again reflected by the popularity of Freidank's Bescheidenheit (c. 1230; "Modesty" or "Moderation") and Hugo von Trimberg's Renner (*c.* 1300; "The Runner").

Prose and drama. Medieval German prose literature was less substantial than that in verse. Berthold von Regensburg's sermons were eloquent vernacular works, and the writing of Mechthild von Magdeburg served as an early example of the mysticism that became important in the 14th century. The *Ackermann* aus *Böhmen* (c. 1400; "Plowman from Bohemia") by Johann von Tepl marked the beginning of the Humanistic tradition in Germany, and the quality of von Tepl's prose was unrivalled until the Reformation. Notable non-religious works were the Saxon Weltchronik ("World Chronicle") and *Sachsenspiegel* (c. 1225; "Mirror of Saxony"), which were representative examples of fanciful historical and legal writing, respectively.

Early works in drama—the fragmentary Easter play of Muri and the St. Gall Passion play (*c.* 1330)—ledontoa growing volume of popular morality and miracle plays, which were part of the church's campaign to instruct and inspire the lower orders. By the beginning of the 15th century, popular drama for secular entertainment appeared in the form of *Fastnacht*, or carnival, plays. These farcical comedies became increasingly popular, finally gaining respectability in the 16th century at the hands of Hans Sachs.

MEDIEVAL DUTCH LITERATURE

Of the earliest inhabitants of the Netherlands only the Frisians have survived, and, although the modern province of Friesland is a small part of the territory once held by them, they have maintained their own separate language and literature from the time of Bernlêf, who lived in the 8th century. The remainder of the Netherlands was overrun and colonized by the Saxons and Franks between the 3rd and 9th centuries, resulting in a predominantly Frankish culture in the south and Saxon or an amalgam of Saxon and Frankish language and culture elsewhere.

Under the less nomadic Franks the south prospered more than the north, and it was there that a literary language first developed, but only very gradually, because of marked differences between the dialects of the east, the centre, and the west (Flanders, with features that linked the coastal dialects with Old English). In the early Middle Ages, when Latin and, later, French were the languages of the educated, the vernacular was largely confined to unrecorded oral legend and folk songs. The earliest text that can claim to contain Old Dutch was the early-10th-century "Wachtendonk Psalm Fragments."

Poetry and prose. The work of the earliest known poet, Henric van (Heinrich von) Veldeke, typified the age's religious zeal, which emanated from the French centres of learning. In addition to a chivalrous rendering of the Roman poet Virgil's Aeneid (c. 1185) and love lyrics, which were important for German poets, Heinrich's *Servatius* was a rendering of a saint's life in the Limburg dialect. Dutch 13th- and 14th-century texts were generally written in the cultural centres of Flanders and Brabant, where for reasons of trade the prevailing influence was French. Throughout Europe the Crusades brought courtly romances into vogue, and Dutch romances, following French models, were written about events from classical history, such as Segher Diergotgaf's Paerlement *van* Troyen ("Parliament of Troy"), about Oriental subjects, or, most popular of all, on themes from Celtic sagas, including the Arthurian cycle. But, while Jacob van Maerlant was writing his ***Merlijns boeck*** ("Merlin's Book") and ***Historie van Troyen*** in the 1260s, chivalry was on the decline; and the titles of his later works bear witness to a late-13th-century reaction against romance. Maerlant's compendia of knowledge, including his ***Der naturen bloeme*** ("The Flower of Nature") and ***Spieghel historiael*** ("The Mirror of History") answered a demand for the kind of self-instructional literature that long remained a characteristic of Dutch literature. The change in social patterns at this time is also evident in two epic tales. ***Karel ende Elegast*** ("Charles and Elegast"), probably an original Flemish chanson de geste of the 12th or 13th century, describes with feudal reverence Charlemagne's adventures in the magic world of folklore. ***Van den vos Reinaerde*** (*c.* 1240; "Reynard the Fox") is the Flemish poet Willem's version of a translation by another Fleming, Aernout, of the French ***Le Plaid,*** which, by contrast, brilliantly satirizes feudal society and the epic manner.

Mystical writing

Mystical writing reached a remarkable lyrical intensity in the poetry and hortatory prose of a Brabantine laywoman, Hadewijch (late 12th, early 13th century), and this inspired later mystics, greatest of whom was Jan van Ruysbroeck, a disciple of the German mystic Meister Eckehart and the Netherlands' greatest medieval prose writer. His most important work was ***Die chierheit der gheestliker brulocht*** (1350; ***The Adornment of the Spiritual Marriage,*** 1916), concerning the soul in search of God. His work was part of a renewed ecclesiastic concern to instruct the laity, resulting in a considerable body of literature: Bible stories, legends, and didactic short stories that were all widely current at this time. Of these, ***Beatrijs,*** an early-14th-century Flemish verse rendering of a popular legend, is told with such humanity and restraint that it still inspires modern versions (*e.g.*, by Maurice Maeterlinck and Pieter Cornelis Boutens).

Songs, drama, and the rhetoricians. The earliest recorded songs suggest a Germanic rather than a Romance tradition, and since the first extant plays, the 14th-century ***Abele spelen*** ("seemly plays"), were entirely secular (and may have been the first of such in Europe), incorporating romantic themes from the earlier songs, there is reason to attribute the emergence of drama in the Netherlands as much to mime and song as to liturgical action. The only evidence of early liturgical drama is the Latin ***Officium stellae*** of the 14th century, after which there is nothing until 1448–55, when the play cycle on the seven joys of Mary was first performed at Brussels. Of the many miracle and morality plays, two deserve special mention: ***Mariken van Nieumeghen*** (late 15th century; "Mary of Nijmegen") and ***Elckerlyc*** (of about the same date). The first anticipates the Renaissance in its psychology and treatment; the second, still entirely medieval in its conception, is the original of the English ***Everyman.*** Both were written by members of ***rederijkers karners,*** or chambers of rhetoric, institutions that spread from the French border in the 15th century and were organized like guilds, with functions similar to those of the French medieval dramatic societies. They were commissioned by the town protecting them to provide the ceremonial and entertainment at religious and secular festivals and were influential in popularizing art and morals. Drama by this time was in the hands of the laity rather than the church, and the introduction of secular themes necessitated the use of stages or carts outside religious buildings. The chambers depended on literary performance in order to survive, and rhetoricians organized national festivals and competitions for drama and poetry.

The "chambers of rhetoric"

SCANDINAVIAN LITERATURE IN THE MIDDLE AGES

The literature of Scandinavia and, in particular, of Iceland has reflected two extraordinary features of the social and cultural history of pagan Europe and of Iceland. The way in which names such as Siegfried, Brunhild, and Attila cropped up again and again in different European literatures has borne witness to the dissemination of legends and traditions, common to the early Germanic tribes of Europe, starting from the great movements westward in the 4th, 5th, and 6th centuries. The literature of Iceland provides not only the most detailed descriptions available of the life-style of early Germanic peoples but constitutes the most complete account of their literature and literary traditions. Although the sagas and poems were first written down by Christian scribes, they present a picture of a pre-Christian European culture that reached its heights in the new settlements in Iceland.

A second feature directly concerns the peoples of Scandinavia. A remarkable characteristic of Scandinavian literature was the accuracy with which it described the geography of northern Europe. This was not to be ascribed to unusually imaginative writing but to accuracy born of actual knowledge. From the late 8th century until well into the Middle Ages, the history of the Norsemen was one of unceasing movement toward western and central Europe. The Norsemen discovered Iceland, as early Icelandic historians had it, through their ships being blown off course about 860. The next century found the Vikings pushing west by way of Britain, Ireland, and France to Spain and then through the Mediterranean to North Africa and east to Arabia. Across land they reached the Black Sea, by sailing north they came to the White Sea, and finally, turning westward again, they reached America long before Columbus.

At worst the literature associated with this period in Scandinavia could reflect some part of these remarkable forays across the Northern Hemisphere; because of its high literary quality, it does in fact offer far more.

The classical period in Iceland. The best known Icelandic literature belongs to the classical period, which was roughly equivalent to the early and medieval periods in west European literature. From. Icelandic manuscripts was derived knowledge of a large part of European myth and legend, which was in part common to all Germanic peoples. Stories of the Norse gods and myths—of Odin, god of war; Balder the Beautiful; Thor, god of thunder; and Valhalla, hall of the slain—formed the nucleus of early Icelandic literature.

Almost all extant early Scandinavian poetry was recorded in Icelandic manuscripts, although some was clearly composed before the Scandinavian peoples reached Iceland in the late 9th century. Much of the oldest poetry was recorded in the Codex Regius manuscript, which contained the ***Poetic Edda*** (*c.* 1270). The poetry is sometimes called Eddaic and falls into two sections: heroic lays, which, broadly speaking, dealt with the world of men; and mythological lays, which dealt with the world of the gods.

The ***Poetic Edda***

The heroic lays. The heroic lays followed the mythological in the Codex Regius and were probably the earlier of the two. Many of the legends on which they were based originated in Germany or even among the Goths. Oldest of all was perhaps the *Hamdismál* ("Lay of Hamdr"), which forcefully expressed the heroic ideals of Germanic tribal life. The story closely resembled one told by Jordanes, a Gothic historian of the mid-6th century, and his account suggested that his source was an even earlier poem about Hamdir. Another of the older lays in the ***Poetic Edda*** was the ***Atlakvida*** ("Lay of Atli"), which referred to events that took place in 5th-century western Germany, Atli (or Attila) being king of the Huns from 434 to 453. Nearly all heroic lays were associated with the story of Sigurd (or Siegfried), the valiant hero, and his ill-fated love for Brunhild, who, too, figured to varying extent in different lays. Many critics hold that the lays concerned with the spiritual conflict of the heroines Brunhild and Gudrun, which tend to be romantic and sentimental, were later compositions than the austere heroic lays. The ***Poetic Edda*** contained only a small portion of the poetry known in Iceland in the Middle Ages and now lost. Fragments of ancient lays appeared in 13th- and 14th-century sagas, such as the *Hlödskvida* ("Lay of Hlöd") in the ***Reidreks saga*** and mention of Danish and Swedish heroes in some fragments that must also have been known to the author of the Old English epic poem ***Beowulf.***

The mythological lays. Mythological lays about the Norse gods made up the first half of the ***Poetic Edda.*** It was unlikely that any of these originated outside Norway, Iceland, and Norse colonies in the British Isles. The *Völuspá* ("Sibyl's Prophecy") was a striking poem on the history of the world of gods, men, and monsters, from the beginning until the "twilight of the gods." Many passages in the poem are obscure, but most modern scholars agree that it was composed in Iceland about the year 1000, when the people were turning from the old religion to the new. An interesting story of the gods was told in the *Skírnismál* ("Words of Skirnir"): sitting in "Gate Tower," throne of Odin, the god Freyr, lord of the world, gazes into the world of giants and falls in love with a giant maiden; to win her, he sends his messenger Skirnir, who first offers gifts and then threatens the maiden until she agrees to make a tryst with Freyr. Scholars have seen an ancient fertility myth in this story, and it was certainly one of the older mythological poems in the ***Edda*** and probably originated in Norway before Iceland was settled by Norwegians.

The mythological poems so far mentioned were all narrative, but many of those in the ***Edda*** were didactic. The *Hávamál* ("Words of the High One"; *i.e.*, Odin) consisted of fragments of at least six poems. In the first section, the god speaks of relations between man and man and lays down rules of social conduct; in other sections he discourses on relations between men and women and tells how love of women may be lost or won; the last two sections are about runes and magic power. Most of the poems were probably composed in Norway in the 9th and 10th centuries. Another didactic poem, the ***Words of Vafþrúdnir***, related a contest between Odin and a giant.

Some important mythological lays appeared in other manuscripts. ***Baldrs draurnaer*** ("Balder's Dreams") described how the god Balder dreamed that his life was threatened and how his father Odin rode to the grave of a prophetess to force her to reveal the fate in store for Balder.

The Eddaic verse forms. Three metres are commonly distinguished in Eddaic poetry: the old story or epic measure, the speech measure, and the song measure. Most narrative poems were in the first measure, which consisted of short lines of two beats joined together in pairs by alliteration. The number of weakly stressed syllables might vary, but the total number of syllables in the line was rarely fewer than four. In these respects it resembled the measure used by Anglo-Saxon and early Germanic poets. The speech measure used in the *Atlamál* ("Words of Atli") differed very little from the epic measure, although its lines generally had a greater number of weakly stressed syllables. The song measure was the most irregular of the Eddaic verse forms, and its origins were the most obscure. It was chiefly in didactic poems and generally consisted of strophes of six lines divided into half strophes of three lines.

Skaldic verse. Norwegians and Icelanders of the 9th to 13th century also composed skaldic poetry (from the Icelandic word *skáld*, "poet"). It was not composed in the free variable metres of the ***Edda*** but was strictly syllabic: every syllable had to be counted and every line had to end in a given form. Like Eddaic lines, the skaldic lines were joined in pairs by alliteration, often using internal rhyme or consonance; but this poetry differed in syntax and choice of expression. Word order is freer than in Eddaic poetry, and a highly specialized poetic vocabulary employed periphrases, or kennings, of such complexity that the poetry resembles riddles. Little is known about skaldic verse forms, but they are thought to have been developed in Norway during the 9th century and could have been influenced by forms and diction of Irish poets of the period. The earliest known poet was Bragi the Old, who probably wrote in Norway in the latter half of the 9th century. Harald I (died *c.* 940) of Norway was eulogized by several poets, among them Thorbjom Hornklofi, whose poem the ***Haraldskvaedi*** ("Lay of Harald) was partly Eddaic and partly skaldic in style.

The distinction between Icelandic and Norwegian literature at this period is difficult to make. Skaldic verse seemed to have originated in Norway and to have been developed by Icelandic poets who either, like Egill Skallagrimsson (910–990), spent much time in Norway or wrote in praise of Norwegian kings, as did Sigvatr (*c.* 995–1045), counsellor and court poet of Olaf II of Norway. Although its complexity means that skaldic poetry is now less appreciated than it deserves, the orally transmitted poems of the 10th and 11th centuries were valuable sources for Icelandic historians in the following centuries.

Prose. Iceland's adoption of Christianity in 1000 opened the way for powerful influences from western Europe. Missionaries taught Icelanders the Latin alphabet, and they soon began to study in the great schools of Europe. One of the first was fsleifr (*c.* 1005–80), who after being educated and ordained as a priest was consecrated bishop. His school at Skálholt in south Iceland was for many centuries the chief bishopric and a main centre of learning. The earliest remembered historian was Saemundr the Wise (1056–1133), but Ari Thorgilsson (*c.* 1067–1148) is regarded as father of history in the vernacular, a short history, ***Libellus Islandorum*** (***Íslendingabók***, or "The Book of the Icelanders"), and the more detailed ***Landna'mabdk*** ("Book of Settlements") being associated with his name. Extant works of the period are few or anonymous. Annals of contemporary events date from the 13th century and the oldest religious manuscripts, consisting of homilies and saints' lives, from c. 1150. Larger collections of religious literature appeared in late-12th- and early-13th-century manuscripts. As elsewhere, the most popular books were often lives of the Apostles and saints.

Influence of Christian missionaries on Iceland

The sagas. The word saga is used in Icelandic for any kind of story or history, whether written or oral. In English it is used to refer to the biographies of a hero or group of heroes written in Iceland between the 12th and 15th centuries. These heroes were most often kings of Norway, early founders of Iceland, or legendary Germanic figures of the 4th to 8th century. The oldest saga is the fragmentary ***Oldest Ólafs saga helga*** ("First Saga of St. Olaf"), written about 1180. In form it is a hagiographic narrative, laying emphasis on miracles worked through the agency of the saint. It was probably written in the monastry of Thingeyrar, which played an important part in cultural life in the late 12th and early 13th centuries.

Several sagas about King Olaf Tryggvason, at whose instigation the Icelanders adopted Christianity, were also written at Thingeyrar, where the work of the monks was fanciful rather than realistic. A more critical style of history was established in the south by Saemundr and Ari, and several notable works were written at Skálholt or nearby in the 13th century, such as the ***Hungrvaka*** ("The Appetizer"), a short history of the bishops of Skálholt from fsleifr to Kloengr. In the late 12th century several short histories of Norwegian kings were brought from Norway to Iceland, where they influenced Icelandic historians. The ***Ágrip***, a summary of the histories, or sagas, of Norwegian kings, written in the vernacular in Norway, was particularly influential. The ***Fagrskinna*** ("Fine Skin") covered the same period in more detail, while the ***Morkinskinna*** ("Rotten Skin"), probably written earlier, covered the period from Magnús the Good (1035–47) to the late 12th century.

The role of Snorri Sturluson

Snorri Sturluson (1179–1241) wrote many kinds of works and played an important role in political wrangles in his time. Among works ascribed to him was the ***Prose Edda***, a handbook of prosody and poetic diction. He twice visited Norway, and a large part of his work consisted of lives of its early kings: he combined his ***Ólafs saga*** with lives of other Norwegian kings to form the ***Heimskringla*** ("Orb of the World"). The value of these as historical sources has long been debated. Snorri was certainly well read in vernacular history and attempted to write faithful accounts of what he had read in earlier records. He did not aim to write scientific history; his work was creative and therefore portrayed his heroes imaginatively. The stirring ***Egils saga*** (on the skald Egill Skallagrumsson) is generally attributed to Snorri.

The Icelanders', or family, sagas. These sagas were about heroes supposed to have lived in the 10th and 11th centuries. Their origins are unclear, and it is debatable whether they were faithful records of history. One theory is that they were composed in the 11th century and transmitted orally until written down in the 13th century; though researchers now reject this view, it is true that the sagas owed much to oral tales and the tradition of oral verse. Their historicity is difficult to verify, since their content and form were shaped both by the sources used and by the author's intentions. While the earlier writers probably intended to write factual history, others relied more on their imagination, anticipating the historical novel by giving speeches to various characters.

It is also difficult to determine the date of many of the sagas. The obviously early works were somewhat crudely structured and express Norse ideals of loyalty and heroism. The ***Gisla saga,*** written before the middle of the 13th century, showed a development of artistic skill and contained rich descriptions of nature and verses of considerable beauty and tragic feeling. The ***Laxdaela saga*** ("Saga of the Men of Laxárdal"), written a few years later, was a delicately worked tragedy in which the author showed an unusual appreciation of visual beauty. One work that was clearly the author's creation was the ***Hrafnkels saga Freysgoda*** ("Saga of Hrafnkell, Freyr's Priest"): despite realistic detail, the saga contained little historical fact. As the century progressed, a taste for fantastic and romantic elements grew. The ***Grettis saga*** ("Saga of Grettir the Strong") included several motifs from folklore and portrayed a hero fighting against trolls and ghosts.

The saga of Njáll and Gunnar

The greatest of Icelanders' sagas, the ***Njáls saga,*** had in fact two heroes, Njáll and Gunnar. Gunnar is young and inexperienced and Njáll is a wise and prudent man endowed with prophetic gifts; he embodies traditional Norse ideals of loyalty and bravery, yet faces his death by burning with the resignation of a Christian martyr.

The heroic sagas. The fantastic element was further developed in the ***fornaldar sogur,*** literally, "the sagas of antiquity," whose heroes were supposed to have lived in Scandinavia and Germany before Iceland was settled. The best known, the ***Völsunga saga (c. 1270),*** retold in prose stories from heroic lays of Sigurd, the Burgundians, and Jormunrekr, and the ***Hrdlfs saga kraka*** incorporated ancient traditions about Danish and Swedish heroes who also appeared in the Old English poems "Widsith" and *Beowulf.* Although not equal to the Icelanders' sagas in literary value, these heroic sagas provide valuable information about early legend and lost heroic poetry.

Many of the works on contemporary history were combined about 1300 in the ***Sturlunga saga,*** including the ***Íslendinga saga*** by Sturla Thórdarson.

Translations from Latin. A quantity of secular literature was translated from Latin between the 12th and 14th centuries. The "Prophecies of Merlin," already translated in verse by a Thingeyrar monk, were combined with a complete translation of Geoffrey of Monmouth's history and titled ***Breta sogur*** ("Stories of the Britons"). In one 14th-century manuscript this was preceded by the ***Trójumanna saga*** ("Story of the Trojans"), translated from Dares Phrygius. A Norwegian translation of the Bible was begun in the reign of Haakon V Magnusson (1299–1319).

Romances. Romances were also translated or adapted from continental romances. Interest in romance began in Norway and soon took root in Iceland. The earliest romance was probably the ***Tristrams saga*** (1226), written for King Haakon IV and taking the Anglo-Norman poet Thomas as its source, while the ***Karlamagnús saga*** ("Saga of Charlemagne") was a collection of prose renderings of French chansons de geste, including a Norse version of the ***Chanson de Roland.*** Romances in Icelandic were numerous, and their effect on the style of later writers is evident in such sagas as the ***Laxdaela saga*** and ***Grettis saga.***

Post-classical literature in Iceland. In the period following the classical age, little was written that attracted attention outside Iceland. Realism and detached objectivity declined, and sentimentality and fantasy gained the upper hand. The decline in literary standards is sometimes attributed to Iceland's loss of independence in 1262 and the changes that followed. Interest in earlier manuscripts continued, and many 14th- and 15th-century manuscript collections of 13th-century material were made. The most beautiful of all Icelandic manuscripts, the ***Flateyjarbdk,*** included versions of sagas of Olaf Tryggvason and St. Olaf, together with texts from other sagas or about heroes associated with Iceland.

Prose. Prose literature of the 14th century included several sagas. Among them were the ***Finnboga saga ramma*** ("Saga of Finnbogi the Strong"), about a 10th-century hero, and another telling the love story of its hero Viglundr. Sagas about bishops, already a theme in the 13th century, became more numerous, as did lives of foreign saints. A large collection of exempla (moral tales) was also made, each short tale illustrating some moral precept.

Ásgrímsson's ***Lilja***

Poetry. Much poetry was written up to the Reformation, and many new forms were devised. The best poems were religious pieces, in honour of the Virgin, the Apostles, or other saints. The well-known *Lilja* ("The Lily") by Eysteinn Asgrimsson, a monk from Thykkvabaer in the south, gave an account of the fall of Satan, the creation, the first sin, and the birth, life, and Passion of Christ. The term *rímur*—rhymes—is used of the narrative poetry developed after 1500 that consisted of mainly four-line strophes: the lines had end rhyme. The metrical forms, although apparently derived from Latin hymns, inherited the alliterative system of earlier poetry. Ballads written in Icelandic never attained the popularity of Danish ballads in Denmark nor achieved the high standard of the Norwegian ***Draumkvaede*** ("Dream Ballad"). Most of those preserved dated from the 14th to the 16th century and were free translations of Danish and Norwegian originals.

Swedish. Swedish literature proper began in the late Middle Ages when, after a long period of linguistic change, Old Swedish emerged as a separate language. The foundations of a native literature were established in the 13th century. The oldest extant manuscript in Old Swedish was the *Västgötalagan* ("Law of West Gotland"), part of a legal code compiled in the 1220s. These legal documents often employed concrete images, alliteration, and a solemn prose rhythm suited to their proclamatory nature, and a number of them have survived.

Early Swedish ballads

The poetry of chivalry was first represented in ***Eufemiavisorna*** ("The Songs of Euphemia"), written in doggerel between 1303 and 1312, which included a translation of Chrétien de Troyes' romance ***Yvain.*** Anonymous ballads probably dating from the 14th and 15th centuries also reflected a new interest in romance and combined courtly themes with native. One of the best known, the ballad of ***Ebbe Skammelsson*** (of Danish origin), told a grim story of love and revenge, tempered with traces of chivalrous and Christian ideals. These ballads, though mostly derived from foreign sources and combining the imported ideals of courtly love with native, pagan themes and historical events, formed the most accessible genre of what can be called Swedish medieval literature.

Danish. Denmark's first literature appeared in the runic inscriptions scratched on stone or carved in metal, mainly epitaphs of warriors, kings, and priests, that occasionally had short, unrhymed alliterative verses in the Viking spirit. Runic inscriptions were used in Denmark from about 250, but most of those preserved date from 800–1100. With the introduction of Christianity, Latin became the predominant literary language, and Denmark's first important contribution to world literature, Saxo Grammaticus' ***Gesta Danorum*** (written between 1185 and 1222; "The Deeds of the Danes"), which contained, for example, the Hamlet story, was written in Latin. The medieval ballads of Denmark are among the most important in Europe; 539 are known in more than 3,000 versions, but nearly all were written down after the end of the Middle Ages, the first printed edition appearing in 1591.

ITALIAN MEDIEVAL LITERATURE

Early vernacular writing. Until the 13th century almost all literary work was written in Latin, was predominantly practical in nature, and was produced by writers trained in ecclesiastical schools. In literary quality and variety it fell short of the standard set by France. Only small fragments of Italian vernacular verse before the end of the 12th century have been found, although a number of legal documents were written in the vernacular. All the surviving 12th- and 13th-century verse reflected French literary influence, the extent of which is indicated by the intermingling of French and Italian subjects and dialects in the work of the Provençal poet Raimbautz de Vaqueiras (Raimbaut de Vaqeyras), one of whose poems was a dialogue between a Provençal lover and a Genoese lady with verses written in Provençal and Genoese dialect alternately.

“Franco-Italian.” French prose and verse romances were popular in Italy from the 12th to the 14th century. Stories from the Carolingian and Arthurian cycles, together with free adaptations from the classics, were read by the literate, while French minstrels recited verse in public places throughout north Italy. By the 13th century a “Franco-Italian” literature had developed; Italians copied French stories, often adapting and extending various episodes and sometimes creating new romances in French about characters from French works. Examples of these new works were the *Entrée d'Espagne*, the *Prise de Pampelune*, the prose *Aquilon de Bavière* ("North Wind of Bavaria") by Raffaele Marmora, and poems on classical themes, such as *Hector, Pharsale,* and *Attila.* In all this literature, though the language used was French, the writers often unconsciously introduced elements from their own dialects, according to their varying knowledge of French. Writers of important prose works such as Martino Canale and Brunetto Latini—who wrote the *La Cronique des Veniciens* and *Li Livres dou trésor* ("The Books of the Treasure"), respectively—were much better acquainted with French, while poets such as Sordello of Mantua wrote lyrics in Provençal revealing an exact knowledge of the language and of Provençal versification. Provençal love lyrics were, in fact, as popular as the French romances, and Italian writers carefully studied anthologies of the verse.

The origins of the vernacular lyric

The Sicilian school. In the cultured environment of the Sicilian court of the Holy Roman emperor Frederick II, who ruled the Sicilian kingdom from 1208 to 1250, lyrics modelled on Provençal forms and themes were written in the vernacular. Poets were careful to eliminate narrowly local elements from their language and used words characteristic of the troubadour tradition. Poetry was considered an escape from serious matters of life, and it is significant that it was the love poetry of Provence—and not the political poetry—that was imitated by the Sicilian school. The Emperor himself, author of at least two poems, seems to have used it in this way, and his chancellor, Piero della Vigna, wrote about political theory in Latin prose and of love in Sicilian verse. The poets did not treat love very seriously, and many poems were stiffly conventional, although Giacomo da Lentinio, who is among those credited with inventing the sonnet, and Giacomo Pugliese wrote with invigorating freshness.

The Tuscan poets. Sicilian poetry continued to be written after the death of Frederick II, but the centre of literary activity moved to Tuscany, where interest in the Sicilian lyric had led to several imitations by Guittone d'Arezzo and his followers. Although Guittone experimented with elaborate verse forms, his language mingled dialect elements with Latinisms and Provençalisms and had none of the beauty of the southern school. The Tuscan poets wrote conventional and dull love poetry. They showed more originality in taking over political themes from the troubadours, reflecting perhaps the situation in central Italy, where the free communes were rent by factions. In addition to lyric verse, Tuscany produced some longer allegorical poems, such as Brunetto Latini's *Tesoretto.*

The "novo" style. While Guittone and his followers were still writing, a new development appeared in love poetry, marked by a concern for precise and sincere expression and a new, serious treatment of love. It became customary to speak of a new school of poets of the *dolce stil novo,* or *nuovo* ("sweet new style"), an expression used by Dante Alighieri in La *divina commedia* ("Purgatorio," Canto XXIV, line 27) in a passage where he emphasized delicacy of expression suited to the subject of love. The major *stil novo* poets were Guido Guinizelli of Bologna, Guido Cavalcanti, Dante (in his poems in *La vita nuova*), and Cino da Pistoia, together with the lesser poets Lapo Gianni, Gianni Alfani, and Dino Frescobaldi.

Dante's La *vita nuova*

These poets seem to have been influenced by each other's work. Guido Guinizelli was best known for his canzone, or poem, beginning “Al cor gentil ripara sempre amore” ("Love always finds shelter in the gentle heart"), which posed the question of the relationship between love of woman and love of God. His poetry was immediately appreciated by Cavalcanti, a serious and extremely talented lyric poet. Most of his poems were tragic and denied the ennobling effect of love suggested by Guinizelli. Dante greatly admired Cavalcanti, but his concept of love, inspired by his love for Beatrice, who died young (in 1290), had much more in common with Guinizelli's. Dante's *La vita nuova (c.* 1293) is the story of his love in poems linked by a framework of eloquent prose: God is the "root" of Beatrice, and she is able to mediate God's truth and love and inspire love of God—but her death is necessary for her lover to reach a state of purification. Cino da Pistoia used the vocabulary of the *stilnovisti,* as these poets were called, in an original and lighthearted way. A comparison of their language with the earlier Tuscan poets reveals extensive refinement of Tuscan dialects. Purely local characteristics were removed, and the standard literary language of Italy had been created.

"Comic" verse. Giocoso, or "comic," verse was a complete contrast to serious love poetry. The language was often deliberately unrefined, colloquial, and sometimes obscene, in keeping with the themes dealt with in the poetry. This kind of verse belongs to a European tradition, owing something to the satirical "Goliard poets of the 12th and 13th centuries, who wrote Latin verses in praise of pleasure or in vituperation of women or personal enemies or the church. The comic poets—whose usual verse form was the sonnet—were all cultured men. The earliest of them was Rustico di Filippo, who produced both courtly love poetry and coarse, sometimes obscene verse of the "realistic" kind. The best known and most versatile was Cecco Angiolieri, whose love poetry often skillfully parodied the *stil novo* writers and whose favourite subject was his father's meanness. Folgore di San Gimignano was known for his sonnets (following Latin models) on the worldly pleasures he considered as suitable to different months and different days of the week. A fervent interest in politics, expressed with great directness and satiric force in works by Pietro dei Faitinelli, was a more serious aspect of the comic poets' work. Dante himself in his *Tenzone* ("Correspondence") with Forese Donati wrote in comic style.

Religious poetry. The beautiful and famous *Cantico di frate sole (c.* 1225; "Canticle of the Sun"), or *Laudes creaturarum*, of St. Francis of Assisi was one of the earliest Italian poems. It was written in rhythmical prose that used assonance in place of rhyme and was in the Umbrian dialect; in it God is praised through all the things of his creation. It is probable that St. Francis also composed a musical accompaniment, and after his death the *lauda* became a common form of religious song used by the confraternities of lay people who gathered to sing the praises of God and the saints and to recall the life and Passion of Christ. The one real poet of the *laude* was Jacopone da Todi, a Franciscan and a mystic. His *laudi* were mostly concerned with the theme of spiritual poverty. Though Jacopone did not write with conscious art, some of the *laudi* were very fine.

The pessimism of north Italian poetry

In north Italy religious poetry was mainly moralistic and pervaded by a pessimism rooted in heretical ideas derived from Manichaeism, which saw this world and the body as being evil and under Satan's control. Giacomino da Verona, a Franciscan, author of *De Jerusalem celesti* and *De Babilonia infernali,* was the liveliest and most imaginative of this group. Two of the most interesting of

Bonvesin da la Riva's works are a dialogue between the Virgin and Satan and one between Soul and Body. All this north Italian verse was written in a language that is distinct from central Italian dialects but that avoids narrowly local or colloquial elements.

Prose. Literary vernacular prose began in the 13th century, though Latin continued to be used for writings on theology, philosophy, law, politics, and science. Much early Italian prose consisted of translations from French and Latin, the latter affecting the sentence structure of the developing prose style and influencing the artistic consciousness of the translators. The earliest original literary prose consists of letters that follow medieval rhetorical rules for Latin prose and use rhythmical patterns for the ending of clauses and sentences, rhymed or assonanced prose, and abundant figures of speech and plays on words.

The founder of Italian rhetorical prose style, Guido Faba, a rhetorician, illustrated his teaching by examples of prose styles in Bolognese. Guittone d'Arezzo, his most notable follower in epistolography, tended toward an extravagant style. In contrast with Guittone's style is the clear scientific prose of Ristoro d'Arezzo's ***Composizione del mondo*** and the simple narrative of the Florentine collection of tales ***Il novellino.*** The masterpiece of 13th-century prose is Dante's La ***vita nuova*** *(c.* 1293; ***The New Life).*** Though not yet completely at ease in vernacular prose and occasionally too involved or repetitious, Dante combined simplicity with great delicacy and a poetic power that derived from the mysterious depth beneath certain key words.

The 14th century. The literature of 14th-century Italy dominated Europe for centuries to follow and may be regarded as the starting point of the Renaissance. Three names stand out: Dante, Petrarch, and Boccaccio.

La divina commedia

Dante. Dante Alighieri is one of the most important and influential names in all European literature, but it was only after his exile (1302) that he set out to write more ambitious works. ***Il convivio*** (*c.* 1304–07; "The Banquet"), revealing his detailed knowledge of scholastic philosophy, was the first great example of a treatise in vernacular prose: its language avoided the ingenuity of popular writers and the artificiality of the translators from Latin. ***De vulgari eloquentia*** ("Of Eloquence in the Vulgar Tongue"), written about the same time, contained the first theoretical discussion and definition of the Italian literary language. Both these works remained unfinished. In a later doctrinal work, ***Monarchia*** (*c.* 1313; "On the Monarchy"), Dante expounded his political theories, which demanded the coordination of the two medieval powers, pope and emperor. Dante's genius found its fullest development in La ***divina commedia*** ***(c.*** 1310–21; ***The Divine Comedy),*** an allegorical poem in terza rima (stanzas of three lines of 11 syllables each, rhyming ***aba, bcb, cdc,*** etc.), the literary masterpiece of the Middle Ages and one of the greatest products of any human mind. The central allegory of the poem was essentially medieval, taking the form of a journey through the world beyond the grave, with, as guides, the Roman poet Virgil and Beatrice, who symbolize reason and faith, respectively. The poem is divided into three ***cantiche,*** or narrative poems: "Inferno," "Purgatorio," and "Paradiso," which are all subdivided into various schematically symbolic areas, and each contains 33 cantos, with one canto as an overall prologue. Dante, through his experiences and encounters on the journey, gains understanding of the gradations of damnation, expiation, and beatitude, and the climax of the poem is his momentary vision of God. The greatness of the poem lies in complex imaginative power of construction, inexhaustible wealth of poetry, and continuing significance of spiritual meanings.

Petrarch. The intellectual interests of Francesco Petrarca (Petrarch, died 1374) were literary rather than philosophical; his political views were more realistic than Dante's and his poetic technique more elaborate though less powerful. Petrarch's influence on literature was enormous and lasting—stretching through the Italian Humanists of the following century to poets and scholars throughout western Europe. He rejected medieval Scholasticism and took as his models the classical Latin authors and the Church Fathers. This convergence of interests is apparent in his philosophical and religious works. Humanist ideals inspired his Latin poem ***Africa*** and his historical works, but the autobiographical ***Secretum meum*** is most important for a full understanding of his conflicting ideals. The ***Rime,*** or ***Canzoniere—a*** collection of sonnets, songs, six-line verses, ballads, and madrigals—gave these ideals poetic expression. Although this collection of vernacular poems intended to tell the story of his love for Laura, it was in fact an analysis and evocation not of present love but of the passion that he had overcome. The main element of this poetry was therefore in the elaboration of its art, even if it always reflected the genuine spiritual conflicts exposed in the ***Secretum.*** In addition to the ***Canzoniere*** Petrarch wrote a vernacular allegorical poem, ***Trionfi*** ("Triumphs"), in the medieval tradition, but it lacked the moral and poetical inspiration of Dante's great poem.

Petrarch's influence

The literary phenomenon known as Petrarchism developed rapidly within the poet's lifetime and continued to grow during the following three centuries, deeply influencing the literatures of Italy, Spain, France, and England. His followers did not merely imitate but accepted his practice of strict literary discipline and his forms, including that for the sonnet—without which the European literary Renaissance would be unthinkable.

The ***Decameron***

Boccaccio. Boccaccio's writings were purely literary, without any ideological implications. His first novel, ***Il filocolo,*** derived from the French romance ***Floire et Blancheflor,*** was little more than literary experiment. Inability to write on an epic scale was evident in his two poems ***Il filostrato*** and ***Teseida,*** while his ***Ninfale d'Ameto*** (1341–42), a novel written in prose and verse, and his ***Fiammetta,*** a prose novel, showed the influence of classical literature on the formation of his style. The ***Decameron*** (1348–53), a prose collection of 100 stories divided into 10 "days," was Boccaccio's most mature and important work. Its treatment of contemporary urban society ranged from the humorous to the tragic. Stylistically the most perfect example of Italian classical prose, it had enormous influence on Renaissance literature.

Boccaccio shared the Humanist interests of his age, as shown in his Latin epistles and treatises. An admirer of Dante, he also wrote a ***Trattatello in laude di Dante*** and a commentary on the first 17 cantos of the "Inferno." He contributed to allegorical poetry with ***L'amorosa visione*** (1342–43).

Popular literature and romances. During the second half of the 14th century, Florence remained a centre of culture, but its literature developed a more popular character. The best known representative of this development was Antonio Pucci (died 1388), whose vast production included the ***Centiloquio,*** a versification of Giovanni Villani's ***Cronica.*** Florentine narrative literature was represented by the ***Pecorone*** ("Dullard"), stories written by Ser Giovanni Fiorentino according to a pattern established by Boccaccio, and Franco Sacchetti's ***Trecentonovelle*** ("300 Short Stories"), which provide colourful and lively descriptions of people and places.

The recasting of the Carolingian and Arthurian cycles continued along lines established during the 13th century. Compilations in prose and verse became commoner, and Franco-Venetian literature gained in literary value. Epic legends were turned into romantic stories, which appealed more to their audiences in town squares and other public places. Novels by Andrea da Barberino, ***cantari*** with legendary subjects by Antonio Pucci, and the anonymous ***Pulzella gaia, Bel Gherardino, Donna del Vergiù, and Liombruno*** were written in the popular style consistent with their practical aim. Prose and verse compilations of classical subjects such as the ***Fatti di Cesare*** ("Deeds of Caesar") and versions of the ***Roman de Troie*** were similar in style and character. The most accomplished of these is perhaps the ***Fiore d'Italia*** ("Flower of Italy") by Guido da Pisa.

Religious, didactic, and historical literature. Religious literature of the period was closely connected with the popular literature. It consisted of sermons, treatises, moral epistles, hagiography, ***laudi,*** and miracle plays. None of the authors of ***laudi*** reached the standard of the earlier Jacopone da Todi. Versifications of stories from the Gospels, such as the ***Passion*** ascribed to Niccolò Cicerchia, were even more popular in style and spirit, but it is the religious prose .of the 14th century that is most interesting. The most important author in this field was Jacopo Passavanti, whose ***Specchio di** vera penitenza* ("The Mirror of True Penitance") was a collection of sermons preached in 1354. Less polished, but of greater literary value, were translations of legends collected in the anonymous ***Fioretti di San** Francesco* ("Little Flowers of St. Francis").

The Florentine chroniclers

Vernacular historiography of this period could be described as popular literature, with Florence as its main centre, whose two principal chroniclers were Dino Compagni and Giovanni Villani. Compagni wrote his chronicle between 1310 and 1312, after having taken part in the political struggles of his town; his dramatic account of the episodes and the liveliness of his prose made it the most original work of medieval historiography. Villani's ***Cronica*** in 12 books was less personal; it followed the medieval tradition by beginning with the building of the Tower of Babel and included many legends. The last six books, which cover the period from Charles of Anjou's Italian expedition (1265) to the author's own time, are of importance to historians. His prose lacked the dramatic power of Compagni's but his work may be described as the greatest achievement of Italian vernacular historiography during the Middle Ages. After Villani's death, the ***Cronica*** was continued by his brother Matteo and by the latter's son Filippo.

From Boccaccio's death to about the middle of the 15th century, Italian poetry suffered a decline. The following period was to be characterized by critical and philological activity rather than by original creative work.

MEDIEVAL LITERATURE OF THE IBERIAN PENINSULA

The medieval literature of the Iberian Peninsula is treated below according to language.

Castilian. ***The beginnings of vernacular writing.*** By the time of the Muslim invasion (beginning in 711), the Latin spoken in the Iberian Peninsula was in process of transformation into Romance. The 10th-century glosses, or marginal notes, to Latin texts in manuscripts belonging to the monasteries of San Millán de la Cogolla and Silos, in La Rioja, reveal traces of a vernacular already substantially developed. The earliest texts in Mozarabic (the Romance dialect of Spaniards living under the Muslims) were recovered from Hebrew and from Arabic *muwashshaḥ*s (poems in strophic form, with subjects such as panegyrics on love and wine), the last strophe of which was the ***markaz*** (theme stanza). They provide evidence of a popular poetry that may have begun as early as the 10th century and explain much in the traditional Spanish lyric types (***e.g.,*** the ***villancico,*** "carpl") of the later Middle Ages and the Renaissance. The ***markaz*** was generally a woman's love song, and, the motif, in Romance, was a cry of passion on which the whole poem was based.

Tales of the Cid

The rise of heroic poetry. The earliest surviving monument of Spanish literature, and one of its most distinctive masterpieces was the *Poema* (or *Cantar)* ***de mío Cid*** ("Song of My Cid"), an epic poem of the mid-12th century (the existing manuscript is an imperfect copy of 1307). It tells of the fall from and restoration to royal favour of a Castilian noble, Rodrigo Diaz de Vivar (*c.* 1043–YY), known by the Arabic title ***sidi,*** "lord." The setting and personages, the topographical detail, the realistic tone and treatment, and the proximity in time of the poet to his hero have led to the acceptance of historical authenticity as a characteristic of the poem and of the Castilian epic in general. The last two of the three sections were, however, wholly imaginative and the mere six lines given to the Cid's taking of Valencia from the Muslims make it clear that the scale of values is subjective and that the work is essentially poetic in conception. It won fame, nevertheless, as a popular embodiment of the Castilian character and lived on in epic, chronicle, ballad, and drama. The only other surviving epic text, a fanciful and decadent ***Cantar de Rodrigo*** ("Song of Rodrigo"), tells of the Cid's early manhood, from which the later legend of the Cid took shape. Frequent allusions in vernacular chronicles to the heroic narratives of minstrels make it clear that many narrative poems must have been lost. The chroniclers accepted these as historically authentic and "prosified" them, so that themes and even fragments of text can be reconstructed. Heroic narratives about which some information is available are ***Los siete infantes de Lara*** ("The Seven Princes of Lara"), ***Fernán González, El cerco de Zamora*** ("The Siege of Zamora"), ***Bernardo del Carpio,*** and others treating themes integral to the feudal history of Castile that were closer to a remote Visigothic past than to any French epic.

The beginnings of prose. A major influence on prose was exercised by Arabic. Oriental learning came to Spain with the capture (1085) of Toledo from the Muslims. The city became a centre of translation from Oriental languages. An anonymous translation from Arabic (1251) of the "beast fable" ***Kalilah wa Dimnah*** was the first example of storytelling in Spanish. An Oriental romance of the Seven Sages, known as the ***Sendebar,*** was translated likewise through Arabic, and other collections of Eastern stories soon followed.

Alfonso the Wise. The middle of the 12th century saw the recovery of Córdoba, Valencia, and Seville by the Christians. A more propitious intellectual ,atmosphere resulted in the founding of universities, and under Alfonso X of Castile and Leon (reigned 1252–84) literature achieved official prestige. Alfonso, in whose chancery Castilian replaced Latin, could be described as the father of Castilian prose. His vast enterprise of translation and compilation aimed at fusing all knowledge-classical, Oriental, Hebrew, and Christian—in the vernacular. The works, often under his personal editorship, included the great legal code Las Siete Partidas ***(The Seven Divisions of Law,*** 1931), a mine of information on the life of the time, and compilations from Arabic sources such as treatises on astronomy, on the magical properties of precious stones, and on games, especially chess. With the ***Crónica general,*** a vast history of Spain, and the ***General estoria,*** an attempt at a universal history from the creation, Alfonso founded Spanish historiography. The former, carried by Alfonso to AD 711 and completed by his son Sancho IV, was the most influential single work of the Spanish Middle Ages. Himself a poet, Alfonso made one of the greatest collections of medieval poetry and music, the ***Cantigas de Santa** María* ("Songs to the Virgin"), in Galician, the then accepted language for lyric.

The ***Crdnica general*** and the ***General estoria***

Learned narrative poetry. A new school of poetry, indebted to France and linked with the monastery and a literate public, became known as the ***mester de clerecia,*** "clerkly craft." Adapting the French Alexandrine in the "fourfold way"—*i.e.,* four-line single-rhyme stanza with a 14-syllable line—it dealt with religious, didactic, or pseudo-historical matter. It was best exemplified in Gonzalo de Berceo (***c.*** 1195–c. 1268), the earliest Spanish poet known by name, who versified in the vernacular the lives of saints, miracles of the Virgin, and other devotional themes with an ingenuous candour and an accumulation of picturesque and affectionately observed detail. Also of the 13th century were the ***Libro de Alexandre,*** much more erudite and pretentious; a version of the Latin romance on Apollonius of Tyre; and the *Poema de Fernán González,* a clerical reworking of a lost epic.

The 14th century. The period of translation and compilation was succeeded by one of brilliant original creation, best represented in the prose of Alfonso's nephew Juan Manuel and in the poetry of Juan Ruiz, archpriest of Hita. Juan Manuel's ***Libro de** los exemplos **del Conde Lucanor** et **de Patronio*** ("Book of the Exemplary Tales of Count Lucanor and Patronio"), a collection of 50 moral tales, still drew on Arabic sources, but its individuality places it high in the beginnings of Spanish fiction.

Under the influence of the Arthurian, or Breton, cycle, which had been circulating in translation, there had appeared (*c.* 1305) the first Spanish romance of chivalry and the first Spanish novel, *El caballero Cifar* ("The Knight Cifar"), on the theme of St. Eustace, the Roman general miraculously converted to Christianity. Cifar's squire, Rihaldo, was a distant precursor of Don Quixote's Sancho Panza and an early token of an impinging of the real on the ideal so characteristic of Spanish literature. About the same time the *Amadís de Gaula,* a chivalric romance related to the Arthurian cycle, whose author may have been a Galician or Portuguese troubadour in exile at the Castilian court, was circulating and was destined to hold the imagination throughout the 16th century through its sentimental idealism, lyrical atmosphere, and supernatural adventure.

The works of Juan Ruiz

Juan Ruiz was the most intensely alert and individual of early poets. His *Libro de buen amor* (1330, expanded 1343; "Book of Good Love"), a collection of disparate elements — Ovid, Aesop, the 12th-century Latin *Pamphilus de amore*, and the liturgy — suggested a Muslim way of thinking in the mingling of eroticism with devotion and an invitation to the reader to intepret for himself equivocal teachings. His Trotaconventos, ancestress of La Celestina, was the first great character in Spanish literature. The Alexandrine metre he handled with new vigour and plasticity, and the text was interspersed with religious, pastoral–farcical, amorous, or satirical lyrics of great metrical variety.

More exotic elements occurred in the verse *Proverbios morales* (*c.* 1355) of Santob de Carrión de los Condes and in an Aragonese version of the story of Joseph, based on the Qur'ān and written in Arabic characters, which was the chief example of "barbarian" literature in Spanish. The *Proverbios* of Santob introduced the grave sententiousness of Hebrew poetry with its extreme aphoristic concision; Santob's sources were the Old Testament, the Talmud, and the Hebrew poet and Arabic philosopher Ibn Gabirol.

Pedro López de Ayala, chancellor of Castile, dominated the poetry and prose of the later part of the century with his *Rimado de palacio* ("Poem of Palace Life"), the last major relic of the "fourfold-way," and his chronicles of Peter I, Henry II of Trastamara, John I, and Henry III of Castile, which stimulated the writing of personal, contemporary history; an early Humanist, he translated and imitated Livy and Boccaccio, Boethius, St. Gregory, and St. Isidore.

The 15th century. The early 15th century witnessed a renewal of poetry under Italian influence. The contrast was strong during the reign of John II between the anarchy of feudalism in its death throes and the cultivation of polite letters, which was becoming a mark of birth and breeding. Collections of poems such as the *Cancionero de Baena,* made for the King by Juan Alfonso de Baena (a converted Jew and a poet), containing 583 poems by 55 poets from the highest nobles to the humblest versifiers, showed not merely the decadence of the Galician–Portuguese troubadour but also the stirrings of more intellectual poetry using symbol, allegory, and classical allusion in the treatment of moral, philosophical, or political themes. Francisco Imperial, a Genoese settled in Seville and a leader of the new poetry, drew on Dante, while the Marques de Santillana, poet, scholar, soldier, and statesman, collected masterpieces of foreign literatures and stimulated translation. His *Proemio e carta al condestable de Portugal* (1449; "Preface and Letter to the Constable of Portugal"), the earliest work of literary history and criticism in Spanish, drew on his reading in contemporary foreign languages and translated classics. Imperial had already sought to acclimatize the Italian hendecasyllable. Santillana's sonnets in the "Italian style" marked the beginnings of the formal enrichment of Spanish poetry. His role as precursor is still acknowledged, though the sonnets and long poems, which reflected his Italian-influenced training, are neglected in favour of his charming rustic songs of native inspiration.

Italian influence on Castilian poetry

Juan de Mena's vast allegorical poem of the drama of history past, present, and to come (*El laberinto de fortuna,* 1444; "The Labyrinth of Fortune"), a more conscious attempt to rival Dante, was weighed down by pedantry and over-Latinization of syntax and vocabulary.

An outstanding anonymous poem of the early part of the 15th century, the *danza de la muerte* ("Dance of Death"), was the finest example of a theme then popular with poets, painters, and composers. Related to the earlier *Danse macabre* at Paris but written with greater satiric force, it introduced characters (*e.g.*, a rabbi) not included in the French cycle. It presented a cross section of society in the form of a dialogue between Death and his protesting victims, and, although not intended for dramatic presentation, formed the basis for later drama.

Catalan. The old Catalan language of northeastern Spain was a branch of peninsular rather than southern Gallo-Romance. Nevertheless it had traces of Provençal, and the literature long used the Languedoc (the dialects of Old French spoken south of the Loire River) and the poetic forms of troubadours north of the Pyrenees.

Poetry. The early Catalan troubadours Guillem de Bergadh, Hug de Mataplana, Guillem de Cervera, and others were genuine Provencal poets. About 100 years later, in the late 14th century, Provençal influence apparently lessened, and poets turned for inspiration to northern France. They took over the long French narratives on romance themes such as the Arthurian cycle and the *noves rimades* metre, a sequence of octosyllabic rhymed couplets. Several poets working in this tradition carried the new interest in the Langue D'oïl (the dialects of Old French spoken north of the Loire) to the extent of incorporating passages of French poetry in their poems.

The great period of Catalan poetry

The great period of Catalan poetry was the 15th century, after John I of Aragon had established in 1393 a poetic academy in Barcelona on the model of the academy in Toulouse with *jocs florals* ("floral games." or poetry congresses), including literary competitions. This royal encouragement continued under Martin I and Ferdinand I and helped to emancipate the literary style from foreign influences. As the century advanced, Valencia emerged as a new focus of literary activity: a school of poetry developing there was noted for its characteristic use of eight-line decasyllabic verses with crossed, or "chained," rhymes and final four-line refrain, illustrating a turning away from French models and a new inspiration from Italy. The *cants d'amor* and *cants de mort* ("songs of love" and "songs of death") by Ausiàs March, an inspired and profound poet, contained the finest verses ever written in Catalan, exerted influence in 16th-century Castile, and continue to influence modern Catalan poets. Jaume Roig's *Lo Spill o llibre de les dones* (*c.* 1460; "The Mirror or Book of Women") was very different — a caustic satire on woman, written in more than 16,000 four-syllable lines, portraying contemporary Valencian life vividly. Johan Roiç de Corella, a great Valencian lyricist, was perhaps the best representative of the Renaissance.

After the union of Aragon with Castile, the Castilian language predominated throughout Spain, spelling a long eclipse of Catalan literature. Nevertheless a Catalan, Juan Boscán Almogáver, inaugurated a new Castilian school of poetry, and Castilians regard him as a landmark in the history of their Renaissance; by the time Boscán's works were published (1543), Catalan poetry had been dead for 50 years.

Prose. Though the oldest document (the text of an oath by a bishop of Urgel) dates from *c.* 1100, literary prose only began at the end of the 13th century. It was written in the everyday speech found in charters from the time of James I's accession to the Aragonese throne in 1213; four great chronicles that survive represent the peak of medieval Catalan prose. The anonymous *Llibre dels feyts del rey en Jacme* ("Book of the Deeds of King James"), compiled after James I's death in 1276, and Ramon Muntaner's account of the Catalan Grand Company's expedition to the Morea in southern Greece and of James II's conquest of Sardinia .were distinguished by skill of narration and quality of language. Bernat Desclot's chronicle deals with the reign of Peter I the Great; though the account of Peter IV the Ceremonious is ascribed to Bernat Descoll, it was planned and revised by the King himself.

The works of Ramon Llull

Ramon Llull was unequalled in his encyclopaedic production, in Catalan, Arabic, and Latin, covering every branch of medieval knowledge and thought. His exhaustive theological treatise *Llibre de contemplacid en Déu* ("Book of the Contemplation of God") began Catalonia's golden age of literature, providing incidentally a mine of information on contemporary society. The *Llibre d'Evast e Blanquerna* founded Catalan fiction. It included the *Llibre d'amic e amat* ("Book of the Lover and the Beloved"), a masterpiece of mysticism, while his *Fèlix* and *Llibre de l'orde de cavaylería* ("Book of the Order of Chivalry") were instructive works in a narrative framework. A century later, Francesc Eximenis, with a similar intellectual amplitude but lacking the originality and driving force of Llull, planned *Lo chrestid* ("The Christian"), an exhaustive treatise on theology, morals, and politics for the layman. It provided vivid pictures of the medieval scene, while his *Llibre de les dones* ("Book of Women"), a manual of devotion and domestic economy, was similarly informative on women's life.

Bernat Metge began the "classical age" by translating Boccaccio's story of Griselda from Petrarch's Latin version and, clothing his scholastic learning with poetic imagination, achieved the stylistic masterpiece of Catalan prose.

The beginnings of drama were represented by a 15th-century Assumption play, *Misteri d'Elch*, which is still performed annually at Elche on the Feast of the Assumption.

Galician. Galician is closely related to Portuguese, and there is no separating the two languages in the three great repositories of medieval verse, the 14th-century *Cancioneiros* ("Songbooks") *de Ajuda, da Vaticana,* and *Colocci-Brancuti.* Indigenous lyric origins were overlaid by Provençal influence, and a dominance of emotion over thought identified Galician with subjective lyricism, so that for over a century Castilian poets made it their medium for lyrics. Macías El Enamorado (flourished mid-14th century), who died for love and became a legend, was the last Galician troubadour; Galicians thereafter wrote in Castilian, and, though there were echoes of their tradition, the Renaissance and Castilian political hegemony together finally ended Galician literature until the 19th-century revival.

Portuguese. *Poetry.* Though no literary documents belonging to the first century of Portugal's history as a nation have survived, there is evidence of the existence of an indigenous popular poetry. A few compositions from before 1200 survive; one, attributed to Sancho I, is the earliest extant *cossante,* or brief, repetitive lyrical poem marked by a wistful sadness, which is never wholly absent from Portuguese literature. The accession in 1248 of Afonso III, who had lived 13 years in France, inaugurated a period of poetic activity illustrated in the *Cancioneiro da Ajuda,* the oldest collection of peninsular verse. Soon literary norms of Provence and of northern France were concurrent, subtlety of form and device taking precedence over thought or emotion.

Diniz' encouragement of literature

The peak of this palace poetry was reached in the early years of the reign of Afonso's son Diniz (or Denis), who had been educated by Aymeric of Cahors, a Frenchman. Diniz founded his country's first university at Coimbra in 1290 and stimulated translation into Portuguese of outstanding works from Spanish, Latin, and Arabic. He founded vernacular prose in Portugal and was also considered the best poet of his age in the Peninsula. To his court came troubadours from Leon, Castile, and Aragon to enjoy the last of a cult dying elsewhere, and about 2,000 poems by its 200 poets were preserved in the three great *Cancioneiros da Ajuda, da Vaticana,* and *Colocci-Brancuti.* Although this poetry lacked ideas and was conventional in form and expression, the language had highly developed musical qualities.

In contrast with the restricted horizons of courtly verse, themes of adventure, war, and chivalry mingled with love, religion, and the sea in a collection of ballad poetry known as the *romanceiro.* Few of these ballads can be dated earlier than the 15th century; they belonged to an anonymous poetry kept alive by oral transmission, with a late artificial flowering from known poets in the 16th and 17th centuries.

Prose. Prose literature in Portugal took much longer than verse to perfect. Religious writings, brief annals of the early kings, and books of descent formed the earliest texts. The *Livro de Linhagens* of Pedro, count of Barcelos, constituted a landmark by going beyond genealogy to history and legend. The early popularity of subject matter based on Celtic tradition is attested in the five songs based on Breton lays with which the *Cancioneiro Colocci-Brancuti* opens: the ideals of chivalry and the spirit of sentimental adventure associated with the knights of the Round Table made strong appeal to the Portuguese temperament, a *História dos Cavaleiros da Mesa Redonda e da Demanda do Santo Graal,* adapted from the French, being the chief relic of considerable activity in this field; from it derived *Amadís de Gaula,* which was the progenitor of a long succession of chivalric romances in Spain and Portugal and whose original was probably Portuguese, at least for the first three books of *c.* 1350, possibly by Vasco de Lobeira.

MEDIEVAL LITERATURE OF THE EASTERN MEDITERRANEAN

Byzantine Greek. Byzantine literature, the literature written in Greek during the Byzantine period (330–1453), is, after Byzantine art, the most direct expression of the self-contained medieval Christian culture that was a continuation of Greco-Roman civilization. Its character was determined by the centralized structure of the Byzantine Empire and particularly by the court life of Constantinople (Byzantium). After Constantine the Great, the Christian Church assumed guardianship of Greco-Roman cultural traditions, and literature written in Greek was affected by a deep linguistic conservatism. It was felt that literary language had to preserve the forms of speech of the Greek New Testament and the Church Fathers and, on the other hand, those of Attic models from the great literary past. Thus literary language was not permitted to display the simplifying influences it underwent in daily use in the mouths of the people. While the language of daily life developed into modern Greek, in the literary field Byzantium clung to speech forms artificially frozen at the Hellenistic Greek stage. Imitation of the ancients became the principle behind all literary forms, and this smothered any sense of originality of content and invention. Though the traditionalism of the Byzantines saved for posterity the best of ancient Greek literature, a consciousness of being in sole possession of the intellectual treasures inherited from the ancients led to an overrating of erudition—hence the pedantic, obtrusively didactic form of many Byzantine writers.

Deliberate archaism of literary Greek

In spite of the admiration that the West entertained for Byzantine culture, any effect of one on the other was small throughout the Middle Ages. Knowledge of Greek was lost in the West during most of the medieval period, and literary contact also suffered from a mistrust that was increased by religious schism and rivalries aroused by the Crusades. In Egypt and Asia Minor Greek literature gave way to writing in the vernaculars (Coptic, Syriac). But Byzantine literature exercised a strong influence on the people of eastern Europe, especially the Slavs.

Non-literary Byzantine works

Much of Byzantine writing was not primarily of literary interest. A vast number of prose works were produced in the fields of theology, hagiography, history, geography, philosophy, military science, and law. In theology, the important writers. all of the 4th and 5th centuries, were Athanasius, the creator of the pattern for Byzantine hagiography; Eusebius, who wrote the first ecclesiastical history; Basil of Caesarea (St. Basil the Great), who organized Eastern monasticism; his brother, Gregory of Nyssa, the earliest Greek ascetical author; and Gregory of Nazianzus, a theologian and poet. In hagiography, mention need only be made of the number of popular saints' lives from which gradually developed a form of religious novel, the best known example of which is the story of Barlaam and Josaphat.

Religious poetry. The oldest surviving religious poetry was composed in ancient Greek metres for use in private devotions. This included the so-called maiden's song in the ***Symposium*** of Methodius of Philippi and the religious poems of Gregory of Nazianzus (written between 381 and 389). More important than this private religious poetry were the Byzantine liturgical hymns. The need for congregational hymns developed to counteract the success that the esoteric religious movement of the Gnostics and the heretic Arius (died 336) had with their religious poetry set to folk music. The model for the poetical development may well have been Syrian church hymns, the most prolific and successful writer of which was a Syrian, Romanos Melodos (6th century). In the 7th century, church poetry entered a new state, characterized by an increase in artistic finish and a falling off in poetical vigour, with the composition of the *kanōn*, a poem built up out of eight or nine lyrics, all differently constructed, of which Andrew of Crete has been regarded as the inventor and of which the most celebrated writers were John of Damascus and Cosmas of Jerusalem (both first half of the 8th century).

Secular poetry. Epic popular poetry began with the vernacular Greek literature (see below); but there were several literary works that can be compared with the epics of the Alexandrian age. Nonnus (flourished c. 400) wrote, while still a pagan, an epic on the god Dionysus and, as a Christian, a lengthy versification of St. John's Gospel. The historic epic was also carefully preserved by the Byzantines, and in the 7th century George the Pisidian described the wars of the emperor Heraclius, though more as an encomiast than as an epic poet. Later, a deacon, Theodosius (10th century), immortalized the capture of Crete by the emperor Nicephorus Phocas.

In the 12th-century revival of culture, some long poems imitated ancient Greek romances: the story of Rodanthe and Dosicles by Theodore Prodromus, the love story of Aristander and Callithea by Constantine Manasses, and one in prose, the story of Hysmine and Hysminias by Eustathius Macrembolites.

The detached and ascetic point of view that dominated the whole Byzantine period was fatal to a dwelopment of secular lyrical poetry. A few poems by John Geometres and Christopher (11th century) of Mytilene and others are placed in this category. The dominant form for all subjective poetry was the epigram, George the Pisidian treating most diverse themes epigrammatically and, in the 9th century, Casia, Byzantium's only poetess, writing several epigrammatic productions and church hymns. Epigrammatic poetry reached its highest development in the 10th and 11th centuries in the productions of John Geometres, Christopher of Mytilene, and John Mauropus, the 10th century also dating the most valuable collection of ancient epigrammatic poems, the ***Anthologia Palatina.***

The epigram

The didactic poem was much loved among the Byzantines, a form in line with their tendency to instruct their fellow men. John Tzetzes wrote metrical commentaries on Homer, Hesiod, Pindar, Aeschylus, Euripides, and Aristophanes.

Dramatic poetry was completely lacking among the Byzantine Greeks. Apart from some moralizing allegorical dialogues (by Theodore Prodromus, Manuel Philes, and others), the only work of the Byzantine period resembling a drama, at least in external form, was the ***Christos paschon,*** which was actually a cento of classical citations.

Vernacular writing. The vernacular literature stood alone, both in form and in content. It showed originality of conception and entirely new medieval matter, and in it poetry took first place. Though a few preliminary attempts were known, longer Greek vernacular works were written down only from the 12th century onward. Toward the close of the 15th century, rhyme came into use. In Constantinople a mixture of the learned and the popular language was first used in poems of admonition, praise, and supplication. To this oldest, 12th-century class of vernacular poetry belonged the admonitory *Spaneas* and (partly, at least) a supplicatory poem composed in prison by a chronicler, Michael Glycas. In the succeeding period, erotic poems began to be written, such as the Rhodian love songs.

Long epic poems, in which were treated such subjects as the legends of Troy, of Achilles, and of Alexander. formed a separate group. To these may be added romances in verse in the artificial classical language—*e.g.*, ***Callimachus and Chrysorrhoe, Belthandrus and Chrysantza***—and those after the Western pattern—such as ***Phlorius and Platziaphlora*** (the old French story of ***Floire et Blanchefleur***), ***Imberius and Margarona,*** and ***Apollonius of Tyre.*** Well-known stories from the ***Physiologus*** ("The Naturalist") about animals, plants, and stones were put into verse in this period, and other animal tales in verse were also popular. There were also poems in which famous heroes and historical events are celebrated: the exploits of Belisarius (Justinian I's general), the fall of Constantinople (1453), and the taking of Athens (1458). The chief of these was the great heroic epic of Digenis Akritas, preserved in several versions, the oldest of which dates from the 13th century. Digenis Akritas represented the bold frontier warrior ***(akrites)*** on the Euphrates frontier in the 8th–10th centuries. His feeling for nature and strong family affections forecast much that is best in modern Greek popuiar poetry.

Heroic epics

Hebrew. In the 6th century, some Jewish groups attempted to enforce the exclusive use of Hebrew in the synagogue, this tendency being part of a Hebrew revival that began in Palestine and spread westward but did not reach Babylonia until the 10th century.

Piyyuṭim. Synagogues began in this period to appoint official precentors, part of whose duty it was to compose poetical additions to the liturgy on special sabbaths and festivals. The authors were called *payṭanim* (from Greek *poiētēs*, "poet"), their poems *piyyuṭim.* The keynote is messianic fervour and religious exuberance. Besides employing the entire biblical, Mishnaic, and Aramaic vocabularies, the *payṭanim* coined thousands of new words. Such poems, presupposing a highly educated audience, abound in recondite allusions and contain exhaustive lists of rites and laws. It is known that the most outstanding poets—Phineas the Priest, Yose ben Yose, Yannai, and Eleazar ha-Kalir or ben Kalir—lived in that order, but when or where in Palestine any of them lived is not known. The currently accepted datings are 3rd century and 5th–6th century AD. Many *piyyuṭim* are still used in the synagogue.

Biblical Hebrew was re-established as the literary idiom about 900 by Saadiah Gaom, grammarian and religious polemicist. The Arabic system of quantitative metre was adapted for Hebrew during this period (900–1000), probably by Dunash ben Labrat. At first the *piyyuṭ* form was retained for religious poems, and the new metres were used only for secular poetry, which closely imitated Arabic models and, like the latter, was chiefly employed for laudatory addresses to prominent people.

Adoption of Arabic metre

The Palestinian tradition in Europe, 800–1300. From Palestine, the Hebrew renaissance soon spread into the Byzantine Empire. In Sicily and southern Italy (which belonged to Byzantium) several important *paytanim* were at work, and before 1000 a secular literature began to arise in Italy: a fantastic travelogue of Eldad the Danite; a historical romance, ***Sefer ha-yashar;*** and ***Josippon,*** a revision of Josephus' ***Antiquities*** filled with legendary incidents—this last-named book was popular until modern times and was translated into many languages. Nathan ben Yehiel completed in 1101 at Rome a dictionary of Talmudic Aramaic and Hebrew, the ***'Arukh,*** which is still used.

In the middle of the 10th century members of the north Italian family Kalonymos brought Talmudic studies and *piyyuṭim* to Mainz, Germany, where the academy became a centre of studies under the direction of Gershom ben Judah, known as "the Light of the Exile." As a poet, he established a distinctive style of European *piyyuṭ* in poems that read very much like early European popular poetry. The greatest alumnus of the Mainz academy was Rashi, an author of complete commentaries on the Bible and on the Babylonian Talmud, himself a poet of note.

Slaughter during the Crusades drove large masses of Jews into eastern Europe, whither they carried their German speech but hardly any literary culture. In Germany accounts of the disaster were written in a new prose style permeated with poetry; liturgical poetry became henceforth mainly a chronicle of persecutions. These sufferings inspired an important mystical movement, largely propagated through stories, of which the chief collections are the *Maʿaseh Book* and the *Sefer Ḥasidim*, the latter attributed to Judah ben Samuel, "the Hasid" of Regensburg (died 1217).

The golden age in Spain, 900–1200. Spanish Jewry began to flourish in Muslim Spain under the caliphate of Córdoba, where Hasdai ibn Shaprut, a vizier, was the first great patron of Hebrew letters. His secretary, Menahem ben Saruq (died *c.* 970), wrote a biblical lexicon, which was criticized by Dunash ben Labrat when the latter arrived in Spain with philological ideas from the East. Samuel ha-Nagid, vizier of Granada (990–1055), himself a poet and philologist, gathered around him a group of poets, most outstanding among whom was Ibn Gabirol. Moses ibn Ezra of Granada (died c. 1139) was the centre of a brilliant circle of poets. Moses' kinsman Abraham ibn Ezra, a poet, philosopher, grammarian, and Bible commentator, attacked the language and style of the early *payṭanim;* he and Judah ben Samuel Halevi were the first to use Arabic metres in religious poems. Dominated by Arab standards of taste, the secular poetry dealt with themes of Arabic poetry and often reproduced Arabic phrases; it was written to be appreciated by a small circle of connoisseurs and declined with the collapse of Jewish prosperity in Muslim Spain. The last major poet in Spain was Judah ben Solomon Harizi, who translated various philosophical works into Hebrew.

The poets of Granada

The use of biblical Hebrew was made possible by the work of philologists. Of great importance was the creation of comparative linguistics by Judah ibn Kuraish (flourished about 900) and Isaac ibn Barun (flourished 1100). Judah Hayyuj, a disciple of Menahem ben Saruk, recast Hebrew grammar, and, in the form given to it by David Kimhi of Narbonne (died *c.* 1235), the new system was taken over by the Christian humanists and through them by modem scholarship. The first complete Hebrew grammar, *Kitāb al-lumaʿ*, was written by Ibn Janāḥ of Córdoba (died 1050).

Jewish medieval philosophers in Spain wrote in Arabic, not Hebrew, until the 13th century. Apart from Isaac Israeli (north Africa, died *c.* 940) few medieval Jews made original contributions to science, but the Spanish Jews shared the best scientific education. Abraham bar Hiyya Savasorda (*i.e.*, "chief of police"; died *c.* 1136) of Barcelona was an original mathematician who wrote in Hebrew works on mathematics, astronomy, and philosophy. When the Almohads expelled the Jews from Muslim Spain in 1148, many learned refugees went to Languedoc and Provence and there translated scientific and philosophical works.

The period of retrenchment. From 1200 to 1750 was the era of the ghetto, during which the area of western Hebrew culture shrank to a remnant in Italy, while an entirely different culture arose in eastern Europe. The appearance of the Hebrew version of Maimonides' *Guide of the Perplexed* (1204), which applied Neoplatonic and Aristotelian philosophy to biblical and rabbinic theology, provoked orthodox circles into opposition to all secular studies. As a result of Maimonides' work, there was a return to Neoplatonist mysticism in a form known as Kabbala. This culminated in the theosophy of the *Zohar*, which is ascribed to Moses de Leon and which exercised an influence comparable only with that of the Bible and Talmud. Hebrew culture, however, was reduced to a miniature scale in the West after the expulsion of the Jews from England (1290), from France (1306), and from Spain (1492). It continued in Italy, where it remained in contact with contemporary Christian thought. The most outstanding figure was the mystical philosopher M.H. Luzzatto, who wrote a work on poetics and three remarkably modern plays.

Coptic. Almost entirely religious in character, literature in Coptic contained, as well as translations from Greek, original writings by the Greek Fathers and founders of Eastern monasticism and texts throwing light on early Gnosticism and Manichaeism (a dualistic gnostic world religion). From the 3rd century, when Coptic began to be widely used as a literary language, to its decline in the 7th and 8th centuries, examples are extant of every kind of religious writing. The earliest original writings in Coptic—letters by St. Anthony of Egypt, first of the "Desert Fathers"—date from the 2nd century; and in the 3rd and 4th centuries many ecclesiastics and monks wrote in Coptic, among them St. Pachomius, whose monastic rule (the first cenobitic rule; for solitary monks gathered in communities) survives only in Coptic; St. Athanasius, the first patriarch of Alexandria to use Coptic, as well as Greek, for didactic homilies; Macarius (the Elder) of Egypt, a famous ascetic desert solitary; and St. Serapion, bishop of Thmuis, whose liturgical texts are a valuable source for early church worship. The first to realize fully the language's literary potentialities was Shenute (Shenoud; *c.* 348–457), abbot of the White Monastery (Deir el-Abyad), near Atripe, Upper Egypt. In sermons, treatises, and homilies he showed mastery of style and the forceful character that made him (though unknown in the West till the 20th century) the most influential personality of his period in Egypt, where he is still regarded as a saint. His works remain the outstanding original writings in Coptic, equalled in intensity only by 7th- and 8th-century hymns sung antiphonally to traditional tunes and written to encourage the Coptic Christians during the persecutions that followed the 7th-century Muslim invasions.

Ethiopian. The literature of Ethiopia (Abyssinia) consists of writings in the classical Ge'ez (Ethiopic) language, constituting very largely a literature of translations from Greek and Arabic and works in Amharic, the modern official language. The earliest extant literary works in Ge'ez are translations of religious writings from Greek, which may have influenced their style and syntax.

Early works in Ge'ez

From the 7th century to the 13th, a period marked by political disturbances, there was no new literary activity, but with the restoration of the "Solomonian" dynasty, *c.* 1270, there began the most productive era of Ge'ez literature, again characterized by translation, not from Greek but from Arabic, though the originals were frequently Coptic, Syriac, or Greek. The subject matter was mostly theological or strongly flavoured by religious considerations. The most interesting work occurring at the beginning of this period was *Kebra nagast* ("Glory of the Kings"), a combination of mythical history, allegory, and apocalypse, the central theme of which is the visit of the Queen of Sheba to Solomon and the birth of a son, Menelik, who became the founder of the Ethiopian dynasty.

Abba Salama, an Egyptian Copt who became metropolitan of Ethiopia in 1350, was not only responsible for a revision of the text of the Bible but translated or induced others to translate several books popular among the Ethiopian faithful. The rhapsodical *Weddase Maryam* ("Praise of Mary") is appended to the Psalter and thus has almost canonical status. In a slightly later period, about the beginning of the 15th century, various separate lives of saints and martyrs, including St. George (the patron saint of Ethiopia), were written. At this time was undertaken a translation of the Arabic Synaxarium, containing lives of saints--one or more for every day in the year.

The early 15th century saw the translation of several apocalyptic books, which inspired two original compositions. *Fekkare Iyasus* ("Elucidation of Jesus") was written during the reign of Theodore I (1411–14); "Mystery of Heaven and Earth" was written somewhat later and is noteworthy for a vigorous account of the struggle between Michael and Satan. This book must not be confused with another original work of the same period, the "Book of Mystery" by Giyorgis of Sagla, a refutation of heresies. The large hymnals and antiphonaries called *Deggua, Mawase'et,* and *Me'raf* also probably dated from this time, though some of the anthems may be older. Another type of religious poetry first composed during the 15th century was the *malk'e* (likeness), consisting generally of about 50 five-line rhyming stanzas, each addressed to a different physical or moral attribute

Apocalyptic works of the 15th century of the saint apostrophized. As a last example of the religious literature of the "golden age" may be mentioned the "Miracles of Mary," translated from Arabic in 1441–42; enormously popular, it went through several recensions, or critical revisions.

Syriac. Apart from its interest to Semitic scholars, Syriac literature is of importance for the study of Syrian Christianity, for the fact that it has preserved Greek Christian texts, and for its role as an intermediary between ancient Greek learning and the Islamic world.

Poetry. The works of St. Ephraem Syrus (4th century) stood at the beginning of Syrian literature and were never surpassed by any later author. Two poetic forms were employed by him, one for spoken speech in metrical form, whether a narrative or didactic epic, the other a more artful composition in strophes to be sung by a choir or double choir. At the beginning of Nestorian Christianity came Narsai (died c. 503), the greatest poet of the East Syrians. The elegance of his poetry and the beauty of his style earned him a reputation as the "harp of the Holy Spirit."

Translations from Greek. The Syrians began their literary activity by making Greek Christian writings available in Syriac. Almost all important authors and documents of Christian Greek were translated by Syrians. This mass of Greco-Syrian translated literature is an essential source for works of Greek Christian literature that have not survived in their original language. The translators also put into Syriac most of the works of Aristotle and other ancient Greek philosophers, together with writings of the chief medical and scientific authors of ancient Greece. These translations formed an important preparatory work for the rise of Islamic civilization, for it was much easier to translate from Syriac into Arabic than directly from Greek into Arabic; thus, for instance, to take the works of Galen alone, 130 were translated into Arabic from Syriac and only nine directly from Greek originals into Arabic. For this reason many works of Greek learning exerted their influence on the Muslim world through the medium of Syriac.

Armenian. There is evidence that a pagan oral literature existed in Armenia before the invention of the alphabet in the 5th century AD, but, owing to the zeal of the early Christian priests, little of this has been preserved. For about a century after their conversion to Christianity (c. 300), the Armenians had to rely on Greek and Syriac versions of the Bible and other religious books. These languages were unintelligible to the common people, and to remedy this St. Mesrop invented the Armenian alphabet (c. 410). The catholicos Sahak the Great and St. Mesrop formed a school of translators who were reputedly sent to Edessa and to Constantinople to procure and translate Syriac and Greek copies of important works.

The importance of religious translations and histories

Much of the literary activity of the 5th century, the golden age of Armenian literature, was devoted to such translations; original works, however, were not wanting, such as the histories of Eghishe and Ghazar of Pharp. The masterpiece of classical Armenian writing is the "Refutation of the Sects" by Eznik Koghbatzi. This was a polemic work, composed partly from Greek sources, in defense of orthodox Christian belief against—and thereby providing valuable information about—pagan Armenian superstitions, Iranian dualism, Greek philosophy, and the Marcionite heresy. Its pure classical style is unsurpassed in Armenian literature. The work of translation of such authors as SS. John Chrysostom and Cyril of Alexandria continued in the 6th–8th centuries. The so-called Hellenistic (Yunaban) school produced excessively slavish translations from Greek grammatical, theological, and philosophical works, including those of Plato, Aristotle, and Philo of Alexandria.

Literatures of the Balkans. Just as the history of the various peoples of modern Yugoslavia has differed over the past thousand years, so the literatures of the Serbs, Croats, Slovenes, and Macedonians have developed independently. From an early date Croats and Slovenes formed part of the central European cultural heritage, whereas Serbs and Macedonians were part of the Eastern, Byzantine, culture and, later, of the Turkish sultanate. Since 1918, when they were united, the Yugoslavs have had three literary languages—Serbo-Croat, Slovene, and Macedonian—and two alphabets, the Latin and the Cyrillic. The Serbs and Croats speak a common language but use the Cyrillic and Latin alphabets, respectively. The Slovenes, with a language akin to Serbo-Croat, use the Latin alphabet and the Macedonians the Cyrillic. The peoples of Yugoslavia continue to develop their national and cultural individualism within the framework of a common federal state.

Serbian. Serbian literature originated in the use of the Old Church Slavonic language in the translation of religious works from Greek into the Glagolitic alphabet (traditionally invented by St. Cyril) by the apostles to the Slavs, St. Cyril (died 869) and St. Methodius (died 885). Originally forming part of South Slav literature, from the 12th century Serbian acquired its own characteristics, as in the illuminated Miroslav Gospel; biblical stories; hagiographies, notably St. Sava's, patron saint of Serbia; and tales such as that of Barlaam and Josaphat.

Croatian. The Croats were converted to Christianity at the turn of the 9th century by disciples of SS. Cyril and Methodius. The first ecclesiastical works were written in the Glagolitic alphabet, and fragments have survived in the Kiev Missal-pieces (11th century). From the 12th century the national language was widely used for inscriptions and legal documents, but from the 14th century Latin was increasingly used. Immediately before the age of Humanism (second half of the 15th century), literature consisted mainly of biblical stories, legends, folklore, and popular stories (of Troy, Alexander the Great. and others). Among other works, a chronicle by a priest, Dukljanin, is outstanding. Parallel with the widening of literary themes was a development of religious drama ("The Torments of Jesus," 15th century) and of religious poetry.

Slovene. The dissolution of the independent state of Carantania (now Carinthia: 821) and a merging of the upper classes into German feudal society retarded any development of Slovene literature. Among medieval manuscripts that have been preserved, the most remarkable are the *Brižinski spomeniki* (Freising manuscripts, c. 1000), containing two forms of general confession and a sermon on sin and penance, which are the first written examples of the Slovene, and perhaps of any Slavic language. Among other important manuscripts is the *Stiški* rokopis (c. 1428, 1440; "The Manuscript of Stična"), which includes a new form of general confession and the first known record of an Easter song in Slovene translation. Besides these few preserved proofs of the use of the Slovene language in catechesis, though not in liturgy, rich folk poetry—lyrical as well as epic—flourished among the Slovene peasants. Ballad motifs prevailed in epic poetry, which has been preserved with all of its archaisms in the remote Alpine valleys of the Venetian Slovenia.

The Freising manuscripts

Macedonian. Like other Slavic literatures, literature in Macedonia began in the 9th century with the translations of religious works by SS. Cyril and Methodius and with the first Slavonic "university" that emerged at Ohrid (then in Bulgaria; see below) under St. Clement (died 916), a disciple of SS. Cyril and Methodius. With this orthodox religious literature, there developed an apocryphal, heretic Bogomil literature. In the 16th century, with a fusion of Old Church Slavonic and vernacular languages, a popular Damascene literature developed: *Damaskini* were translations from the works of Damascenus the Studite. This lasted until the beginning of the 19th century, when books written in modern Macedonia were first published.

MEDIEVAL LITERATURE OF EASTERN EUROPE

Bulgarian. Slavo-Bulgarian literature stems from the Christianization of the Slavs beginning with Khan (Tsar) Boris I's adoption in 870 of the Eastern Orthodox faith for his court and people. This political decision, combined with geographical proximity to Byzantium, determined a key role for Bulgaria in the Balkan development of the first Slavic written language and its corpus of ecclesiastical writings known as Old Bulgarian literature.

Following this religious conversion, the pupils of Cyril and Methodius were soon to establish the first Slavic literary school (893–971), under patronage of the royal court of Preslav, capital of Tsar Simeon (died 927) and Tsar Peter (died 969). Arising from its mission to Macedonian Devol and Ohrid, an illustrious, first Slavic "university" was also founded by St. Clement, patron of Bulgaria's modern Sofiiski Universitet "Kliment Ohridsky." Among monastic centres at Preslav and Ohrid were St. Panteleimon foundations. In this Golden, or Old Bulgarian, period, medieval Bulgarian culture aspired to rival even the "Imperial City" (called Tsarigrad by the Slavs), Constantinople itself, as suggested by John the Exarch's *Shestodnev* ("Hexameron"; *i.e.*, "Six Days [of Creation]"). Tsar Simeon's own name is closely linked with work on his *Simeonov sbornik* ("Simeon's Collection [of Gospel Commentaries]") and with the *Zlatostruy* ("Golden Stream"), the first Slavic version from the Greek of St. John Chrysostom. The predominant role played in this early Slavic literature by translating from (and so Slavicizing) the Greek reflects the determination of these Bulgarian writers to promote and mature the Slav dialect and to convey, in structure and spelling, all the complexities and sophistication of Byzantine thought.

The 13th–14th-century Middle Bulgarian, or Silver, age of the Asen and Shishman dynasties excelled in sheer graphic virtuosity (script, layout, illumination, binding) of its manuscripts, such as the Vatican *Manasses Chronicle* of 1345 and the London *Tsar Ivan Aleksandur Gospel* of 1356. In content, too, Byzantine influences and translations from the Greek continued to yield, as in the Old Bulgarian period, abundant literary resources. The Asenid taste for historical and temporal themes was succeeded in the late 14th century by the mystical doctrines of Hesychasm, with its quest for "inner light." This was the doctrine of Theodosius of Turnovo (noted for his Kilofarevo monastery school) and his most celebrated pupil, Patriarch Evtimy (died c. 1404). Both were leading figures in the Turnovo literary school, famous for its endeavours to standardize and purify the Old Church Slavonic (OCS) tradition as closely as this could still feasibly be related to its pristine 9th–10th-century forms. The withdrawal, if not the close, of Bulgarian medieval literature features a curious genre of contemporary travelogue—the conveyance of saints' relics from Turnovo to Tsar Ivan Stratsimir's Bdin (Vidin) or farther west. Such Bulgarian scholars as Grigory Tsamblak and Konstantin of Kostenets ("the Philosopher") also migrated westward, taking with them literary skills and tradition. With these last scions of the early and late medieval Bulgarian literatures often went, too, the actual heritage of Old Church Slavonic.

Important as this sophisticated ecclesiastical literary tradition proved to be in its Bulgarian (and wider Balkan and Slav) contexts, it never displaced the other, less refined, and certainly quite uncanonical medieval literary streams that flowed more broadly over biblical, historical, and even heathen grounds to fill Balkan legends with classical, Christian, and apocryphal themes. Such fiction, with its moral didactic intentions, includes an early-10th-century Bulgarian short story (probably the earliest Slavic example in this prolific genre) of Simeon's credulous cavalryman's "miraculous encounter" with the Magyars ("Chudo s bulgarina") and the tales of "Mikhail the Warrior," "Teofana the Innkeeper," "Stefanit and Ihnilat," and, based probably on Boris I's daughter Praksi, "Bulgarian Queen Persika." Here also belong, very definitely outside the ecclesiastical and royal pale, the body of writings, doctrinal and apocryphal, of the Bulgarian Bogomil heresy, provoking from official quarters those lively and informative reactions of Presbyter Kozma and of Tsar Boril's synodal convocation of 1211. Finally, on ground common to both the Old Bulgarian ecclesiastical and the more popular literatures, are the accounts of St. John the Rila, the traditional patron saint of Bulgaria and founder of the first monastery. Notable here, alongside the popular versions, are Patriarch Evtimy's "The Life of Our Most Blessed Father Yoan of Rila" and Vladislav Gramatik's "The Rila Story: Conveyance of St. Yoan's Relics to the [Re-founded] Rila Monastery."

Russian. *Old Russian.* The earliest works of Old Russian literature arose in the 11th century. Since the development of literature was connected from the first with the acceptance of Christianity by St. Vladimir in 988 and the country's Christianization, Old Russian literature was primarily religious and didactic? and didactic works were influenced by Byzantine literature in translations from the Greek. Secular literature was influenced by oral folk poetry, which had existed among the East Slavs long before Russia had a written literature. The most valuable heritage of Old Russian folklore was an epic or heroic folk song, the *bylina,* already being composed in the 10th century and later ranging widely in theme and subject; these folk songs were produced in Kiev, Novgorod, and Galicia.

The Kievan period. The first phase of Old Russian literature, from the 11th to the early 13th century, is called the Kievan period or, sometimes, because of the Tatar Mongol invasion that followed it, the pre-Mongol period. The chief concentration of literary activity was in southern Russia, mainly in Kiev, although isolated works were written in other towns in the south and in the north. The language of both northern and southern writers was basically the same—the old literary language of the East Slavs, which had absorbed to a greater or lesser extent elements of Old Church Slavonic.

The literature of the Kievan period possessed from the first many translations, mainly from Greek but some also from Latin. Among them were service books, mostly in Old Bulgarian, that included prayers and chants. Several kinds of biblical books appeared, among them "Aprakos" Gospels, in which readings from the Gospels were arranged by days of the week as part of the church service; the earliest surviving manuscript of an "Aprakos" Gospel is *Ostromir's Evangelium,* copied in 1056–57, and the oldest known manuscript of the Four Gospels was the *Galician Gospel* (1144). Other translated works included apocryphal literature, saints' lives, chronicles, books about the creation of the world, an Alexander romance, and versions of the Troy legend and various Byzantine romances.

Old Russian chronicles and sermons

One of the oldest literary genres, widely developed in Old Russian, was the chronicle. As early as the mid-11th century, compilations of chronicles were beginning to be made, and by the early 12th century a compilation of outstanding historical and literary significance was taking definite shape. This was the *Povest vremennykh let (The Russian Primary Chronicle),* the first version of which was compiled *c.* 1112, probably by Nestor, a monk of the Kiev-Pechersk Monastery.

Sermons occupied an important place in the literature of the Kievan period, the best example being *Slovo o zakone i blagodati* ("The Discourse on Law and Grace") by Ilarion, written before his appointment as the first Russian metropolitan in 1051. In the 12th century, the second Russian metropolitan, Kliment Smolyatich, and, in particular, Kirill, bishop of Turov, were outstanding exponents of the art of ecclesiastical oratory in Kievan Rus. Of early hagiographical works, the anonymous *Skazaniye* ("Legend") of the princes Boris and Gleb, the sons of St. Vladimir, merits particular attention. It resembles the historical legend rather than the traditional Byzantine saint's life; and it is full of lyrical laments, monologues, prayers, and meditations.

In the first quarter of the 13th century a work was written that later was called *Kievo-Pechersky paterik* ("The Paterikon of the Kiev-Pechersk Monastery"). It is based on correspondence between Simeon, bishop of Vladimir, a former monk of the Kiev-Pechersk Monastery, and Polikarp, a monk at the monastery. The literature of pilgrimage to the Holy Land originated in Kievan Russia. The most noteworthy work of this kind was a *Khozhdeniye* ("Pilgrimage") to Palestine in 1106–08 of Daniil, prior of a southern Russian monastery.

The most outstanding work of Kievan Russia was *Slovo o polku Igoreve* ("The Lay of Igor's Host"), an account of the unsuccessful campaign in 1185 by Prince Igor of Novgorod-Seversky and the princes allied with him against the Polovtsians. It was written between 1185 and 1187 and was preserved in a single manuscript, discovered in 1795 and lost in the burning of Moscow during Napoleon's invasion of Russia in 1812. The copies that were made before it was lost have been authenticated, and the events narrated in the "Lay" were probably written shortly after they took place. Deeply patriotic in spirit and remarkable for its literary mastery, the lay ranks as an outstanding medieval epic.

Development of regional literatures. During its brief existence Kievan Russia created a literature distinguished for both artistry and ideas. From the mid-12th century, nevertheless, a gradual decline set in, accentuated in the 13th century by the Tatar Mongol invasion, and there was a marked falling off of literary activity.

Works recounting the Tatar invasion

The Tatar invasion was reflected as a great disaster in a number of 13th-century works. As literature, most noteworthy is ***Povest o razoreni Ryazani Batyem*** ("The Story of the Destruction of Ryazan by Batu Khan"). The eloquent, fiery sermons of Serapion, archimandrite of the Kiev-Pechersk Monastery, are full of grief at the Tatar invasion. At the end of the 13th or in the early 14th century, a life of Prince Alexander Nevsky was written in the traditional form of a hagiography but closely akin to a war narrative; it was dominated by the epic figure of Alexander. An outstanding literary work of 13th-century southern Russia was "The Galich-Volynsk Chronicle," describing events from 1201 to 1292, which formed part of a copy of a compilation of chronicles called the Hypatian collection, written in the 1420s.

The regional character of Russian literature resulted from increased lack of communication between the territories of Russia and was still apparent in the 14th and 15th centuries. But by the beginning of the 15th century the literature of Moscow began to predominate, for Moscow had, from the mid-14th century, played the part of a unifying centre for Great Russian nationalism. Tendencies toward unification are particularly clear in the tales of *Mamayevo poboishche* ("The Rout of Mamay"), composed at the end of the 14th century about the Battle of Kulikovo (1380). The conquest of Constantinople by the Turks in 1453 inspired Nestor-Iskander's ***Povest o vzyati Tsargrada*** ("Tale of the Taking of Tsargrad"—*i.e.,* Constantinople), mainly written in the style of a traditional war narrative. In the late 15th and early 16th centuries other stories were written in Moscow on the theme of Russia's political succession to the inheritance of Byzantium.

Also of the 15th century was the *Khozhdeniye* ***za tri morya*** ("Journey Across Three Seas") by Afanasy Nikitin, a merchant from Tver who from 1466 to 1472 travelled in India and Persia. The "Journey" was written in lively, colloquial Russian, with an admixture of Persian, Arabic, and Turkish words. Regional traits are also absent from ***Povest o Petre i Fevroni*** ("The Tale of Peter and Fevronia"), which may be ascribed to the 15th century.

Of other 15th-century regional literatures, that of Novgorod was particularly highly developed, producing a succession of literary works written in defense of Novgorod's political and ecclesiastical independence. In the literature of Tver, which vied with that of Moscow, the most outstanding work was a eulogistic discourse on the Orthodox Great Prince Boris Aleksandrovich, written c. 1453. The most significant work of the literature of Pskov was ***O pskovskom vzyati,*** describing the subjugation of Pskov by Moscow in 1510.

The surge of literary activity beginning at the end of the 14th century was connected with the influx from Bulgaria and Serbia of clerics into Russia after the conquest of their countries by the Turks. As well as in retranslations from the Greek, the content and style of South Slavonic literature was brought to Russia largely in hagiographical works.

Literature of unified Russia. The establishment of a unified Russian state dated from the reign of Ivan III the Great. In 1480 the liberation of Russia from two and a half centuries of Tatar domination was finally achieved, and autocratic power centralized in the Moscow state. In about the middle of the 16th century, Ivan IV the Terrible (died 1584) began an energetic campaign against the powerful feudal princelings, the boyars. This struggle was reflected in literature. The most notable spokesman for the new nobility, created by Ivan in opposition to the boyars, was Ivan Peresvetov, whose work included several propagandist tales and two petitions to Ivan the Terrible. The boyars were represented by a publicist of exceptional literary gifts, Prince Andrey Mikhaylovich Kurbsky. In his ***Istoriya o*** *velikom* ***knyaze moskovskom*** (written in the 1560s and 1570s; "History of Ivan IV"), Kurbsky severely castigated Ivan for his persecution of innocent boyars.

Effects of the campaign against the boyars

Georgian. The literary history of Georgia, like that of neighbouring Armenia, began with the conversion to Christianity in the 4th century and a need to invent an alphabet for the propagation of the Scriptures in the vernacular. The Gospels were translated in the 5th century from Armenian, and the acts of the Apostles and Psalms were translated soon afterward, as well as works of early Greek Fathers.

Original Georgian literature began with lives of saints, the first one being that of St. Shushanik (c. 480), and other early works include an account of the conversion of Georgia by St. Nino and stories about King Vakhtang I Gorgasal, a 5th-century hero. The Georgians were the first to give a Christian colouring to a Buddhist legend of Barlaam and Josaphat, which reached them in the 9th century through an Arabic version.

Ecclesiastical literature was rooted in Byzantine Greek culture, and romance and epic blended the civilization of Persia with original elements of Caucasian folklore. Important prose romances were the *Visramiani* ("Visramiani: The Story of the Loves of Vis and Ramin"), adapted from an Iranian romance going back to Parthian times, and ***Amiran-Darejaniani,*** a cycle of fantastic tales. Secular poetry started with Ioann Shavteli and Chakhnt-khadze, who wrote odes in honour of King David II (died 1125) and Queen Tamara (died 1213). The supreme achievement of Georgia's golden age was Shota Rustaveli's epic ***Vep'khis-tqaosani*** ("The Man in the Panther's Skin") on the themes of ideal comradeship, courtly love, and heroic endeavour. His poetic tradition, broken by Mongol invasions, was renewed in the 17th century by the royal poets Teimuraz I and Archil II.

Czech. The earliest preserved texts in the Czech language were written in the latter part of the 13th century. The courts of the Přemyslids kings Wenceslas (Vaclav) I (died 1253) and Otakar II (died 1278) gave encouragement to courtly literature in German; but there is no certain evidence of a significant body of literature in Czech before the end of the century.

The 14th century saw a continuous stream of Czech literary production. A group of verse "Legends," written in the first quarter of the century, already showed considerable maturity of technique. The earliest secular work was the epic ***Alewandreis,*** a life of Alexander the Great, showing contact with a German Alexander epic by Ulrich von Eschenbach. The Czech poem is a minor masterpiece of courtly literature. Another striking epic, "The Dalimil Chronicle," portrayed the history of the Czechs in a vivid, forthright style. Under the emperor Charles IV learning and literature flourished in Bohemia in Czech and German. The greatest achievement of this period was a verse "Legend of St. Catherine."

From about 1350 prose genres began to be cultivated, first lives of saints and chronicles and then, no doubt for a growing urban middle class, versions of several popular medieval tales. From the last part of the century dated a group of verse satires and didactical poems showing great originality and character, especially the anonymous satires of the Hradec manuscript and a political allegory. ***Nova rada*** ("The New Council"), written by Smil Flaška of Pardubice to defend the rights of the Bohemian nobility against the power of the crown.

A preoccupation with social and moral questions coloured much Czech writing in the last part of the 14th century. Thomas of Štítný wrote treatises in the vernacular that anticipated one of the main tendencies of the Hussite movement. This reform movement, inspired by Jan Hus, dominated the life of Bohemia for the first half of the 15th century; amid the religious controversy and civil strife of this period it was natural for vernacular writing to be directed largely to practical and polemical purposes. Hus's own importance lay not only in his vernacular sermons and in his letters but also in the reform of the Czech orthography advocated in the treatise De *orthographia Bohemicu.* A direct product of the reform movement was the simple and moving hymns composed by the Hussites, while the events of the Hussite wars were reflected in the poems of the Bautzen manuscript and in a number of chronicles and other prose works.

Influence of the Hussite movement

Among the successors of Hus the most original was Petr Chelčický, who developed his radical social ideas and his pacifism in a series of striking treatises and homilies. It was from Chelčický's ideas that the Bohemian Brethren drew their inspiration: this sect, prototype of the Moravian Church, was to be one of the most important sources of vernacular Czech literature in the next two centuries.

Polish. As in other European countries, Latin was at first the only literary language of Poland, and early writings included saints' lives and annals and chronicles written by monks and priests. Two of them, the first and the last, are most important: the *Chronicon,* compiled c. 1115 by an anonymous Benedictine of French origin known as Gallus, and the *Annales seu cronicae inclyti Regni Poloniae*, written *c.* 1480 by Bishop Jan Długosz. The two works introduced Polish history and literature into the culture of Europe. Use of the vernacular was allowed by the church where Latin could not meet particular needs—in prayers, sermons, and songs; and the oldest surviving text of poetry in Polish is traditionally but erroneously considered to be a song in honour of the Virgin Mary, "Bogurodzica," in which language and rhythm were used with high artistic craftsmanship. Preaching in Polish became established toward the end of the 13th century; the earliest known example of Polish prose, the *Kazania świetokrzyskie* ("Sermons of the Holy Cross"), dating from the end of the 13th or the beginning of the 14th century, was discovered in 1890. Part of a translation of the Bible, made in 1455 by Jdrzej of Jaszowice for Queen Sofia, widow of Władysław II Jagiełlo, has also survived.

Secular works began to appear toward the end of the 15th century. There was a satire on the Pope (c. 1449) by Jędrzej Galka, a follower of reformers John Wycliffe and Jan Hus, and a high literary standard was achieved in a morality verse dialogue, *Rozmowu mistrza Polikarpa ze Snzierciq* ("Dialogue Between Master Polycarp and Death"). The style of the medieval period lasted late in Poland. Marcin Bielski represented this late medievalism in his *Kronika wszystkiego świata* (1551; "Chronicle of the World"), the first general history in Polish.

The best examples of Polish literature imply a continuous literary tradition. Although its themes were those of the common European heritage, Polish writing of this period could be intensely personal within its anonymous framework. The groundwork for a leap into the ranks of major literatures was laid during the reign of Casimir III the Great, who in 1364 founded the University of Cracow.

Hungarian. The first written traces of the Hungarian language consisted of Hungarian words embedded in the Latin text of legal or ecclesiastical documents. The 13th century saw many translations from Latin, only one of which has come down, a free version of a poem by Godefroy de Breteuil. The 14th century also produced translations of the legends of St. Margaret and of St. Francis of Assisi. The 15th century saw the first translations from the Bible. A great part of the vocabulary, created for the purpose, is still in use. Other translations included the first Hungarian drama, "On Three Christian Virgins," and a translation of the Song of Solomon.

Hungarian literature was not entirely religious. The existence of a history of the Trojan war and of a Hungarian version of the Alexander romance can be inferred from South Slavonic translations. In the 14th century, conditions for the development of secular literature greatly improved, as more laymen went to study at universities abroad and on their return tried to introduce the literary forms they had learned.

Finnish. From the mid-12th century until 1809, Finland belonged to Sweden, and Swedish remained the language of the upper classes until the end of the 19th century, when use of Finnish as a cultural medium increased. The oldest literature, however, consisted of folk poetry in Finnish, orally transmitted and first recorded in the late 18th century. This was rich, with a wide range of forms. The metre, a trochaic four-stress line, was well suited to Finnish and to the memorizing of long passages. Most of the poems were epic sequences or short songs. The epics centred upon mythical events or the deeds of a hero. Into this category fell an early poem about the creation of the world and those on the quest of the Sampo, the adventures of Lemrninkainen, and the tragic legend of Kullervo. In many, the central character is Väinämöinen, regarded by some as mythical, by others as historical. These stories also formed part of the national epic, the *Kalevala.* The lyrics, many by women, dealt with everyday griefs and joys; many were love songs or described the moods of a solitary soul, and the setting was rural—a landscape of forests and lakes, with glimpses of a village community.

Finnish folk poetry

In oral transmission many poems have changed or become confused with one another. The oldest were mythological, dating from pagan times (lasting in remoter areas till the 12th century), and may be thousands of years old. Many were of medieval origin and contained datable events and persons. Systematic collection of folk poetry began in the second half of the 18th century, and the best known collector, Elias Lonnrot, concluded that the epic poems could be composed as a continuous folk epic. He joined a number together with connective material of his own and imposed a unifying "plot": the result was the *Kalevala* (final form, 1849), which was based on folk material but in structure was Lönnrot's creation. Lonnrot also published a selection of the lyrics, *Kanteletar,* and their influence and that of the *Kalevala* as a whole on the Finnish national consciousness, art, and culture has been immense.

III. The Renaissance

CHARACTERISTICS OF THE AGE

The term Renaissance (literally meaning Rebirth) is given to the historical period in Europe that succeeded the Middle Ages. The awakening of a new spirit of intellectual and artistic inquiry, which was the dominant feature of this political, religious, and philosophical phenomenon, was essentially a revival of the spirit of ancient Greece and Rome; in literature this meant a new interest in and analysis of the great classical writers. Scholars searched for and translated "lost" ancient texts, whose dissemination was much helped by developments in printing in Europe from about 1450.

The characteristic features of the age gathered under the label Renaissance are generally agreed to be those that are most evident in the greatest works of art and literature, which reached a level unattained in any previous period. What these characteristics were in individual authors or works is harder to define than the sources of the themes and concerns reflected in them. Of these characteristics, the three most important were undoubtedly, first, the new interest in learning, mirrored by the classical scholars known as Humanists and instrumental in providing suitable classical models for the new writers; second, the new form of Christianity, initiated by the Protestant Reformation led by Martin Luther, which drew men's attention to the individual and his inner experiences and stimulated a response in Catholic countries summarized by the term Counter-Reformation: third, the voyages of the great explorers that culminated in Christopher Columbus' discovery of America in 1492 and had far-reaching consequences on the countries that developed overseas empires, as well as on the imagina-

Humanism, Protestantism, and exploration tions and consciences of the most gifted writers of the day.

To these may be added many other factors, such as the developments in science and astronomy and the political condition of Italy in the late 15th century. The new freedom and spirit of inquiry in the Italian city-states had been a factor in encouraging the great precursors of the Renaissance in Italy, Dante, Petrarch, and Boccaccio. The flowering of the Renaissance in France appeared both in the poetry of the poets making up the group known as the Pléiade and in the reflective essays of Michel de Montaigne, while Spain at this time produced its greatest novelist, Miguel de Cervantes. Another figure who stood out above his contemporaries was the Portuguese epic poet Luis Camões, while drama flourished both in Spain and Portugal, being represented at its best by Lope de Vega and Gil Vicente. In England, too, drama dominated the age, a blend of Renaissance learning and native tradition lending extraordinary vitality to works of Christopher Marlowe, Ben Jonson, John Webster, and others, while Shakespeare, England's greatest dramatic and poetic talent, massively spanned the end of the 16th century and the beginning of the 17th. The impetus of the Renaissance carried well into the 17th century, when John Milton reflected a spirit of Christian Humanism to be linked with Desiderius Erasmus. The name Erasmus was perhaps the one most often linked with Humanism, the spirit of critical enquiry, regard for classical learning, intolerance of superstition, and high respect for man as God's most intricate creation.

An aspect of the influence of the Protestant Reformation in literature was the number of great translations of the Bible into vernacular languages during this period, setting new standards for prose writing. In eastern Europe the 16th century was remarkable as a golden age of both Hungarian literature, as represented by a passionate cavalier poet, Balint Balassi, and Polish literature, where the father of Polish writing, Mikołaj Rej, was quickly joined by many rivals and imitators, especially the satiric, lyric, and tragic poet Jan Kochanowski.

The following account has kept in view the importance of national, political, and language divisions while at the same time following the general "spread" of the Renaissance from Italy outward and drawing comparisons and parallels demanded by political or linguistic proximity, as, for example, in the case of Spain and Portugal.

Beginning with an account of the "international language," Latin, the Renaissance's inheritance from the Middle Ages, which was so crucial to a dissemination of the ideas of men such as Erasmus, the account then treats the flowering of vernacular literatures, beginning with the Romance languages of Italian, French, Spanish, and Portuguese and following with the Germanic (starting with English) and Scandinavian groups and then finally with the literatures of eastern Europe.

RENAISSANCE LATIN LITERATURE

The term Renaissance Latin is associated, for 14th-century Italy, mainly with Dante, Petrarch, and Boccaccio, though mention should also be made of the Florentine historian Leonardo Bruni and the Humanist scholars Albertino Mussato, Coluccio Salutati, and Aeneas Silvius Piccolomini (Pope Pius II). In verse there was a general return to classical models and elegance, while in prose Latin was still a necessary medium for the abundant Humanistic, scientific, philosophical, and religious literature that was a mark of the new age.

In Italy there were three main centres of learning and literature in the 15th and 16th centuries: Florence, Rome, and Naples. Each of these centres had its own circle of writers and scholars. The Florentine group was chiefly noted for the Platonist philosophers Poggio Bracciolini, Marsilio Ficino, Giovanni Pico della Mirandola, and a poet and scholar, Angelo Poliziano. Rome was the centre for a grammarian, Pietro Bembo, and for Marco Vida, author of a Latin epic on the redemption, while Naples was the home of many poets and scholars, notably Giovanni Pontano, Jacopo Sannazzaro, Lorenzo Valla, and Girolamo Fracastoro.

Germany and the Low Countries also made a large contribution in prose and verse to Latin literature in the 15th and 16th centuries. Many Humanists owed their early education to the Brethren of the Common Life, a Dutch Christian community that laid great emphasis on the classics. Among these was Erasmus, the greatest figure of the northern Renaissance. Bred in the rhetorical tradition of literary Humanism, he had little interest in the scientific premonitions of the age. As an editor and expositor of classical texts and the writings of the Church Fathers, as a commentator on the ecclesiastical conflicts of his time, and as a scholar, wit, and satirist, he was unsurpassed by any Humanist in northern Europe. A German abbot, Johannes Trithemius, was a historian. and scholar with an immense range of interests and knowledge; Conradus Celtis was conspicuous as a Humanist and poet, while Petrus Lotichius wrote elegant verse.

Influence of the Brethren of the Common Life

Spanish Humanism was best seen in the scholar and friend of Erasmus, Juan Vives, while in England the statesman and scholar Sir Thomas More was the outstanding figure. Polydore Virgil, an Italian, brought the new methods of historical writing into England, though a poet and historian, Tito Livio Frulovisi, had written a life of Henry V that influenced later English writers. Among many Latin poets should be mentioned George Buchanan and John Barclay, both Scots. The strong English tradition of classical verse composition in the schools was shown in the Latin poems of such 17th-century poets as John Milton, Henry Vaughan, Richard Crashaw, and Abraham Cowley.

In France, where, as in England, the Renaissance came late, some members of the group of writers known as the Pléiade wrote Latin verse. Despite the eventual triumph of the French vernacular, Latin poems continued to be written, and several hymns composed in classical forms were included in church services in the 17th and 18th centuries.

Until the early 18th century, Latin was recognized as the best medium for historical and scientific work if it were to reach a European audience. For this reason Marsilio Ficino and Pico della Mirandola, Erasmus and More, and, later, the scientists and philosophers Francis Bacon, Hugo Grotius, René Descartes, Benedict de Spinoza, and Sir Isaac Newton used what was still an international language.

The Humanists were the first to see the literature of a people in its totality as a cultural achievement. They were not content merely to write classical Latin. Desiring to equal the cultural triumphs of antiquity, they attempted every classical genre, thus giving imitation a range it had not had before. And this was also an age when the national literatures of Europe were attaining their maturity so that imitation was no longer limited to Latin. It was practiced in national languages and with striking success.

By 1550 the Humanist tradition, embodied in the rhetorical works of Erasmus, had a firm grip on education, but Latin, on which that education rested, was no longer the unique medium of education, and the vernaculars had absorbed all they could throngh imitation of classical models. A new approach was required. As early as the end of the 15th century the Dutch Humanist Rodolphus Agricola recommended that writers should develop their subjects with reference to species, causes, effects, similarities, and opposites and suggested 30 such headings, analogous to the "topics" of classical rhetoric. In the 16th century in France Petrus Ramus (Pierre de la Ramée) popularized Agricola's ideas and had an enormous influence. Ramus detached invention and disposition from rhetoric, leaving it only with style and delivery. Rhetoric was thus severely limited in scope by losing all concern for content.

Petrus Ramus and the revival of rhetoric

ITALIAN LITERATURE OF THE RENAISSANCE

The age of Humanism. The European Renaissance had really begun in 14th-century Italy with Petrarch and Boccaccio. The 15th century, devoid as it was of major poetic works, was nevertheless of very great importance because it was the century in which a new vision of human life, embracing a different conception of man and life as well as more modern principles of ethics and politics, gradually found its expression. This was the result, on the one hand, of the rediscovery of classical antiquity, and, on the other, of political conditions quite different from those of previous centuries. With regard to the second point, nearly all Italian princes competed with each other in the 15th century to promote culture by patronizing research, offering hospitality and financial support to literary men of the time, and founding libraries. As a consequence, their courts became centres of research and discussion, thus making possible the great cultural revival of the period. The most notable courts were that of Florence, under Lorenzo de' Medici, "the Magnificent"; that of Naples, under the Aragonese kings; that of Milan, first under the Visconti and later the Sforza family; and finally the papal court at Rome, which gave protection and support to a large number of Italian and Byzantine scholars. As for the first point, the search for lost manuscripts of ancient authors, begun in the second half of the previous century, led to an extraordinary revival of interest in classical antiquity: in particular, much research was devoted to Plato and Greek philosophy in general, a fact that was to have profound influence on the thinking of the Renaissance as a whole.

Italian Humanist view of man

By and large, the new culture of the 15th century was a revaluation of man. Humanism opposed the medieval view of man as a being with relatively little value and extolled him as the centre of the universe, the power of his soul as linking temporal and spiritual, and earthly life as a realm in which the soul applies its powers. These concepts, which mainly resulted from the new interest in Plato, were the subject of many treatises, the most important of which were Giannozzo Manetti's *De dignitate et excellentia hominis* and Pico della Mirandola's *Oratio de hominis dignitate.* The Humanist vision evolved during this period condemned many religious tenets of the Middle Ages still widely prevalent: monastic ideals, for example, were attacked by Leonardo Bruni, Lorenzo Valla, and Poggio Bracciolini. Forthright though these attacks were, Humanism was not essentially anti-Christian, for it generally remained faithful to Christian beliefs, and the papal court itself regarded Humanism as a force to be assimilated rather than defeated.

In the first half of the century Humanists, with their enthusiasm for Latin and Greek literature, had a disdain for the Italian vernacular. Their poetic production, inspired by classical models and written mostly in Latin or occasionally Greek, was abundant but of little value. Writing in a dead language and closely following a culture to which they had enslaved themselves, they rarely showed originality as poets. Among the few notable exceptions are Giovanni Pontano, Michele Marullo Tarcaniota, Angelo Ambrogini Poliziano (Politian), and Jacopo Sannazzaro. These poets sometimes succeeded in creating sincere poetry in which the conventional themes of 15th-century lyrics were expressed with new, original intimacy and fervour.

The rise of vernacular literature. Toward the middle of the 15th century Italian began to oust Latin as the literary language. The Certame Coronario, a public poetry competition held in Florence in 1441 with the intention of proving that the spoken language was in no way inferior to Latin, marked a definite change. In the second half of the century there were a number of works of merit inspired either by Carolingian legends or by the new Humanist culture.

The chivalrous epic of Carolingian legends, which had degenerated into clichés, was given a new lease on life by two poets of very different temperament and education: Matteo Maria Boiardo, whose *Orlando innamorato* (1483; "Orlando in Love") reflected past chivalrous ideals as well as contemporary standards of conduct and popular passions; and Luigi Pulci, whose *Morgante,* published before 1480, was pervaded by a new bourgeois and popular morality.

The new ideals of the Humanists were most complete in Angelo Ambrogini Poliziano (Politian), Jacopo Sannazzaro, and Leon Battista Alberti, three outstanding figures who combined a wide knowledge of classical antiquity with a personal and often profound inspiration. Politian's most important Italian work was *Stanze per la giostra* (1475–78; "Stanzas for the Joust"), which created a mythical world in which concepts of classical origin were relived in a new way. The same could be said of Sannazzaro's *Arcadia* (1504), a largely autobiographical work in verse and prose that remained widely influential up to the 18th century. A more balanced view of contemporary reality was given in Alberti's literary works, which presented a gloomy picture of human life, dominated by man's wickedness and the whims of fortune. As for Lorenzo de' Medici, patron of many men of letters, he himself had a vast poetic output, though this is more notable for documentary than for literary value.

Pietro Bembo's view of language

Pietro Bembo of Venice published his *Prose della volgar lingua* (one of the first grammars of Italian) in 1525. In this work Bembo demanded an Italian literary language based on 14th-century Tuscan models, particularly Petrarch and Boccaccio. He was opposed by those who thought that a literary language should be based on existing linguistic developments, particularly by Gian Giorgio Trissino, who developed Dante's theories on Italian as a literary language. In practice the problem was both linguistic and stylistic, and there were in the first half of the 16th century a great number of other contributors to the question, but Bembo's theories did finally triumph in the second part of the century. This was due to a large extent to the activities of the Florentine Accademia della Crusca, and this more scientific approach to the language question resulted in the academy's first edition of an Italian dictionary in 1612.

During the first decades of the 16th century, treatises on poetry were still composed according to Humanist ideas and the teachings of the Roman Augustan poet Horace. It was only after 1536, when the original classical Greek text of Aristotle's *Poetics* was first published, that a gradual development became apparent in aesthetic theory. The traditional principle of imitation was now better analyzed, emphasis being given to the imitation of classical authors rather than to that of nature. The three unities of tragedy (time, space, action) were among the rhetorical rules then re-established. The classical conception of poetry as a product of imagination supported by reason was at the basis of 16th-century rhetoric, and it was this conception of poetry, revived by Italian literature, that triumphed in France, Spain, and England during the following century.

Political and moral literature. Niccolò Machiavelli's works reflected Renaissance thought in its most original aspects, particularly in the objective analysis of human nature. Machiavelli has been described as the founder of a new political science: politics divorced from ethics. His own political experience was at the basis of his ideas, which he developed according to such general principles as the concepts of *virtù* ("power") and *fortuna* ("chance"). He considered *virtù* to be power with a practical aim that should struggle against *fortuna,* which represented the forces of violence and irresponsibility. His famous treatise *Il principe (The Prince),* composed in 1513, revealed the author's prophetic attitude, based on observation of contemporary political affairs. Its description of a model ruler became a code for the wielding of absolute power throughout Europe for two centuries. Machiavelli's *Discorsi . . . sopra la prima deca di Tito Livio* (*c.* 1513–21: "Discourses on the First Decade of Titus Livius"), showed the same realistic attitude: public utility was placed above all other considerations, and political virtue was distinguished from moral virtue. His seven books on *L'arte della guerra* (1519–20; *The Arte of Warre,* 1560), concerning the creation of a modern army, were more technical, while historical works, including the *Zstorie fiorentine* (1520–25, "Florentine History"), exemplified theories expounded in his treatises. Machiavelli also holds a place in the history of literature, above all for *La Mandragola* (1518), one of the outstanding comedies of the century.

The Prince

Although more of an individualist and pragmatist than Machiavelli, Francesco Guicciardini was the only 16th-century historian who could be placed within the framework of the political theories he constructed. He drew attention to the self-interest of those involved in political actions and made Machiavelli's theories appear idealistic by contrast. One of Guicciardini's main works, the ***Ricordi*** (1528, 1530; "Remarks"), was one of the most lucid and original literary works of the century. Among other historical works of the period an autobiographical ***Vita*** ("Life") of the sculptor and goldsmith Benvenuto Cellini was interesting for its spontaneous style and use of popular Florentine language.

Castiglione's *Il cortegiano*

The highest moral aspirations of the Renaissance are expressed in Baldassare Castiglione's ***Il cortegiano*** (published 1528; ***The Courtyer,*** 1561), which deals with the perfect courtier, the noble lady, and the relationship between courtier and prince. It became one of the most influential books of the century. Giovanni della Casa was the author of another famous treatise, the ***Galateo*** (*c.* 1551–54; Eng. trans., 1576), a book on manners in which the author's witty mind and the refinement of contemporary Italian society found full expression. The life of the period was also vividly reflected in the work of Pietro Aretino, who was called "the scourge of princes" by Ariosto. His *Ragionamenti* (1534–36; "Discussions") were written in a spontaneous style and showed a sensuous and unscrupulous nature.

Poetry. Lyric poetry in the 16th century was dominated by the model of Petrarch mainly because of the acceptance of the Renaissance theory of imitation and the teaching of Bembo. Almost all the principal writers of the century wrote lyric poems in the manner of Petrarch. Some originality was to be found in Della Casa's poems, and Galeazzo di Tarsia stood out from contemporary poets by virtue of a vigorous style.

The tradition of autobiographical, humorous, and satirical verse was kept alive during the 16th century, when it reached some real stature with Francesco Berni, whose burlesque poems, mostly dealing with indecent or trivial subjects, showed his stylistic skill. Didactic poetry, already cultivated by Humanist writers, was also continued during this period, chiefly by Giovanni Rucellai, who recast in ***Le api*** ("The Bees") the fourth book of the Roman poet Virgil's ***Georgics,*** and Luigi Alamanni in six books about rustic life called ***Della coltivazione.***

The most refined expression of the classical taste of the Renaissance was to be found in Ludovico Ariosto's ***Orlando furioso*** (first version published 1516; Eng. trans., 1591), which embodied many episodes derived from popular medieval and early Renaissance epics; but the poem's unique qualities derived from Ariosto's sustained inspiration and technique and detached ironical attitude to his characters. The ***Orlando furioso*** was the most perfect expression of the literary tendencies of the Italian Renaissance at this time, and it exercised enormous influence on later European Renaissance literature. Ariosto also composed comedies that, by introducing imitation of Latin comedy, marked the beginning of Renaissance drama in the vernacular.

There were also attempts to renew the epic by submitting the tradition of chivalry to Aristotle's rules of composition. Gian Giorgio Trissino, a theorist on language, wrote according to the strictest Aristotelian rules, while Luigi Alamanni tried to focus the narrative on a single character in ***Girone il cortese*** ("Girone the Courteous") and ***Avarchide*** (an imitation of the ***Iliad*** of Homer). Giambattista Giraldi, more famous as a tragic playwright, was a literary theorist who tried to apply his theories to his own poem ***Ercole.***

Fidenziano and macaronic poetry

Two burlesque medley forms of verse were invented during the century. ***Fidenziano*** poetry derives its name from a work by Camillo Scroffa, a poet who wrote in a combination of Latin words and Italian syntax. Macaronic poetry, on the other hand, is a term given to verse consisting of Italian words used according to Latin syntax. Teofilo Folengo, a Benedictine monk, was the best representative of macaronic literature, and his masterpiece was a poem in 20 books called *Baldus* (first version 1517). A tendency to parody, ridiculing the excesses of Humanist literature, was present in both the *fidenziano* and macaronic verse.

Torquato Tasso was the last great poet of the Italian Renaissance and one of the greatest of Italian literature. In his epic ***Gerusalemme liberata*** (completed 1575, first published 1581; "Jerusalem Delivered") he summed up a literary tradition typical of the Renaissance: the classical epic renewed according to the spiritual interests of his own time. His work shows a conflict between a desire to express himself according to classical ideals and a tendency to moralize. Some of Tasso's works, however, reflected his spontaneous inspiration. *L'Aminta* (1573), a joyous and uninhibited drama, was the best example of Tasso's youthful poetry and belonged to the new literary genre of pastoral (dealing with idealized rural life). ***Gerusalemme liberata*** was the result of a balance in the poet's conflicting aspirations: a Christian subject dealt with in a classical way. In the subsequent ***Gerusalemme*** *conquistata* (1593; "Jerusalem Vanquished"), Tasso recast his poem according to strict Aristotelian rules and the ideals of the Catholic Church's reaction against the Protestant Reformation, known as the Counter-Reformation. Tasso's conflict had ended in the victory of the moralistic principle: poetically the new poem was a failure. Tasso also wrote shorter verses throughout his life, including religious poems, and his prose works show a style no longer exclusively dominated by classical models.

Drama. Trissino's ***Sofonisba*** (written 1514–15, first produced 1562) was the first regular tragedy of Italian vernacular literature; its structure derived from Greek models, but its poetic qualities were somewhat mediocre. Toward the middle of the 16th century Giambattista Giraldi (Cinzio) reacted against imitation of Greek drama by proposing the Roman tragedian Seneca as a new model, and in nine tragedies and tragicomedies—written between 1541 and 1549—he showed some independence from Aristotelian rules. He greatly influenced European drama, particularly the English of the Elizabethan period.

Italian comedies

Italian comedies of the century, inspired by Latin models, possessed greater artistic value than tragedies, and they reflected contemporary life more fully: they could be considered as the starting point for modern European drama. To the comedies of Ariosto and Machiavelli should be added a lively play, the ***Calandria*** (first performed 1513), by Cardinal Bernardo Dovizi da Bibbiena and five equally amusing comedies written by Pietro Aretino. Giordano Bruno, a great Italian philosopher who wrote dialogues in Italian on his new cosmology and anti-Humanist ideas, wrote also a comedy, *Il candelaio* (1582; "The Candlemaker").

Narrative. The classicist trend established by Pietro Bembo also affected narrative literature, for which the obvious model was Boccaccio's ***Decameron.*** Originality and liveliness of expression were to be found in the 22 stories called ***Le cene*** ("The Suppers"), of the Florentine apothecary Anton Francesco Grazzini. The worldly monk Agnolo Firenzuola produced several stories, including the fable ***Asino d'oro,*** a free version of Apuleius' ***Golden Ass.*** A cleric and short-story writer, Matteo Bandello, started a new trend in 16th-century narrative with 214 stories that were rich in dramatic and romantic elements while not aiming at classical dignity. This trend was partly followed also by Giraldi Cinzio in his collection of 112 stories called *Ecatommiti* (1565).

FRENCH LITERATURE OF THE RENAISSANCE

The 16th century was one of France's richest and most varied literary periods. Three names—François Rabelais, Pierre de Ronsard, and Michel de Montaigne—tower above a host of others. It was the century that saw the birth of the Renaissance in France and the passing of the Middle Ages; it was also a period when the movement for religious reform, at first closely connected with the rebirth in learning, finally came into conflict with it.

Drama. The break with the Middle Ages was most marked in drama. The religious mysteries and miracle plays (relatively simple presentations of well-known religious themes) began to die out by the middle of the 16th century, largely because of clerical opposition. Some medieval forms of drama continued to be performed until the end of the century, but in Paris, at least, the Renaissance swept away all other vestiges of the drama of the Middle Ages. The change came from Italy, mainly because of the importation of Italian translations of the classics. New translations were made of other works, including the *Electra* of Sophocles from the classical Greek, and French Humanists hastened to write original tragedies on the classical model. Etienne Jodelle wrote the first modern tragedy, *Cléopâtre captive* (performed 1552 or 1553), and the first modern comedy, *Eugène* (1552). The chief tragic playwright of the period, Robert Garnier, chose his subjects from classical, sacred, and medieval literature. His most famous play, *Les Juives* (1583; "The Jews"), was a biblical tragedy; and he wrote the first French tragicomedy, *Bradamante,* based in part on Ariosto's *Orlando furioso.* Garnier's best plays were lyrical, eloquent, and quite moving. The chief comic playwright of the century was Pierre de Larivey, whose best play was, perhaps, *Les Esprits* (1579; "The Ghosts").

Poetry. There was an enormous variety and wealth of poetry in this period. The first important poet was Clément Marot, in whom were to be found forms and themes typical of both the Middle Ages and the Renaissance. Marot composed some allegorical and mythical poetry. His poems were remarkable for their easy and polished style, their light and graceful wit, and an elegance new in French literature. His patroness, Margaret of Angoulême, was a poet who wrote in a religious as well as an amorous vein, her chief work being *Les Marguerites de la marguerite des princesses* (1547; "The Pearls of the Pearl of Princesses").

At Lyons there flourished a group of able poets, chief of whom was Maurice Scève, famous for a long series of love verses dedicated to *De'lie* (1544). A most talented poet was Louise Labé, whose three elegies and 24 sonnets told a passionate love story in accents of purest poetry.

Ronsard and the Pléiade

The principal poet of the French Renaissance was Pierre de Ronsard, whose chief studies were carried out under a Humanist and poet, Jean Dorat. These two, together with five other writers—Joachim du Bellay, Jean-Antoine de Baïf, Rémy Belleau, Pontus de Tyard, and Étienne Jodelle—made up the famous group known as La Pléiade, from the seven stars of the Pleiades star cluster. Their object was to create a French literature to rival that of the Renaissance Italy, first, by bringing the French language to a level with the classical tongues and, second, by reproducing in French regular tragedy such classical verse forms as the Pindaric and Horatian odes and the Petrarchan sonnet. The Pléiade's manifesto was written by du Bellay and called the *Dkfense et illustration de la langue Françoise* (1549; *The Defence and Illustration of the French Language,* 1939). The men of the Pléiade achieved much that they aimed to achieve, and they often produced important results. They occasionally went too far and thus were responsible, to some extent, for the reaction led by François de Malherbe at the beginning of the 17th century. Ronsard wrote an enormous amount—sonnets, odes, elegies, philosophical and reflective poems, and even the beginning of an epic. His odes and sonnet sequences were particularly fine, and he produced excellent poetry from the late 1540s to the early 1580s. Next to him in stature (but far less in range) came du Bellay, whose sonnets were written with consummate skill, especially those in the sequence *Les Regrets* (1558). Belleau was polished, light, and gracious and wrote a great deal, notably a vast *Bergerie* (pastoral sequence).

The poets of the Pléiade, in particular Ronsard, had many imitators. Jacques Peletier du Mans championed the theories of the Pléiade and himself wrote several worthy lyric poems after the manner of Ronsard. Among many who were influenced by Ronsard, Jacques Tahureau wrote *Sonnets de l'admirée* (1554; "Sonnets of the Admired") and Olivier de Magny *Les Soupirs* (1557; "Sighs"); Jean Vauquelin de La Fresnaye wrote several graceful poems, including *ldillies* (1605; "Idylls") and a fine verse, *Art poktique;* and Philippe Desportes, a favourite court poet and seen as a successor to Ronsard by his contemporaries, occasionally produced a really good poem such as the sonnet about Icarus.

The second half of the 16th century was a period of bitter civil strife, and many poets gave vent to vehement political and religious feeling. Perhaps the most notable was Théodore Agrippa d'Aubigné, whose main work was *Les Tragiques.* D'Aubigné was both a religious (Calvinist) and a violently satirical poet. Guillaume du Bartas, also a Protestant, achieved great fame for his epic work, *La Semaine* (1578; "The Week"), in which he set out to describe the whole process and product of the creation. Jean de Sponde, rediscovered by 20th-century scholars, was a Protestant who became converted to Catholicism. His *Essay de quelques poèmes chrestiens* (1588; "Sample of Some Christian Poems") was written in an impassioned and ornamental style. Jean-Baptiste Chassignet, who wrote a long sonnet sequence containing several remarkable similes and metaphors, was another modern rediscovery. Among other minor talents, Jacques du Perron and Gabrielle de Coignard were religious poets who both wrote several good poems.

The contributions of Rabelais

Prose. In quantity, 16th-century prose fiction was not considerable, but its quality was enhanced by one of the most outstanding French writers of all time—François Rabelais. His great work was written in several parts—*Pantagruel* (1532), *Gargantua* (1535), the *Tiers Livre* (1546; "Third Book"), the *Quart Livre* (1552; "Fourth Book"), and a fifth book—the authenticity of which has been contested—published in 1564 after his death. With immense erudition, representing almost every aspect of contemporary knowledge, with untiring invention, and above all with acute observation and satire, Rabelais reached a depth of insight and poetical imagination that was unique. His work was an exposition of and commentary on the thoughts, feelings, aspirations, and knowledge of a particular time and nation expressed in a most attractive literary form, and with the richest, imaginative language—by a man who was unusual in combining practical knowledge of the world with great power of literary expression.

Other storytellers worthy of note included Margaret of Angoulême, more famous as a poet, whose collection of short stories, the *Heptamkron,* was an imitation of Boccaccio's *Decameron.* By contrast, the tales of Bonaventure des Périers were shorter, simpler, and more homely, showing more wit and less refinement. An original storyteller was Noel du Fail, whose work revealed direct observation of peasants and their way of life.

In history and memoirs the writers of the 16th century hardly compared with those of the Middle Ages. Perhaps the best known is Pierre de Brantôme, a courtier and soldier whose memoirs contained lives of great captains and soldiers, as well as scandalous episodes of the French court. The chief political writer of the period was Jean Bodin, who in the protracted argument between the royalists and the reformers took the royalist side in his *Rkpublique* (1576), arguing in favour of unlimited monarchy. Perhaps the most able writer supporting the opposite view was Étienne de La Boétie, who protested against the monarchical theory in his brilliant essay *Contr'un* (1576; "Against One"). Religion was closely linked to politics at this time, and this was particularly evident in the *Satire Ménippée* (1594), a prose and verse satire written by a number of authors against the Holy League, the Catholic political party.

The religious controversy

Of the many theological works, the great textbook of the French Reformation was Calvin's *Institutes of the Christian Religion* (Latin, 1536; French, 1541). This was the first great composition in argumentative French prose and influenced thought in France and other countries. Even before the *Institutes* was written, a Humanist, Jacques Lefèvre d'Étaples, who first translated the whole Bible into French, and his disciple Guillaume Farel had seen the importance of using French for religious and other prose works. Another follower of Lefèvre, Pierre Viret, helped Calvin and himself wrote a large number of theological and moral dialogues. The pattern of religious polemical writing on both Protestant and Catholic sides continued throughout the century, though some religious writers, such as the Protestant Philippe Duplessis-Mornay, achieved a wider outlook. On its side, the Catholic Church had St. Francis de Sales, bishop of Geneva, whose ***Introduction à la vie dkvote*** (1608; "An Introduction to Devout Life") remains a popular classic. Pierre Charron, finally, should be mentioned for ***Les Trois Vkrities*** (1593; "The Three Truths") and the ***Trois Livres de la sagesse*** (1601; "Three Books of Wisdom"), in which he discussed the evidence for knowledge of the existence and nature of God.

Three outstanding writers of scientific and learned works were Jacques Arnyot, Henri Estienne, and Étienne Pasquier. Amyot, bishop of Auxerre, was best known for his translations of Latin classics, and the essayist Michel de Montaigne declared that his contemporaries owed to Amyot their knowledge of how to write and speak French. His masterpiece is a translation (1559; 1565 ff.) from the ancient Greek of Plutarch's ***Lives,*** which, perhaps more than any other classic, coloured the thoughts and writings of the 16th century. Estienne, a printer and lexicographer, wrote several pungent and eloquent treatises on the French language, while Pasquier was one of the first great French literary critics to write in French, and his *Recherches* ***de la France*** ("French Studies") is a masterpiece of balanced criticism and good French. Among minor writers in this field, the famous surgeon Ambroise Paré achieved literary importance for the use of French in surgical treatises and in his autobiography.

Montaigne's *Essais*

The essayist Michel de Montaigne stands alone. Literature received one of its greatest acquisitions in his famous *Essais* (1580–88; ***Essays).*** The principle of this work was not skepticism but egotism, and, its author being by nature skeptical, this quality of mind appeared naturally rather than intentionally. The ***Essais*** expressed the attitude of mind of the latter part of the century as completely as Rabelais expressed that of the earlier part. Inquiry and protest, ebullience, and imagination run riot had given way to a placid conviction that little of importance remains to be discovered and that it does not greatly matter. Though neither business nor pleasure was decried, both were regarded as useful pastimes incidental to a man's life rather than with the eager appetite for them characteristic of the Renaissance. The style is remarkable for its absence of pedantry, its rich vocabulary, and picturesque brilliance. It is one of the best loved books in the French language, an infinite source of pleasure, to be read at leisure, to be put down and taken up again at will.

Provencal. During the late 15th century, Provençal literature had waned without fading entirely, and the town of Toulouse never wholly ceased to supply funds of some kind to its College of Rhetoric. Fresh prizes for poetry had been introduced; but in 1513 poems in French were admitted to the competitions, and (after 1679) under Louis XIV these were alone held eligible. In the two centuries that followed a glorious medieval period there was a succession of works in Provençal, chiefly of a didactic and edifying character, that nevertheless served to keep alive some kind of literary tradition.

In the 16th century there were signs of a strong movement toward a Provençal Renaissance. The Gascon Pey de Garros, author of the remarkable ***Eglogues,*** raised his native dialect to the rank of a literary language with translations of the Psalms of David and ***Poesias*** *gasconas* (1567; "Gascon Poems"). At the other end of oc-speaking southern France, in Provençal proper, the movement toward poetry in native dialect began with Louis Bellaud de la Bellaudière.

SPANISH LITERATURE OF THE RENAISSANCE: THE SIGLO DE ORO

The beginning of the "Golden Age." The unification of Spain in 1479 and Columbus' discovery of the New World (1492), following the introduction of printing (1474) and concurrent with the cultural traffic with Italy (Naples had been since 1443 a dependency of Aragon), may be taken as opening the era of the Renaissance in Spain. In this period—known as the Siglo de Oro, or "Golden Age"—Spanish literature reflected the wealth of new experience born of overseas adventure and the detached questioning attitude to the rediscovered norms of classical authority. The early Spanish Humanists included the first grammarians and lexicographers of any Romance tongue. Juan Luis Vives, a figure of European eminence, the brothers Juan and Alfonso de Valdés, and others were followers of Erasmus, whose writings circulated in translation from 1536 and whose influence was seen in the Counter-Reformation figure of St. Ignatius of Loyola, founder of the Society of Jesus, and, later, in a religious writer and poet, Fray Luis de León.

La Celestina

The masterpiece of the early Renaissance is the ***Comedia de Calixto y Melibea*** (1499), a novel in dialogue form published anonymously but attributed to a converted Jew, Fernando de Rojas. The dominant character, a procuress called Celestina, depicted with a realism unsurpassed in Spanish letters, gave this work the title by which it is most commonly known, ***La Celestina.*** The analysis of passion and the dramatic conflict that its pursuit involved were worked out with such psychological intensity as to make this the first masterpiece of Spanish prose and Spain's first realistic novel.

Poetry. Spanish ballads ***(romances)*** link medieval heroic epic to 20th-century poetry and drama; they lay at the heart of the national consciousness; and their expansion and capacity for survival, from Salonika to Chile and from the Low Countries to North Africa, reflect the far-flung boundaries of Spain's prestige in its age of greatness. The earliest datable ***romances*** (mid-15th century) treated of frontier incidents or lyrical themes. The ballads on medieval heroic themes had importance because they formed everyman's source book on national history and character. Traditional ballads were collected in the Antwerp ***Cancionero de romances*** ("Ballad Songbook") and in the ***Silva*** *de* ***varios romances*** ("Miscellany of Various Ballads"), both ***c.*** 1550, and thereafter repeatedly. Soon the form (octosyllabic, alternate lines having a single assonance throughout) was exploited for lyrical purposes by the most famous poets of the age, and it has remained the chosen medium for popular narrative verse.

The rebirth of the lyric

The earlier attempt to Italianize Spanish poetry had failed because Spanish language and verse techniques were still incapable of sustaining the burden. The Catalan Juan Boscán Almogárer, reintroducing Italian metres, prepared the way for a much greater poet, Garcilaso de la Vega, with whom the lyric was reborn. To his mastery of poetic technique derived from medieval and classical poets he added an intense personal note in the use of characteristic Renaissance themes. His short poems, elegies, and sonnets largely determined the course of lyric poetry throughout the Siglo de Oro.

Fray Luis de León, adopting some of Garcilaso's verse techniques, typified the "Salamanca school," with its emphasis on content rather than form. The poet and critic Fernando de Herrera headed a contrasting school of Seville, which, deriving equally from Garcilaso, was concerned rather with subtleties of refined sentiment; in a quartet of remarkable odes he gave vibrant expression to topical heroic themes. A defense of the short native metres, associated chiefly with the satiric poet Cristbbal de Castillejo, was reinforced by the aforementioned traditional ballad collections ***(romanceros)*** and by the evolving drama.

For epic poetry the models were Ariosto and Tasso, but the themes and heroes were those of overseas conquest and expansion or defense of the empire and the faith. Alonso de Ercilla y Zúñiga came nearest to real achievement with his ***Araucana*** (published 1569–90), telling of native resistance to the Spanish conquest of Chile. Other typical examples of the attempt at epic include Lope de Vega's ***Dragontea*** (1598), a verse history of Sir Francis Drake's last voyage and death, and ***Corona*** *trágica* (1627; "The Tragic Crown"), on Mary, Queen of Scots.

Early drama. The origins of drama in Spain are to be sought, as elsewhere, in the church. The ***Auto de*** *los* ***reyes*** *magos* ("Play of the Three Wise Kings"), an incomplete play of an Epiphany cycle, is the only surviving text of medieval Spanish drama. The characterization of the Magi and of Herod and his advisers was realistic, and the medley of metres foreshadowed one aspect of later development of the drama in Spain.

A reference in King Alfonso X's legal code suggested also the existence of some form of secular drama in the 13th century, but no texts have survived. These *juegos* (short satiric entertainments given by travelling players) were the forerunners of the short plays and interludes—*pasos, entremeses, sainetes—that* formed one of Spain's main contributions to dramatic genres.

Juan del Encina marked the emancipation of the drama from ecclesiastical ties by performances for a noble patron. His *Cancionero* (1496; "Songbook") contains pastoral–religious dramatic dialogues in a rustic dialect, but he soon turned to secular themes or vivid farce. Some of his plays showed a conception of drama changed by his long stay in Italy, in which native medievalism was transformed into bold Renaissance experimentation.

Encina's Portuguese disciple Gil Vicente, a court poet at Lisbon who wrote also in Castilian, showed a great advance in naturalness of dialogue, acuteness of observation, and sense of situation. His weakness lay in plot, his distinction in a lyricism expressed in some of the finest Spanish poetry of the period.

The emergence of the drama from court to marketplace and the creation of a public were largely the work of Lope de Rueda, who toured Spain with his modest troupe and a repertoire of his own composition. His four prose comedies were clumsy, but the ten *pasos* (comic interludes between the acts of longer plays) showed his dramatic merits, and he had the distinction of fathering the one-act play, which may be regarded as the most vital and popular dramatic form in Spain.

The first dramatist to realize the possibilities of the ballads in the theatre was Juan de la Cueva. His comedies and tragedies were mostly taken from classical antiquity, but in ***Los siete infantes de Lara*** ("The Seven Princes of Lara"), ***El reto de Zamora*** ("The Challenge of Zamora"), and ***La libertad de España por Bernardo del Carpio*** ("The Liberation of Spain by Bernardo del Carpi~"), all published in 1588, he turned to early heroic stories already familiar in ballad; and thus, although not an accomplished dramatist, he helped to found a "national" drama.

Prose. ***Historical writing.*** Prose before the advent of the Counter-Reformation produced some notable dialogues, especially from Alfonso de Valdés in ***Dia'logo de Mercurio y Carón*** (1528; "Dialogue Between Mercury and Charon"). His brother Juan wrote a ***Dia'logo de la lengua*** ("Dialogue About the Language") of great critical value. History continued to be cultivated, patriotism waxing higher as Spain's greatness cast its shadow over Europe; its last flowering was seen in Juan de Mariana's own translation into Spanish (from 1601) of his Latin history of Spain, which marked the triumph of the vernacular for all literary purposes.

Narratives by the explorers

But the landmarks in historical writing came from the New World and showed the transmitting of vital experience into literature with a vividness unknown in Spain. Columbus' letters and accounts of his voyages, the letters and accounts to Charles V of the famous explorer Hernán Cortés, and many other narratives by more humble conquistadores ("conquerors") opened up new horizons to the reader and, in the attempt to capture exotic landscapes in words, enlarged the resources of the language. Most engaging, though not published until 50 years after his death, was the ***Historia verdadera de la conquista de la Nueva España*** (1632; "True History of the Conquest of New Spain"), by an explorer and chronicler, Bernal Díaz del Castillo, who had a phenomenal memory and an infallible eye for detail. Bartolomé de Las Casas, the "Apostle of the Indies," wrote a ***Brevísima relacidn de la destruccidn de las Indias*** (published 1552; "Very Brief Account of the Destruction of the Indies"), criticizing Spanish colonial policy and the ill-treatment of the native population, which gave rise among Spain's enemies to the famous *leyenda* ***negra,*** or black legend.

The novel. Popular taste in the novel was dominated for a century by the progeny of the medieval courtly romance ***Amadís de Gaula.*** These interminable romances kept alive certain ideals of medieval chivalry, but, having lost touch with life at every point, they represented pure escapism and in due course evoked such literary reactions as the pastoral novel. Naturalized from Italy and filled with nostalgia for an arcadian golden age, its shepherds were courtiers and poets who, like the knights-errant of chivalric romance, turned their backs on reality.

A more positive reaction was in the picaresque novel initiated in 1554 with the brief, anonymous ***Lazarillo de Tormes.*** This genre, native to Spain and influential in many other literatures, had as its hero a picaro ("rogue"), essentially an antihero, living by his wits and concerned only to stay alive, who, passing from master to master, depicted life from underneath. Important in guiding fiction back to direct observation of life, the picaresque formula had little to contribute to development of the novel as an art form.

Don Quixote

Miguel de Cervantes, the culminating figure in Spanish literature, produced in ***Don Quixote*** (part *1*, 1605; part 2, 1615) the prototype of the modem novel. Nominally a satire on the moribund chivalric romance, Cervantes' conception allowed the presentation of reality on two levels: the "poetic truth" of Don Quixote and the "historic truth" of his squire, Sancho Panza. In the constant interaction of these rarely compatible attitudes, Cervantes revealed the novel's scope as a philosophical commentary on existence, and in the interplay of the two characters he established psychological realism in contrast with the static characterization of previous fiction. In the *Novelas ejemplares* (1613; "Exemplary Tales") Cervantes particularized his claim to be the first to write *novelas* (short stories in the Italian manner) in Spanish by differentiating between those that interest by the action and those whose merit lies in the mode of telling.

Mystical writings. The great period of Spanish mysticism coincided with the Counter-Reformation. though it had antecedents, particularly in the expatriate Spanish Jew León Hebreo, whose *Dialoghi di amore* (1535; "The Dialogues of Love") exercised a profound influence on 16th-century and later Spanish thought. The literary importance of the mystics derives from the consideration that in trying to transcend limitations of language they were liberating previously untapped resources of expression. In the writings of St. Teresa of Avila, particularly in her autobiography and letters, there were the gifts of a great novelist in embryo. A Humanist and theologian, Fray Luis de León, in his prose as in his poetry, showed passionate devotion, sincerity, and profound feeling for nature in a style of singular purity. St. John of the Cross achieved pre-eminence through three poems expressing in exalted style the experience of mystic union.

The works of Lope de Vega

Later drama. The drama achieved its true splendour in the genius of Lope de Vega. Its manifesto was Lope's own treatise, ***Arte nuevo de hacer comedias en este tiempo*** (1609; "New Art of Writing Plays at This Time"), which rejected the classical and Neoclassical "rules," opted for a blend of comedy and tragedy and metrical variety, and made public opinion the arbiter of good taste.

The new ***comedia*** ("drama") was therefore a "social" drama, ringing changes on the triple foundation of society: respect for crown, church, and the human personality, the last symbolized in the "point of honour" that Lope commended as the best theme of all. This was a matter of convention, "honour" being equivalent more or less to reputation. It was a drama less of character than of action and intrigue, which rarely grasped the essence of tragedy. What the great Spanish playwrights did possess was a remarkable sense of stagecraft and the ability to make the most intricate plot gripping.

Lope, who claimed authorship of more than 1,800 ***comedias,*** towered over his contemporaries. He had an unerring sense of what could move an audience to respond to a reflection on the stage of some of the ingredients of its country's greatness. Through Lope the drama became "national" in the truest sense. The two main categories of his work are the native historical and the ***capa y***

espada ("cloak-and-sword") plays of contemporary manners. Lope ransacked the literary past for heroic themes, chosen to illustrate aspects of the national character or of that social solidarity on which Spain rested.

The cloak-and-sword play, which dominated drama after Lope, was pure diversion, with much use of disguise, falling in and out of love, false alarms about honour, and a duplication of the plot by the humorous lackey and the lady's maid. It gave pleasure by its dexterity, sparkling dialogue, and entanglements in relations between the sexes, and showed the pleasures of the moment in an irresponsible world.

The greatest of Lope's immediate successors was Tirso de Molina, whose ***Burlador de Sevilla*** (1630; "The Trickster of Seville") presented the Don Juan legend on stage for the first time. La ***prudencia en la mujer*** (1634; "Prudence in Woman") figured among the greatest of Spanish historical dramas, as did ***El condenado por desconfiado*** (1635; ***The Doubter Damned,*** 1956) among theological, while Tirso's cloak-and-sword comedies were among the liveliest of the type. Mexican-born Juan Ruiz de Alarcón y Mendoza struck a distinctive note. His 20 plays were sober, studied, and imbued with serious moral purpose, and his ***Verdad sospechosa*** (1634; "The Truth Suspected") inspired the great French dramatist Pierre Comeille's ***Merzteur.*** Comeille's famous ***Cid*** similarly had its source in the conflict between love and honour posed in ***Las mocedades del Cid*** ("The Youthful Exploits of the Cid") by Guillén de Castro y Bellvís. These and many others with more or less distinction and dwindling originality bridged the transition from Lope to Calderón.

PORTUGUESE LITERATURE OF THE RENAISSANCE

The influence of the Aviz dynasty

The 15th century. Under King John (João) I (1385–1433), founder of the new dynasty of Aviz, the Portuguese court became once again a literary centre. The King himself wrote a treatise on hunting. His son Duarte (Edward) collected a rich library of the ancients and of medieval poems and histories and composed a moral treatise, ***Leal Conselheiro*** (*c.* 1430; "Loyal Counselor"), which revealed a conscious stylist. His brother Pedro's ***Tratado da Virtuosa Benfeitoria*** was a version of Seneca's ***De beneficiis.*** But the historical chronicle distinguished the age, with credit to King Edward, who in 1434 created the office of ***Cronista mor do reino,*** or "chief chronicler of the realm," and appointed Fernão Lopes, father of Portuguese historiography and author of chronicles of the first ten kings of Portugal, of which only those of Pedre I, Ferdinand I, and John I survive. Vividness of style combined with serious documentation to produce in Lopes the finest achievement of medieval Portuguese prose.

His successor in office, Gomes Eanes de Zurara, continued the chronicle on a lower level of artistry. His chief work was ***Crdnica*** do ***Descobrimento e Conquista da Guiné*** ("Chronicle of the Discovery and Conquest of Guinea"). Rui de Pina, if likewise not of the stature of Lopes, was free from the rhetorical defects of Zurara, and his chronicles were characterized by a notable frankness.

The *Cancioneiro Geral*

Poetry was cultivated in the mid-15th century after a long eclipse, but much had changed. The dominant influence came now from Spain, and Portuguese poets initiated the long chapter of allegiance to Spain. Apart from the ballads, popular poetry had disappeared along with that of the troubadours. The constable Dom Pedro de Portugal initiated the fashion of writing in Castilian. As one of the first to adopt the new Spanish trend toward allegory and the cult of classical antiquity derived from Italy, his influence on his compatriots was doubly important. His own poems were inspired by deep feeling and much reflection on life, and he was one of almost 200 poets represented in an anthology of poetry, the ***Cancioneiro Geral*** (1516), of the chronicler Garcia de Resende, covering the preceding three-quarters of a century. The main subjects of these 1,000-odd poems, in Portuguese and Castilian, were love, satire, and epigram. Resende was a better poet than most of his contributors.

Gil Vicente and early drama. The emergence of the modern play may be traced in the works of Gil Vicente, court dramatist and the greatest name in the Portuguese theatre. Eleven of his 44 plays were written wholly, and another 17 partly, in Spanish. The trilogy of the ***Barcas*** (1517–19) revealed his dramatic power and a fondness for comic relief; in this lay his strength, and in construction lay his weakness. The phenomenon of a potential national theatre, however, died with its founder, and his real influence was felt in Spain.

The Inquisition, introduced into Portugal in 1536, early declared war on the popular theatre on the charge of grossness. Vicente's own plays, which figured on the Spanish Index of 1559, were reduced in number to 35 and were sadly mutilated in the second edition of 1586; no new edition appeared for another 250 years.

The Renaissance in Portugal. The Renaissance reached Portugal both indirectly through Spain and directly from Italy, with which country there had been close cultural relations through the 15th century. In the following century many famous Humanists took up residence in Portugal. In 1547 John III reformed the University of Coimbra, and distinguished Portuguese teachers returned from abroad to assist the King in his task. At home Portugal produced scholars of note, including André de Resende, author of ***De antiquitatibus Lusitaniae*** (1593; "Of the Antiquities of Portugal"), and Francisco de Holanda, painter, architect, and author of ***Quatro Didlogos da Pintura Antiga*** ("Four Dialogues on Ancient Painting").

Luís de CamBes and *Os Lusíadas*

The Italianate school of poetry and drama. The return in 1526 of the poet Francisco de Sá de Miranda after a six years' stay in Italy initiated a literary reform of far-reaching effect. Like his contemporary Garcilaso de la Vega in Spain, he introduced the new poetic forms of sonnet, canzone, ode, and epistle, and he gave fresh vigour to the national verse forms, mainly through his ***Satires.*** His chief disciple, António Ferreira, a convinced classicist, wrote sonnets superior in form and style. Other poets continued the erudite school, which definitely triumphed with Luis de CamBes. CamBes, for all his eminence in the epic, is, in the opinion of some, greater still as a lyric poet. Here, a profound classical education combined with perfect mastery of his instrument and a lifetime of varied experience to produce in sonnets, eclogues, odes, elegies, and *canções* the greatest poetry in the language. ***Os Lusíadas*** (1572; ***The Lusiads,*** 1952), his epic narrative of Portuguese achievement in the East, is the national poem ***par excellence*** and the greatest of all Renaissance epics after the pattern of the Roman poet Virgil. Of the many who were moved to emulation by its success, none approached CamBes in inspiration or poetic gifts.

In the drama Sá de Miranda and his followers substituted prose for verse. Taking the Roman dramatist Terence as their model, they produced not Portuguese characters but Romano-Italian types. This revived classical comedy was to be short-lived. Sá de Miranda, avowedly to combat the school of Vicente, wrote ***Os Estrangeiros*** ("The Strangers"), the first prose comedy, and ***Vilhalpandos,*** both actions being set in Italy. Antonio Ferreira, a greater dramatist, likewise attempted both kinds: ***O Cioso,*** Italian even to the names of the personages, came nearer to being a comedy of character, but his fame rests chiefly on ***Castro*** (*c.* 1557), which treated one of the most moving tragic themes in the nation's history—the murder of Inês de Castro—by reference to the ancient Greek dramatists Sophocles and Euripides. From the comic playwright Jorge Ferreira de Vasconcelos came a "new invention" of another kind with ***Eufrósina,*** written under the influence of the Spanish novel La ***Celestina.*** This and his other plays, ***Ulissipo*** and ***Aulegrafia,*** resembled novels in dialogue and contained a treasury of popular lore and wise and witty sayings introduced with a moral purpose.

Prose. ***16th-century history.*** Discovery and conquest in Africa, Asia, America, and on the ocean inspired historians as well as poets. Their records gained in vividness what they might have lost in scientific detachment. In the three "Decades" of his ***Asia,*** the chronicler João de Barros told in vigorous language the deeds of compatriots overseas down to 1526. His first "Decade" undoubtedly influenced Camões, and together, one by his prose and the other by his verse, these two fixed the written language. This work, continued by the more critical and inclusive Diogo do Couto, can rank as the noblest historical monument of the century. In his *Soldado* ***Prdtico,*** Diogo do Couto added acute observations on the causes of Portuguese decadence in the East. Ten years of investigation in India underlay the ***Histdria do Descobrimento e Conquista da fndia pelos Portugueses*** (1552–54 and 1561) of the chronicler and notary Fernão Lopes de Castanheda, a work that ranks close to those of Barros and do Couto.

Literature of the Portuguese Empire

From this spate of writing on expansion overseas attention returned, by way of chronicles of the monarchs who presided over the work, to the history of Portugal itself. Damião de Góis, diplomat, traveller, Humanist, and intimate friend of Erasmus, possessed an encyclopaedic mind and was one of the most critical spirits of the age: his ***Crdnica de D. Manuel*** (1566–67) was most valuable where the author's own feelings of experience came into play. Its chief fault was that it—like other chronicles of the time—allowed Eastern affairs to overshadow events at home.

Works of travel abounded, and their authors were often the first Europeans to visit the countries in question. Among the more noteworthy was a much-translated classic, ***Histdria da Vida do Padre Francisco Xavier*** (1600), by Padre João de Lucena. Important both as history and as human documents were the *cartas,* or "letters," written home by Jesuits in China and Japan. An anonymous ***Descobrimento da Frolida*** ("Discovery of Florida") and a ***Tratado Descritivo do Brasil em 1587*** of Gabriel Soares de Sousa were reminders that Portugal was also present and active in the New World to the west. In all this literature of travel the palm is still held for curious interest by the ***Peregrinação*** ("Peregrination"), which Fernão Mendes Pinto, prince of adventurers throughout the East, composed in his old age, and for tragic pathos by the ***Histdria Trágico-marítima,*** a collection of contemporary narratives, told by survivors or based on their accounts, of the more notable disasters that befell Portuguese ships between 1552 and 1604.

The novel and other prose. The poet Bernardim Ribeiro, whose five eclogues introduced pastoral poetry to Portugal, was equally an innovator in the pastoral novel with his ***Livro das Saudades*** (published 1554–57; "Yearnings"), better known by its opening words "Menina e Moça" ("Child and Damsel"). This tale of rustic love and melancholy with chivalric elements transferred themes and emotions previously found only in poetry to a new medium. From it a musician and poet, Jorge de Montemor, drew some part of his inspiration for ***Diana*** (c. 1559), which, written in Spanish, started a fashion subscribed to by Cervantes and Lope de Vega among many others and represented one of the outstanding contributions of Portugal to the neighbouring literature. Both countries shared in the new enthusiasm of the 16th century for the romance of chivalry, in an age when imperial enterprise to east and west was such as to blur the dividing line between fact and the most improbable flights of the imagination. The first work of João de Barros, historian of empire, was a ***Crdnica do Imperador Clarimundo*** (1520), concerning the adventures of a fictitious progenitor of the kings of Portugal. The dramatist Jorge Ferreira de Vasconcelos kept alive memories of the Arthurian cycle with ***Sagramor,*** or ***Memorial . . . da Segunda Távola Redonda*** (1567).

Among moralists and theologians, three ranked as masters of prose style: Frei Heitor Pinto, for his ***Imagem da Vida Cristã*** (1563); Bishop Amador Arrais for ten ***Diálogos*** on religious and other topics; and Frei Tomé de Jesus for a mystic and devotional treatise, ***Trabalhos de Jesus*** (1602–09). The work of scientists included that of a cosmographer and mathematician, Pedro Nunes, and of a botanist, Garcia de Orta, whose ***Coldquios dos Simples e Drogas*** (1563) was the first book to be printed in the East (at Goa in India).

ENGLISH LITERATURE OF THE RENAISSANCE

Effects of Renaissance culture. Humanism was disseminated in England from 1500 on one hand by such followers of the Dutch Humanist Desiderius Erasmus as John Colet, Sir John Cheke, Sir Thomas Elyot, Thomas Wilson in his ***Arte of Rhetorique*** (1553), and Roger Ascham, especially in ***The Scholemaster*** (1570): they put forward the classics both as the true staple of a liberal education and as the pattern for vernacular prose, an ideal furthered by a series of classical translations and the rapid growth of the universities at Oxford and Cambridge. Sir Thomas More's ***Utopia*** (Latin, 1516; Eng. trans., 1551) exemplified classical disciplines as a source of new and independent thinking.

Qn the other hand, the Renaissance in England was complicated by the transmission of classical authors through Italian and French Humanists, so that an Englishman's knowledge of Seneca and classical literary theory, for example, was derived from continental Senecan tragedies. Throughout Elizabeth I's reign strong native traditions of thought and expression worked with and modified the new ideas from Europe—the ideal of the "complete man," as exemplified in the person of Leonardo da Vinci and the hope that the mind of man provided infinite promise of achievement and hope of stability—and produced a fruitful and vigorous culture. Until 1580, however, this blending showed little prose or verse of any permanent value. Such theological works as the Great Bible (1539) of William Tyndale and Miles Coverdale, Archbishop Cranmer's ***Book of Common Prayer*** (1549), and John Foxe's ***Book of Martyrs*** (1563) are the most important inheritance from the first part of the 16th century.

Prose and poetry. ***Poetry before 1580.*** A sudden flourish of lyric poetry at the court of Henry VIII was the only true Renaissance achievement of the first half of the century. Leaders of the movement were Sir Thomas Wyat and Henry Howard, earl of Surrey. Wyat's short lyrics, mostly on conventional subjects, and sonnets showed a genuine originality of handling. In a very large number of poems, a single line reveals at once his own poetic genius and the potentialities of the language. Surrey showed, in contrast, greater technical assurance, but his pages rarely glow with Wyat's sudden splendour. His great metrical achievement was his use of blank verse in his translation of books II and IV of the ***Aeneid.*** His importation of the metre from Italy marked the introduction of the cardinal measure of English verse.

The miscellanies

The poems of Wyat and Surrey were not printed until 1557, in Richard Tottel's ***Miscellany,*** the first of a series of collections of lyric verses. Its successors included ***A Handeful of pleasant delites, The Paradyse of daynty devises,*** and ***A gorgious Gallery, of gallant Inventions.*** The publication of miscellanies continued throughout the century; one of the best, Francis and Walter Davison's ***Poetical Rhapsody,*** appeared in 1602. The one large literary project of the age, ***A Myrrour for Magistrates,*** was not successful. It originally contained 19 legends from English history "wherein may be seen by example of other, with howe grevous plages vices are punished," and it was issued with enlargements several times until 1621: the 1563 edition contained a famous "Induction" by Thomas Sackville, an accomplished poem in itself. Otherwise the years between Surrey and Sir Philip Sidney showed little that recalled Wyat or anticipated Edmund Spenser, except delightful single poems by such generally undistinguished writers as Richard Edwards, George Turberville, George Gascoigne, and George Whetstone.

Prose before 1580. Spoken English changed slowly in the 16th century. It is possible to learn how the language was spoken, with its gradual assimilation of classical influences, from early plays, state documents, the sermons of Hugh Latimer, letters and memoirs, and such works as William Bullein's ***Dialogue . . . against the fever Pestilence.*** In more formal prose writing, Latin ideals were adopted more rapidly but always at first with the intention of strengthening English. Humanists such as Thomas Wilson, Cheke, and Ascham were constantly vigilant against pedantic Latinisms and Italian affectations. Some of the best prose of this period was to be found in writers on rhetoric and in works of churchmen and such translators as Sir Thomas Hoby, who translated Castiglione's ***Il cortegiano (The Courtyer)*** in 1561. The finest example of a prose rich in Ciceronian cadences and yet unmistakably native was the English ***Book of Common Prayer*** (1549, 1552, 1559), while the successive versions (1525, 1535, 1537, 1539, 1560, 1568) of the Scriptures provided a noble and enduring pattern for vast heroic narrative and sublime lyrical prose. The chroniclers, antiquarians, and voyagers were an important group. The chronicle of Edward Hall and Raphael Holinshed's ***Chronicles of England, Scotlande, and Irelande*** (1577) used plain English style. Imaginative prose, however, did not flourish: the collections by William Painter, Sir Geoffrey Fenton, and George Pettie of stories from classical and continental authors were more valuable as sources of stage plots than as models for English novelists.

Euphues and its sequel

The novel, 1580–1600. The first English novels, ***Euphues*** (1578) and its sequel, ***Euphues and His England,*** by John Lyly, were far livelier than the novellas translated from Italian and met with immediate success. They presented an idealized picture of polite society, embodied many fragments of philosophy, natural history, or sociology, were of a faultless moral tone, and were written in a circuitous style displaying the most elaborate artifices of the rhetorician, which came to be called euphuism.

Sir Philip Sidney's ***Arcadia*** (published 1590) was a picaresque pastoral romance, written in adorned prose with grandiose phrases used to dignify a simple act, that under cover of fiction offered a guide to conduct. ***Arcadia*** and ***Euphues*** were in style and matter extremely influential, although their successors, such as Robert Greene's ***Pandosto*** and Thomas Lodge's ***Rosalynde,*** were much below them.

Of more realistic novels the best were ***The Unfortunate Traveller*** (1594) by Thomas Nashe, a brutal and realistic tale of adventure, narrated with speed and economy, and ***Jacke of Newberie*** (probably 1597) by Thomas Deloney. After this brief burst from 1578 to 1600 the English novel had to wait a century before it again achieved even minor importance.

Other prose, 1580–1600. English prose was naturally used for works of literary theory and accounts of voyages. Sidney's ***Defence of Poesie*** (1595) was the first work to introduce in English the ideas of the Neo-Platonic theorists of the Renaissance. He wrote in a lofty but clear style, commenting acutely on the state of contemporary English literature and expressing noble ideas of the poet's gift of creating an ideal "golden" world that would convey in its beauty "the form of goodness which seen we cannot but love." The discoverers were well represented in the ***Principall Navigations*** of Richard Hakluyt, a wide-ranging collection of travellers' accounts by various authors, including Sir Walter Raleigh, but the capabilities of English prose were better dispalyed in the great theological controversies. The two branches of the new Protestant Church of England—the Puritans and the more moderate Episcopalians—were in violent conflict, particularly about the relation of civil and ecclesiastical powers. They conducted their disputes in a series of long and learned works, culminating in the work of an Episcopalian bishop, Richard Hooker, whose ***Of the lawes of ecclesiasticall politie*** (1593–1662) expressed a liberal view of the nature of the Church of England in a Latinized but limpid style.

The religious pamphleteers seem almost to parody the erudite disputations of the great divines by conducting the same arguments in a series of scurrilous pamphlets. On the Puritan side the most effective were the seven anonymous Marprelate Tracts, which attacked the bishops in a sharp colloquial prose. The replies to these were also made anonymously, but the authors of the replies probably included Lyly and Nashe. Nashe (with Greene, until his death in 1592) also exchanged insulting pamphlets at intervals from 1590 to 1596 with the brothers Gabriel and Richard Harvey. Many lesser disputants joined issue in a similar way. Other writers produced pamphlets on an extraordinary variety of topics. The most lively were those that described in an appropriately informal style the low life of London, partly in admonition and partly in admiration: those of Greene are justly the most celebrated.

The Marprelate Tracts

Even among such diverse writers a general development may be traced from an ornamental Latinate rhetorical style, often florid but often eloquent, to a simpler, more flexible prose, much more closely related to English speech and capable of expressing with equal ease the commonplace and the sublime. The prose passages in Shakespeare's plays provide a good specimen of the later style.

The poetry of Spenser and Sidney. Edmund Spenser and Sir Philip Sidney were the first considerable poets of Elizabeth's reign. Sidney was the greatest of all those Elizabethans who regarded poetry only as one of the necessary accomplishments of the complete courtier; yet in his sonnets and in ***Astrophel and Stella*** (1591), his sonnet sequence, he attempted and achieved more than did any of his predecessors. His sonnets, many of which comment on contemporary fashions in poetry, were compressed but rarely contorted.

Spenser's popular ***Shepheardes Calender*** (1579) was notable especially for its deliberate use of archaic language and dialect. Spenser was most successful when his form or his content imposed limits on the fluency of his genius, as in the "April" and "November" sections of ***The Shepheardes Calender*** or in ***Epithalamion*** and ***Prothalamion,*** these latter two being poems of intricate and sustained melody.

For his unfinished masterpiece, ***The Faerie Queene*** (1590–1609), Spenser chose an infinitely looser form, the long poetic romance made popular by such writers as the Italian poet Ludovico Ariosto, and broadened it to accommodate an extended allegory in the medieval manner. Spenser's explanatory letter to Raleigh declared that he designed 12 books, portraying "the twelve private moral virtues, as Aristotle hath devised." The title virtues of the extant books are holiness, temperance, chastity, friendship, justice, and courtesy. The heroic adventures were of a kind with which English readers would be familiar from Arthurian legend—the rescue of maidens from enchanters, the capture of the strongholds of the wicked, tournaments and single combats, and the use of magic weapons. Blended with this medieval matter and Renaissance aim of fashioning "a gentleman or noble person in virtuous and gentle discipline" was a third level of meaning using such contemporary events as the subjugation of Ireland and such personages as, inevitably, Queen Elizabeth herself, who was represented as Gloriana, the "Faerie Queene."

Arthurian legend in ***The Faerie Queene***

Spenser's handling of rhymed verse was unsurpassed in English but, like his story, moved slowly and without excitement. His nine-line stanzas gave him room to amplify every incident; to this he was so often tempted that the current of his story, never fast, sometimes checked and stood motionless:

> The waies, throgh which my weary steps I guide,
> In this delightfull land of Faery,
> Are so exceeding spacious and wyde,
> And sprinkled with such sweet varietie,
> Of all that pleasant is to ear or eye,
> That I nigh ravisht with rare thoughts delight
> My tedious travell doe forget thereby;
> And when I gin to feele decay of might,
> Its strength to me supplies, and cheares my dulled spright.
> (***The Faerie Queene,*** Book ***VI,*** lines 1–9)

The successors of Spenser. The most original of Spenser's successors were the brothers Giles and Phineas Fletcher. Giles Fletcher in ***Christs Victorie, and Triumph*** (1610) retained Spenser's control over a long stanza but anticipated Milton in the burning solemnity of his vision of heaven. Phineas Fletcher's ***Purple Island*** (1633) was a long and detailed allegory of the human body and soul. Fletcher's fine metaphysical intuitions and his command of a condensed version of the ***Faerie Queene*** stanza made ***The Purple Island*** the most successful of the poems inspired by Spenser. These poets wrote with more fervour than "grave morall Spenser" and to win the reader not to the classical virtues but to an active Christian belief. In this they looked forward to the great religious poems of the 17th century.

Michael Drayton may be taken as the type of the Elizabethan poet. He followed Spenser in pastoral, the fashion of the age in the sonnet, Samuel Daniel in chronicle and legend, and Christopher Marlowe in mythological story and yet remained himself. The work of Drayton most often mentioned and least read was his *Polyolbion* (1612–22), a vast and pious effort to versify the scenery, legends, customs, and particularities of every English county. In his "Epistle to Henry Reynolds" Drayton gave a brief survey in easy-paced verse of the state of English poetry until his own day. In it he said of Samuel Daniel, "His rimes were smooth, his meeters well did close, But yet his manner better fitted prose." This is a just account of Daniel's long chronicle poems, such as his *Civile Wares* (1595–1609), but in his sonnets and shorter poems he displayed a cultivated and equable artistry.

Sonnets and mythological poetry. The English love sonnet, introduced by Wyat and revived by Sidney, enjoyed its greatest popularity between 1591 and 1598. Lovely lines, passages, and complete sonnets abound in such sequences as Samuel Daniel's *Delia,* Thomas Watson's *Tears of fancie,* Barnabe Barnes's *Parthenophil and Parthenophe,* Henry Constable's *Diana,* Spenser's *Amoretti,* and Bartholomew Griffin's *Fidessa.* In most of these sonnet series the form was that of three separate sections of four lines with the closing couplet for emotional and dramatic climax. The sense of personal revelation in these cycles was strong but often misleading. While, for example, Shakespeare's *Sonnets* were addressed to definite persons and were charged with sincere poetic emotion, the three strands of imitation, invention, and reminiscence were inseparably intertwined, and it would be extremely dangerous to treat them as autobiographical records.

Shakespeare's *Sonnets*

The amorous mythological tale in verse derived ultimately from Ovid, whose *Metamorphoses* and *Heroides* furnished descriptions, decorations, and many tales, so that within ten years, between 1589 and 1598, appeared Thomas Lodge's *Scillaes Metamorphosis,* Shakespeare's *Venus and Adonis* and *The Rape of Lucrece,* Thomas Edwards' *Cephalus and Procris,* Drayton's *Endimion and Phoebe,* and Marlowe's *Hero and Leander.* Marlowe's unfinished poem promised to be a major achievement; the common fault of the others was that the love situations were used as convenient occasions for the display of technical virtuosity.

Lyric poetry. As has been said above, the first English lyric poet of the 16th century was Sir Thomas Wyat. His translations and adaptations from the Italian embodied one feature that became almost a convention in all subsequent lyrics. The favourite subject was the poet's sorrow —such as the pain of parting or the death of a lover—yet the poems are light and exuberant. This was not a lack of sincerity: the strength of many of the lyrics lay in this very tension between subject and manner. Further characteristics inherited from Wyat were immediacy of subject and concreteness of imagery in the lyrics. They reflected every side of Elizabethan life; poets were not deterred by fear of seeming unpoetical from using the stuff of everyday affairs as material for lyrics.

The introduction of Italian music, with its favourite metrical schemes, powerfully schooled and coloured lyric. Italian poems were often translated in their own measure, line by line, and the musical setting was retained: these tunes were then coupled with new and original poems. Both kinds of lyric appeared together in the songbooks of the lutenist John Dowland, William Byrd, Thomas Morley, and especially those of Thomas Campion. Much of the best lyric verse of the period appeared in songbooks or in the miscellanies, which, after Tottel, were published regularly. *The Phoenix Nest* (1593), *England's Helicon* (1600), and Francis Davison's *Poetical Rhapsody* (1602) contained poems of a surprisingly high and even quality.

A third source of lyric was in the plays of the period. From the time of Lyly it had been customary for playwrights to diversify their action with a song, and, although for their full effect the songs require their context, even in isolation they are of singular power and beauty. Shakespeare was here, of course, pre-eminent, but Lyly, George Peele, Greene, and Thomas Dekker had an art as fine, though slighter.

Everywhere from 1580 to 1600, and intermittently thereafter, the nation found words for its lovely airs and madrigals. The poems of men such as Lodge, Raleigh, Nicholas Breton, Watson, Nashe, Donne (in his earlier poems), and Constable are among the happiest and most durable monuments of the Renaissance in England. After 1600 disillusionment with the material prosperity of England, with the slack restraints of the Reformed Church, and with certain Humanistic ideals influenced poetry deeply, but there were few satirists concerned with the state of society. Wyat had, as so often, shown the way and established the manner, but his successors in the 16th century were heavy-handed.

Despite the large number of writers who turned their hands to lyric and the chance of immortality, the general impression of Elizabethan verse of the great period is that it was uniformly high. No doubt the lyric became gradually overconventional, and its diction lost immediate force; yet at its matchless best it fittingly expressed the temper of Elizabethan England, half springtime and half golden age.

Elizabethan and Jacobean drama. *The transition from medieval drama.* Out of the medieval morality play, with its edifying personification of vices and virtues, had grown up the shorter interlude, usually a debate in a realistic setting between characters representing different types or trades and often lightened by comic plays about a "vice" or incarnation of grotesque roguery. The value of the interlude was most clearly seen when its separate elements broke away to form new dramatic compounds. For instance, when the interlude dealt with issues such as man's duties to his king, it clearly involved affairs of state and suggested a possible way of writing a historical play. When it drew a moral lesson from the fall of a king or a great man through ill fortune or as retribution for ill deeds, it clearly approached different kinds of tragic pattern; yet if it showed the good rewarded and the wicked confounded, it was closer to comedy. The vice himself was a distant ancestor of the Machiavellian prince (such as Richard III), the ogre-like villain (such as Marlowe's Jew of Malta), and the scheming servant whose intrigues spin the plot of many later comedies.

Senecan imitations

The simple medieval idea of tragedy as the rise and fall of a great man was complicated by the Senecan plays of Italy and by those of Seneca himself, all moral melodramas of violent passions marked by revenge, usually bloodthirsty, ghosts, and Stoic philosophy. Such direct English imitations as *Gorboduc* (performed 1561) by Sackville and Thomas Norton were poor things; yet the Senecan form was only barren in England, as the heroic tragedies of Corneille and Racine in France, also derived from Seneca, show. The Senecan tradition was, however, important for later playwrights, bequeathing to them the pattern of high tragical action swayed by large ideas of fate and retribution and also such points of stage mechanics as the ghost, the chorus, and the play within the play. At the same time the use of iambic pentameter blank verse (unrhymed line of five feet, each foot consisting basically of an unstressed syllable followed by a stressed) became standard thereafter in English verse tragedy and epic poetry.

From about 1550 various academic writers produced comic plays in English, accepting disciplines, mainly through the classical Latin playwright Plautus, such as the five-act structure and unity of action, but choosing native themes and treating them with realism. The most celebrated comedies of the middle of the century were *Ralph Roister Doister* by Nicholas Udall and the broad rustic farce *Gammer Gurton's Needle.*

By 1580 a number of chronicle plays and historical interludes such as John Bale's *Kynge Johan* had already appeared, and thus all the major dramatic forms had been foreshadowed. As Sidney pointed out in his *Defence of Poesie,* however, they were not clearly distinguished and were often ruinously entangled. It was to be a characteristic of the English drama to prefer in general mixed modes to a classical purity of tone.

The "university wits" and Kyd. Of the playwrights who flourished between 1580 and 1595, the "university wits" (who included Lyly, George Peele, Greene, Nashe, and Marlowe) were the most notable. Lyly, the most polished, wrote renderings of classical legends or fanciful superstitions, such as *Campaspe* and *Endimion,* with a close regard for unity of action and employed with assurance many devices that later became conventional—the ready introduction of light but poised lyric, the handling of a comic subplot as a foil to romance, and above all the use for high comedy of a highly wrought prose. Peele and Greene wrote lively blank, or unrhymed, verse and contributed to the tone of idyllic romance that characterized later comedy—Peele in *The Araygnement of Paris* and *The Old Wives Tale* and Greene in *The Honorable Historie of Frier Bacon, and frier Bongay*.

The Spanish Tragedie (*c.* 1590) by Thomas Kyd was as popular as it was influential. Drawing freely on Seneca and Machiavelli, Kyd established the theme of revenge that was to be the staple of much later tragedy. The play is a series of terrifying theatrical shocks, until the wronged Hieronimo takes a spectacular revenge and dies in his moment of triumph.

In the plays of Christopher Marlowe, *Tamburluine, Doctor Faustus, The Jew of Malta,* and *Edward II* (all written between about 1587 and 1592), which were in the widest sense morality plays, the heroes aspired to boundless dominion, forbidden knowledge, ideal beauty, or limitless wealth, inevitably overreaching themselves and ending in destruction. The plays are borne irresistibly forward in the theatre by Marlowe's verse, which, though sometimes rhetorical and bombastic, rose often to a classical serenity.

The histories

Shakespeare. Perhaps the first poet to combine appreciation of the freedom of the Renaissance with a profound intuition of the spiritual needs and potentialities of man was William Shakespeare. But, apart from *Titus Andronicus,* a tragedy of blood strongly influenced by Kyd, his first plays were little different from those that already held the London stage. *Henry VI* (three parts) broke no new ground and lacked the vigour of its successor, *Richard III.* These plays formed Shakespeare's first historical cycle. His later histories, *Richard II, Henry IV* (two parts), and *Henry* V, showed his increasing power to transcend the mere subject of a play, reaching beyond it to a larger view of the world. In the eight plays, Shakespeare used the tragic histories of the Plantagenet kings to show the transmission of evil from one generation to the next. The plays differed sharply in dramatic technique. While *Henry V* had much in common with a patriotic pageant and *Richard II* with a morality play, *Henry IV* was something more. The creation of Falstaff and the way his life in Eastcheap is compared to the King's marked a new complexity in the drama and showed Shakespeare exercising upon the past of England the creative powers of a poet. He wrote two more history plays: *King John* and *Henry VIII,* his last play, a complimentary but dignified attachment of the Plantagenet cycle to the reign of Elizabeth.

The comedies

Shakespeare's early comedies showed his willingness to follow any fashion that seemed profitable. In *Love's Labour's Lost* he displayed admiration for the courtly comedy of Lyly; yet he excelled him in human interest and diversified diction. Similarly, since Greene in *Frier Bacon, and frier Bongay* had successfully mixed sentiment, comedy, sensation, and mystery, Shakespeare continued the style in *The Two Gentlemen of Verona. The Comedy of Errors* was of the same type as *Ralph Roister Doister;* yet it was unmistakably the work of a poet. The first play Shakespeare wrote in a form new to the English stage was *A Midsummer Night's Dream,* developing a pattern of gentleness and magnanimity in the main plot set against the humbler virtues of honesty and conviviality in the subplot, a pattern that was repeated in *The Merchant of Venice* and his three mature comedies, *Much Ado About Nothing, Twelfth Night,* and *As You Like It.* Each of these had a unified tone: the threat of tragedy or the uncertainty of fate induces contemplation of the transience of happiness, while the robust subplots prevent sentimentality. Above all, what distinguished these plays from earlier and later English comedy was that they were written mainly in blank verse. Shakespeare varied it with expert prose, courtly or vulgar as the action required, but his staple was verse of unprecedented flexibility.

By 1600 the gradual darkening of the tone of literature had begun: Shakespeare's later comedies, *All's Well That Ends Well* and *Measure for Measure,* strain continuously against the conventionally happy ending of their plots. These plays and *Troilus and Cressida* closely resembled the bitter but powerful *Timon of Athens* in their cynical view of fate and human nature.

The tragedies and the last plays

Shakespeare's second tragedy, *Rotneo and Juliet,* written at about the same time as *A Midrummer Night's Dream,* shared that play's tone of youth and hope, but Shakespeare's conception of the purposes of tragedy deepened before *Hamlet, Othello, Macbeth, King Lear, Anthony and Clropatra,* and *Coriolanus.* These tragedies were not merely studies of the flawed virtue of their heroes but were a complete vision of the human condition explored with the burning curiosity of the Renaissance. In them Shakespeare's consummate command of blank verse gave the poetry unity of thought and poetic expression without loss of dramatic impact.

From 1608 Shakespeare showed an increasing interest in plays in which the action contained all the potentialities of tragedy but in which tragedy was eventually averted, usually through supernatural powers. *Pericles, Cymbeline, The Wittier's Tale,* and *The Tempest* abounded in violent emotions, crimes, and disasters; yet in each play characters thought to be dead were found to be living, parents and children were reunited.

Although Shakespeare's plays dealt profoundly with cosmic themes, he never lost sight of the dramatic needs of the theatre or the relevance of contemporary events. His plays were in the widest sense topical, so that when a Scottish king ascended the throne of England, as James I (James VI of Scotland) did, he wrote a play about another Scottish king, Macbeth. His ability to combine dramatic power with poetic power raised him to his unquestioned position of greatness in literature.

Jonson. Shakespeare's only rival among his contemporaries was his friend Ben Jonson, whose sardonic comedies exercised a lasting influence on English drama. Jonson condemned romantic Elizabethan comedy for its fairy-tale plots, amorality, and mingling of dancing, music, and drama, since he had learned dramatic practice from the works of the Roman dramatist Plautus and from Aristotle's *Poetics.* His early comedies, *Every Man In his Humour* (performed 1598) and *Every Man Out of his Humour,* were based on the theory that the humours, fluids in the body, determined the characteristics of men; but in *Volpone* (1606) and *The Alchemist* (1610) Jonson modified the theory, allowing more scope for development of character and anatomizing and correcting the vices of his own time. Unlike many Elizabethan dramatists, Jonson carefully created a series of plots of great ingenuity. He divided men into cheats and gulls (dupes). His plots demonstrated the villainy of the cheats and then laid bare with a deadly expertness the true nature and motives of the hypocrites, misers, and braggarts who were caught by the bait. His Roman plays, *Sejanus his fall* (1603) and *Cataline his Conspiracy* (1611), appeared formal and contrived, but his later comedies, such as *The Devil is an Ass* and *Bartholomew Fair* (1614), were abundant in life. Jonson was very much an Elizabethan in the delight he took in language, and he achieved a distinctive metallic splendour, admirably fitted for his hard and energetic comedy.

Jacobean tragedy. Apart from Jonson, Jacobean drama was chiefly tragedy. A sombre tone was evident in the satiric comedy of Jonson and in the tragedies and dark comedies of Shakespeare but was even plainer in the

Influence of Machiavelli

work of the other Jacobean dramatists, many of whom were fascinated by the doctrines of Machiavelli. Such a view of humanity set its own problems. It was when their characters were at the point of death that the Jacobean tragedians were confronted with the dilemma implicit in all Renaissance literature. They searched for a compromise between Humanism and a desire for immortality but found none. Their heroes thus died in tragic defiance, bewildered doubt, or at best Stoic resignation. The plays are fairly termed revenge tragedies. Crime was heaped on crime, and the last act saw the spectacular death of all those contaminated by sin. The plays emphasized man's viciousness or helplessness.

Only John Webster in *The White Divel* (*c.* 1612) and *The Dutchesse of Malfy* (*c.* 1613) offered more than his material of murder, madness, torture, and treachery. Yet each of his plays ended in despair, with only the vaguest hints that men are anything but "the stars' tennis balls, struck and bandied/Which way please them." The most positive thing in the plays was Webster's poetry, which, even when its subject and theme were most despairing, glowed with an iridescent light, illuminating the dark places of the mind and offering some promise that beauty could exist even in a charnel house.

Although Cyril Tourneur's *Revengers Tragedie* (published 1607) was concerned with the same theme as Kyd's *Spanish Tragedie,* that of the wronged man, Tourneur, from the opening scene in which Vendice apostrophizes his dead mistress' skull, gave revenge a much more sinister interpretation. Everything in the play is vicious: Vendice, set to cleanse this world, is himself sinning in taking his murderous revenge. By contrast, *The Atheist's Tragedie* (published 1611) was pervaded by the idea that private revenge defeats its own ends. In both plays Tourneur wrote verse rich in macabre imagery; in neither did he offer a view of man that was more elevated than that of his corrupt villains.

Thomas Middleton, similarly, in *The Changeling* (performed 1622), in which William Rowley collaborated, saw the dramatic possibilities of a series of murders and the horror that strikes to the soul at the prospect of further crimes and degradation. His heroine and her hireling lover were portrayed in all the complexity of their relationship with an insight and psychological thoroughness unique in the literature of the period.

The work of John Marston was patently that of a less gifted poet. His most famous work, *The Malcontent* (1604), is singular since the wronged man kills none of his victims but preserves them in suspense for the denouement, when the minor villains reform and the archvillain is expelled with ridicule. George Chapman's plays were crammed with poetry of the spacious majesty to be expected from the translator of Homer. His historical tragedies on French themes were the highest expression in English drama of the Stoic philosophy. Philip Massinger, on the other hand, was a craftsman of the theatre. His tragedy *The Roman Actor* (1626) is a demonstration of what happens when the Machiavellian has attained supreme dominion. John Ford maintained, until the Puritan ordinance closed the theatres in 1642, the tradition of tragedy wrung from human emotion and human depravity. *The Broken Heart* and *Tis Pitty Shees a Whore* (both published 1633) depicted the inevitability of fate and the pettiness of man's defiance, even while they recall Webster's ability to extract beauty from despair.

Jacobean comedy. A similar change of tone could be traced in comedy, where *The Shomakers Holiday* (first performed 1599) by Dekker was, apart from the romances, perhaps the last play without overt social criticism. With Dekker and Thomas Heywood should be named the professional journeymen playwrights who did not lack power but wrote usually in collaboration, such as Henry Chettle, Michael Drayton, and William Rowley, who supplied underplots of rough, lively comedy or tragedy. A more considerable comic writer was Middleton, whose play *A Chast Mayd In Cheape-side* (1613?) offered a particularly penetrating examination of matrimony that was, like the bulk of Elizabethan and Jacobean comedy, not for the polite. Most of the humour was obscene, a taste of the time, but many of the jokes were plays upon words of which one meaning has now become unfamiliar. The shifts of vulgar speech have scarcely affected Massinger's famous comedy *A New Way To Pay Old Debts* (published 1633), which was as firmly built as any play of Jonson's. Massinger showed here and in *The City-Madam* (1632) his ability to imply moral values through the action of his plays.

Beaumont and Fletcher

The work written in collaboration by Francis Beaumont and John Fletcher ran through all available forms from riotous topical burlesque, as in *The Knight of the Burning Pestle* (1607), to pathetic melodrama, as in *The Maides Tragedy*. The tradition of tragicomedy persisted in the fluent writings of James Shirley. He was happiest in the realm of light comedy, where his best work, such as *The Gamester* and *The Lady of Pleasure* (1635), looked forward to the comedy of the Restoration.

Minor plays and masques. Many minor plays, powerful domestic tragedies, chronicle plays, melodramas, citizen comedies, academic entertainments, pastoral plays in the Italian style, and Senecan dramas intended only for the study showed that the great vitality and esteem of the drama was not confined to major playwrights. The masque, a symbolic entertainment involving lavish sets, music, and dancing, called for the most exquisite combination of lyric, melody, and setting. Masques were usually written and performed as private entertainments for the nobility. From Jonson onward the country was enriched by a series of masques that stand as a graceful memorial to the Renaissance.

GERMAN LITERATURE OF THE RENAISSANCE

From Middle High German to Baroque. The fall of Constantinople to the Turks (1453) created a fear of the Turk in western Europe that was often reflected in literature of the 16th and 17th centuries, while territorial rulers in Germany grew more independent at the expense of the empire. The invention of printing with movable type around 1440 revolutionized printing and literature. The old art forms continued, however, beyond 1440—medieval allegory, folk songs, songs of love and nature, drinking and conviviality, and ballads survived in profusion. Hans Folz introduced more liberal rules to revive the art of the Meistersinger (members of guilds for cultivating singing and poetry). In the south, *Fastnachtsspiele* ("Shrovetide plays") became a vehicle of satire and broad humour, while in the north they were more restrained. Collections of comic anecdotes (*Schwänke*) evolved, often grouped around a single hero; anecdotes connected with Till Eulenspiegel, a 14th-century peasant jester, acquired a European reputation.

From about 1450 a new bourgeois realism that rejected the medieval knightly tale in favour of Humanism and new classicism evolved.

The courtly revival. At some princely courts the Burgundian "renaissance" fostered a revived interest in medieval chivalry. At Innsbruck, Duchess Eleonore of Austria wrote a prose version of a French chivalrous romance, *Pontus und Sidonia;* at Rottenburg, Jacob Piiterich von Reichertshausen's work fostered a cult of the medieval poet Wolfram von Eschenbach. Medieval poems were printed, either in their original verse form or in prose versions of little literary distinction. The allegorical works of the Holy Roman emperor Maximilian I, *Weisskunig* ("The White King") and *Theuerdank* ("Noble-Mind"), represented probably the last attempt to live by medieval chivalrous ideals.

Humanism. In sharp contrast to this courtly renaissance stood the new element in German literature after 1450: the Humanist movement. The Humanists between 1450 and 1480 drew their inspiration from Italy. Albrecht von Eyb, Heinrich Steinhowel, Niklas von Wyle, "Arigo" (probably the pseudonym of Heinrich Schlüsselfelder), and Antonius von Pforr were chiefly translators who produced versions of many Latin classics, some Greek and Indian ones, and Italian works, including Boccaccio's *Decameron.*

German Humanism presented quite a different picture after 1480. Conradus Celtis, Eobanus Hessus, and others were closely associated with university circles and wrote almost entirely in Latin. A strong patriotic and political strain was demonstrated in the ***Germania*** of Jakob Wimpfeling and the diplomatic activities of Konrad Peutinger. A lively interest in German history appeared in the work of Beatus Rhenanus, the ***Chronik der Äbbebte von St. Gallen*** (1533; "Chronicle of the Abbots of St. Gall") of Joachim Vadianus (pseudonym of Joachim von Watt), and the ***Bayerische Chrortik*** ("Bavarian Chronicle") of Aventinus. But the masses were not influenced as they were to be later by the Lutheran Reformation.

Revitalization of drama by Humanism

The Humanist impact was more direct in the drama. A lively dramatic tradition was established as part of university courses. Roman comedies and Latin plays by modern authors were performed for moral instruction and to teach eloquence. These plays contrasted sharply with the medieval dramatic types, and from them 16th-century German drama developed. Later this Latin drama was turned to the service of the Reformation,

The concentration on Latin by the most gifted writers of the age goes far to explain why vernacular literature between 1490 and 1520 was so scarce. Satirical or didactic works predominated; the most famous work of the 15th century, ***Das Narrenschiff*** (1494; "The Ship of Fools"), by Sebastian Brant, reviewed all the major and minor vices of the age. In the Low German ***Reynke de Vos*** (1498; "Reynard the Fox"), the verse was accompanied by an extensive prose commentary applying the satirical episodes to contemporary evils.

The Reformation. The Reformation and the vigorous protest of Martin Luther against ecclesiastical abuses affected life in Germany; it left little room for purely aesthetic considerations. The Humanists, who favoured reform without doctrinal changes, mostly held aloof, though with certain notable exceptions: Ulrich von Hutten, hoping for a political reformation too, supported Luther's cause in a series of pamphlets; Philipp Melanchthon, a Greek scholar, introduced the Humanist tradition into Protestant schools. But apart from the Humanists, almost every writer of note was preoccupied with the Reformation.

From the torrent of printed works that swept over Germany a few had literary merit: Luther's own many and various writings, his eloquent Reformation pamphlets of 1520, for example, or his terrible condemnation of the peasants' revolt; ***Die* 15 *Bundsgenossen*** (1521–23; "Fifteen Comrades"), in which Johann Eberlin von Giinzburg attacked religious abuses; Thomas Murner's ruthless satires in defense of the old religion; the fierce anti-Catholic satire in the ***Fastnachtsspiele*** of Niklas Manuel of Bern; and the writings of Sebastian Franck, an independent radical thinker.

Luther's influence

In three respects the Reformation had a lasting effect on German literature. First, Luther's Bible translation (New Testament, 1522; Old Testament complete, 1534) was not only based for the first time on Hebrew and Greek texts but also achieved a vigorous, popular German style. It has exercised an incalculable influence on the style and ideas of later German writers. Second, Luther established congregational hymn singing as an essential part of the Protestant service and himself wrote several hymns. His example established a tradition of hymn writing that has been a major German contribution to the Christian world. Third, Luther had commended certain biblical subjects as suitable for plays, encouraging many dramatists—of whom Sixtus Birck, Paul Rebhuhn, Joachim Greff, Burkhard Waldis, and Jorg Wickram were the most outstanding—to write plays in German as vehicles of Lutheran teaching. Modelled closely on the Latin school drama, these plays marked a break with medieval types of drama.

The Reformation, despite its importance, did not account for all the vernacular literature after 1520. In scholarly works, for example, such as the treatises of the Swiss doctor, alchemist, and scientist Paracelsus (Theophrastus von Hohenheim), German was gradually beginning to replace Latin. But satirical or didactic works remained predominant, as in the collection of proverbs by Johann Agricola. The fable enjoyed revival: Luther's own translations from Aesop had no special merit, but Erasmus Alberus in his ***Fabeln*** (1534) and Burkhard Waldis in his ***Esopus*** (1548) turned the fable into a lively minor genre. ***Grobianus*** by Friedrich Dedekind, a satirical Latin guide to table manners, was translated into German by Kaspar Scheidt.

The works of Hans Sachs

Apart from Luther, the most prolific and characteristic writer of the century was Hans Sachs of Niirnberg. His works, didactic in aim but entirely unpolemical, reflected ideals of the devout, upright, industrious Lutheran townsman and artisan. By common consent the greatest of the Meistersinger, he brought the form to perfection and virtually to an end, despite attempts at revivals by later poets. Sachs was at his best in comic verse anecdote and *Fastnachtsspiel.*

Light literature. As the tide of Reformation polemics receded, works of pure entertainment came into greater prominence. The popular chapbooks were often prose versions of medieval verse romances or were adapted from foreign sources. The stories of Fortunatus, Magelone, Melusine, and many others were favourite reading far beyond the 16th century. Men found distraction also in collections of anecdotes, a type of work that went back at least to Eulenspiegel. The *Schimpf und Ernst* (1522: "Jest and Earnestness") of Johannes Pauli had didactic aims, but entertainment was the sole aim of Jorg Wickram's ***Rollwagenbiichlein*** ("Coach Book") and of similar collections by others. About this time the modern novel began in Wickram's more substantial prose narrative—*e.g.*, ***Der Goldfaden*** (1557; "The Gold Thread") —and in the German translation of the Spanish chivalrous romance ***Amadís de Gaula.***

The close of the 16th century. The Council of Trent (1545–63), which attempted to reform the Roman Church, and the advent of the Catholic Society of Jesus gave religious controversy a new turn. The Jesuits adapted the Latin school drama for their own educational purposes, but no outstanding dramatists appeared until the 17th century. Meanwhile, the Protestant academy in Strassburg, especially under a Christian Humanist, Johannes Sturm, developed a rich educational and dramatic tradition.

Source of the Faust story

In the last quarter of the century the most outstanding author was Johann Fischart, many of whose works were directed against the Jesuits, the Counter-Reformation, and notably a brilliant preacher, Johannes Nas. In all Fischart's writings, the didactic aims of the age were combined with an interest in literary form. ***Das glückhafft Schiff*** (1576; "The Ship of Good Fortune") was, formally and stylistically, one of the most distinguished poems of the century. Fischart's comtemporary Philipp Nikodemus Frischlin was the last notable Latin dramatist of the century. On the other hand, the versions of Psalms by Paul Schede Melissus and Ambrosius Lobwasser, on the Calvinist model of Clément Marot, added a new element to the rich tradition of Protestant hymn writing, while an anonymous chapbook, ***Historia von Dr. Johann Fausten*** (1587), gave rise to the whole European Faust tradition. A noteworthy feature in the last years of the century was the arrival in Germany of troupes of English actors whose repertoire included versions of contemporary English plays. They influenced Jakob Ayrer and Duke Heinrich Julius of Brunswick; the latter composed 11 plays. With this new influence German literature moved into the Baroque age.

Swiss–German literature of the Reformation. The activities of the religious reformer Huldrych Zwingli had only an indirect influence on literature. Zwingli himself wrote mainly in Latin. The so-called Ziirich Bible of 1529 was gradually adapted to conform with the Luther Bible, which tightened the connection between Swiss and German writings. An important result of Zwingli's work was that the Protestant majority of German-speaking Switzerland established a permanent connection with Protestant parts of western Switzerland and with Protestant countries abroad. The anonymous play about William Tell from the canton of Uri was a forceful and popular expression of Swiss patriotism. Gilg Tschudi's ***Chroni-***

con Helveticum, covering the years 1000–1470 in Swiss history, endured as literature.

Austrian literature of the Renaissance. In Austria the vast expansion of the Habsburg Empire brought the imperial chancellery from the Prague of the Luxembourg dynasty to the Vienna of the Habsburgs, carrying with it a double heritage: the German literary language and a Humanist disposition and care for style. The emperor Maximilian I, himself a writer, made Vienna, through its chancellery and university, the leading Humanist city in Germany, helped by Conradus Celtis. Johannes Spiessheimer (Cuspinian), and Joachim Vadianus (von Watt).

Development of the theatre in Vienna

Vienna acquired a theatre at the Renaissance with the court plays of Conradus Celtis and the monastery plays of the abbot of the Schottenkloster ("Scottish," *i.e.*, Irish, monastery), Benedictus Chelidonicus. The German plays of the schoolmaster Wolfgang Schmeltzl were first performed in the Schottenkloster. From 1554 onward the Jesuits created out of the dynastic alliance of Austria and Spain a community of religious conviction and mental inclination. They began to perform a considerable repertoire of plays, and at the same time the Italian opera developed and was patronized especially by the emperors. The Viennese Baroque theatre became one of the artistic triumphs of Europe.

The individual regions flourished too. Tirol had its own courts from time to time. From 1430 there was an abundance of sacred and secular drama. A century later, in Sterzing, Vigil Raber was collecting and rewriting the plays of the people. About 1570 Archduke Ferdinand of Tirol led a poets' court, whose preacher was Johannes Nas, while the monk Laurentius von Schnüffis, from being theatrical producer to a duke, became a mystical lyricist. Salzburg had a life of its own, shining through a Humanist, Paul von Hofhaimer, who set its tone, through the theatre of the archbishop's court, and through the Benedictine university founded in 1620. Styria had a wide range of artistic interest in the great monasteries such as Kremsmünster, where Simon Rettenbacher wrote religious plays. Christoph van Schallenberg, Wolfgang Helmhard von Hohberg, and Katharina von Greiffenberg were poets of the nobility. It was the emperor Ferdinand II, sometime ruler of Styria, who decided the battle against the Reformation, first in Styria and then in his other lands. The voice of the triumphing church was uplifted in Vienna by Abraham a Sancta Clara, Habsburg court preacher and master of Baroque German prose.

DUTCH LITERATURE OF THE RENAISSANCE

The literature of Flanders and Holland must be considered as a whole until about 1582, when the fall of Antwerp marked the final rift between the Protestant north and the Catholic south. The 16th century was characterized by a diversity of literary forms, which reflected an age in decline confronted by the first articulation of the new art coming to the Netherlands from Italy through France. The Renaissance in the Netherlands began with several writers such as Lucas de Heere, who had fled from the Catholic southern provinces for religious reasons. Chapbooks, containing prose versions of medieval romances, folk songs, and *rederijkers* ("rhetoricians' ") verse; Reformation propaganda, marching songs of the Calvinist revolt against Spain; these and the first sonnets, the first dissertations in the vernacular, and the first grammars of the Dutch language displayed the restlessness of an age of change. So, while the Catholic Anna Bijns was fulminating against Lutheranism in her glowing satirical versa, which was countered later by the Calvinist Marnix van Sint Aldegonde in his polemical attack on the Catholic Church, the echoes of classical antiquity were reaching the Netherlands in the odes, sonnets, and translations of Jan van der Noot and Jan van Hout. Carel van Mander, painter and poet, introduced scholarly vernacular prose writing, though the Latin prose of Erasmus had been famous throughout Europe for nearly a century.

Van der Noot's collection of Petrarchan sonnets in the manner of the French poet Pierre de Ronsard was published in London, where he was then in exile on account of his participation in an insurrection in 1567. The two great moderates of the age were the Erasmians J.H. Spiegel and Dirk Volkertszoon Coornhert, both liberal Humanists with a social, undogmatic Christian ethic. Spiegel's poetry is generally more intellectual than Coornhert's prose, which was influenced by Montaigne and the Bible, with a remarkably supple and lucid, even entertaining, style. It was Coornhert and his successors, in particular the translators of the Dutch authorized version of the Bible, who laid the foundations of the standard language.

SCANDINAVIAN LITERATURE OF THE 16TH CENTURY

Swedish. Two dates mark the beginning of modern Swedish history: 1523—the breach with Denmark and Gustav I Vasa's accession; and 1527—the breach with Rome and the establishment of a national Lutheran Church. The political revolution that eventually brought Sweden to the position of a European power had no considerable effect on literature until a century later, but the Reformation wholly dominated Swedish letters during the 16th century.

The Swedish translation of the Bible

The most important literary event of this period was the translation of the Bible in 1541, which inaugurated modern Swedish and provided an inexhaustible source for poets of subsequent times. Closely involved in the Bible translation were the apostles of the Swedish Reformation, Olaus Petri and his brother Laurentius. Olaus Petri's vigorous approach was revealed in his published sermons and in a Swedish chronicle, the first historical Swedish work based on critical research. Olaus Petri may also have written the biblical *Tobie comedia* (published 1550), the first complete extant Swedish play.

As a consequence of the Reformation, two of Sweden's most distinguished scholars of the period were driven into exile. In his history of all the kings of the Goths and Swedes, Johannes Magnus provided Sweden with a number of valiant kings unknown to critical historians. Johannes' brother Olaus Magnus wrote the first geographical and ethnographical account of Scandinavia, *Historia de gentibus septentrionalibus* (1555; "History of the Northern Peoples").

Meanwhile, academic learning was in decline. The University of Uppsala, founded in 1477, was closed in 1515 because of political unrest and was not revived until late in the century. Associated with it was Johannes Messenius, who wrote several mythical-historical plays.

Danish. In 1536 the Lutheran Reformation was carried through in Denmark, and the beginning of the 16th century was characterized by many pamphlets for or against the Roman Catholic Church. European Humanism and the Renaissance made their influence felt also in Denmark, where Christiern Pedersen was the most prominent Humanist who supported the Reformation. He edited *Gesta Danorum* by the 13th-century historian Saxo Grammaticus. translated the New Testament, adapted Martin Luther's pamphlets into Danish, and above all participated in a translation of the Bible (1550). Poul Helgesen was the most gifted opponent of the Lutheran Reformation and Hans Tausen its most talented spokesman. The *Visitation Book* by the Lutheran bishop Peder Palladius is an important literary document. The two most important historians (both belonging to the transition between the 16th and 17th century) were Anders Sørensen Vedel and Arild Huitfeldt.

Sixteenth-century Danish poetry was religious or polemical, with fine love poetry and hymns. The earliest plays date from the beginning of the century. The most important playwright of the period was Hieronyrnus Justesen Ranch, whose farce *Karrig Nidding* ("The Miserly Rascal") was his best play. The full force of the Renaissance in literature was not felt till the next century.

Icelandic. The chief political figure and poet of the Reformation was Jon Arason, last Catholic bishop of Hólar, beheaded in 1550. By his life Jon showed that he was a Viking as well as a martyr, although most of his surviving poetry is religious.

The effect of the Reformation on Icelandic learning and literature was that Catholic poetry was discarded and attempts were made by the first Lutheran bishops to replace it with hymns poorly translated from Danish and German.

Lutheran teachers instructed the people in Protestant dogma, and several translations of sermons and books of instruction by German Lutherans were printed in Icelandic from as early as 1540. Gudbrandur Thorláksson was the most energetic and wisest of the Lutheran teachers. In translating the Bible he used earlier Icelandic versions of some books of the Old Testament and Oddur Gottskálksson's Icelandic translation of the New Testament. In his psalmbook he showed appreciation of Icelandic poetic tradition and adhered to Icelandic alliteration and form.

RENAISSANCE LITERATURE OF EASTERN EUROPE

Hungarian. Until the 16th century, Latin remained the language of serious literature in Hungary, a result partly of Humanist influences. In the second half of the 14th century, Italian influence had become predominant in Hungary, and in 1367 the first Hungarian university was founded. In the middle of the 15th century, with the accession of King Matthias I Corvinus, the court became a centre of Humanism. The King established a university and the first Hungarian press and was famous for his library. Latin literature reached its peak in Janus Pannonius. the greatest Humanist of Hungary, who was educated in Italy. He wrote numerous elegies, inspired by the troubles of a responsible statesman, difficulties in combatting the Turks, love of country, and personal sorrows.

Writers of the Hungarian Reformation

The 16th century brought changes, under the shadow of the Turkish wars. After the Battle of Mohács (1526) the country was split into three, and centres of Humanist and literary thought were destroyed. There followed the dawn of a new era—the Reformation—with which Hungarian national literature really began. Mention must be made of three forerunners of this new era: Benedek Komjáti, Gábor Pesti, and János Sylvester, who were disciples of Erasmus. With methods infinitely superior to those employed 100 years earlier by the Hussite teachers, they translated parts of the Bible with philological accuracy. Pesti made a very readable translation of Aesop's ***Fables*** and published a Latin–Hungarian dictionary. Sylvester published the first Hungarian grammar and, to show the adaptability of the vernacular to classical verse forms, wrote the first Hungarian poem in couplets. In 1541 he published a translation of the New Testament.

With the spread of the new religion, translations were made of the Psalms and of hymns. András Szkhárosi Horvát wrote didactic poems, versified sermons, and paraphrased biblical stories, and he was perhaps the greatest Hungarian satirist,

The second half of the 16th century saw the beginnings of Hungarian drama. ***Balassi Menyhárt árulásáról irott komédia*** ("Comedy on the Betrayal of Menyhárt Balassi"), a satire by an unknown author, was perhaps the most interesting literary achievement of the Reformation. Péter Bornemisza's literary activity consisted mainly in sermons, but he gave an entrancing view of Hungarian life, teeming with fresh observations, vivid descriptions, and original comments. His volume ***Ördögi Kísértetek*** ("On the Temptations of the Devil") offered an interesting consideration of moral and sexual problems in the 16th century. A poem of farewell, written on leaving the country, was one of the gems of early Hungarian poetry.

Perhaps the greatest single literary achievement of the Hungarian Reformation was a translation of the Bible by Gáspár Károli (1590); this translation played a role in the development of Hungarian similar to that of the Authorized Version in English.

Up to the 16th century, religious literature seems to have fared better than secular literature partly because secular literature was not written down. The late-16th-century minstrels were more learned than their predecessors, often driven to their profession by difficult economic conditions. Perhaps the most important was Sebestyén Tinódi, by temperament more historian than poet. He described the wars against the Turks with remarkable accuracy, but his verse was monotonous. Peter Ilosvai Selymes was the author of a romance, ***Miklds Toldi*** (written 1574), which achieved great popularity in Hungary and served as a basis for a masterpiece by János Arany in the 19th century. The 14th-century hero, a man of immense strength, was a popular figure in Hungarian literature. This romance was the one original piece in the flow of the mere entertainment literature characteristic of the 16th century, whose principal genre was the ***széphistoria*** ("beautiful story"), adapted from western European originals. Perhaps the best was the ***Argirus historiája*** ("The Romance of Argirus") by Albert Gergei, from an Italian original but interwoven with Hungarian folklore.

Originality of Bálint Balassi

A great poet emerged in Bálint Balassi (1554–94), who was highly original and comparable with the later Cavalier poets of English literature in dash and frivolity. But he had a depth that they lacked. Balassi enjoyed fighting, love, and the good things in life, but his poetry was shot through with sudden and deep religious impulse. Technically, his verses were unmatched in early Hungarian literature, and he had many imitators and followers.

Polish. Although the Renaissance reached Poland comparatively late, it was the golden age of Polish literature. External security, constitutional consolidation, and the Reformation contributed to this flowering: the first permanent press was established at the turn of the 15th century, and the first book printed entirely in Polish was issued in about 1513. The modern literary language dates from this time.

The first generation of writers influenced by the Italian Humanists wrote in Latin, and it included Jan Flachsbinder (Joannes Dantiscus), author of incidental verse, love poetry, and panegyric; Andrzej Krzycki (Cricius), an archbishop who wrote witty epigrams, political verse, and religious poems; and Klemens of Januszkowo (Janicki, Janicius), a peasant who studied in Italy and became poet laureate and the first original Polish poet.

Mikołaj Rej of Nagłowice combined medieval and Renaissance aspects. Self-educated, he was the first idiomatically Polish talent and widely read writer of his time, being known as "the father of Polish literature." He wrote satirical epigrams, but of more importance were his prose works, especially ***Postilla,*** a collection of Calvinist sermons, and the ***Żywot cztowieka poczciwego*** (1558; "Life of an Honest Man"), a description of an ideal nobleman.

The Humanist poet Jan Kochanowski

The second generation of Humanist poets, and indeed the whole Renaissance period, was dominated by Jan Kochanowski. The son of a country squire, he wrote in the vernacular and was the first Polish writer to attempt both satirical poetry and classical tragedy, but his lyrical works surpassed these experiments. His crowning achievement, the first Polish work to equal the great poems of western Europe, was ***Treny*** (1580; ***Laments,*** 1920), inspired by despair after the death in 1579 of his infant daughter, to be succeeded by a final recovery of his spiritual harmony. In Kochanowski's poetry the archaisms still apparent in Rej's work had almost disappeared; language and idiom were modern. The flexibility and assurance of his poetic genius were immediately recognized as a sign that the literary language had attained its maturity.

The most notable of Kochanowski's followers was Szymon Szymonowic (Simonides). He introduced in his ***Sielanki*** (1614; "Eclogues") a poetic genre that was to retain its vitality until the end of the 19th century. These pastoral idylls exemplified the processes of imitation, adaptation, and assimilation by which Renaissance writers brought foreign models into the native tradition.

The numerous poems, in Latin and Polish, of Sebastian Klonowic were of interest for their description of contemporary life. ***Worek Judaszdw*** (1600; "Judas' Sack") was a satirical poem on the plebeian life of Lublin, of which he was mayor.

The prose of the 16th century ranked with its poetry in vitality and range. The most eminent writer in Latin was Andrzej Frycz-Modrzewski. In ***Commentariorum de republic~emendanda libri quinque*** (1551, 1555), he evolved a bold social and political system, based on the principle of equality before God and the law. Another notable writer in Latin was Marcin Kromer, scholar, Humanist, historian, and Catholic apologist. Of his controversial works, most interesting is the Polish *Rozmowy dworzanina z mnichen* (1551–54; "Dialogues Between the Courtier and the Monk"), a defense of Catholicism. Many historical and polemical writings and translations of the Bible were also published during this period; the Catholic translation by Jakub Wujek is an outstanding literary work.

Polish literature had become a national literature, reflecting Poland's position as a great power with far-flung boundaries, the evolution of the nobility as a ruling class, and the country's economic prosperity. Its influence spread east, above all to Moscow, and to the west its culture was represented by men of such high repute as the scientist and astronomer Nicolaus Copernicus (Mikołaj Kopernik).

Russian. About the middle of the 16th century Moscow's conception of itself as a centre of learning, of Orthodoxy, and of political authority began to take shape. There appeared a succession of literary undertakings aimed at exalting and strengthening Muscovite ecclesiastical and political traditions and at demonstrating that these had come down from the very beginning of Russian statehood. In 1552 there appeared the grandiose compilation known as the *Velikiye Minei-Cheti* ("Grand Minei Cheti"), by the metropolitan of Moscow, Makary, which collected numerous works of original and translated ecclesiastical literature. It was followed bv the *Stevennava Xniga* ("Book of Degrees"—*i.e.*, generations), containing biographies of Russian princes and ecclesiastical figures and completed in 1563 by Makarios' successor. At this period also were collected the important Muscovite chronicle compilations. An enormous collection, the *Litsevoy svod* ("Illuminated Compilation"), was produced, which begins with the creation and ends in the 1560s. Other works such as the *Domostroy* ("Household Management") and the Stoglav (Council of a Hundred Chapters) aimed at strengthening moral, social, and political standards of conduct. The introduction of printing in Moscow did much to help Russian unification, and the first dated book, *Apostol,* was printed in 1564.

Czech. In the 16th century there was a mass of prose literature in which didactic and scholarly writing predominated. The influence of Humanism was apparent in the elegant style of the legal historian Viktorin Kornel of Všehrdy. Of considerable interest is the small group of travel books from this period. The development of Czech prose culminated in the translation of the Bible made by a group of scholars of the Unitas Fratrurn and known as the Kralice Bible (1579–93). The language of this version became the model of classical Czech.

Croatian. In the 15th and early 16th centuries the most outstanding Croatian writer was Marko Marulić, author of the epic *Judita* (1501), in which he encouraged his countrymen in their struggle against the 'Turks. Under the influence of classical and Italian literature and Croat folklore, the Petrarchan lyric was developed in Dalmatia and in Dubrovnik. One of the most prominent poets was Hanibal Lucić, author of *Robinja* ("The Slave Girl"), the first Yugoslav secular play. Among the best known writers of Old Croat literature was Marin Držić, who wrote pastoral dramas and comedies portraying Renaissance Dubrovnik. His comedy *Dundo Maroje* (written *c.* 1551) was performed throughout western Europe.

Slovene. At the end of the 1620s, Protestant activity began in Slovenia. The demand to replace Latin with a liturgical language that all could understand gave impetus to literary activity. The Protestant reformer Primož Trubar, who as a supporter of Lutheranism had to flee to Saxony, published the first printed Slovene book *Abecedariutn* (1550), a catechism (1551), parts of the Bible, the constitution of the Slovene church (1564), and polemic writings, and with his followers founded a rich religious literature. This activity reached its height in 1584, when an outstanding translation of the Bible by Jurij Dalmatin and the first Slovene grammar, *Articae horulae successivae de latinocarniolana literatura*, written by educationalist Adam Bohorič were published. The Protestant writers also published seven books of songs (1551, 1563, 1567, 1574, 1579, 1584, 1595) intended for liturgical needs, as well as the German–Latin–Slovene–Italian Dictionary (1592, 1608), compiled by Hieronym Magiser. The basis thus created for literary Slovene, together with the peasant risings of this time, formed the starting point of Slovene national consciousness.

Romanian. The first book printed in Walachia in 1508 was a Slavonic liturgical book. A certain Deacon Coresi printed translations of the Acts of the Apostles (1563). Others of his publications that survive are the *Talcul Evangheliilor şi Molitvenic* ("Sermons and Book of Prayers") and *Evanghelia cu invătatura* (1581; "Commentary on the Gospels"), and these all encouraged the use of Romanian.

In this period some secular literature was also produced, but it consisted mainly of translations from Greek, Slavonic, Byzantine, and Oriental books (*e.g.*, a history of Alexander III the Great, *Aesop's Fables,* and *The Thousand and One Nights).*

CELTIC LITERATURE IN THE RENAISSANCE

Welsh. After the pre-eminence of Dafydd ap Gwilym and his followers in the *cywydd,* there arose for a short time a school of literary formalists. The chief of these was Dafydd ab Edmwnd: the main characteristic of his poetry was its ingenuity. His poetic heirs were Tudur Aled (died 1526), a bard of great craft, and Gutun Owain (flourished *c.* 1460–1500).

Effects of the Reformation and the accession of the Tudors

The Reformation broke the hold of the Catholic educational system in Wales, which resulted in a diminution of literary appreciation and in the standards of the language. The Tudor policy of encouraging the spread of English at the expense of Welsh and of inducing the Welsh aristocracy to emigrate to England almost destroyed the old Welsh culture, which was completely bound up with the language. Yet fine work was written by the satirists Siôn Tudur and Edmwnd Prys. Other masters of the *cywydd* deserving of mention were William Llŷn and Siôn Phylip.

The rise of modern prose. As printing made its appearance in Wales in the 16th century, the traditional prose was abandoned by the Humanists of the Renaissance. The new prose they fashioned was based on bardic language and classical authors, enriched by new formations and borrowings. The first Welsh printed book, *Yn y lhyvyr hwnn* (1547; "In This Book"), so named because of its opening words, consisted of extracts from the Scriptures and the prayer book: from this time Welsh prose began to take definite form.

The Reformution. The most important figure of the Reformation was William Salesbury, who translated the New Testament by 1567. Marred though it was by philological foibles and the mechanical means used to make the language intelligible in every part of Wales, it was nonetheless a fine piece of translation. In the same year was published the Welsh Prayer Book, also translated mainly by Salesbury in collaboration with Richard Davies, bishop of St. David's. The Welsh Bible translated by William Morgan, bishop of St. Asaph, aided by Edmwnd Prys, was published in 1588. The revised version, published in 1620, is still used. It would be difficult to exaggerate the importance of these three translations in the development of Welsh prose. From 1588 onward there was no break in the production of Welsh prose books. The first were translations from English and Latin aimed at grounding the Welsh nation in the principles of the Reformation.

The Counter-Reformation. During the years in which the reformed religion was being established in Wales, Welsh society and the Welsh language were at their lowest ebb. The Catholic writers of the Counter-Reformation regarded the new religion as something imported from England—and they thought that the way to preserve the old religion was to insist on the old Catholic culture. As a result there appeared *Dosparth Byrr,* the earliest printed primer of Welsh grammar, the work of Gruffudd Robert (*c.* 1522–c. 1610), and also several religious works, many of which were published on the Continent.

The Welsh Renaissance. Just as Italy and other European countries under the Renaissance turned to the Latin and Greek classics, so Wales turned to its own classical tradition of bardism. In addition to Gruffudd Robert's primer mentioned above, there appeared a set of rules for bardic poetry and principles of the Welsh language compiled by Siôn Dafydd Rhys and a dictionary and a grammar by John Davies of Mallwyd.

Breton. Breton literature. like the ianguage, divides into three periods—old, middle, and modern. Old Breton (8th to 11th century) is only found in lists and glosses in documents or as names in Latin books and charters. From the Middle Breton period (11th to 17th century) the 11th- to 15th-century compositions were mainly oral, and little except a few scraps of verse is extant until the late 15th century, when there appeared the *Catkolicon* of Jean Lagadeuc, a Breton–Latin–French dictionary printed in 1499, and Quiquer de Roscoff's French–Breton dictionary and conversations (printed 1616).

The waulking songs

Scottish Gaelic. Some Gaelic poetry from the 16th century survived in oral tradition until the mid-18th century, when it was written down, such as *An Duanag Ullamh*, composed in honour of Archibald Campbell, 4th earl of Argyll, or the lovely lament *Griogal Cridhe,* composed *c.* 1570. It is certain that the poetry recorded in the *Book of the Dean of Lismore* was not an isolated outburst; much must have been lost, both of the work of the professional bards and of the popular songs. Songs in the nonsyllabic, accented measures survived, again orally, from the earliest years of the 17th century. This was the tradition that produced the work songs—*e.g.*, waulking songs used as an accompaniment to the fulling of cloth (a process of thickening and shrinking), better known from 17th- and 18th-century examples.

In 1567 the first book printed in Gaelic in Scotland appeared: Bishop John Carswell's *Foirm na Nurrnuidheadh*, a translation of John Knox's liturgy, in classical common Gaelic.

IV. The 17th century

CHARACTERISTICS OF THE AGE

The 17th century was a period of unceasing disturbance and violent storms, no less in literature than in politics and society. The great question of the century, which confronted serious writers from Donne to Dryden, was Michel de Montaigne's "What do I know?" or, in expanded terms, the ascertainment of the grounds and relations of knowledge and faith and reason and authority in religion, metaphysics, ethics, politics, economics, and natural science.

The questioning attitude that characterized the period is seen in the works of its great scientists and philosophers: Descarte's *Discourse on Method* (1637) and Pascal's *Pensées* (written 1657–58) in France; Bacon's *Advancement* of *Learning* (1605) and Hobbes's *Leviathan* (1651) in England. The importance of these works has lain in their application of a skeptical, rationalist (and thereby scientific) mode of thought not only to scientific problems but to political and theological controversy and general problems of understanding and perception. This fundamental challenge to both thought and language had profound repercussions in man's picture of himself and was reflected in what T.S. Eliot described as "the dissociation of sensibility," which Eliot claimed set in in Enland after the Civil War, whereby, in contrast to the Elizabethan and Jacobean writers who could "devour any kind of experience," later poets in English could not think and feel in a unified way.

In any division as arbitrary as one that begins and ends with the turn of a century, some acceptance of the overlapping of movements and influences is required. This is particularly true of the early part of the 17th century, which in many countries, particularly England, was still under the influence of the Renaissance. Thus, for the general convenience of the division, there have been included here those writers who belong chronologically to the 17th century (such as Donne, Milton, and Marvell) and who are yet a late flowering of Renaissance Humanism. From the point of view of influence alone, it has been necessary to discuss the plays of Jonson in the section on dramatists in *English literature of the Renaissance,* above, and yet to analyze in some degree Jonson's effect on poets of the early 17th century such as Robert Herrick.

If in most literatures the 17th century was ushered in by the Renaissance, real understanding of the forces at work in the period must take into account the way in which the 16th-century revival of interest in the classics and philosophy and the mighty challenge delivered to accepted dogma by the Protestant Reformation in religion prepared a receptive environment essential to the dissemination of the ideas of the new science and philosophy—of the work of Bacon, Descartes, and Hobbes.

The effects of war on 17th-century literature

A true picture of the period must also take into account the enormous effect of social and political upheavals, reflecting the struggle between the constituent "estates" of most emerging national states and the growing power of a centralized administration, during the early and middle parts of the century, which from many points of view provide a convenient breaking-off point for a new age. In England, where the literary history of the period is usually divided into two parts, the break seems to fall naturally with the outbreak of the Civil War (1642–51), marked by a closure of the theatres in 1642, and a new age beginning with the restoration of the monarchy in 1660. In France the bitter internecine struggle of the Fronde (1648–53) similarly divided the century and preceded possibly the greatest period of all French literature —the age of Molibre, Racine, Boileau, and La Fontaine. In Germany the early part of the century was dominated by the religious and political conflicts of the Thirty Years' War (1618–48) and thereafter by the attempts of German princes to emulate the central power and splendour of Louis XIV's French court at Versailles. The Netherlands was also involved in the first part of the century in a struggle for independence from Spain (the Eighty Years' War, 1568–1648) that resulted not only in the achievement of this but also in the "Golden Age" of Dutch poetry—that of Heinric Spieghel, Daniel Heinsius, and Gerbrand Bredero.

Baroque literature

The civil, political, and religious conflicts that dominated the first half of the century were in many ways also the characteristic response of the Catholic reaction to the Protestant Reformation known as the Counter-Reformation. The pattern of religious conflict was reflected in literary forms and preoccupations. One reaction to this —seen particularly in Italy, Germany, and Spain but also in France and England—was the development of a style in art and literature known as Baroque. This development manifested itself most characteristically in the works of Giambattista Marino in Italy, Luis de Góngora in Spain, and Martin Opitz in Germany. For long regarded by many critics as decadent, Baroque literature is now viewed in a less condemning light and is understood to denote a style the chief characteristics of which are elaboration and ornament, the use of allegory, rhetoric, and daring artifice.

If Baroque literature was the characteristic product of Italy and Germany in this period, Metaphysical poetry was the most outstanding feature in English verse of the first half of the century. This term, first applied by Dryden to John Donne and expanded by Dr. Johnson, is now used to denote a range of poets who varied greatly in their individual styles but who possessed certain affinities with Baroque literature, especially in the case of Richard Crashaw.

The conflict between the classics and the new science

Perhaps the most characteristic of all the disputes of the 17th century was that in which the tendency to continue and develop the Renaissance imitation of the classics came into conflict with the aspirations and discoveries of new thinkers in science and philosophy and new experimenters with literary forms. In France this appeared in a struggle between the Anciens and Modernes, between those who thought that literary style and subject should be modelled on classical Greek and Latin literature and supporters of native tradition. In Spain a similar conflict was expressed in a tendency toward ornament, Latinization, and the classics (*culteranismo*) and that toward a more concise, profound, and epigrammatic style *(conceptismo)*. This conflict heralded through the Modernes in France and the idea of *conceptismo* in Spain a style of prose writing suitable to the new age of science and exploration. The Modernes in France were largely, therefore, followers of Descartes. In England a similar tendency was to be found in the work of the Royal Society in encouraging a simple language, a closer, naked, natural way of speaking, suitable for rational discourse, paralleled by the great achievements in prose of John Milton and John Dryden.

ENGLISH LITERATURE IN THE 17TH CENTURY

The disturbed and violent climate of the 17th century in England resulted from many converging and conflicting ideas. It was often a conflict between tradition and innovation, as in the Anglican appeal to established order and expediency against Puritan simplicity and adherence to a literal interpretation of the Bible, or in King James's conception of the divine right of kings opposed to the many new theories of democracy. It is also seen in Thomas Hobbes's sacrifice of democratic liberties to the necessities of order as against John Milton's Puritan republicanism. But the predominant tone throughout all this is one of questioning examination and analysis.

Early prose. Most books were works of edification and instruction, so that the theory and practice of prose style moved from the rhetorical but disciplined flexibility achieved in the 16th century by writers such as Richard Hooker to the scientific and realistic style inaugurated in the 17th century by Francis Bacon. Whereas a balanced, rotund style imitative of Cicero was the natural vehicle for accepted truth, the skeptical new age found the stylistic irregularities of Seneca and Tacitus more suitable. While Elizabethan writers had thought and felt in metaphor, in prose as well as poetry, in the new age science promoted a breakdown of the old unity of thought and feeling. The scientific mind tended to outlaw metaphor and image, except as illustration, and to seek precisely denotative language. At all times plain prose was being written by travellers, men of affairs, and pamphleteers, who, innocent of theories, wrote in the pungent language they spoke.

The King James Version of the Bible

Translations and books of travel. This was the great age of translation. Philemon Holland, translating chiefly from the classics, combined scholarship with colour. John Florio's translation (1603) of the ***Essays*** of Michel de Montaigne, Thomas Shelton's ***Don Quixote,*** and Sir Thomas Urquhart's ***Works*** by Rabelais were done with a racy gusto that has made them English classics. Greatest of all English translations, the King James (Authorized) Version of the Bible, was mainly the work of a succession of 16th-century translators (and hence was somewhat archaic in 1611), yet Jacobean translators must share the glory through their taste in phrase and rhythm. Modem versions embody advances in learning and accuracy but suffer immeasurably by comparison. The cultural effects of the century's saturation in the English Bible are impossible to estimate.

Observation of foreign countries was a duty of the Renaissance Humanist that was to make the Grand Tour of Europe an obligatory part of the education of young aristocrats. Those at home found diversion in books of travel by such men as the eccentric Thomas Coryate, William Lithgow, Fynes Moryson, and Peter Mundy. Of the value of such narratives to imaginative writers ***The Tempest*** of Shakespeare is one example; and George Sandys' ***Relation of a Journey*** left traces in Milton's ***Paradise Lost.*** Among the absorbing pictures of American settlements were those of the amazing Capt. John Smith. For travels of all kinds, Samuel Purchas' collection ***Hakluytas Posthumas or Purchas his Pilgrimes*** (1625) revealed him a successor to the 16th-century voyager Richard Hakluyt.

The essay and "character." Increasing analysis of self and society was manifested in the development of the essay and the "character" (modelled on the brief sketches of ethical types by the ancient Greek author Theophrastus). Early essayists, such as Sir William Cornwallis, Ben Jonson in ***Timber: or, Discoveries,*** and Owen Felltham in ***Resolves,*** carried on the ethical preoccupations of Renaissance Humanism. Francis Bacon's ***Essayes*** (1597, much enlarged 1612 and 1625) were an integral part of a scheme for the advancement of knowledge, and his mundane scale of values was that of an objective social psychologist examining motives and behaviour. James Howell's *Epistolae* ***Ho-Elianae*** (1645–55) was a lively reminder of the kinship between the letter and the familiar essay. Robert Burton's ***Anatomy of Melancholy*** (1621) was a learned treatise on the related maladies of body and mind, but no other psychological treatise has been a treasured bedside book. This cloistered scholar was the master of a colloquial style, an ironic observer of the human comedy, and a wise, compassionate counsellor.

The best known books of "character" are Joseph Hall's ***Characters of Virtues and Vices*** (1608); the lighter and wittier collection of Sir Thomas Overbury; the ***Microcosmographie*** (1628–33) of John Earle, who combined wit, penetrating irony, and humanity; and Thomas Fuller's edifying and amusing conduct book ***The Holy State (The Profane State)*** (1642).

History and biography. Classical and modern foreign historians emancipated English historians from the naïve methods of the older chroniclers. The first really critical historian was William Camden, whose ***Britannia*** (1586) was translated from Camden's Latin by Philemon Holland in 1610. Sir Walter Raleigh's ***History of the World*** (1614) is apocalyptic, revealing history as a working out of God's providential will. Raleigh showed poetic vision in his brooding on the mortality of men and empires and military sagacity in his discussion of logistics. The nonapocalyptic tradition of Machiavelli and Tacitus was behind Bacon's study of statecraft, ***Historie of the Raigne of King Henry the Seventh*** (1622), whereas Thomas Fuller's ***Church-History of Britain*** (1655) was a moving story of the church's trials.

Memoirs and letters

The leading biographical writers were the clerical Fuller and a businessman, Izaak Walton, friend and eulogist of clerics, and the unclerical John Aubrey. Fuller, essayist, humorist, and antiquary, showed his love of anecdote and oddity in ***The Holy State (The Profane State)*** and in ***The Historie of the Worthies of England*** (1662), the first dictionary of national biography. Izaak Walton turned into a semi-professional biographer; his lives of Donne, Sir Henry Wotton, Hooker, George Herbert, and Robert Sanderson were printed during 1640–78; they have remained delightful and valuable through their author's sweetness of temper and his concern for truth and style. Fulke Greville's monument to Sir Philip Sidney was less of a biography than a political treatise. The period yielded a wealth of memoirs and letters that are nearly all highly readable. Perhaps the most famous were the poet-philosopher Lord Herbert of Cherbury's account of himself in a swashbuckling role; ***Eikon Basilike*** (1649), a stained-glass image of the royal martyr—the joint work of King Charles and John Gauden—which became a potent weapon of propaganda; and that idyllic "picture of my own disposition," Walton's ***Compleat Angler*** (1653). Among collections of letters, those of Dorothy Osborne stand first for their humour, pathos, and charm.

The influence of religion. In the national arena, religious issues were a more or less central force interwoven with politics and economics, and writers engaged in controversies that accompanied war and revolution.

In this century the sermon was the only literary form familiar to all classes of people. Although preaching was much more stressed by Puritans than by Anglicans, the names that chiefly survive are Anglican. Among the best are the "witty" and minute critical analysis and austere exaltations of Lancelot Andrewes and the large body of Donne's sermons. Jeremy Taylor is known for his richly pictorial images from nature and earnest moral counsel. He was in the classical tradition, having nothing of Donne's agonized involvement and little rhetorical violence. The best sellers were manuals of piety by Arthur Dent and Lewis Bayly and books of meditation and guidance, including Andrewes' ***Private Devotions*** (1647); Donne's very individual ***Devotions upon Emergent Occa-***

sions (1624); Taylor's *Rule and Exercises of Holy Living* and *Rule and Exercises of Holy Dying* (1650–51); the Puritan Richard Baxter's popular *Saint's Everlasting Rest* (1650); *The Whole Duty of Man* (1658) by Richard Allestree; and Thomas Traherne's radiant visions, which circulated in manuscript and were much later printed as *Centuries of Meditations* (1908).

Religion, science, and philosophy. The conflict between religion and science first took shape in the 17th century. The great Elizabethan exposition of divine and human reason and law had been the first book of Richard Hooker's *Of the lawes of ecclesiasticall politie*, which sought to adapt the traditional study of Aristotle, Augustine, and Aquinas to the Anglican Church's situation. This classical Christian philosophy of order and harmony encountered three agents of disintegration: Calvinism, with its emphasis on the arbitrary omnipotence of God and the impotent depravity of man; the Baconian doctrine of scientific progress, which tended to enthrone "things" and bestow power upon man; and traditional skepticism and naturalism fortified by scientific discovery and thought in the mechanistic philosophies of René Descartes in France and Hobbes in England.

Influence of Francis Bacon

Francis Bacon, an eminent lawyer, judge, and statesman, was the prophet and philosopher of a new conception of progress. In *Advancement of Learning* (1605), the broadest statement of his grand scheme and a monument of his massive prose, he arraigned traditional modes of education and inquiry and proposed new lines of exploration. His prime interest was science as the achievement of a mastery of nature for the use and benefit of man. His inspiring vision and his inductive method were set forth in *Novum Organum* (1620). His dream of cooperative research in *New Atlantis* (1627) to some degree played its part in the formation of the Royal Society.

The disruption of traditional religious thought proceeded rapidly. The dualism of Descartes led either to the philosophy of Idealism, with its concentration on God and the mind, or to Materialism (or Pantheism), which emphasized that matter was the only reality. For Hobbes, already a thoroughgoing Materialist, mind was, like everything else, a system of bodies moving in space and time; and his bare, trenchant style, as in *Leviathan* (1651), suited his thought. The deterministic philosopher was attacked on all sides. In the face of irrational Calvinism and irreligious Materialism, the Cambridge Platonists reasserted traditional Christian Humanism in partly new terms. Henry More and Ralph Cudworth were philosophically important in trying to accommodate science to liberal Christianity and vindicate metaphysical and ethical Idealism. More influenced Sir Isaac Newton and, in the 20th century, W.B. Yeats.

Sir Thomas Browne's *Religio Medici* (1642) was the work of a young scientist and a Christian Platonist conscious of a Rationalist climate. In this central book his rhetorical Senecan style gives foretastes of his later splendours. *Pseudodoria Epidemica* (1646) was a soberly scientific attempt to fulfill Bacon's desire for a sifting of popular errors, especially in natural history. Browne's chief later works were the. sombre *Letter to a Friend, The Garden of Cyrus*, and *Hydriotaphia, Urne-Buriall* (1658). In the last two, antiquarian in impulse with richly imaginative meditations on life and death, time and eternity, Browne blended Christian and pagan ideas with simple Saxon and complex Latinate words: he was the supreme metaphysical poet in prose.

Early poetry. Poets who spanned the end of the 16th century, such as the followers of Spenser, notably Giles and Phineas Fletcher and Michael Drayton, were in the mainstream of Elizabethan thought and are treated under *English literature in the Renaissance,* above. Samuel Daniel, in "Musophilus" (1599) and some weighty epistles, and George Chapman, in *The Teares of Peace,* upheld culture and Stoic integrity and wisdom against sensual or intellectual distractions. But, while Daniel wrote with a smooth half-prosaic clarity, the texture of Chapman's expression was metaphysical in its complexity. Fulke Greville's treatises in verse and his philosophical sonnet sequence had a sinewy toughness that is nearer Chapman than Daniel; but his dark religious view of man's corruption could not share the optimism of their Humanistic creed or the scientific optimism of his fellow statesman, Bacon.

Verse translations. The 17th century opened with some famous translations, including George Chapman's of Homer (1598–1615), Edward Fairfax's *Godfrey of Bulloigne* (1600), and George Sandys' translation of Ovid's *Metamorphoses* (1626). Chapman's elaborate metaphors and moralizing interpolations were rather obtrusive, but his translations were truly heroic poems in the Renaissance tradition. Fairfax's translations of the 16th-century Italian poet Tasso, in Spenserian style, has remained a standard English version. Joshuah Sylvester's *Divine Weekes and Workes* (1605), a translation of du Bartas' *Création du monde*, was the Protestant epic of creation until displaced by John Milton's *Paradise Lost.* Some of these translations, notably Sandys', contributed to a development of the closed, balanced couplet, which culminated in Dryden's translation of Virgil's *Aeneid* and in *Fables Ancient and Modern* (1700). In the vocabulary they developed to render "philosophic" poems, Sylvester and Sandys illustrate the beginnings of the "poetic diction" of the Augustan Age of Queen Anne, and Sandys' allegorical commentary on Ovid, one of the last flowerings of a potent tradition that could not be assimilated by Augustan reason, was to revive in the mythological imagination of John Keats.

Donne and the Metaphysical poets. The revival of general interest in the poetry of John Donne and the Metaphysicals in the 20th century and the effects of this upon modern poetry and criticism have altered the present perspective. Some elements of their poetry are "wit"; a forcible yoking of apparently unrelated ideas and things; organic rather than merely illustrative images; concrete particulars; fusion of thought and feeling, or an intellectualization of feeling; tension between opposed ideas and emotions, and hence complexity, obliquity, irony, paradox; colloquial language and rhythms; and an expressive rendering of experience. Many poets, from Shakespeare down, were more or less Metaphysical; but Donne carried into the poetry of love and religion the technique that Elizabethan decorum reserved for satire. In his love poetry he could be cynical or sensual or outsoar Petrarchan hyperbole. He was concerned with the relations of body and soul and, in some of his best poems, found in love the only reality. His explorations were presented with explosive force and originality; they were witty and serious, dialectical and passionate. The religious poems, including the holy sonnets, were mostly written in the years 1607–12. Although Donne was ordained only in 1615, he had always been a Christian and re-enacted in his soul the drama of human guilt and divine love.

Characteristics of Metaphysical poetry

Donne's poems circulated in manuscript and exerted a strong influence on secular and religious poetry. Of the chief religious writers, Herbert, Crashaw, Henry Vaughan, and Traherne, only the first really wrote Metaphysical poetry. Herbert's richly symbolic wit was demonstrated in a poet-priest's celebrations of Christ and church and in personal poems in which he strove to subdue his will or quicken his zeal. Tensions made impact through his subtle craftsmanship, organic structure and unified imagery, his understatement, colloquial language, and variety of rhythm. Crashaw's work, that of a Catholic convert, was the very embodiment of Baroque poetry, and the Italianate and Spanish style of his devotional ecstasies had little of the English Metaphysical. Vaughan, a Christian Neoplatonist who, but for intimations of God in nature, felt an exile on earth and longed for eternity, was a poet of contemplative vision.

Secular poets included Lord Herbert, Henry King (author of the splendid "Exequy" on his young wife), the fantastic and satirical John Cleveland, and Abraham Cowley, author of clever amatory verse, the unfinished epic *Davideis,* and loose, irregular *Pindarique Odes* (1656) on public themes. The finest of them was Andrew Marvell. In "To His Coy Mistress," "The Garden," "The Picture of Little T.C.," "A Dialogue Between the Soul and Body," "The Coronet," and "An Horatian Ode upon

Marvell's poetry

Cromwell's Return from Ireland," all written about 1650–53, Marvell combined complex metaphysical intelligence and sensibility with the grace of a Cavalier and a structural and stylistic symmetry, economy, and clarity that may be called classical.

Ben Jonson and his influence. In the golden age of English music, before and after 1600, numberless musicians created a wealth of song. Thomas Campion is the most noted example of the composer-poet. Along with elaborate lyricism went the simpler popular tradition; the songs of Jonson and Shakespeare illustrate both. Jonson's classicism extended from the ethical and Stoic ideals of Renaissance Humanism (embodied in reflective poems) to the artistic ideals that molded his songs and epigrams. He cultivated disciplined impersonality, symmetry, and clarity; his style was rational and urbane, without shocks or extravagance.

Among the many writers who felt Jonson's influence, perhaps the purest disciple and the most original was Robert Herrick. His Metaphysical wit, his instinct for the concrete particular, for ambiguity, paradox, and surprise, was half concealed by his classical elegance, which included an exquisite feeling for word and rhythm. Herrick could live in the pagan world of his fancy or in the Christian world of his belief. The Cavalier poets Thomas Carew, Sir John Suckling, and Richard Lovelace were also heirs of Jonson. Carew, the most studied artist, praised Donne's originality but showed little of Donne's passion and intellectualism. He preferred a classical pattern, one controlling idea and image. If Carew elevated the sensual game of love into a courtly ritual, Suckling speaks with the nonchalance of a courtier and man-about-town. Lovelace lives in a few songs of the ideal Cavalier, but even in these simplicity is touched with complexity.

John Milton. Milton's career, one of continuous development, fell into three periods: the early poems in Latin and English; the years 1641–60, given to public affairs (and occasional sonnets); and the last phase of his three major poems. The volume of 1645 included the ode of Baroque beauty "On the Morning of Christ's Nativity" and the two companion pieces "L'Allegro" and "Il Penseroso," which develop the Jonsonian manner; his masque, now known as *Comus*, diversified in its stylistic texture, unified in its Christian Platonism; and "Lycidas," one of the most complex poems in the language, a first effort to justify God's ways to men, in which religious, political, and personal tensions are powerfully stated in the metaphor of classical pastoral. The Metaphysical current left the young classicist almost untouched; he began in the Spenserian line but was from the start a master of original power.

Milton dreamed of writing the great modern heroic poem, but his sense of responsibility impelled him to give 20 years to the Puritan struggle for liberty. He wrote five tracts against episcopacy; four on divorce; "Of Education"; "Areopagitica" (1644), a plea for freedom of the press; and, on the eve of the Restoration, a defiant plea for a republic. They record his increasingly radical thought, his hopes of a reformation, and his successive disillusionments.

Paradise Lost

When he came to compose ***Paradise Lost*** (1667), Milton was no longer a militant revolutionary; his three major works deal with the temptations, defeats, and victories of individual man. His early ardour gave way to reasserting in the face of science and oppressive power the Christian Humanist view of God, man, and nature through an inspired use of the biblical myth of the Fall of man. Eve, in her desire for the godlike knowledge Satan promises, and Adam, in his curiosity and his following of Eve, re-enact the sin of Satan, pride and rebellion. His heroic villain, Satan, is a personality of grandeur, but in balancing this with his corruption Milton could rely upon his audience's reaction to evil.

He wrote as the conscious heir of the ancients, revivifying the conventions and details of the classical epics. Images blend the general with the particular, and his bold, idiosyncratic style is elevated above common speech, though ornate stylization includes both simplicity and complex density of suggestion and overtone. The use of blank, or unrhymed, verse (10-syllabled lines of alternately stressed syllables) for a long poem was a radical novelty, and Milton's handling of it enlarged English prosody.

Paradise Regained and ***Samson Agonistes*** (both 1671) show still further artistic developments. The former is in a style of almost biblical simplicity; the latter, the only English drama on the Greek model that can stand with those of the ancients, is in ruggedly irregular verse that comes close to speech. The hero of ***Paradise Regained***, in a conflict of mounting tension, Christ displays the obedience and integrity rhat Adam and Eve have lacked. In the drama of Samson, a sinful human hero—isolated, like the half-human Christ, even from human sympathy—achieves inward regeneration. These three poems of faith and fortitude stand out against the rest of Restoration literature.

The Restoration period. The restoration of the monarchy in 1660 heralded a political and cultural break, but new developments came from the ferment and confusion of new and old. The year 1667 saw the appearance not only of the first edition of the old Milton's ***Paradise Lost*** but also of a modish poem, the young Dryden's ***Annus Mirabilis.*** Isaac Newton in physics and mathematics, Robert Boyle in chemistry, and developments in biology strengthened radical change in the picture of the physical world, propagated by the newly formed Royal Society, which profoundly affected the intellectual temper of the age. A new political feeling emerged, in some ways cynical, corrupt, and self-seeking yet marked by increasing toleration and humanity. It allowed the growing consciousness of the self that is found from Francis Bacon onward and that produced the diaries of Samuel Pepys and John Evelyn.

The court returned from exile with many French tastes and fashions. London became a European capital of high civilization. The theatre, reopened under royal patronage, became a preserve of the upper class and flourished. The literary forms of the modern world—the novel, biography, history, travel writing, and journalism—had their beginnings in the Restoration period, when English literature could be seen remaking itself in the face of new scientific and philosophic concepts and new social and economic conditions that caused the disappearance of patronage, the emergence of mass readerships, and the formation of the literary market.

Prose. The year 1660 marked a development in English prose. Men became increasingly interested in things rather than in words as science came to play a greater part in human affairs and the custom of writing emotional prose disappeared and a more familiar diction took its place. The Royal Society played some part: its origins were much earlier than 1662, when it received its first charter. The society was interested in prose style and soon after its foundation appointed a committee "for improving the English language." Thomas Sprat, in his ***History of the Royal-Society of London*** (1667), really a work of propaganda, laid down rules for a plain English style, though many scientific works, such as those by the great naturalist John Ray, were in Latin. Robert Plot's ***Natural History of Oxfordshire*** (1677) was in a simple style, but he still thought it necessary to write:

The Royal Society

> I intend to deliver as succinctly as may be, in a plain easie and unartificial Stile, studiously avoiding all ornaments of language, it being my purpose to treat of things, and therefore would have the Reader expect nothing less than words.

Hobbes, in his ***Leviathan*** (1651), had scorned and derided "obscure, confused and ambiguous Expressions, also all metaphorical Speeches, tending to the stirring-up of Passion," but probably the rise of journalism was as influential as any other cause. During the Civil War and the Commonwealth, pamphlets had been abundant. They were plain and simple, as a political pamphlet must be if it is to be effective.

The sermon, too, became simpler after 1660. John Wilkins, bishop of Chester, advocated a plain style in his discourse on preaching, ***Ecclesiastes*** (1646), and in his ***Essay towards a Real Character and a Philosophical Lan-***

guage (1668) he made a gallant attempt to substitute ciphers for words so that their associations should be lost. But it was probably the influence of a few individual writers rather than vague tendencies of the time that made a simpler prose the usual method of expression. John Tillotson was a lucid and logical preacher. He almost belonged to the 18th century in manner, and his influence was for a time very great. Dryden considered Tillotson a writer akin to himself.

Dryden's prose

Dryden was the most influential prose writer from 1660 to the end of the century. His first long piece of prose, the "Letter to Sir Robert Howard" prefixed to *Annus Mirabilis*, was followed by *Of Dramatick Poesie, an Essay* (1668) and some dedications and prefaces and, in 1700, by his masterpiece, the preface to *Fables Ancient and Modern*. Dryden's was a style both lucid and conversational. Under all his seeming ease he took great care and actually revised the syntax of *Of Dramatick Poesie* in the second edition. There is no doubt that the subject matter with which he and his contemporaries were dealing led to a change in style.

No longer were problems such as the transience of life or death itself, which lent themselves to rhetorical treatment, discussed, but rather attention was turned to matters of immediate concern, such as criticism or politics. Innumerable books were written showing disapproval of pedantry. The most amusing and popular was John Eachard's *Grounds and Occasions of the Contempt of the Clergy*. In 1672 Marvell introduced a vein in prose called "drolling" (or "banter"). In *The Rehearsal Transpros'd* and *The Rehearsal Transpros'd: The Second Part*, he ridiculed Samuel Parker, later bishop of Oxford, for his bigotry. In 1675 Herbert Croft, bishop of Hereford, argued that Protestants differed about nothing essential, a view furiously attacked. Marvell took Croft's side in *Mr. Smirke; Or the Divine in Mode*, producing an amusing satire. Marvell helped to make prose lighter and more familiar.

In 1678–79 there was an outpouring of pamphlets over the Popish Plot, supposedly a conspiracy to recover England for the Roman Catholic Church but in fact a fabrication of Titus Oates. One writer stood out: George Savile, marquess of Halifax, who wrote *A Letter to a Dissenter* to try to heal differences in the state and *The Character of a Trimmer* to defend his own position. His masterpieces were the *Maxims and Character of King Charles the Second*.

Bunyan's works

John Bunyan belonged to the world of artisans, wage earners, yeomen, and shopkeepers whose culture was founded on Bible reading and the Puritan sermon, ballads, songs, newsbooks, and romances. A tinker by trade, he was one of many "mechanick preachers" of the period who printed sermons and religious discourses. His first notable book was the spiritual autobiography *Grace Abounding* (1666). His masterpiece, *Pilgrim's Progress* (first part, 1678), grew directly out of *Grace Abounding*. It is a translation into allegorical terms of the soul's journey to salvation. It gives a vivid picture of the life of the high road and small country towns of the reign of Charles II. In the second part (1684), the realistic and humorous elements are more prominent. Among Bunyan's other works the most memorable are *The Life and Death of Mr. Badman* (1680) and *The Holy War* (1682), an allegory on a grand scale. Bunyan can be regarded as the last medieval allegorist and as the first great master of realistic and psychological fiction.

The writings of the Cambridge Platonists (a group of philosophico-religious thinkers who hoped to reconcile Christian ethics and Renaissance Humanism) were among the influences that led to simplification of prose style. Most notable were John Smith's *Select Discourses*, and the *Select Sermons* (1698) and *Moral and Religious Aphorisms* (1753) of Smith's master, Benjamin Whichcote, and *A Discourse of the Freedom of the Will* (1675), by Peter Sterry, the most imaginative and poetical of the group. A great Puritan leader, Richard Baxter was one of the most important religious writers, remembered for *The Saint's Everlasting Rest, A Call to the Unconverted* (1657), *A Breviat of the Life of Margaret . . . Charlton*, and *Reliquiae Baxterianae* (1696), his autobiography, a fascinating account of his attempts to heal the differences between Christians of different persuasions.

Restoration interest in biography

One of the most important developments in Restoration prose was an interest in biography: the word seems to have been first used in English *c.* 1660. Earlier biographies had tended to be either eulogy, in the tradition of funeral oration, or hagiography, deriving from lives of saints. The famous *Lives* by Izaak Walton, though still containing hagiography, were landmarks in the development of English biography. They were based on observation and research and contain invaluable records of personal traits and conversations and examples of correspondence. Almost equally important is Roger North's *Lives of the Norths*. The life of John North is of particular interest in that it successfully describes an extremely neurotic man. The Earl of Clarendon's *History of the Rebellion* was first published in 1702–04, and his autobiography appeared in 1759. Bishop Gilbert Burnet's *History of His Own Time* was published in 1724–34, although lives of the Earl of Rochester, Sir Matthew Hale, and Bishop William Bedell, with an "essay in memory" of Queen Mary II, were published much earlier. Despite their publication dates, however, both these great histories were written during the Restoration period. *Athenae Oxonienses* (1691–92; "Oxonian Athens"), by an Oxford antiquary, Anthony à Wood, was the first serious attempt at an English biographical dictionary. It was partly the work of John Aubrey, whose own *Brief Lives* is vivid and entertaining. It is hardly necessary to mention the diaries of John Evelyn and Samuel Pepys except to remind the reader that, whereas Evelyn's *Diary* extended from 1631 to 1706, Pepys's ended in 1669. As Evelyn often wrote up his diary some days after the events recorded, his dates cannot always be trusted. Neither was published until the early 19th century.

During the Restoration, the problem of translation troubled many. No doubt a more exact scholarship than that of the Elizabethans raised in an acute form the question of what a translation should be: whether it should be an exact reproduction of the original or a paraphrase. The problem was not, and never can be, solved. Dryden produced admirable versions of portions of Lucretius, while his Virgil (1697), if not Virgil, has its own splendours, and his work helped to form the poetic diction of Augustan verse. Hobbes made complete translations of the *Iliad* and the *Odyssey*, which Dryden called "bald," adding that Hobbes began studying poetry as he did mathematics, when it was too late. The taste for translations must have been keen, as they are so often found in the miscellanies of verse of the time.

Critical essays in verse and prose ranged over literature in general. The most outstanding work was Robert Wolseley's preface to the Earl of Rochester's version (1685) of John Fletcher's tragedy *Valentinian*. It attempted to solve the problem of the relation of art to morals, and, with the exception of Dryden's prefaces, it is the most significant piece of Restoration criticism. *A Short view of the Profaneness and Immorality of the English Stage* (1698), by Jeremy Collier, effected its purpose to some extent in making plays more decorous. Collier's violence gave him an advantage over William Congreve, Sir John Vanbrugh, and others who attempted replies.

Poetry. The impulse to the writing of ceremonial verse was given by the restoration of Charles II. Robert Wild's *Iter Boreale* ("Northern Journey") greeted the King's return. Dryden was soon in the field with *Astraea Redux*. In 1661 he wrote a panegyric on the coronation and in the next year addressed a poem to Clarendon. These poems helped to establish the use of the heroic couplet (rhymed pair of 10-syllabled lines), as they were technical improvements on the poems in couplets that had preceded them. But all these thin folios could not compare in popularity with Samuel Butler's mock-heroic satire *Hudibras* (first part 1663). Much of this remarkable work had been written a good while before the Restoration. The poem delighted all Royalists, but it was less an attack on the Puritans than on the folly and pedantry of mankind in general. Though never finished, it is a great work and remained popular until the 19th century. ***Hudibras*** was constantly imitated, notably by Swift.

Dryden's poetry

Dryden presided over the age. Apart from his prose, he wrote nearly 40 plays as well as a great deal of verse. His most famous poems were nearly all satires. ***Mac Flecknoe,*** his attack on Thomas Shadwell, was published surreptitiously in 1682, though written earlier. His best known poem, ***Absalom and Achitophel*** (1681), had the supreme merit of satire in being reasonably fair and was immensely popular. ***The Medall,*** also a political poem, followed in 1682, as did his ***Religio Laici,*** which indicated the evolution of Dryden's religious opinions. ***The Second Part of Absalom and Achitophel,*** although by Nahum Tate, contains about 200 of Dryden's most telling lines. In 1687 he published ***The Hind and the Panther,*** an able argument for Roman Catholicism. Among Dryden's greatest achievements are his Pindaric odes in irregular rhyming verse. He took this form from Abraham Cowley and gave it a new strength in his superb elegaic ode "To the pious Memory of the accomplisht Young Lady Mrs. Anne Killigrew" and in memorable poems written for a London musical society, "A Song for St. Cecilia's Day" (1687) and "Alexander's Feast" (1697), in which colourful rhythm and magnificence of imagery combine with touches of wit to produce effects unique in English poetry.

At the Revolution of 1688, Dryden lost his posts of poet laureate and historiographer royal. With unflagging vigour, late in life, he poured out a series of works, including some fine plays, his translation of the Roman poet Virgil, and ***Fables Ancient and Modern,*** containing some of his most mature writing, including three verse tales founded on stories by the 14th-century Italian Boccaccio, which were admired and imitated by Lord Byron and John Keats. His alliance with the publisher and anthologist Jacob Tonson was a literary event of importance. The translation of Virgil published by Tonson in 1697 was the first large-scale publishing venture in England and probably the first to bring considerable profits to an author.

Poets surviving from the reign of Charles I, such as Abraham Cowley, Sir John Denham, and Edmund Waller, continued to produce competent verse. More typical was the poetry of the group of Restoration "court wits," which included Rochester, Sir Charles Sedley, Charles Sackville, and Sir George Etherege. Sedley, Sackville, and Etherege wrote admirable light verse continuing the courtly tradition of Carew and Suckling and at times catching the note of the contemporary popular street ballad. Rochester was the one considerable poet among the group. His best lyrics had an intense feeling and a purity of melodic form. His verse satires were among the most notable of the century, ranging from the devastating political comments of "The History of Insipids" (1676) to the powerful attack on facile optimism in ***A Satyr Against Mankind*** (1675), his masterpiece. Indeed, these works were apparently written before Dryden had published his satires, and there is little doubt that he owed a considerable debt to Rochester's example.

Drama. The Restoration period is chiefly remembered for its many plays. The King and the Duke of York gave their patronage to the drama and re-established the London theatres, closed in 1642. Drury Lane, known as the King's house, and Dorset Garden, known as the Duke's, were founded soon after the King's return.

Heroic plays

There were few real tragedies, their place being taken by heroic plays, mostly dealing with the conflict between private life (love) and duty (political power or status). These offsprings of heroic romance and French tragedy were a mixture of bombast with occasional fine passages. Dryden's ***Conquest of Granada*** (performed 1670–71) *is* the best known, and ***All for Love*** (1677) can be read with enjoyment if it is remembered that it is not meant to be an imitation of Shakespeare's ***Antony and Cleopatra*** but a re-ordering of Shakespeare's themes. Two of his comedies, ***Marriage-A-la-Mode*** (published 1673) and *Amphitryon* (published and performed 1690), are excellent.

Of other playwrights, Thomas Otway wrote well-constructed plays. His masterpiece, a topical political melodrama, was ***Venice Preserv'd*** (performed 1682), which has a tide of passionate language and bold delineation of character. Real comedy began with Sir George Etherege, who, in ***The Man of Mode,*** mirrored the gaiety and insolence of the world he knew. William Wycherley was a writer of comedy with a serious purpose: his ***Country-Wife*** (published 1675) has great power, while ***The Plain-Dealer*** (1677) is a skillful transplantation of Molière's ***Misanthrope.*** Wycherley scavenges in the most callous good spirits and with careful cynicism, and his prose is excellent. The apex was reached by William Congreve, who, had his plots been coherent, would have been one of the world's great dramatists. His dialogue in ***Love for Love*** (performed 1695) and ***The Way of the World*** (1700) has hardly been surpassed for sheer wit. Sir John Vanbrugh wrote ***The Provok'd Wife*** and ***The Relapse*** (both published 1697); and though he was a less stylish writer than Congreve, his plays are full of vigour and humour. The coarse, powerful comedies of Thomas Shadwell formed a link between the art of Ben Jonson and the realistic fiction of the age of Henry Fielding.

Scottish literature. With the union of the crowns in 1603 and the transference of the court to London, Scottish writing became increasingly Anglicized. Although James VI showed that he could turn out a lyric in Scots, the new poets wrote almost exclusively in English. The best of these, William Drummond, was "well-languaged," but his lyrics lack intensity and force. Vernacular poetry was kept alive in popular songs and ballads, such as Robert Sempill's ***Life and Death of Habbie Simson.*** Whether grim and tragic, as in the older ballads, or amusing, as in ***Habbie Simson,*** this tradition was of the heart rather than the head, and the absence of an intellectual and critical quality has been a besetting weakness in later Scots verse.

The first original literary prose appeared in the theological writing of John Ireland (*c.* 1435–c. 1500). The prose of these works had not yet developed great variety or flexibility of construction. Early in the 16th century there was an attempt at a less Latinized prose in Murdoch Nisbet's version of Wycliffe's New Testament, but the language is too southern to be quite successfully idiomatic. Some English influence is seen also in the 16th-century historical writings of John Bellenden, John Leslie, and Robert Lindesay of Pitscottie and to a greater extent in works such as the Scottish Reformation leader John Knox's ***Historie of the reformaiioun of religioun within the realm of Scotland*** (1586). These writers, however, had a good command of narrative, and Pitscottie and Knox had also the gift of style. Almost as vivid as Knox's writing was the autobiographical ***Historie of the Lyff of James Melvill,*** which covered the second half of the 16th century. The remarkable anonymous ***Complaynt of Scotlande*** (1548–49) is interesting both as the work of a well-read Scottish patriot and as the first work of a Scottish experimenter in prose style. The need to reach English readers is seen in the writings of James VI: the manuscript of his ***Basilikon Doron*** was in Scots, but, when published in 1599, it had been translated into English. William Drummond's ***Cypresse Grove*** (1623) was in careful, grave, mannered English; and not much Scots prose occurred after that date. A general justification for this turning to English may well be seen in Sir Thomas Urquhart, "the free'st spoke Scot of any," whose translation of François Rabelais and extraordinary original works displayed an ornateness that would have cracked the molds of Scots as it did those of English.

Anglo-Irish literature. Until the 17th century the language of Ireland was Irish. The Anglo-Irish were a small colony, a garrison holding a perilous outpost, constantly yielding linguistic as well as physical ground to the resurgent Gael; and even among the colonists English had to struggle for supremacy with Latin and French. A few manuscripts have survived, showing the existence of a vernacular poetry of no great originality or distinctiveness. The most striking example is the 13th-century satire ***Land of Cockaygne.***

The first notable Irish writer in English was Richard Stanyhurst, who contributed sections on Ireland to Raphael Holinshed's *Chronicles* (1577) and attempted to write English verse on Latin models. As the colony grew and the conquest of the whole country was attempted, the stream of Anglo-Irish writing increased, and Gaelic as well as English names began to appear. In the next century there were Sir John Denham, author of the much admired *Cooper's Hill* (1642), and Lord Roscommon, Nahum Tate, and Nicholas Brady, among minor poets. Sir John Temple (1600–77) may be mentioned among minor historians, but there was as yet no major name.

CELTIC LITERATURE IN THE 17TH CENTURY

Welsh. *Prose.* So far, writers of Welsh prose had contented themselves with translation, until Morgan Llwyd produced his religious works. A Puritan, he made an original contribution to Welsh religious thought, chiefly in *Llyfr y Tri Aderyn* (1653; "The Book of the Three Birds", a disquisition on the theory of government and on religious liberty, and *Llythur ir Cymru Cariadus* (*c.* 1653; "Letter to the Beloved Welsh"), which expounded a mystical gospel. During this period and well into the 19th century, there still appeared many translations, mostly of theological works and often inspired by the Society for Promoting Christian Knowledge. Among the clergy who produced these works were Edward Samuel; Moses Williams, a diligent searcher into manuscripts; Griffith Jones, the father of Welsh popular education; and Theophilus Evans, author of *Drych* y *Prif Oesoedd,* which, like the *Llyfr y Tri Aderyn,* is established as a classic in the annals of Welsh literature. Ellis Wynne o Lasynys is often regarded as the greatest of Welsh prose writers. His two great works were *Rheol Buchedd Sanctaidd,* a translation of Jeremy Taylor's *Rule and Exercises of Holy Living,* and *Gweledigaetheu y Bardd Cwsc,* an adaptation of a translation of the *Sueños* of the Spanish poet Quevedo.

Poetry. When Henry VII came to the throne, the old Welsh gentry turned toward England for recognition and preferment, and interest in their own country waned. Soon, poets of the older school had no audience, and the only ones to keep the old tradition were the rich gentlemen farmers. A new school, however, was rising that combined a vast store of folk song, previously despised and unrecorded, with imitation of contemporary English popular poetry and sophisticated lyrics. The first landmark of this new development was the publication of Edmund Prys's metrical version of the Psalms and Rhys Prichard's *Canwyll y Cymry* (1659–72; "The Welshman's Candle"). These were written in the so-called free metres like those of English poetry and not in the strict metres peculiar to Wales. The metrical Psalter is of particular importance, as it contained the first metrical hymns. Prichard's work consisted of moral verses in the metres of the old folk songs (*penillion telyn*). Many other poets wrote in these metres, but they were generally crude until handled by the greatest poet of the period, Huw Morus, who was particularly famous for his love poems. Later came Lewis Morris, the inspirer and patron of Goronwy Owen and thus a strong link with the next period, which Goronwy Owen initiated and which was to be one of the most productive in Welsh poetry.

The folk
element in
Welsh
poetry

Breton. A 17th-century collection entitled *Cantiques bretons* (1642) names several Breton airs. All the remaining works of the middle period (11th to 17th centuries) were religious and mostly in verse. Three mystery plays were probably the most significant products of the period: *Buez santez Nonn* ("Life of St. Nonn"), *Burzud bras Jezuz* ("The Great Mystery of Jesus"), and *Buhez santes Barba* ("Life of St. Barbara"). Three long poems, "Tremenvan an itron gwerches Maria" ("The Passing of the Virgin Mary"), "Pemzec levenez Maria" ("The Fifteen Joys of Mary"), and "Buhez mabden" ("The Life of Man"), were all probably based on French versions. A book of hours in verse, a prose extract from the Leon missal, and a prose catechism also belong to this period, as does the prose *Buhez an itron sanctes Cathell, guerches ha merzeres* ("Life of St. Catherine, Virgin and Martyr"). *Am Mirouer an Confession* ("The Mirror of Confession") and *Doctrin an Christenien* ("Christian's Doctrine") are translated from the French. A collection of carols, *An Nouelou ancient ha devot,* appeared in 1650, and a book of metrical meditations appeared in 1651. In general, the Middle Breton literature lacked originality, and the indigenous culture of Brittany seems to have been entirely neglected by the educated classes, who introduced an enormous number of French words into the preponderantly religious works published.

Scottish Gaelic. The 17th century was a period of great interest in Gaelic Scotland. The political, ecclesiastical, and social structures of Scotland were gradually changing, and so was the relationship between the central government and the Gaelic area. Enough survives of the Gaelic poetry to show that there was a large number of poets of great talent, and there seems to have been a diffusion of artistic talent scarcely matched in any other period in Scottish Gaelic history. It was the great age of the work songs (or waulking songs) and of the classical music of the bagpipes. Some of the poetry and prose was contained in three 17th-century manuscripts. The first two were the *Black* and the *Red Books of Clanranald,* written by members of the MacMhuirich family, who were latterly hereditary bards to the MacDonalds of Clanranald. They were probably written for the most part in the 17th century but contained poems by earlier representatives of the family. The other important document was the Fernaig manuscript, compiled between 1688 and 1693, containing about 4,200 lines of verse, mostly political and religious. The poets represented belonged to the 16th and 17th centuries, although little work by the major 17th-century poets was given.

The two best known poets of the 17th century were Mary Macleod and Iain Lom. The former, known as Màiri Nighean Alasdair Ruaidh, was closely associated with the households of the Macleods of Harris and Dunvegan. Her poems show deep personal emotion, and her style is fresh and natural. She inherited the imagery of the bardic poets but placed it in a new setting, and her metres were the comparatively new strophic ones (having repeating patterns of lines) rather than the strict syllabic metres of the classical bards. John Macdonald, known as Iain Lom, took an active part in the events of his time. His life spanned an eventful period in Highland history, and his poetry reflected this. He composed poems about the Battles of Inverlochy and Killiecrankie, a lament for the Marquess of Montrose, a poem on the restoration of Charles II in 1660, several poems dealing with the Keppoch murder of 1663, and a song bitterly opposing the union of the Parliaments in 1707. The breadth of his interest in national affairs was notable. His versification was less melodious than Mary Macleod's, but he had that compression and concentration of expression lacking in so much later Gaelic poetry. A significant body of bardic verse by Cathal MacMhuirich survives.

Mary
Macleod
and
Iain Lom

Other 17th-century poets of whose works interesting fragments have survived include Donnchadh MacRaoiridh, whose best known poem consists of four calm, resigned verses composed on the day of his death. The work of Alasdair MacKenzie and of Murdoch Mackenzie was represented in the Fernaig manuscript. Roderick Morison, known as An Clàrsair Dall (the Blind Harper), became harper to Iain Breac Macleod of Dunvegan. Little of his work has survived, but the strong texture and poetic intensity of his *Oran Mór MhicLeòid* and his *Creack na Ciadain* were remarkable. Dorothy Brown and Silis na Ceapaich were women poets of great talent, and Dorothy Brown's poem on Alasdair Mac Colla and Silis na Ceapaich's laments for Lachlan MacKinnon and for Alasdair of Glengarry were moving and artistic utterances.

Four other poets mark, in various ways, the transition from the poetry of the 17th century to that of the 18th. They were Lachlan MacKinnon (Lachlann Mac Thearlaich Oig), John Mackay (Am Piobaire Dall), John Macdonald (Iain Dubh Mac Iain 'Ic Ailein), and John Maclean (Iain Mac Ailein). All four were born *c.* 1655–65. Much of their work lacked the economy of the best 17th-century poetry and also its richness of imagery.

John Macdonalds' ***Oran nam Fineachan*** was not the worst of those tedious jingles in praise of the various clans that were popular with 18th-century poets. John Maclean showed a great interest in early Gaelic legend, and he composed amusing verses in a mock-heroic style. John Mackay's ***Coire an Easa*** was significant in the development of Gaelic nature poetry. Finally, bardic poetry continued to be composed into the 18th century by two members of the MacMhuirich family, Niall and Domhnall.

Irish. During the reign of Elizabeth I, English policy succeeded in breaking the power of the Irish and Anglo-Irish nobility, and the process was completed by the middle of the 17th century. The dispossession of this nobility entailed the disappearance of the bards who were its dependents and propagandists. Their eyes fixed on the past, the bards had shown no interest in the renaissance of learning; with their elimination the old tradition was doomed. From this time Irish itself began to slip from its position as the common speech of the country.

Verse. Hardly any correct bardic verse was written in Ireland after 1650, but just as the bards had taken over from the ***filid***, the first professional seers or poets, 500 years before, so new poets took over from the bards. And just as the bardic measures had been in preparation for centuries before they established themselves as canonical, so the song metres that replaced them had existed for centuries among the people. The new poets abandoned the syllable counting by lines with a fixed number of stresses; the stressed vowels rhymed in patterns that might be very simple or, later, bewilderingly intricate, but simple vocalic assonance took the place of earlier rhyme. The language of poetry moved toward that of the people. While they had little patronage, there was at least an increasing supply of paper, so that their works, still barred from the printing press, were able to circulate. The tone of verse throughout the 17th century was of passionate defiance of the new regime. In it is found the first coherent expression of patriotism conceived as devotion to an abstract ideal rather than as loyalty to an individual, but much of the verse represents a mere nostalgia for the past.

Structural changes in poetry

The greatest poet of the song metres was Dáibhidh Ó Bruadair, who still had some patrons. After him the poetic tradition, while remaining subtle in craftsmanship, was preserved only among the common people. But it had real vigour in satire and ***pastourelle,*** and peasant poets maintained the tradition until the end of the 18th century.

Prose. During the 17th century a good deal of prose was produced, mainly antiquarian and of inestimable value today, for the writers drew on sources now lost. The most important is ***Annála Rioghachta Éireann*** ("The Annals of the Four Masters"), a compilation of all available material on the history of Ireland to 1616, directed by Michael O'Clery. Geoffrey Keating, writing dignified, simple prose based on bardic language, produced the first historical (as opposed to annalistic) work in his ***Foras Feasa ar Eirinn*** ("Foundation of Knowledge of Ireland") as well as some fine verse in both old and new metres and two spiritual treatises.

An interesting development in prose style was the satire ***Pairliment Chloinne Tomáis*** ("Parliament of Clan Thomas"). It appears to be by a representative of the bardic order, for it attacked with equal savagery the new ruling class and the native peasantry, using a style close to that of the earlier *crosánacht* but with prose now predominating over verse. It found several imitators, but the old tradition was by this time too attenuated for so aristocratic an attitude to be maintained. The imaginative prose was of a more popular kind; it consisted of developments of the Fenian themes or of romances in which themes from Irish and foreign medieval literature were mingled with folklore and ***fabliau*** (short metrical tale) elements. As in the case of the song metres, these romances had a considerable tradition before they appeared in writing. Nevertheless, the public was growing apathetic and less able to read Irish, so there was little real hope for prose production.

AMERICAN LITERATURE IN THE 17TH CENTURY

American literature at first was naturally a colonial literature, by authors who were Englishmen and who thought and wrote as such. John Smith, a soldier of fortune, is credited with initiating American literature. His chief books included ***A True Relation of . . . Virginia . . .*** (1608) and ***The generall Historie of Virginia, New England, and the Summer Isles*** (1624). Although these volumes often glorified their author, they were avowedly written to explain colonizing opportunities to Englishmen. In time, each colony was similarly described: Daniel Denton's ***Brief Description of New York*** (1670), William Penn's ***Brief Account of the Province of Pennsylvania*** (1682), and Thomas Ashe's ***Carolina*** (1682) were only a few of many works praising America as a land of economic promise for English colonists.

Such writers acknowledged British allegiance, but others stressed the differences of opinion that spurred the colonists to leave their homeland. More important, they argued questions of government involving the relationship between church and state. The attitude that most authors attacked was jauntily set forth by Nathaniel Ward of Massachusetts Bay in ***The Simple Cobler of Aggawam in America*** (1647). Ward amusingly defended the status quo and railed at colonists who sponsored newfangled notions. A variety of counter-arguments to such a conservative view were published. John Winthrop's ***Journal*** (written 1630–49) told sympathetically of the attempt of Massachusetts Bay Colony to form a theocracy—a state with God at its head and with its laws based upon the Bible. Later defenders of the theocratic ideal were Increase Mather and his son Cotton. William Bradford's ***History of Plymouth Plantation*** (through 1646) showed how his pilgrim Separatists broke completely with Anglicanism. Even more radical than Bradford was Roger Williams, who, in a series of controversial pamphlets, advocated not only the separation of church and state but also the vesting of power in the people and the tolerance of different religious beliefs.

Theocratic writings

The utilitarian writings of the 17th century included many sorts—biographies, treatises, accounts of voyages, and sermons. There were few achievements in drama or fiction, since there was a widespread prejudice against these forms, but more in the field of poetry. Bad but popular poetry appeared in the ***Bay Psalm Book*** of 1640 and in Michael Wigglesworth's summary in doggerel verse of Calvinistic belief, ***The Day of Doom*** (1662). There was some poetry, at least, of a higher order. Anne Bradstreet of Massachusetts wrote some lyrics published in ***The Tenth Muse*** (1650), which movingly conveyed her feelings concerning religion and her family. Ranked still higher by modern critics is a poet whose works were not discovered and published until 1939: Edward Taylor, an English-born minister and physician who lived in Boston and Westfield, Massachusetts. Less touched by gloom than the typical Puritan, Taylor wrote lyrics that showed his delight in Christian belief and experience.

The poet Edward Taylor

All 17th-century American writings were in the manner of British writings of the same period. John Smith wrote in the tradition of geographical literature, Bradford echoed the cadences of the King James Bible, while the Mathers and Roger Williams wrote bejewelled prose typical of the day. Anne Bradstreet's poetic style derived from a long line of British poets, including Spenser and Sydney, while Taylor was in the tradition of such Metaphysical poets as George Herbert and John Donne. Both the content and form of the literature of this first century in America were thus markedly English.

FRENCH LITERATURE IN THE 17TH CENTURY

For many readers, the 17th century is still the central and supreme period of French literature, but the attempt to characterize it or distinguish its main movements and limits is a complex task. The death of one great king (Henry IV) in 1610 and of another (Louis XIV) in 1715 provide convenient points for a beginning and end (though the period after 1685, the date of the revocation of the Edict of Nantes, can also be regarded as one of transition to the 18th century), but it was not the kings who were decisive influences in creative writing or even in establishing the type of society in which writers lived. Royal encouragement was only one among many influences: of greater importance were economic prosperity, the leisure of a rising middle class, fashion in reading, and, perhaps most of all, an absence of long-continued civil or foreign war.

The example of Malherbe

The first half of the century. ***Poetry and prose.*** French society in the 17th century had not achieved stability; it was energetic and unregulated, and death and personal violence were intrusive. The polish and elegance of poems and novels were escapes from anarchy and reflected a desire for stability in a changing world. This was especially true of the early years of the century. A virtual end of civil war in 1598 and a return to peaceful conditions were shown by the appearance of collections of verse in which the poetry was carefully defined and restrained. This tendency was embodied in an individual, François de Malherbe, who from 1605 worked in Paris, enjoyed royal support, and became increasingly appreciated. Nicolas Boileau rejoiced that at last there had appeared a French poet who satisfied the modern taste for regular and smooth verse—a fair picture of what most Frenchmen in the 17th century thought Malherbe had accomplished. In addition to Malherbe's imitators and disciples, the generation born around 1590 produced a number of poets of distinction who all inclined to free thought: a nature poet, Thhophile de Viau; a dramatist and literary agent of Richelieu, François de Boisrobert; and Girard de Saint-Amant, best known in his own days as author of a biblical epic, ***Moïse sauvk*** ("Moses Delivered"). The last of these continued to write throughout the 1630s, when the most popular name was that of Vincent Voiture, a writer of social verse and a leading light of the salon of the Hôtel de Rambouillet (see below).

The poetic language of 1630 was probably more varied and interesting than it seems now and included even a metaphysical element in the religious poets Jean de Sponde and Jean de La Ceppbde, whose poetry had a personal force rarely found in classical verse, and a satirical movement, found chiefly in Mathurin Régnier, who, like many poets of the period, was esteemed more in his own time than today.

The new poetry appealed to a smaller circle, however, than did the new fiction, especially when a brilliant writer emerged in Honor6 d'Urfé, whose ***Astrée*** (1607–27; ***Astrea,*** 1657–58) was more than just a fashionable success. Many of Urfé's characters were for many generations what those of Scott or Dickens were for Victorian England, and his work inspired a host of imitators at home and abroad.

The acadkmies and salons. The literary activity of the time was not confined to the production of verse and novels: under Louis XIII small groups of people, spoken of as ***académies*** or salons, developed a taste for literary discussion. According to recent research, ***acadkmies*** were of two types: the first, a 16th-century development of centres of learned discussion for men; the second, scientific societies concerned with new knowledge and experiments. One of the ***acadkmies*** was taken over by Richelieu and became the Acadhmie Française. The salons were more general social and literary assemblies, an invention of the Renaissance that played a fluctuating part in French life until the 19th century. In many cases, notably that of the Hôtel de Rambouillet, presided over by Catherine de Vivonne, marquise de Rambouillet, their function was simply to provide a friendly meeting ground for upper class persons interested in talk and manners.

As well as the presiding figure, usually a fashionable woman such as Mme des Loges, the salons tended to have one major literary figure: in the case of that of Mme des Loges, this was Guez de Balzac, a keen critic and a master of expression, whose letters and literary quarrels gave him a great reputation. Another critic, Claude Favre de Vaugelas, acquired great authority in discussions of language; his ***Remarques sur la langue françoise*** (1647) became famous as a plea for intelligent usage. Most of these critics were possibly more successful than they knew in creating a standard of literary taste and in making the French language a supple intrument of expression.

The influence of Descartes

The ***académies*** that undertook scientific research were all influenced by or in some way associated with the great philosopher and mathematician René Descartes (1596–1650), whose influence seems to have pervaded the entire century. His views on vortices, doubt, innate ideas, or proofs of God were not, it would seem, accepted widely or for very long. It seems rather to have been his cast of mind and temper that suggested a new attack on old problems, a new way of thinking and arguing. His vision of science leading to mastery of nature found expression in the final section of the ***Discours de la mkthode*** (1637; ***Discourse on Method,*** 1950). His effect on the way intellectuals thought was summarized by one of his followers: "Before Descartes argument was easier; past ages are fortunate not to have known him" (Bernard de Fontenelle).

The theatre. Another prominent feature in the first half of the century was a rise in prestige of the Parisian theatre. It is hard to say where the new vigour and skill came from. In 1625 there was still only one licensed theatre in Paris. A dramatist of talent, Alexandre Hardy wrote about 600 plays that offered both Renaissance pathos and real dramatic suspense. Gradually the whole public attitude to the theatre changed; it became a polite entertainment, and, as with so many other social habits, this was partly caused by the active encouragement of Louis XIII's chief minister, Cardinal de Richelieu. A new generation of authors appeared: Jean Mairet, who appeared at a decisive moment, was to achieve most. His ***Sylvanire*** (performed 1630) was the first successful "new" play using the Italian convention of the unities of time, place, and action and thereby concentrating the action within strict limits. Other young dramatists, notably Boisrobert, in ***Pyrande et Zisimène*** (1633), and the great Pierre Corneille, in ***Clitandre*** (1630–31), showed how the new fashion could admit the adventures that the public wanted. Some resisted the new style, but eventually fresh new talent appeared with Jean de Rotrou's ***Hercule mourant*** ("Hercules Dying") and Jean Mairet's ***Sophonisbe*** (both 1634 or 1635). Corneille's famous ***Cid*** (1637) was therefore part of a movement and at first indistinguishable from the rest.

Corneille and the Classical theatre

The great success of ***Le Cid*** and the famous polemic connected with it (which used to be attributed to the jealousy of Richelieu but seems rather to have been caused by the vanity of Corneille) have caused the play to be seen in a false perspective. ***Le Cid*** is not the solitary emergence of French Classicism; it is simply a brilliant play. In a sense it was what followed that demonstrated its greatness. Corneille took three years over his next play, ***Horace*** (1640), a work of great dramatic power. Corneille used the Roman historian Livy's famous story of the Horatii and the Curiatii to give force to a tragic split within a family between patriot and pacifist. With incredible inventiveness he produced three plays of similar brilliance within a few years: ***Cinna*** (1640 or 1641), a drama of political conspiracy; ***Polyeucte*** (1641 or 1642), the social tragedy of a Christian martyr; and ***Rodogune*** (1644 or 1645), a powerful study of a virago. Thus, by supplying a body of plays that were treated within a generation as classics, Corneille almost singlehandedly created the French Classical theatre. Rotrou, in ***Le Véritable Saint Genest*** (1645 or 1646), wrote a martyr play as interesting as ***Polyeucte*** and, in ***Venceslas*** (1647), a tragedy of politics; he adapted Plautus in ***Les Sosies*** ("The Doubles") and Lope de Vega in ***Laure persécutée.*** Thus at the death of Richelieu (1642), France seemed to be in full literary revival with an abundance of talent, a new public, and a rapidly forming tradition.

The classical period. The civil war known as the Fronde (1648–53) cuts the century in two, sociologically no less than by date. There is no clear evolution from the best literature of the 1640s to the classicism of the 1660s. Certain features persisted, but there was a change of attitude. The war left its mark on the middle classes, which supplied the authors and most of the reading public; it left the writers, and no doubt their readers, more sophisticated, more cynical, less inclined for heroics. It should be remembered that there was no recognized profession of letters; writers lived by patronage, and the Fronde produced a new type of patron. Both Boileau and Molitre started their careers as hack writers for Louis XIV's chief minister, Nicolas Fouquet, and it was Fouquet who induced Corneille to return to the theatre and attempt a version of the chief tragic theme of antiquity, the Oedipus story. ***Oedipe,*** Corneille's first play since ***Pertharite*** (1652), was played in 1659.

Pascal's ***Pense'es***

Blaise Pascal, the greatest mind of the age, wrote within the orbit of the Jansenist religious movement—a movement that never succeeded in overcoming the King's suspicion that it was subversive. The new spirit is clearly seen in the series of pamphlets written by Pascal known as ***Les Provinciales*** (1656–57). Here, used to brilliant effect, are all the novelties of Guez de Balzac, with a relevance, force, and polemic gift that he could not command. Pascal devoted his last energies to planning an apology for the Christian religion. It got no further than the plan and a mass of partly arranged fragments, but, known as the ***Pense'es (Thoughts),*** these formed the most profound work of French Classicism. They influenced later thinkers even more than they impressed the first readers of 1670. Left in note form, and with the vigour, even violence, of expression that seems to have been natural to Pascal, this was the only French Classical work to have survived in its primitive state—a marvel of compression and annotation.

The new realism appeared in the plays of the poet and wit Paul Scarron and the later volumes of the novels of Madeleine de Scudéry. The existence of a public alive to intellectual subtleties allowed Molière to bring his theatrical company back to Paris and to introduce to a royal audience, after a performance of Corneille's tragedy *Nicomède,* an entertainment of his own that was a prelude to 15 years of intense dramatic activity, the results of which are now called Classical comedy.

It was the same spirit that animated a great work unpublished during the 17th century: the ***Me'moires*** (published 1717) of Jean-François-Paul de Gondi, cardinal de Retz. One of the chief participants in the Fronde, he embodied its spirit in these caustic and brilliant pages. Retz could brilliantly convey an awareness of things and an uninhibited attitude to human behaviour. This quality was outstanding, too, in the collection of epigrams, known as the ***Maximes,*** published in 1665 by the moralist La Rochefoucauld. These mordant and brilliant epigrams offer, within a severely classical form, a pitiless analysis of the characteristics of his time—the self-seeking attitude to morality, the cynical pessimism, and the discovery of what is now called the subconscious. They attacked individualism, idealism, stoicism, and intellectualism with wit and force.

Classical works of the 1660s and early 1670s. These years have some claim to be the most creative decade in the history of French literature. Moiltre's first attempt at a five-act comedy, *L'École des femmes* (1662; ***School for Wives,*** published 1948), aroused a controversy as animated as that of ***Le Cid.*** Boileau's first satires and the ***Maximes*** circulated about the same time. Jean Racine's first masterpiece, ***Andromaque,*** appeared in 1667, and Jean de La Fontaine's first collection of ***Fables*** in 1668. All these were distinguished by a new sense of the dignity of the language, by powerful imagination devising new forms for literary communication, and by an intellectual ability to renew the commonplace and the ordinary.

The comedy of Molibre

There was apparently no intention to create "Classical" works and very little collusion among the creators. The emergence of this body of great writing after 1660 has never been explained. It is possible to trace the course of a slow apprenticeship in each of the writers, but they did not belong to a single generation. The most impressive talent to watch at work in the 1660s and early 1670s is probably that of Molibre. Modern research has shown that he was, in fact, fighting to keep a dramatic company on its feet and that when he began writing his main concern was to fill a blank in its repertory; it was only when the public unmistakably favoured his own pieces that he assumed his full creative role. It seems that he succeeded best with his occasional shorter plays, such as ***Le*** *Médecin malgré lui* ***(The Doctor in Spite of Himself)*** and ***Le Bourgeois gentilhomrne (The Prodigious Snob).*** He spent much time on ***Le Misanthrope (The Misanthrope)*** and *L'Avare* ***(The Miser),*** where the success was short-lived, and on ***Tartuffe,***which took him five years to put on the stage. But to watch, in retrospect, the way his plays mount in range and intensity gives one some idea of what it meant at that time to run a Paris theatre and supply it with plays. In the season of 1666 Molitre wrote five new plays for his company; six years later, in 1671, the company put on 15 of his plays, three of which were new. Fame came to him fairly soon, but he appears to have had neither leisure to write nor freedom from financial worry; as late as 1669 the company was saved from collapse only by the fantastic success of ***Tartuffe.*** Most of the plays were hurriedly composed to please a discriminating public. Yet this haste devised characters of astonishing comic force, compared with which all previous French comedy looks fumbling and feeble. As Voltaire said: "Molitre has created comedy from chaos."

One of the neglected features of French Classical literature is its poetry. Nicolas Boileau, in the famous ***Art*** *poétique* of 1674 and in his epistles and satires, handled certain styles with entire mastery: puns, rhymes, paradox, topical allusion, all are easily and forcefully expressed. Each of his works was an experiment and there can be few poets who have exercised influence in such different directions.

The poet and scholar Jean de La Fontaine commanded a greater range of both irony and fantasy than did Boileau. Steeped in the ancient poets, he added his own distinctive and imaginative expression to his works. The delight given by the ***Fables*** cannot be accounted for otherwise than by their poetic renewal of the details of common life and by the relationship of the author to such material.

Racine's tragedies

Apart from these two poets the real achievements of Classical poetry in these years were in tragedy. What the great tragedian Jean Racine achieved was no doubt due to the competition of many authors now forgotten—not only, as is so often claimed, to that of Corneille. They, no less than he, attempted to recover the ancient Greek themes in ballet, opera, and tragedy. But Racine's recoveries of ancient tragic suspense and emotion were impressive, not least by their frequency. Several good judges think that his first success, ***Andromaque,*** was never surpassed. The limits of concentration would seem to have been reached in *Bérénice,* which confined the tragic subject to the farewell of a queen to two lovers and to the poignant opposition of public and private interests. Yet in 1672 Racine produced a profound study of primitive violence in *Bajazet,* to be followed a year later by ***Mithridate,*** where an equally violent character was opposed by one of great integrity. *Iphigénie* (1674) and ***Phèdre*** (1677) show an intentional return to themes where gods and men form a more complex symbolism. In the supreme achievements of ***Phèdre*** many have seen the classic 17th-century statement of the tragic gulf between passion and purity. The final irony of this unique career was that after a silence of 20 years he wrote for a girl's school two biblical plays that have evoked as much admiration as any of the others ***(Esther*** [1689] and *Athalie* [1691]).

Writings of churchmen. Although the role of the church in 17th-century life was immense, its contribution to literature was very limited; but no fair account of writing of this time can omit two great ecclesiastical figures who were colourful, contrasting personalities, Jacques Bossuet and François de la Mothe Fénelon. Bossuet, older than Fénelon by 24 years, was an orator and bishop in a traditional mold—a man of integrity, intelligence, and fabulous industry. Though most of the ideas he defended came to be rapidly discredited, he is always readable, always direct, forceful, and capable at any moment of sustained beauty of language. His funeral orations, such as that on Queen Henrietta Maria in 1669, have been unjustly called "false masterpieces"; Bossuet recovered for the church what in ancient times was the distinction of statesmen and lawyers—the power to express in noble language the praise of the dead. His attempts to show the plan of universal history in his ***Discours sur l'histoire universelle*** or to refute the Protestants and Quietists are now seen to be the weakest part of his work. He was so wedded to the hierarchical and orthodox view of things that he condemned all independent thought and even defined a heretic as "anyone who holds an opinion." But his expositions of religion remain unsurpassed.

By contrast, Fénelon, the temperamental opposite of Bossuet, was able to see both sides and to convince each that he was of their view. (His work and that of other churchmen of the end of the century is treated under *French* ***literature in*** *the* ***18th century,*** below, as representative of a period of transition.)

Women in the literary world. One mark of this period was the success of women in the literary world, both in fostering the discussion of literature and as writers, though as the latter they had to use pseudonyms. The greatest woman writer of the time, the Marquise de Sévigné, did not publish anything at all, but her ***Letters*** to her daughter in Provence made hers one of the great names of French culture. Intelligence, wit, imagination, argument, and temper were displayed, as in Pascal's notes, with an abandon that intended publication would have ruined, and she conveyed aspects of her time that no precise documentary evidence could suggest.

Her friend Mme de La Fayette, a more careful and a more limited artist, was the author of a single masterpiece, ***La Princesse*** *de Clèves* (1678). This novel's psychology and restrained pathos make it as impressive to read as a tragedy by Racine.

The satire of La Bruyhre. Jean de La Bruyère is a lonely figure in literary history. Yet he wrote in the popular tradition of satirical moralizing, and he was chiefly responsible for the fact that his contemporaries are forgotten—his portraits seem more alive than their models.

He sheltered his reflections at first behind a translation of the "characters" of Theophrastus (1688), but his pen pictures proved such a success that he greatly increased his book in later editions; keys to the characters, claiming to say who was meant by each satiric portrait, were circulated.

Les Caractères of La Bruyère

It was not primarily as satirical portraiture that ***Les Caructbres*** won the later admiration of Gustave Flaubert and the Goncourts; they praised La Bruyère as a stylist. His biting sketches of French upper class immorality, of the effects of money and fashion, of human vanity and social climbing took many forms, from the epigram to the rudimentary short story.

La Bruyère is the youngest of a group of writers that modern research has not made easier to describe or sum up. The old conception of them as illustrating Cartesian notions of literature under the chairmanship of Boileau has had to be abandoned, and with it what used to be called "the Decline of the Age of Louis XIV." As the king got older, times certainly got harder; society and, by inference, letters give the impression of decline. But that is not so in kindred fields: in scholarship, in ideas, in science, in polemic. The readers of Fontenelle and Pierre Bayle were as intelligent and as well versed as were those of Corneille and Pascal. The artistic vigour of the middle years may seem to have abated, but this is to say no more than that one or two of the supreme artists are now missing. But the duc de Saint-Simon continued to write, and, thanks to the secrecy in which he composed his memoirs, we are able to read pages of unexampled vigour and power.

The Anciens and Modernes. The dispute between those who maintained that the classics were the only models for good literature (the Anciens) and those who challenged the supremacy of classical writers and recommended subjects from romance legends and the Bible (the Modernes) was a major feature of French literature during the 17th century from about 1659.

Of the main writers of the period, Boileau was supported by La Fontaine and La Bruyère, and Charles Perrault chiefly by Fontenelle. Eventually and inevitably the controversy waned, but not before the bitter exchanges ended with Boileau's apology to Perrault in 1700, recognizing the virtues of the 17th century as equal in some ways to those of antiquity. The influence of the quarrel was felt in England, chiefly in inspiring Jonathan Swift to write his "Battle of the Books" in 1704.

Provençal. In Gascony, an entire school developed after Pey de Garros, of which Guillaume Ader, author of an admirable poem on Henry IV ***(Lou Gentilorne*** *gascoun*, 1610), is the chief representative. But the most conspicuous among the poets of the south during this period was certainly Pierre Goudelin of Toulouse, "the Malherbe of Oc," who, had he written in French, would have ranked with the best poets of the period. Nicolas Saboly excelled in the popular form known as the ***noel.***

Other 17th-century Provençal writers worth mentioning are the playwrights Claude Brueys, remarkable chiefly for comedies, and Gaspar Zerbin. The *avocat* ("barrister") Bonnet, author of the best among the open-air plays that were annually performed at Béziers on Ascension Day, the *Jugement de* ***Paris,*** and Nicolas Fizes of Frontignan, author of the vaudeville ***Opéra de Frontig****nan* (1679). both show real literary gifts.

A number of Toulouse poets, mostly from the academy, may be termed followers of Goudelin, and the classical revival that may be noted about this time is also generally ascribed to Goudelin's influence. Its most distinguished representative was Jean de Vales of Montech, who translated and wrote a brilliant burlesque of the ancient Roman poet Virgil. The best of the pastoral poets was François de Cortète of Prades, whose comedies, *Ramounet* and ***Miramoundo*** (published 1684), were written with such true feeling and in so pure a style that they can still be read with pleasure.

Walloon. At the beginning of the 17th century, Wallonia—particularly the district of Liège—became conscious of the literary possibilities of dialect, and from then on the number of writings increased. An "Ode" in the Liege dialect appeared in 1620, and *pasquèyes* (*paskeyes*, *paskeilles*), poems describing local life and history, enjoyed a vogue.

GERMAN LITERATURE IN THE 17TH CENTURY

German 17th-century literature, widely known as Baroque literature, was a product of the times and of the social situation in Germany. Ushered in against a background of witch burning, religious strife, social and political uncertainty, the early period was dominated by the Thirty Years' War and the shadow it cast over life. The local context was one of princely absolutism, each ruler emulating Louis XIV's court of Versailles and champions of the Reformation struggling against supporters of the Counter-Reformation. The new scientific discoveries were set in a society overshadowed by the influence of Luther, Machiavelli, Petrarch, and the Renaissance figures of the preceding century.

The affinities of the 20th-century situation with that of the 17th century, together with the modern rejection of concepts of the importance of confessional elements in literature (an inheritance of the great German 18th-century writer Goethe), account for modern appreciation of Baroque literature. It is dominated by writers who were guided by a sense of occasion rather than by the value of experience. They did not express themselves subjectively but used a medium of conventional images and strict formal patterns that heightened the emotional tensions of the themes. The themes were those common to European literature of the period and appeared in the lyric, epic, and drama: "fickle fortune" presides over a world ruled by chance, worldly pleasures are illusory, and man's duty is to fulfill his actor's role on the stage of life. Baroque literature presented these roles as examples of how man may escape the toils of fortune and cheat time: the martyr scorns the world and aspires to eternity, the statesman uses the present to his advantage, the stoic endures steadfastly, the hermit flees, while the pastoral figures escape to make-believe and the mystic seeks union with God. The writers themselves secured relative immortality through their work.

Martin Opitz's influence

Lyric poetry. Lyric poetry quickly reached a high standard, particularly in areas where poets congregated, such as at Heidelberg, Niirnberg, and Konigsberg and in Silesia. In critical works such as the *Buch von der deutschen Poeterey* (1624; "Book on German Poetry"), Martin Opitz demonstrated an elevated literary style in German, formulated rules such as the observance of stress in lines, and made other introductions and suggestions that he further demonstrated in his own poetry. His contemporaries quickly adopted his ideas, and thus Opitz had an important formative influence on later and superior poets such as Georg Weckherlin and Paul Fleming.

Given the basic similarity of many of its themes, the variety of Baroque verse is surprising. Paul Fleming's love poems and sonnets revealed emotional depth and lyric power, Andreas Gryphius was simultaneously pessimistic and religious, while formal dexterity in the patriotic poet and novelist Philipp von Zesen and verbal exuberance in the nature poetry of Johann Klaj illustrated an urge to experiment, often with an effect of creative spontaneity. Baroque poets wear masks: Gryphius regally expressed the existential problems of the day; Christian von Hofmannswaldau was a virtuoso of form who was intellectually flamboyant or soberly sincere according to his theme. In an age concerned with ultimate values, poets wrote both secular and religious verse, and in both this variety was present. The outstanding Lutheran poet Paul Gerhardt wrote hymns of quiet simplicity and warmth, the Jesuit Friedrich von Spee wrote religious pastoral lyrics, while Quirinus Kuhlmann was a mystic visionary.

The intellectual temper of the age found admirable expression in the epigram, which was developed to a fine art by Daniel Czepko, Angelus Silesius, and Friedrich von Logau. The variety of Baroque lyric poetry was matched by its volume. The surprising thing is that it maintained the high standard it did.

Drama. In the first half of the century, vernacular drama was relatively uninspired. Heinrich Julius, duke of Brunswick, introduced plays that reflect the influence of English players—revealing how backward the language of vernacular drama was. Later, with Daniel Caspar von Lohenstein, a dichotomy between excess of feeling and acuteness of intellectual perception was expressed in full-blown rhetoric, and there were two notable dramatists—Jakob Bidermann and Gryphius. Bidermann was a successful Jesuit who dealt with themes of mutability, worldly vanity, and the urgency of salvation with great dramatic effectiveness. Gryphius' tragedies and comedies dealt with stoicism and personal integrity. His work was literary rather than popular and so had little permanent influence, despite its artistic merit. The increasing popular masques, ballets, and operas were to lay the foundations of the later, very different drama of Germany.

The achievement of *Simplicissimus*

The novel. The development of the novel was considerably influenced by Spanish, French, and neo-Latin sources. A host of foreign translations followed the appearance of that of Mateo Alemán's *Guzmán de Alfarache* in 1615. Many were large, diffuse works, amalgams of moral and didactic elements with fantasy and amusement designed to satisfy the thirst for knowledge of the times. Some read like formal exercises in a particular form, but two writers stood out: Johann Beer (1655–1700), who foreshadowed 18th-century realism, and H.J.C. von Grimmelshausen, whose *Abentheuerliche Simplicissimus Teutsch* (1668–69; *The Adventurous Simplicissimus,* 1912) was one of the great novels of German literature, incorporating the principal themes of the age with metaphysical depth and religious insight.

Philosophy and criticism. The Silesian mystic Jakob Bohme was outstanding as an original and influential philosopher whose doctrines inspired several religious separatist movements. G.W. Leibniz's achievements in philosophy summed up the age and pointed toward the literature of the 18th century, when Baroque was to be rejected until its rediscovery in the 20th century.

DUTCH LITERATURE IN THE 17TH CENTURY

The Spanish hold on the Catholic south of the Netherlands during and after the Eighty Years' War (1568–1648) caused a decline in Brabant and Flanders; by contrast there was a spectacular expansion in Holland, to which artists, intellectuals, and financiers had fled from the Spanish armies. The rise of Amsterdam and The Hague as emergent capitals of an empire and the birth of civic pride in the writers of the "Golden Age" symbolize the final passing of a medieval age belonging to Ghent, Bruges, Liège, and Antwerp.

Henric Laurenszoon Spieghel was the greatest of a generation straddling the old and the new, and he wrote for both the burgher and the scholar. His *Nieu jaar liedekens* ("New Year Songs") and *Lieden op 't Vader Ons* ("Songs on the Lord's Prayer") continued a medieval tradition in a Renaissance style echoing Erasmian moderation; at the same time his learned *Twe-spraack vande Nederduitsche letterkunst* (1584, "Dialogue on Dutch Literature") was intended to popularize the use of a national language. His most scholarly work, the unfinished *Hert-spieghel* (1614, "Mirror of the Heart"), was particularly abstruse because it was a first attempt at philosophizing in the vernacular and in poetry.

The dichotomy inherent in the Renaissance—between popular religious revival and Humanism—was particularly marked in Holland because of the incompatibility of Calvinistic principles with the ideals of pagan antiquity. This caused a tense ambivalence in many writers of the 17th century who took both their religion and art seriously. Daniel Heinsius, a celebrated Humanist at the University of Leiden, wrote Latin plays, but he also added to the vernacular by writing *Hymnus oft lof-sanck van Bacchus* (1614; "Hymn in Praise of Bacchus") and an equally devout *Lof-sanck van Jesus Christus* (1615).

A poet, playwright, and painter, Gerbrand Adriaenszoon Bredero took his material from the life of the commoner; his medium was the folk song, farce, or comedy. His secular songs in medieval style and devotional songs in Renaissance verse told of a passionate devotion to women and yearning for religious moderation. His three tragicomedies were not successful, and he wrote three farces that marked the zenith of the medieval genre, but it was contemporary life in Amsterdam that provided material for his two comedies, including his masterpiece, *De Spaanschen Brabander* (performed 1617).

Amsterdam was also the home of the poet and dramatist Joost van den Vondel. Like Bredero, he was a self-educated man, and he resolved the conflict between artistic and religious leanings only when he entered the Roman Catholic Church at the age of 54. This was a courageous act of faith at a time when Catholics formed an unpopular minority. It is a measure of Vondel's indomitable personality that his attitude toward contemporary people and events, of which he was a fearless chronicler, still prevails even when history has recorded a different view. His plays, however, are too austere to find favor now, although in his Sophoclean *Jeptha* (1659) and in his Baroque masterpieces, *Lucifer* (1654) and *Adam in ballingschap* (1664; *Adam in Exile,* 1952), he was as great an artist of the Counter-Reformation as his contemporary, the Flemish painter Rubens.

Italian Baroque influence

If Bredero was the noble commoner and Vondel the noble citizen, Pieter Corneliszoon Hooft was certainly the noble "aristocrat." He was one of a fortunate few in Holland to bring the refinements of the new art direct from Italy. He lavished an Italianate flourish alike on his sonnets, his plays, on the studied prose of his letters, and on a monumental unfinished history of the war against Spain. His castle at Muyden became a centre for the entertainment of artists and scholars attracted by mutual interest in poetry, music, and learning and by the charm of such gifted young women as the Roemer Visscher daughters.

Anna Visscher in verse, like her father Roemer in prose, popularized ethics in the manner that was to bring Jacob Cats such unmerited fame. Cats's prolix moralizing, pedestrian doggerel, and patronizing tone force their way into the literature of his country if only because of the disastrous influence they had on the taste of the middle classes.

A more harmonious individual, Constantijn Huygens had all the qualities to which the Dutchman of his day might aspire. Of strict Calvinist principles, he was an able diplomat who wrote trenchant, shrewd, and witty verse and made excellent translations of John Donne's poetry.

There were also five poets, three of whom were clerics, who contributed religious verse of considerable merit. The Roman Catholic Joannes Stalpaert van der Wiele's ***Den schat der geestelycke lofsangen*** (1634; "The Treasury of Devotional Praise") contained songs of medieval simplicity and devotion. Jacobus Revius, stern poet of Calvinist orthodoxy, was a master of the Renaissance forms and the sonnet, using them to express the vigour and sincerity of his dogmatism. By a strange fate Revius, teaching the faith of many in a language they could understand, was almost unread, whereas Dirk Rafaelsz Camphuysen, who was removed from his parish because of his unorthodoxy, satisfied a widespread demand for personal, devotional poetry in ***Stichtelycke rymen*** (1624; "Edifying Poems"). Equally popular were the introspective collections of mystical poetry by the ascetic Jan Luyken, a layman who had started by writing hedonistic songs in ***De Duytse lier*** (1671; "The Dutch Lyre"), containing fine love lyrics. The last important poet and playwright of the Baroque period was a Fleming, Michiel de Swaen, whose religiously inspired verse compares favourably with that of most writers of the time. In general, however, the decline that had set in earlier in the south became noticeable throughout Dutch literature of the late 17th century.

ITALIAN LITERATURE IN THE 17TH CENTURY

The 17th century in Italian literature is usually described as a period of "decadence" in which writers who were devoid of sentiment resorted to exaggeration and tried to cloak an utter poverty of matter beneath an exuberance of form. In this period, it is said, freedom of thought was fettered by the Accademia della Crusca of Florence (the aim of which was to maintain the purity of the Tuscan tongue), the Counter-Reformation, and the political supremacy of Spain. This was not, however, simply an Italian phenomenon. It was at this time that "Gongorism" (the ingenious, metaphorical style of the poet Luis de Gdngora) flourished in Spain and the witty, figurative verse of the Metaphysical poets in England. Far from being exhausted, indeed, this was an extremely vital period, so much so that a new and more comprehensive understanding of the literature of the Italian Baroque has been formulated by scholars conversant with the changing attitude toward this phase of civilization in Germany, France, and England.

Poetry and prose. The popularity of satire was a reaction against prevailing conditions. Prominent in this genre was Salvator Rosa, who attacked in seven satires the vices and shortcomings of the age. Alessandro Tassoni acquired great fame with La ***secchia rapita*** (1622; ***La Secchia Rapita; or, The Rape of the Bucket,*** 1825), a mock-heroic poem that is both an epic and a personal satire. The greatest poet of the period was Tommaso Campanella, a Dominican friar, less well-known for his rough-hewn, philosophical verses than for the ***Città del sole*** (1602; ***Campanella's City of the Sun,*** 1885), a vision of political utopia, in which he advocated the uniting of humanity under a theocracy based on natural religion.

The influence of Marino

The principal representative of Italian writing during this period was Giambattista Marino, author of a large collection of lyric verse (***La lira*** [1608–14; "The Lyre"] and La ***sampogna*** [1620; "The Syrinx"]), and a long mythological poem, ***Adone*** (1623). Marino derived inspiration from the poetry of the late 16th century, but his aim—typical of his age—was to excite wonder by novelty. His work is characterized by "conceits" of fantastic ingenuity, far-fetched metaphor, sensuality, extreme facility, and a superb technical skill. His imitators were innumerable, and most 17th-century Italian poets were influenced by his work.

Gabriello Chiabrera, more sober in style than Marino, was successful in imitating the metres of classical poetry (especially of the Greek Pindar) and excelled in the composition of musical ***canzonette*** (short lyrical verses). Toward the end of the century a patriotic sonneteer, Vincenzo da Filicaia, and Alessandro Guidi, who wrote exalted odes, were hailed as major poets, though Guidi's verse is now seen as little more than rhetoric.

Among prose writers the satirist Traiano Boccalini stood out with ***Ragguagli di Parnaso*** (1612–13; ***Advertisements from Parnassus,*** 1656) in the fight against Spanish domination. A history of the Council of Trent (which defined Catholic doctrines in reaction to the Reformation) was written by Paolo Sarpi—advocate of the liberty of the Venetian state against papal interference—and a history of the rising of the Low Countries against Spain by Guido Bentivoglio. The Venetian novels of Girolamo Brusoni are still of interest, as are the travels of Pietro della Valle and the tales of Giambattista Basile. All the restless energy of this period reached its climax in the work of Galileo, a scientist who laid the foundations of mathematical philosophy and earned a prominent place in the history of Italian literature through the vigour and clarity of his prose.

The **music drama; Accademia dell'Arcadia.** With the rise of the music drama and the opera, Italian authors worked to an increasing extent with the lyric stage. Librettos written by poets such as Ottavio Rinuccini were planned with dramatic and musical artistry. During the 17th century a popular spirit entered the opera houses: intermezzi (short dramatic or musical light entertainments) were required between the acts; this practice undermined the dramatic unity of the performance as a whole, and toward the end of the century every vestige of theatrical propriety was abandoned. The spread of Marino's influence was felt by many to be an abuse. In 1690 the Accademia dell'Arcadia was founded in Rome for the express purpose of eradicating "bad taste." The purpose of Arcadia was in tune with a genuinely felt need. Many of its members were rationalist followers of Descartes with severe classical sympathies, but their reaction consisted mainly in imitating the simplicity of the nymphs and shepherds who were supposed to have lived in the golden age. Thus a new artifice replaced an old one: extreme sophistication and a flight from reality took the place of bombast and rhetoric. Typical exponents of the Arcadian lyric were Giovan Mario Crescimbeni, a founder member of Arcadia, Faustina and Giambattista Zappi, Paolo Rolli, and Pietro Metastasio, the 18th-century reformer of the operatic libretto.

SPANISH LITERATURE IN THE 17TH CENTURY

In poetry and prose the early 17th century in Spain was marked by the rise and spread of two interrelated stylistic movements. That known as ***culteranismo,*** the high-flown style of which Luis de Gdngora y Argote was archpriest, resumed attempts to ennoble the language by re-Latinizing it. This process was not only carried to extremes in vocabulary, syntax, and word order, but expression was garbed in classical myth and allusion and complicated by every subtlety of metaphor. It was an attempt to achieve a poetry that, if rediscovered after hundreds of years, could be still fresh through the immortality of its Latinism. His example (***Soledades,*** 1613; "Solitudes") invited many untalented imitations. The other stylistic movement, ***conceptismo,*** played on ideas as ***culteranismo*** did on language. Aiming always at the semblance of profundity, the style was concise and epigrammatic and thus belonged to prose, especially satire, since it was much concerned with stripping appearances from reality. Its best outlet was the essay. Francisco Gómez de Quevedo y Villegas, the greatest satirist of his age and one of Spain's masters of language, was, in ***Sueños*** (1627; "Dreams"), an outstanding exponent. Baltasar Gracián y Morales reduced its refinement to an exact code in his ***Agudeza*** y ***arte de ingenio*** ("The Conceit, a Treatise on Style"), as he also tried to codify in a series of treatises the art of living. Gracián was certainly a thinker, and his allegorical novel ***El criticón*** ("The Animadverter"), a pessimistic interpretation of life as a "daily dying," was to influence La Rochefoucauld, Schopenhauer, and Nietzsche.

Pedro Calderón de la Barca spanned the century of Spain's decline, and his drama constituted a studied ex-

Calderón's plays of the Counter-Reformation

pounding of patterns of behaviour or belief. He wrote 80 ***autos sacramentales*** (one-act allegorical plays on the Holy Sacrament), and these, with ***La vida es sueño*** (written 1635; ***Life Is a Dream,*** 1959), ***La devocidn de la cruz*** (*c.* 1625; ***Devotion to the Cross,*** 1961), and ***El mágico prodigioso*** (1637; ***The Wonder-Working Magician,*** 1961), which gave what dramatic tension is possible to a highly intellectualized exposition of Catholic dogma, made him the perfect literary representative of the Counter-Reformation. In the plays about questions of honour, such as ***El mkdico de su honra*** (1635; ***The Surgeon of His Honour,*** 1960), ***A secreto agravio, secreta venganza*** (1635; ***Secret Vengeance for Secret Insult,*** 1961), ***El pintor de su deshonra*** (*c.* 1645; ***The Painter of His Own Dishonour,*** 1906), ***El mayor monstruo los celos*** ("The Greatest Monster, Jealousy"), the cold logic with which an unreal code of behaviour is pursued to the extreme becomes repugnant. ***El alcalde de Zalamea*** (*c.* 1640; ***The Mayor of Zalamea,*** 1958), a reworking of a play by Lope de Vega, was perhaps his most successful play. ***La vida es sueño*** raised profound philosophical issues and gave him great fame abroad. After Calderón's death Spanish drama lay dead for a hundred years. ***Culteranismo*** and ***conceptismo,*** though symptoms rather than causes of decline, played their part in stifling imaginative literature, and by the close of the century little remained of its former greatness.

PORTUGUESE LITERATURE IN THE 17TH CENTURY

The influence of the Spanish poet Góngora

From a literary and political point of view, the 17th century found Portugal in a state of decadence. Before the loss of independence to Spain in 1580, Spanish influence had introduced the Inquisition and, with it, censorship of books and the preparation of an Index of Forbidden Books. Between 1552 and 1555 the Jesuits gained control of higher education. The cult of classical Latin was already present, in the work of Camões and others, before the example of the Spanish poet Luis de Góngora y Argote was felt; but with the exhaustion of the national spirit that underlay political eclipse the influence of Góngora penetrated deeply. Its extent may be seen in the five volumes of the ***Fénix Renascida*** (1716–28; "Phoenix Reborn"), which anthologizes the poetry of the preceding century and reveals the futilities to which good talents could devote themselves. The trend survived the throwing-off of the Spanish yoke in 1640; Portuguese editions of Góngora continued to appear.

The foremost literary figure of the age was the encyclopaedic Francisco Manuel de Melo, a classic of Spanish and—with his series of historical episodes, ***Epanáforas*** de ***Vdria Histdria Portuguesa*** (1660), and dialogues on literary and social topics, ***Apdlogos Dialogais--of*** Portuguese literature who strove, more successfully in prose than in verse, to free himself from subservience to Spanish form and style. Most lyricists of the period remained steeped in Gongorism. Epic poets continued active, but few of their productions were more than rhymed chronicles.

Frei Bernardo de Brito, beginning his ponderous ***Monarquia Lusitana*** (2 parts, 1597, 1609) with the Creation, reached only to the founding of the monarchy. His work was a mass of legends without foundation or critical sense, but here and in the ***Crdnica de Cister*** (1602) he proved himself a great stylist. Frei Antonio Brandão, a man of high intelligence and a learned and conscientious historian, continued the ***Monarquia Lusitana*** from Afonso Henriques to John II and was the best of the four continuers of the work. Frei Luis de Sousa, another monastic chronicler, won fame as a stylist with his ***Vida do Arcebispo D. Frei Bartolomeu dos Mártires*** (1619) and the ***Histdria de São Domingos*** (3 parts, 1623, 1662, 1678). Another notable biography, one of the best known works of the century, was Jacinto Freire de Andrade's ***Vida de Dom João de Castro*** (1651). A Jesuit, António Vieira, missionary and diplomatist and regarded for his ***Cartas*** ("Letters") and ***Sermões*** ("Sermons") as one of the greatest of all writers in Portuguese, repeated his triumphs of Bahia and Lisbon in Rome, which proclaimed him the prince of Catholic orators. Padre Manuel Bernardes, a priest and recluse whose sermons and oratorical works ***Luz e Calor*** ("Light and Warmth"), ***Nova Floresta*** (published 1706–28; "New Forest"), and ***Exercícios Espirituais*** (1686; "Spiritual Exercises") breathed a serenity and naturalness alien to Vieira and exerted a powerful influence in freeing the language from outworn "conceits." Padre Ferreira de Almeida's translation of the Bible had considerable linguistic importance.

The popular theatre lived on obscurely with mostly anonymous plays that were never printed. Those that survived were mainly religious and showed the common Gongoristic abuse of metaphor and simile. All through the century most dramatists who aspired to be heard wrote in Spanish. The court after 1640 preferred Italian opera, French plays, and Spanish operettas—to the detriment alike of native drama and of acting.

LATIN-AMERICAN LITERATURE IN THE 17TH CENTURY

By 1600 the day of the conqueror was past and the New World began a less spectacular development for two more centuries under Spanish–Portuguese dominion. The first printing press in the Americas was established in Mexico City about 1539, another in Lima in 1584, and both cities were granted university charters in 1551. In Brazil no university was founded or printing permitted until the close of the colonial period. But everywhere in the Americas a growing leisure class devoted more time to intellectual and artistic pursuits, and close ties with the homeland encouraged the development of parallel literary patterns. Life became a reflection of trends across the Atlantic, and there was little in these two centuries that had originality.

The theatre, emerging from medieval forms at the time of the conquest, served during the 16th century as a missionary medium in the Indian tongues for the conversion of the natives. But for the Creoles and the growing mestizo class, the dramatic repertoire was mainly Spanish and Portuguese, and even plays written in America were patterned after the drama of the Golden Age; those of the greatest of these native dramatists, the Mexican Juan Ruiz de Alarcón, more properly belong to Spain.

Fiction, banned by the Spanish crown, only emerged with independence, leaving the field in prose to historical and biographical works. One of the best was the travelogue ***El lazarillo de ciegos caminantes*** (*c.* 1773; ***A Guide for Inexperienced Travelers,*** 1965), written by a Spanish postal official, Alonso Carrió de la Vandera (better known as Concolorcorvo), based on a journey between Buenos Aires and Lima.

The writings of the Mexican nun Sor Juana Inés de la Cruz

Poetry was the most popular form of literature, and a contest held in Mexico in 1585 attracted more than 300 versifiers, indifferent disciples of the cult of Luis de Góngora y Argote. Their poems are in the curious anthology ***Ramillete de varias flores pokticas*** (1675; "A Nosegay of Poetic Flowers"), gathered by Jacinto de Evia of Ecuador. One later poet stood out above all: the Mexican nun Sor Juana Inks de la Cruz. Her admirers all over the Western world kept her in the limelight, where she added further to her fame as Mexico's first feminist and as a polemist bold to enter the lists even against the most eloquent Jesuit missionary of the day, António Vieira. Sor Juana's plays, religious tracts, a revealing autobiographical piece, and above all her sacred and profane lyrics of simple beauty and emotional appeal, free of Gongorist excesses, are still held among the greatest literary achievements of Latin America.

As the colonial period ended, an occasional voice protested against the decadence of viceregal life and letters; the Peruvian Juan del Valle y Caviedes and Gregório de Matos Guerra, the epigrammatic "devil's mouthpiece" of Brazil, were both influenced by the Spanish satirist Francisco Gómez de Quevedo y Villegas. Their verse circulated widely and was extremely popular.

SCANDINAVIAN LITERATURE IN THE 17TH CENTURY

Swedish. In the first half of the 17th century, Swedish literature remained limited in scope and quantity. A unique contribution, however, was made by Lars Wivallius, whose lyrics, many of them written in prison, re-

Georg Stiernhielm and his followers

vealed a feeling for nature new to Swedish poetry. With its intervention in the Thirty Years' War, Sweden established itself as a European power, and this led to a development of national pride and culture, as revealed in literature of this epoch. The outstanding work was the allegorical epic ***Hercules,*** by Georg Stiernhielm, which reflected many of the social and political problems of the time. Stiernhielm was the foremost literary figure of the 17th century and was often in attendance at the court of Queen Christina, an outstanding patroness of the arts. His followers included the two brothers Columbus, one of whom, Samuel, wrote ***Odae sueticae*** (1674; "Swedish Odes") and the prose ***Mdl-roo eller roo-mdl,*** a charming collection of anecdotes that illumine Stiernhielm's character. A rival to Stiernhielm was the unidentified "Skogekar Bargbo," whose ***Wenerid*** (1680) was the first sonnet cycle in Swedish. Several of Stiernhielm's admirers, including Urban Hiarne, formed a coterie at Uppsala in the 1660s. Hiarne led a troupe of student actors who, in 1665, performed his play ***Rosimunda*** before the young Charles XI. He aiso wrote iyrics and the prose pastoral *Stratonice* (1666–68), the first "psychological" narrative in Swedish.

Stiernhielm aimed at an integration of Sweden's cultural heritage with the accepted ideals of continental classicism. His ***Hercules*** is full of old Swedish words that he was eager to revive. Columbus also demanded a more vigorous, flexible language as did "Skogekär Bargbo" in ***Thet swenska språkets klagemål*** (1658; "The Lament of the Swedish Language"). National pride and religious feeling are combined in the works of the bishops Haquin Spegel and Jesper Swedberg, father of the Swedish mystic Emanuel Swedenborg. Spegel contributed to Swedberg's new hymnbook of 1695, which became the poetry book of the Swedish people and was of lasting influence. Even Lucidor was represented in it, giving intense expression to the contrasting moods of the period: in his love songs and, above all, in his drinking songs, he was as pagan and reckless as he was devout in his hymns and funeral poems.

At Uppsala, meanwhile, the scholar Petrus Lagerlof attempted to impose purer classical standards on native literature, and Olof Verelius edited and translated Icelandic sagas. It was Olof Rudbeck, however, who became interested in Verelius' work and enthusiastically developed a theory that Sweden was the lost Atlantis and had been the cradle of Western civilization. These theories he launched in a massive work, ***Atland eller Manheim*** (1679–1702), which translated into Latin as *Atlantica,* attained European fame.

Baroque and classicist tendencies ran parallel in late-17th-century Swedish literature. Gunno Eurelius (Gunno Dahlstierna) wrote an elaborate epic, ***Kungaskald*** ("Hymn to the King") for King Charles XI's funeral in 1697. Simpler in style was Johan Runius, who expressed a Christian stoicism of the kind found among Swedes during the disastrous early decades of the 18th century. Jacob Frese was a gentler and more intimate poet; his lyrics and hymns contained some of the emotional pietism that became a feature of 18th-century thought.

Danish. In the 17th century the literary Renaissance reached Denmark and led to a strict adherence to classical patterns and blind belief in authority in political, religious, and literary matters. In religious literature Latin dogmatics and pamphlets reflecting the superstitions of the century were dominant.

The revival of interest in Scandinavian antiquity

It was, however, a great era of science and scholarship. Ole Worm is famous for his book on the runic inscriptions, *Monumenta Danica* (1643). Thormod Torfaeus and Árni Magnússon introduced the study of Old Norse literature, and Peder Hansen Resen edited and translated some of the poetry of the Old Norse ***Edda;*** Erik Pontoppidan and Peder Syv introduced the linguistic study of Danish.

Danish poetry in the 17th century tended to follow the classics slavishly, and the favourite forms were the hexameter, the Alexandrine, and the sonnet. Simplicity is deliberately avoided; the style is precious; allegories, euphemisms, and metaphors abound. Anders Arrebo translated the Psalms and wrote ***Hexaemeron*** (1661), a Danish version of the 16th-century French poet Guillaume du Bartas' ***La Semaine.*** The century was rich in occasional poetry celebrating weddings and birthdays or mourning for a deceased paragon of virtue; didactic and pastoral poems were also common. Anders Bording, an interesting exponent of Danish Baroque poetry, was also the founder of the first Danish newspaper, ***Den danske mercurius*** (from 1666), in which the news appeared in rhymed alexandrines. The only truly great poet was Thomas Kingo, a supreme master in almost every kind of poetry. His hymns reflect a violent, passionate character, worldly and yet deeply religious. Among the few playwrights, Mogens Skeel deserves mention.

Of special interest among Danish prose works of the 17th century were the memoirs of Leonora Christina, daughter of Christian IV, a fascinating document, discovered in 1869, about her 20 years' imprisonment in the Blue Tower of Copenhagen.

Norwegian. Political union between Denmark and Norway started in 1380, and the Danish language eventually became the official and the literary medium; Copenhagen, with its university, established itself as the cultural capital of the two countries. Not until after the Reformation were there signs of renewed literary activity in Norway itself; *e.g.,* in the nostalgic apologia for Norway, ***Om Norgis rige*** (written 1567), by Absalon Pederssøn Beyer. The most original and most conspicuously Norwegian writer of this age was Petter Dass, whose ***Nordlands trompet (The Trumpet of Nordland)*** gives a lively picture in verse of the life of a clergyman in northernmost Norway; although probably completed before the turn of the century, this work was not printed until 1739.

Icelandic. In Iceland the foremost poet of the 17th century was Hallgrímur Pétursson, a Lutheran pastor who struggled against poverty and ill health. His ***Passíusálmar*** (1666; "Hymns on the Passion") is among the most popular books in Iceland. Another interesting poet was Stefan Ólafsson, remembered both for religious and secular works, the latter notable for exuberantly humorous portrayal of contemporaries and satirical observation of manners and customs.

As in other countries, interest in antiquity was stirred during the 17th century, and modern learning may be said to date from that period. Arngrímur Jónsson called the attention of Danish and Swedish scholars to Icelandic traditions and literature in a series of works in Latin, some containing abstracts of sagas now lost. Later in the century Árni Magnitsson systematically collected the early Icelandic manuscripts. Although an age of learning, the 17th century was also an age of superstition, as shown in the *Píslarsaga* ("Passion Story") of the pastor Jón Magnitsson, describing his persecution by magicians; his complaints led to the burning of two unfortunate victims in 1656. The story's lurid language is an interesting example of the period's prose.

EASTERN EUROPEAN LITERATURE IN THE 17TH CENTURY

Literature of the Time of Troubles

Russian. The end of the 16th century and the beginning of the 17th in Russia were marked by stormy political events characterized by an intense anti-feudal struggle, into which were drawn both the peasants, who were impoverished by serfdom, and the boyars, who had been defeated by the feudal nobility. The situation was complicated by the intervention of the Swedes and Poles, who were repulsed only by the united efforts of the Russian people. The events of the Time of Troubles were described in a succession of tales, echoing the social and political upheavals and commemorating the surge of patriotism against foreign intervention. The elegance and artistry of the Muscovite literary tradition was shattered; new writers and readers farther down the social scale introduced secular themes, realism, and folklore into literature. Examples are the stories of the capture of the town of Azov in 1637 by the Don Cossacks and its siege in 1641 by the Turks, which are outstanding literary works. The international relations of the state of Muscovy helped the flow of secular literature from the West. The secular translated story, the tale of chivalry, and the humorous tale replaced Byzantine religious didactic literature or existed alongside it, and folklore began to intrude noticeably.

The original Russian secular story began to appear from about the second half of the 17th century. Side by side with tales in which the conservative tradition of the church still predominated, there appeared comic tales such as *Povest o Frole Skobeyeve* ("The Tale of Frole Skobeyeve") and *Povest* o *Karpe Sutulove* ("The Story of Karp Sutulov"), both free from accepted moral and religious principles. Satire and parody of the court, church, and legal procedure also began to be written. The autobiography of the head of the Old Believers, the archpriest Avvakum, written between 1672 and 1675 *(Zhitiye protopova Avvakuma, im samim napisannoye* ["The Life of the Archpriest Avvakum by Himself"]) turns the canonical form of a typical hagiography into a polemically incisive narrative written in vivid conversational speech.

At the beginning of the 17th century, mainly under the influence of Polish models, verse began to develop. At first it was clumsy and rhythmically unorganized, but later syllabic verse was introduced and developed. In 1672, with the help of a German Lutheran pastor in Moscow, Iogann Gotfrid Gregori, a secular court theatre was created, with a basically western European repertory. Almost at the same time in Moscow Simeon of Polotsk's first attempts at school drama appeared. His *Komediya pritchi o bludnom syne* (published 1685; "Comedy of the Parable of the Prodigal Son"), in which the biblical story was adapted to treat the subject of fathers and sons in Russia, is particularly interesting. Thus it can be seen that Russian literature in the second half of the 17th century was transitional between the old literature and the literature of modern times, which dates from the beginning of the 18th century.

Polish. The Baroque period began very early in Poland. In 1564 Poland invited the Jesuits to settle in the country, and from about 1570 Protestant influence began to wane. The Baroque style and outlook were congenial to the Polish spirit; the period was one of considerable literary output, in spite of almost incessant wars. Indeed, perhaps it mirrored, in its stylistic tension, the external strife characteristic of the 17th century.

A forerunner of Baroque poetry was Mikołaj Sęp Szarżyński, who wrote predominantly religious poetry akin to that of the English Metaphysical poets. In this period satire and pastoral were the most popular forms. Foremost among satirists was Krzysztof Opaliński. His *Satyry* (1650) is bitter, pessimistic, and wide-ranging. The pastoral was represented by Samuel Twardowski, author of *Daphnis* (1638) and a romance *Nadobna Pasqualina* (1655; "Enchanting Pasqualina"), a tale of sacred and profane love, in which Polish Baroque achieves its most finely wrought splendour. The *Roxolanki* (1654), a collection of love songs by Szymon Zimorowic and the *Sielanki* (1663; "Pastorals"), written by his brother Bartlomiej, introduced topical dramatic elements into the traditional pastoral lyric; images of war and death are superimposed upon the pastoral background, with macabre effect and typical Baroque incongruity. A parallel but less formalized rustic genre produced much verse celebrating rural life. One of the more successful examples is the *Votum* by Zbigniew Morsztyn, whose finest achievement, however, was in religious poetry. In contrast was the work of his cousin, Jan Andrzej Morsztyn, whose language was marked by the extravagant style of 16th-century Italian. The formal complexity and skill of his verse was unsurpassed; and his translation of the French dramatist Pierre Corneille's *Cid* has remained the standard Polish version.

Polish Epics

The age was characterized by ambition to write heroic epics—a preoccupation to be explained perhaps by historical events: wars with Sweden, Russia, and Turkey, internal revolts, and attempts to introduce constitutional reforms. The Italian poet Torquato Tasso's *Gerusalemme liberata,* brilliantly translated by Piotr Kochanowski, inspired attempts at epics on national themes, notably the vigorous *Wojna chocimska* (written 1670; "The War of Chocim") by Wactaw Potocki. Another epic, the *Psalmodia polska* (1695), by Wespazjan Kochowski, was written under the impact of John III Sobieski's victory over the Turks at Vienna in 1683, at which Kochowski was present. The work was the first example of a theme developed by writers of the Romantic movement—the messianic interpretation of Poland's destiny.

The prose of the period did not rise to the level of its poetry, though there was a wealth of diaries and memoirs. Outstanding were the memoirs of Jan Chryzostom Pasek, a country squire and soldier. The period was also notable for the emergence of the letter as a literary form. The letters of John III Sobieski to his wife are remarkable for their passion and tenderness and for their day-by-day account of his experiences of battles and diplomacy. Another interesting development was the rise of a popular anonymous literature, exemplified by the *komedia rybattowska* ("ribald comedies"). These were generally popular satiric comedies and broad farces written mainly by playwrights of plebeian birth. One of the few whose names are known was Piotr Baryka, author of a carnival comedy. *Ž chtopa król* (performed 1633; "The Peasant-King"), which developed the theme of the induction to Shakespeare's *Taming of the Shrew.* Some 30 anonymous examples of this type of comedy survive, and they include realistic depictions of popular customs and grotesquely humorous situations, parodying the lofty themes of "official" literature and so expressing an indirect protest against social inequality.

The last stage of Baroque literature (*c.* 1675–c. 1750) displayed a long process of decline, marked only by the emergence of the first women writers and by the major figure of Stanislaw Konarski, a reformer of education, literature, and the political system, who wrote poetry in Latin, a tragedy in Polish (*Tragedia Epaminondy*), and political treatises.

It was not until the mid-20th century that the literature of the Baroque period was fully appreciated. It may well be regarded as the most enduring of Polish styles, for many of its features have recurred in the Romantic period and in modern writing.

Armenian. There were signs in the 17th century that the Armenians were emerging from the cultural decadence of the preceding centuries. The deeds of Turkish and Persian overlords figured prominently in histories by Araqel of Tabriz, Zaqaria the Deacon, and Eremia Chelebi Komiirjian, but there was some contact with Western scholars and works in Latin. Oskan of Erevan, born in 1614 in the newly founded trading colony of New Julfa, Isfahan, collaborated with the Dominican Pirandelli and printed the first Armenian Bible in Amsterdam in 1666. From the 13th century imaginative writing had been represented by a succession of popular troubadours, the most famous of these being Nahapet Kuchak (16th century), one of the rare Armenian poets to sing of physical love; Hovnatan Naghash (1661–1722); and in the following century, most famous of all, Aruthin Sayadian, called Sayat-Nova.

Hungarian. In the 17th century Hungary was still divided into three parts. The first, under Turkish rule, played no part in the development of Hungarian literature. The second, under Habsburg rule, was open to Italian and German and Catholic influence; the third part, Transylvania, was in close relationship with Dutch and English Protestant thought. The leading Protestant scholar and writer of the 17th century was János Apáczai Csere. His chief work, a Hungarian encyclopaedia in which he endeavoured to sum up the knowledge of his time, marked a development in technical vocabulary.

Effects of the Counter-Reformation

By the end of the 16th century the Counter-Reformation was gaining momentum in western Hungary. A Jesuit cardinal, Péter Pázmány was outstanding as a master of Hungarian prose, an orator and essayist of the first order. Vigour of style, clear though far from simple, use of popular expressions and solidity of argument were the chief qualities of his work, which was all controversial. His *Isteni igazságra vezérlő kalauz* (1613; "Guide to Divine Truth") was a refutation of non-Catholic religious doctrines and a masterpiece of Baroque prose.

Under the influence of the Jesuits, many Hungarian aristocrats returned to the Catholic faith and sent their sons to the Austrian Catholic universities and to Rome, where they became acquainted with the Baroque style, which can be seen in the work of Miklós Zrinyi, a great Hungarian statesman and general of his time. Most of his prose work was an exposition of political and strategic ideas. His greatest literary achievement was an epic, *Szigeti Veszedelem* ("The Peril of Sziget"), in 15 cantos, on the siege in 1566 of Szigetvdr, defended against the Turks by Zrínyi's great-grandfather. Though the influence of classical epics is clear, the work remains profoundly original and Hungarian. Zrinyi succeeded in imbuing the historical facts with a symbolic value, which conveyed his own political ideas. Another epic poet of this time, Istvdn Gyongyosi, composed many epithalamia, or nuptial poems, was inventive and handled rhyme with ease, and was much read during the 17th and 18th centuries. During this period, writing poetry seems to have become a fashion among all classes. The ***kuruts*** (Hungarian partisans fighting against the Germans and the Habsburgs) wrote poems with gripping descriptions of their poverty and misery, their joys, and the deeds of great figures of the insurrection. Many Hungarian poets have sought inspiration in these ***kurut*** songs, which have high artistic merit.

Czech. The defeat of the Protestant Bohemian forces by the Habsburgs in 1620 at the Battle of the White Mountain brought the integration of the Czech provinces into the Habsburg empire and the eradication of Protestantism from the country. The nobility was replaced by newcomers with little knowledge of Czech. It was only among the émigrés that Czech literary traditions continued. Among these exiles Jan Amos Komenský (John Amos Comenius) was pre-eminent. His Latin works on educational and theological problems and his works in Czech revealed him as a writer and thinker of European stature. His ***Labyrint světa a ráj srdce*** (1631; ***"Labyrinth of the World and Paradise of the Heart")*** was one of the supreme achievements of Czech prose literature. Some of the new Jesuit literature was of high quality, notably the poetry of Bedřich Bridel and the Latin works of Bohuslav Balbin, whose Czech patriotism expressed itself in studies of his country's past and in a remarkable defense of the Czech language.

By the beginning of the 18th century, Czech had practically ceased to be used for higher literary purposes. Rough vernacular dramas and chapbooks were produced, together with popular poetry, which included not only folk songs but forceful expressions of peasant discontent.

Croatian. In Croatian literature of the 17th and 18th centuries, the most important position belonged to Ivan Gundulić, author of a stirring epic, ***Osman,*** describing the Polish victory over the Turks in 1621. Ivan Bunić-Vučićević wrote a number of good lyrics, and the poet Ignjat Durdević and Antun Kanižlić wrote poetry in the Baroque style. Andrija Kačić Miošić treated historical themes in the popular verse forms of the travelling minstrel in a cycle on south Slav history from Alexander the Great (regarded as a Slav hero) to his own time (1756). Tito Brezovački wrote comedies and typified in his satirical verse the spirit of this age.

Gundulić's epic

Slovene. The half century of Protestant literary output in Slovene from the mid-16th century was, save for the Bible, destroyed by the Counter-Reformation. Only in about 1700 was there a revival of religious literature, with collections of Baroque sermons by Janez Svetokriški and Rogerij and the plays for Passion Week, which were written by Romuald in a comparatively pure style. The most interesting secular work was the learned patriotic description in German of Carniola by baron Janez Vajkard Valvasor.

Romanian. The printing of Romanian religious books was continued in the 17th century and was given new impetus in Transylvania, Walachia, and Moldavia by the controversy resulting from the Reformation. In Moldavia, where monasteries were centres of theological scholarship, reaction against the Reformation found expression in the *Răspuns* (1645; "Reply") of a metropolitan Varlam to the Calvinist catechism printed in 1642. A Moldavian metropolitan, Dosoftei, a great scholar and theologian, fled to Poland during the fighting between Poland and Turkey and, in 1673, published there the first Romanian metrical psalter, which was also the first poetry to be written in Romanian. He returned to Moldavia in 1675 and in 1679 translated the liturgy from the Greek. His other outstanding contribution to Romanian literature was his *Vieaţa şi* ***petrecere*** *sfinţilor* (1682–86; "Lives of the Saints"), in which he introduced popular idioms and encouraged development of a more flexible prose style.

Toward the end of the 17th century the monastery of Snagov, near Bucharest, became a centre of literary activity, and books were printed in Romanian, Greek, Slavonic, and Arabic. A Romanian printer, Mihai Stefan, introduced the press into the Caucasus, where he printed the first Georgian books. Religious literature reached its climax with a translation of the Bible (1688) that became the basis for all later translations.

Historiography was at its height with the Humanist historiographers of 17th-century Moldavia, whose leader was Miron Costin. He wrote in Polish a chronicle of Moldavia and a poem on the history of his country. The chronicle was continued by his son Niculae, who also pioneered the collection of folklore and legends. Dimitrie Kantemir, prince of Moldavia, a great linguist and historiographer, wrote Latin histories of Romania, of Moldavia, and of the Ottoman Empire. A special place among Moldavian historians is occupied by Nicolae Milescu, who wrote theological, historical, and travel works. The Walachian chroniclers were on the whole less original and more personal.

Romanian historiography

V. The 18th century

THE CHARACTERISTICS OF THE AGE

To call the 18th century the Age of Reason is to seize on a useful half-truth but cause confusion in the general picture, because the primacy of reason had also been a mark of certain periods of the previous age. It is more accurate to say that the 18th century was marked by two main impulses: reason and passion. The respect paid to reason was shown in pursuit of order, symmetry, decorum, and scientific knowledge; the cultivation of the feelings stimulated philanthropy, exaltation of personal relationships, religious fervour, and the cult of sentiment, or sensibility. In literature the rational impulse fostered satire, argument, wit, plain prose; the other inspired the psychological novel and the poetry of the sublime.

The cult of wit, satire, and argument is evident in England in the writings of Alexander Pope, Jonathan Swift, and Samuel Johnson, continuing the tradition of John Dryden from the 17th century. The novel was established as a major art form in English literature partly by a rational realism shown in the works of Henry Fielding, Daniel Defoe, and Tobias Smollett and partly by the psychological probing of the novels of Samuel Richardson and of Laurence Sterne's *Tristram* ***Shandy.*** Reflective nature poems like James Thomson's ***Seasons*** were partly scientific and partly embodiments of sensibility, while the Gothic novel combined the empirical fruits of the study of the past, which was becoming fashionable, with the modish sentiment of the supernatural and the frightening. In France the major characteristic of the period lies in the philosophical and political writings of the Enlightenment, which had a profound influence throughout the rest of Europe and foreshadowed the French Revolution. Voltaire, Jean-Jacques Rousseau, Charles de Montesquieu, and the Encyclopédistes Denis Diderot and Jean d'Alembert all devoted much of their writing to controversies about social and religious matters, often involving direct conflict with the authorities. In the first part of the century, German literature looked to English and French models, although innovative advances were made by the dramatist and critic Gotthold Ephraim Lessing. The great epoch of German literature came at the end of the century, when cultivation of the feelings and of emotional grandeur found its most powerful expression in what came to be called the *Sturm* ***und*** *Drang* ("Storm and

Stress") movement. Associated with this were two of the greatest names of German literature, Johann Wolfgang von Goethe and Friedrich Schiller, both of whom in drama and poetry advanced far beyond the turbulence of *Sturm und Drang.*

ENGLISH LITERATURE IN THE 18TH CENTURY

After the civil wars of the 17th century, the primary desire in England was for calm and consolidation. There was a practical interest in the norm, which became the basis of satire and much historical and travel writing, which appealed to all reasonable men, whose cultivated judgment could be recognized as taste. All of this gave a solid social support for artistic achievement in writing and thought, as well as in the public arts of architecture and painting. The complex of ideas collected under "Nature" was the bridge between order and the life of the heart. A taste for landscape gardening, "Nature methodized," existed simultaneously with a taste for wild scenery, ruins, "Gothic" architecture, and chinoiserie, or decorative art in the Chinese style. In literature the "horrors" of the Gothic novel and the extravagance of the Oriental tale were counterparts of the controlling influence of reason. Among philosophers, the property-based social speculation of Thomas Hobbes and John Locke was complemented by the philosophical Idealism of George Berkeley and the work of David Hume, a great thinker of the later 18th century, who carried rational introspection as far as it would go, until he met the intractable passions within man.

Poetry of the Augustan Age. The virtues of order and balance and respect for classical Latin and Greek authors were most prized in England at the end of the 17th century, when the work of John Dryden was modelled on and sometimes directly imitated Virgil and other Latin poets of the age of the emperor Augustus, and the same was to be true of Alexander Pope, the greatest satirical and moral poet of the 18th century. The name Augustan Age was applied to the reign (1702–14) of Queen Anne by Oliver Goldsmith (died 1774); it had already been applied in France to the age of Racine, Corneille, and Molière.

Pope's early work

The early poetry of Alexander Pope is mainly literary in inspiration. In the poems collected in his first volume (1717) he handled the genres of "local poetry" ("Windsor-Forest"), didactic *(An Essay on Criticism),* heroic epistle ("Eloisa to Abelard"), and mock-heroic *(The Rape of the Lock)* with remarkable sensitiveness and dexterity. Although his ten-syllable rhyming heroic couplets have less majesty than Dryden's, they display more warmth and variety and as much brilliance and wit. The influence of classical authors can be seen throughout his life—in translations (of Homer's *Iliad* [1720] and *Odyssey* [1725–26]) and in imitations *(Imitations of Horace* [1733–38]). *The Imitations of Horace* are distinguished by spirited attacks on figures of his time, presenting Pope's judgment of the literary and public scene; the irony and invective are made more weighty by Pope's careful statement of his point of view as a satirist in moving autobiographical passages that often celebrate the virtue of loyal friendship. Pope's brilliance is seen at its height in *The Dunciad* (1728), a satire on dullness in three books, to which a fourth was added later, which deepens the effect of the satire by employing reminiscences of the classical epics and the great cosmological poems of Dante and Milton.

The qualities of rational intelligence, sensitivity, and poetic dexterity are evident in Pope's later poetry on moral and social themes. In the "Moral Essays" and *An Essay on Man* (1733–34) he presented his vivid and imaginative response to the intellectual attitudes of his time. The two aspects of Pope—the satirist and the moral poet—exemplify the Augustan balance between reason and feeling.

The Scriblerus Club

Pope was a member of a Tory and literary group, the Scriblerus Club, an informal gathering that included Jonathan Swift, John Arbuthnot, John Gay, and Thomas Parnell and was dedicated to writing the memoirs of an imaginary pedant named Martinus Scriblerus. Some of the satires produced by the club were collaborative, but these writers were remarkably diverse. Many of Swift's poems are distinguished by a playful fantasy and racy idiom. John Gay's reputation rests on his offbeat masterpiece *The Beggar's Opera* (1728), but his *Trivia; or; The Art of Walking the Streets of London,* a mock Georgic, and his verse *Fables* (1727) deserve recognition. Thomas Parnell, in his *Night-Piece on Death* (published posthumously), is an important early influence on the poetry of feelings, which found its most important expression in Edward Young's *Night-Thoughts* (1742–45) and the graveyard poets. Pope had many imitators but few successors. Oliver Goldsmith, in *The Deserted Village* (1770), shows the displacement of Pope's tough argument by pathos. After Pope, the one great satirist was Samuel Johnson, whose *London* (1738) and *Vanity of Human Wishes* (1749) have all of the melancholy force of their originals in Juvenal's Latin, with a finality of phrasing that is an irreplaceable contribution to English verse.

Newspapers and periodicals

Essayists and pamphleteers. The remarkable development of English prose in the 18th century owed much to the spread of popular journalism. The lapse of the restrictive Licensing Act (1695) whetted public appetite for controversy but increased the risk of prosecution for treason or sedition and so turned writers to fiction, allegory, and irony as loopholes of escape from punishment. Nearly a score of newspapers were in publication during the first decade, along with countless polemical tracts; and, although few deserved the name of literature, much of *The Craftsman* (1726–47)—a periodical run by Bolingbroke and Chesterfield attacking the prime minister Sir Robert Walpole—and Daniel Defoe's *Review of the Affairs of France* (1704–13) assuredly did. The clarity and sense of contact with the reader demanded by argument in print are common in the prose writing of the period.

The convergence of journalism and literature is best seen in the work of Defoe, whose rough but workmanlike and clear prose surpassed immediate commercial or political concerns. In works such as his *Journal of the Plague Year* (1722), he showed an astonishing ability to project himself into a situation of which he had only read or been told. When, at the age of nearly 60, he turned to novels, he showed the same shrewdness and exploited his ability to assume another's voice in his narrative for powerful imaginative effect.

The influence of Addison and Steele

Within the fictional framework of the periodicals *The Tatler* (1709–11) and *The Spectator* (1711–12), the imaginary members of the Spectator Club are mouthpieces for their authors' ideas about society. The fiction elaborated a social microcosm within which to establish a tone at once grave, good-humoured, and flexible. The authors were free to consider any topic, either behind a mask, as in Sir Richard Steele's account of Sir Roger de Coverley's views on marriage, or straight, as in Joseph Addison's papers on John Milton's poem *Paradise Lost. The Tatler* and *The Spectator* steered serious discussion clear of religious and political partisanship and made it the pastime of the leisured middle class. The new periodicals created the public for which the novelists wrote and ensured that the new prose was both entertaining and, as in the "essays" interspersed in Fielding's novels, essentially serious and improving. Addison brought not only literature but philosophy and science to tea tables and coffeehouses. It was Steele, in *The Tatler,* who provided a moral impetus, and his writing has a friendliness that afforded needful emotional support for Addison's more dispassionate style. Their collaboration was more than the sum of its parts: separated, neither was again at his best.

The genuine *Spectator* tradition persists in Johnson's *Rambler* (1750–52) and in Goldsmith's *Citizen of the World* (1762). Intellectually, Johnson's essays are more impressive than Addison's, but they lack the charm with which *The Spectator* meets the reader on his own level without condescension. Goldsmith more freely depicted humanity's odd and touching aspects. He was more the exponent of personal sensitivity than Addison but less so than the later essayist Charles Lamb.

During the second half of the century the periodical essay was gradually absorbed by the newspaper. Johnson's "Idler" essays appeared in ***The Universal Chronicle,*** Goldsmith's ***Citizen*** (as "Chinese Letters") in ***The Public Ledger.***

Swift's irony and satirical methods

The satire of Swift. Jonathan Swift, Irish-born of English parents and educated in Ireland, became both a dominant literary figure in England and the idol of Ireland for his championship of its wrongs. He was a man of subtle wit and wide reading, much of whose work was devoted to deriding the arrogant claims of the unsupported intellect. He ridiculed whatever he thought pretentious in religion, philosophy, or science by solemnly pretending to agree with his opponents and then assuming that they would agree with him in his logical reduction of their case to absurdity. The title of his brilliant tract "An Argument To prove, That the Abolishing of Christianity in England, May, as Things now stand, be attended with some Inconveniences and perhaps not produce those many good Effects proposed thereby" (1711) demonstrates the method. Here the irony is at its most deft. In A ***Modest Proposal For Preventing the Children of Poor People From being a Burthen to their Parents, or the Country*** (1729)—by rearing them as food for the tables of the rich—a pamphlet devoted to shaming the public into awareness of Irish poverty, it is at its most trenchant. The "Drapier's Letters" (1724–25) raised pamphleteering to the status of literature. ***Gulliver's Travels*** (1726) is a "true narration" like Defoe's ***Robinson Crusoe;*** in "The Battle of the Books" (1704), he makes use of the mock-heroic allegory, a form popular in political pamphleteering, to savage the complacent insensitivity of "modem" critics and thinkers egotistically interested in progress at the expense of a proper awareness of human frailty and sin; in his most varied and ingenious work, A ***Tale of a Tub*** (1704), all these forms are exploited. He used man's less agreeable physical functions to symbolize moral shortcomings, the allowed method of satire in all ages. Disappointed ambition may have chastened the sheer brilliance of Swift's cutting wit, yet that he did not see himself as righteous and the rest of mankind as vile is evident from the fourth book of ***Gulliver's Travels,*** in which Gulliver's position between depraved and virtuous creatures (represented, respectively, as apelike Yahoos and noble horselike Houyhnhnms) is a representation of Swift's own. His invention was inexhaustible, and he could write brilliantly on any subject.

Swift's friend George Berkeley, bishop of Cloyne in Ireland, is remembered chiefly as a philosopher, though his tracts and Platonic dialogues are also of some literary importance. Like Swift, he was a lover of his country and expressed strong views about its welfare in his ***Querist.***

Criticism. Criticism, appearing for the first time as a continuous and widespread literary activity, placed the reputations of the major English poets—Shakespeare, Milton, and, to a lesser degree, Chaucer and Spenser--on a firm critical basis in historical perspective. The foundations of modem texts of Shakespeare were laid by Johnson (1765) and, most notably, by Lewis Theobald, whose ***Shakespeare restored*** (1726) and an edition of the plays (1733) represented pioneer work in textual criticism.

Shaftesbury's influence

Individual critics such as John Dennis, Addison, Pope, and Johnson worked within the frame of Neoclassical rules expounded by such French critics as Nicolas Boileau, but bent them to accommodate their feelings and personal vision. Critical theory dealt largely with the nature of poetic inspiration and aesthetic taste, an increasingly psychological account of the literary process. Lord Shaftesbury's ***Characteristics of Men, Manners, Opinions, Times*** (1711; revised and enlarged 1714, 1723) deprecates enthusiasm or imagination in religion and insists on its importance in literature to exalt the mind to serene contemplation. The preoccupation is with sublimity, a concept only vaguely apprehended but strongly felt not only in the century's criticism but also in its poetry. Shaftesbury posited a "moral sentiment" that draws man naturally to what is good and beautiful. He sought to lift criticism above the level of religious or political squabbling, and Addison, a brilliant popularizer of difficult theory rather than an original thinker, had a balanced approach to literature that made the way easier for later critics who developed this kind of judgment.

Johnson condemned those who judge by precept rather than by perception and insisted that literary judgment must be related to experience. He believed that literature was valuable only if it clarified human experience. His greatness as a critic lies in his wide, humane reading and in the fact that his grounds for criticism are always apparent and seldom capricious. These grounds are, in a broad sense, moral. When his judgment is at fault the failure is rather from inadequacy of experience than want of critical perception. For Johnson, as for most critics of his day. singularity implied deformity; nature, as experience common to all men, was the business of literature. Johnson's position in English criticism may be inferred from the words of his preface to his edition of Shakespeare: "Let the fact be first stated. and then examined."

Johnson's ***Dictionary***

What made Johnson a great critic made him also the great conversationalist depicted in James Boswell's ***Life of Johnson*** and the great prose stylist of ***Rasselas*** (1759) and ***The Rumbler.*** Johnson's ***Dictionary of the English Language*** (1755), though not the first English dictionary, introduced methods of illustrating the use of words that have changed little since it was first published. As a lexicographer he was always concerned with accuracy, and, like Sir Thomas Browne, on whom he modelled his style, he found in Latin derivatives the most precise and compact form of denotation.

The revolution of taste. During the second half of the century, the trend intensified toward a psychological expression of literary judgment based on unique, personal feeling. Sir Joshua Reynolds' ***Discourses*** (1769–91) uphold Neoclassical traditions of imitation, but Edward Young's ***Conjectures on Original Composition*** (1759) put original genius above learning. Young, like the Neoplatonist Shaftesbury, had a great influence in Europe.

Thomas Warton's ***History of English Poetry*** (1774–81) and ***Observations on the Faerie Queene*** (1754) and Bishop Richard Hurd's ***Letters on Chivalry and Romance*** (1762) all demonstrated a growing interest in medieval and Renaissance literature. It is seen too in James Macpherson's ***Ossian*** (1762), the "medieval" poems of Thomas Chatterton, and collections of earlier poetry, notably Bishop Thomas Percy's ***Reliques of Ancient English Poetry*** (1765). Another important pointer to changing critical opinion was the first part of Joseph Warton's "Essay on the Writings and Genius of Pope," with its emphasis on sublimity as the mark of "true" poetry and on freedom of expression. His praise of "poetical enthusiasm" and the poet's "creative imagination" indicates the basis on which the Romantic poets and critics built. All these works stress the importance of fancy or imagination in poetry, but their originality lies in their emphasis rather than in a revolutionary approach. Perhaps the most revolutionary work was Thomas Tyrwhitt's "Essay on the Language and Versification of Chaucer" (1775), which destroyed an earlier illusion of Chaucer as a clumsy metrist. William Blake's marginal comments on Reynolds' ***Discourses*** were more indignant outbursts against academic precept than revolutionary criticism, although when taken with Blake's other work they represent a consistent literary theory.

The novel. The growth of the English novel is as indigenous as the development of the periodical. Daniel Defoe's picaresque heroes had prototypes in the Elizabethan fiction of Thomas Nashe, Robert Greene, and Thomas Deloney, but ***Moll Flanders*** (1722), ***Colonel Jack,*** and ***Captain Singleton*** were also imaginative heightenings of actual rogue histories of the time, and Defoe's most famous novel, ***Robinson Crusoe*** (1719), was based on contemporary fact. The novels of Samuel Richardson drew on original sources. *Pamela* (1740) was based on his own ***Letters Written to and for Particular Friends,*** a collection of model letters to enable "country readers" to "endite for themselves," prescribing "the requisite Style and Forms" and "how to think and act justly and prudently, in the Common Concerns of Human Life." In *Clarissa* (1747–48) Richardson produced a seminal work of prose fiction and psychological analysis of sombre power that has had a profound influence on the development of the novel throughout western Europe.

Richardson and Fielding

The success of *Pamela,* with its concentration on moral respectability and neglect of recognized literary genres, irritated Henry Fielding, a man of refinement and extensive social experience, and he ridiculed it in *Shamela.* As the work advanced, turning into the substantial autonomous work *Joseph Andrews* (1742), Fielding obviously found more in the form, and especially in the character of Parson Adams, than he had expected. His later description of the episodic *Joseph Andrews* as the "comic epic poem in prose" better fits the tightly organized *Tom Jones* (1749), which showed him more clearly as the conscious artist. Introductory chapters to each book endow the novel with a critical rationale and constitute an important technical advance, for in them the author himself appeared as the shaping spirit of the story. Hitherto in realistic fiction authors (even Richardson) had commended their stories by claiming historical accuracy; Fielding asserted the author's right to manipulate his narrative openly in the interests of artistic truth. To suggest that his achievement is chiefly technical would be to ignore the unity of his artistic and philosophic outlook.

Smollett and Sterne

Another journalist who found fame as a novelist was Tobias Smollett, who, in his life as in his novels, resembled Defoe. Like him he was multifarious—naval surgeon, unsuccessful medical practitioner, hack writer, translator, historian—and he too knew prison from the inside. Such life, varied and arduous, was the source of *Roderick Random* (1748). Smollett's good-hearted déclassé heroes move through loosely connected episodic adventures, much as Defoe's; but whereas Defoe's characters derive their individuality from their surroundings, Smollett's are creations in their own right, an individuality expressed often by brilliant caricature. Smollett's writing lacks Richardson's sense of tragic predicament or Fielding's awareness of human comedy. His picture of life is crude but a telling expression of a feeling for the brutality and brutalization of implacable fortune. His flair for making haunting imaginative use of peculiarities and idiosyncracies of human behaviour, especially in *Humphry Clinker* (1771), has been matched only by Charles Dickens in the 19th century.

The development of the novel was rapid and diverse. In the 40 years after *Robinson Crusoe,* it retained the framework of a more or less chronological sequence of events leading to an outcome foreseen by the omniscient author. Laurence Sterne radically altered this chronological framework in the creation of the first impressionistic novel and one of the greatest works in the language—*Tristram Shandy* (1760, dated 1759)—which wittily presents life as a flux of events without relationship except in the consciousness of the person experiencing them. His true successors in developing these possibilities of the novel were to be found in the 20th century.

Horace Walpole's *Castle of Otranto* (1765) is important as the first in the line of Gothic novels, titilating the reader by stimulating fear and horror at tales of the supernatural in an exotic scene located in the past. In this tradition were Clara Reeve (*The Champion of Virtue,* 1777), Ann Radcliffe (*The Mysteries of Udolpho,* 1794), and Matthew Lewis (*The Monk,* 1796). The sentimental romance was another aspect of the cult of sensibility to be seen in Goldsmith's *Vicar* of *Wakefield* (1766), Henry Mackenzie's *Man of Feeling* (1771), Henry Brooke's *Fool of Quality* (1764–70), and in Sterne's *Sentimental Journey* (1768), which, by describing the author's feelings and not what he sees, had considerable influence on conceptions of "sentiment" and the cult of the "feeling heart." The exotic possibilities of the Gothic tale were further developed in William Beckford's Oriental romance *Vathek* (1786). Beckford's *Dreams, Waking Thoughts, and Incidents* (1783) had shown a developing sensibility in describing personal experiences during his travels that also had its place in the revolution of taste.

Fanny Burney's *Evelina* (1778), her most successful novel, manifests the qualities that made her a great diarist—spontaneous interest in life and cool appraisal of social convention. She learned something from Smollett's portrayal of eccentricity but had a firm sense of obvious structure.

Later poetry. *Reflective poetry.* The cult of solitude and the contemplation of nature's solemnities was explicit in Shaftesbury's writings, and as early as 1726 James Thomson's *Winter* was winning general approval. *The Seasons* (1730), a longer poem that developed from *Winter,* retaining the poet's sensibility as the sole unifying thread, is not only vivid in its detail but dramatic in its portrayal of observed natural phenomena and landscape. Goldsmith's *Deserted Village* (1770) vindicated the Augustan theory of generalized description in its depiction of both human and nonhuman nature. William Cowper's *Task* (1785) was perhaps the last in this line of succession. Cowper turned to the hills and snowbound woods of his home at Weston in Buckinghamshire as a relief from religious terror. *The Task,* discursive and with a placid tone and masterly yet unostentatious handling of language, had the comforting tranquility of poetry written for its own sake, seldom touched by the ratiocinative urge of Thomson or the sensuous brilliance of Christopher Smart, the author of *A Song to David* (1763) and *Hymns,* poems that glorified the particularities (rather than the universal plan) of nature as symbolizing God's love for man. William Collins, a fine lyric poet, devoted himself to the ode, a poetical form that for the Augustans exalted the connective power of the poet's imagination over the skeleton of rational argument. Thomas Gray's "Elegy Written in a Country Church Yard" (published 1751), in the universality of its themes and its dignified yet intimate tone, epitomized the Augustan ideal of reflective poetry.

Gray's "Elegy"

The poetry of Burns and Blake, The two great lyric poets Robert Burns and William Blake lie, in some measure, outside the traditions of the century. Burns's *Poems, Chiefly in the Scottish Dialect* (1786) had a contemporary success only in part due to the increasing interest in "primitive" poetry, though this interest was often scholarly or dilettante. Burns's vocabulary and rhythms, however, sprang not from any poetic theory but from the vernacular speech of his native Ayrshire and from the wealth of song of the Scottish people, for whose old tunes he composed and restored many sets of words. Burns was a lively and successful personal and social satirist. Because his background was not academic, his reading nourished his genius without subduing it to the requirements of "learning" (which hampered even such poets as Gray), and he is one of the few writers of the age who has retained a worldwide popular image. Blake's *Songs of Innocence* (1789) and *Songs of Experience* (1794) are, indeed, descendants of the hymns of Isaac Watts; but what Watts perceived by the light of moral doctrine, Blake perceived intuitively and, by a unique, original technique, fashioned into precise, moving poems marked by a strangely personal vision, "the lost traveler's dream." Pressing his private associations, Blake deepened his isolation from any responsive audience, and his work moved farther into the complex, mystical symbolisms of the later "prophetic" books, such as *The First Book of Urizen* or *The Book of Los* (1795), which he did not "publish" in the generally accepted sense but laboriously engraved by a process that frustrated much possibility of immediate free communication, though it resulted in the superb illuminated books in which poem and illustration are fully complementary. Like Burns, Blake was in revolt against human limitations, and it is this that distinguished these two poets from others of their age.

Blake's "prophetic" books

Imaginative and learned prose. The increase of travel and a new sense of leisure and connoisseurship among the wealthier families speedily united with the increasing importance of individual observation to produce in England a crop of those elaborate collections of letters and memoirs that earlier had reached their summit in France with those of Mme de Sévigné and the Duc de Saint-Simon. In 1709 began the letters of Lady Mary Wortley Montagu; 17 years more saw the commencement of Lord Hervey's *Memoirs of the Reign of George 11,* while Lord Chesterfield and Horace Walpole both began their series of *Letters* about 1740. Less modish but pleasing in their unostentatious charm, the letters of Gilbert White, published in 1789 as *The Natural History and Antiquities of Selborne*, show that the countryman's loving eye for detail was not only active in the Romantic period. William Cowper also produced a collection of beautiful, perceptive, and graceful private letters. Stylistically at the opposite extreme is Chesterfield, who in treating of manners and social amenity deliberately sought a form of expression that is the perfection of tact, good order, and savoir faire. The most frivolous and also the most pungent of letter writers was Horace Walpole, whose writings are an epitome of the taste, history, and biography of the Georgian era. He was in some ways a corrective to the complacency of his generation, a vast dilettante, a lover of Gothic, of curios, costly printing, old illuminations, and stained glass.

Long known for his celebrated *Life of Johnson*, Boswell was also established as a master of English prose. When his journals were published in the 20th century, he became known as a pioneer in the exploration of the human unconscious and an informative commentator on the 18th-century London scene. *The Diary and Letters of Mme d'Arblay* (Fanny Burney) and the *Autobiography* of the historian Edward Gibbon are in their different ways also representative of 18th-century excellence.

It was not until the second half of the century that English historical writing escaped its reputation of lagging behind France, Italy, and Germany. David Hume published the first volume of his *History of England* in 1754. William Robertson's *History of Scotland* appeared in 1759 and his *History of the Reign of the Emperor Charles V* in 1769, adding to the growing contribution from Edinburgh of philosophical and sociological writing. Edward Gibbon's *History of the Decline and Fall of the Roman Empire* came in 1776–88. This monumental work is notable for its indefatigable attention to detail and its shrewd judgments. Gibbon's style was portentous in its dignity; he combined with the Skepticism of Hume and the method of Robertson an ordered harmony of thought and prose and a pleasing malice all his own.

Gibbon's *Decline and Fall*

In the philosophical and political writing of the latter part of the century, free thought was beginning to sap the defenses of the religious apologists, in the writings of Thomas Paine, Joseph Priestley, William Godwin, and Sir James Mackintosh. The greatest champion of the continuity of tradition and conservation and the perfect master of emotional prose was Edmund Burke, one of the most commanding intellects in the field of political letters—and in this respect a striking contrast to "Junius," the pseudonym of an unidentified author, whose journalistic talent for invective had a quite ephemeral value.

Drama. An age of great actors, such as Samuel Foote and David Garrick, the 18th century in England was poor in drama. With the exception of Fielding, a playwright of some originality, and Goldsmith (*She Stoops to Conquer*, 1773), the men of literary genius, though many wrote plays, were not men of the theatre. Toward the end of the century, in *The Rivals* (1775) and more especially *The School for Scandal* (1777), Richard Brinsley Sheridan showed that he alone possessed both the literary and theatrical qualities for great drama in the orthodox sense.

Sheridan's plays

Scottish literature in English and Scots. While Scottish prose writers were preparing to challenge English on its own terms, and while the union of 1707 was suggesting an even closer cultural binding of the two countries than had been possible a century before, a contrary impulse was making itself felt in poetry. Almost in the year of the union appeared James Watson's first volume of *Choice Collection of Comic and Serious Scots Poems* (1706), which contained (among English pieces) vernacular poetry, from *Christis Kirk on the Green* and *The Cherry and the Slae* to recent verse like *Habbie Simson* and *The Blythesome Bridal*. This was followed by Allan Ramsay's *Ever Green* (1724) and *Tea-Table Miscellany* (1724 ff.) and by the later collections of David Herd, John Pinkerton, James Johnson, and George Thomson. These anthologies testify to a new national consciousness that began with a deliberate invocation of past achievements and eventually produced original work in the tradition of these achievements. As political identity was lost, cultural differences were increasingly recognized as significant. Educated speech was slowly but surely following written prose into anglicization, but in poetry the heart of the language was still felt to beat, however faintly, as it was also in common speech and in the old songs and ballads that were being reprinted. Ramsay's own poetry, like that of Robert Fergusson (1750–74) and Robert Burns (1759–96), flourished on a union of these elements: the racy vigour of Scots speech, the musical lilt of the songs, and the poetic forms and techniques, such as the Habbie Simson stanza, the humorous elegy, and the verse epistle, with which recent poets like the Sempills and William Hamilton of Gilbertfield had helped to solder the vernacular tradition. This new poetry lacked the range and also the intellectual power of the makaris' verse. Nevertheless, its qualities were notable and infinitely attractive: in the love songs "Duncan Gray" and "O, Wert Thou in the Cauld Blast"; in the voice given to the joy and pathos of "randie gangrel bodies" in "The Jolly Beggars"; in the unembittered descriptive tartness of Fergusson's "Auld Reikie" and "Butterfly." Burns excelled in both satires and love lyrics; in the former he carried on where Dunbar left off, and in the latter he crystallized the floating anonymous songs in forms at once popular and authoritative. Many traditions meet in Burns, and his poetry reaffirms, perfects, and consolidates existing poetic modes and themes in a way that even in his own lifetime made it apparent that he was going to be a national bard. If any poet has deserved the epithet "life-giving" it is Burns, and yet by one of the ironies of literary history he proved a liberator for the poetry of England but not for that of his own country.

Among minor poets of the period, some of the best were women: Lady Grizel Baillie, Jane Elliot, Lady Anne Lindsay, and Lady Nairne. Of writers in English, James Thomson in *The Seasons* (1726–30) shares with Gavin Douglas the ability to make winter more impressive than summer, and produces some heavily drawn pictures of "Caledonia, in romantic view." Robert Blair's *Grave* (1743) shows a grim, macabre relish not uncommon in Scottish poetry. William Falconer's *Shipwreck* (1762), admired by Burns, deserves mention; and James Beattie's *Minstrel* (1771–74) has historical importance as an early Romantic poem. Here must also be noted three popular and influential works of the time: John Home's tragedy *Douglas* (1756), Scottish in theme but not in language; the cloudy grandeurs of James Macpherson's Ossianic fragments, which presented Europe with a version of the noble Gael; and that "bosom favorite" of the young Burns, Henry Mackenzie's tear-strewn novel *The Man of Feeling* (1771).

AMERICAN LITERATURE IN THE 18TH CENTURY

In America in the early years of the 18th century, some writers, such as Cotton Mather, carried on the older traditions. His huge history and biography of Puritan New England, *Magnalia Christi Americana*. in 1702, and his vigorous *Manuductio ad Ministerium*, or introduction to the ministry, in 1726, were defenses of ancient Puritan convictions. Jonathan Edwards, initiator of the Great Awakening, a religious revival that stirred the eastern seacoast for many years, eloquently defended his burning belief in Calvinistic doctrine—of the concept that man, born totally depraved, could attain virtue and salvation only through God's grace—notably in the philosophical treatise *Freedom of Will* (1754). He supported his claims by relating them to a complex metaphysical system and by reasoning brilliantly in clear and often beautiful prose.

But Mather and Edwards were defending a doomed cause. Liberal New England ministers such as John Wise and Jonathan Mayhew moved toward a less rigid religion. Samuel Sewall heralded other changes in his amusing *Diary*, covering the years 1673–1729. Though sincerely religious, he showed in daily records how commercial life in New England replaced rigid Puritanism with more worldly attitudes. The *Journal* of Mme Sara Knight comically detailed a journey that lady took to New York in 1704. She wrote vividly of what she saw and commented upon it from the standpoint of an orthodox believer, but a quality of levity in her witty writings showed that she was much less fervent than the Pilgrim founders had been. In the South, William Byrd of Virginia, an aristocratic plantation owner, contrasted sharply with gloomier predecessors. His record of a surveying trip in 1728, *The History of the Dividing Line,* and his account of a visit to his frontier properties in 1733, *A Journey to the Land of Eden,* were his chief works. Years in England, on the Continent, and among the gentry of the South had created gaiety and grace of expression, and, although a devout Anglican, Byrd was as playful as the Restoration wits whose works he clearly admired.

William Byrd's prose

The wrench of the American Revolution emphasized differences that had been growing between American and British political concepts. As the colonists moved to the belief that rebellion was inevitable, fought the bitter war, and worked to found the new nation's government, they were influenced by a number of very effective political writers, such as Samuel Adams and John Dickinson, both of whom favoured the colonists, and Loyalist Joseph Galloway. But two figures loomed above these—Benjamin Franklin and Thomas Paine.

Franklin, born in 1706, had started to publish his writings in his brother's newspaper, the *New England Courant,* as early as 1722. This newspaper championed the cause of the "Leather Apron" man and the farmer and appealed by using easily understood language and practical arguments. The idea that common sense was a good guide was clear in both the popular *Poor Richard's* almanac, which Franklin edited between 1733 and 1758 and filled with prudent and witty aphorisms purportedly written by uneducated but experienced Richard Saunders, and in the author's *Autobiography,* written between 1771 and 1788, a record of his rise from humble circumstances that offered worldly wise suggestions for future success.

Franklin's self-attained culture, deep and wide, gave substance and skill to varied articles, pamphlets, and reports that he wrote concerning the dispute with Great Britain, many of them extremely effective in stating and shaping the colonists' cause.

Thomas Paine's "Common Sense"

Thomas Paine went from his native England to Philadelphia and became a magazine editor and then, about 14 months later, the most effective propagandist for the colonial cause. His pamphlet "Common Sense" (January 1776) did much to influence the colonists to declare their independence. *The American Crisis* papers (December 1776–April 1783) spurred Americans to fight on through the blackest years of the war. Based upon Paine's simple Deistic beliefs, they showed the conflict as a stirring melodrama with the angelic colonists against the forces of evil. Such white and black picturings were highly effective propaganda. Another reason for Paine's success was his poetic fervour, which found expression in impassioned words and phrases long to be remembered and quoted.

The Federalist papers

The new nation. In the postwar period some of these eloquent men were no longer able to win a hearing. Thomas Paine and Samuel Adams lacked the constructive ideas that appealed to those interested in forming a new government. Others fared better—for example, Franklin, whose tolerance and sense showed in addresses to the constitutional convention. A different group of authors, however, became leaders in the new period—Thomas Jefferson and the talented writers of *The Federalist* papers, a series of 85 essays published in 1787 and 1788 urging the virtues of the proposed new constitution. They were written by Alexander Hamilton, James Madison, and John Jay. More distinguished for insight into problems of government and cool logic than for eloquence, these works became a classic statement of American governmental theory. At the time they were highly effective in influencing legislators who voted on the new constitution. Hamilton, who wrote perhaps 51 *Federalist* papers, became a leader of the Federalist Party and, as first secretary of the treasury (1789–95), wrote messages that were influential in increasing the power of national government at the expense of the state governments.

Thomas Jefferson was an influential political writer during and after the war. The merits of his great summary, the Declaration of Independence, consisted, as Madison pointed out, "in a lucid communication of human rights . . . in a style and tone appropriate to the great occasion, and to the spirit of the American people." After the war he formulated the exact tenets of his faith in various papers but most richly in his letters and inaugural addresses, in which he urged individual freedom and local autonomy—a theory of decentralization differing from Hamilton's belief in strong federal government. Though he held that all men are created equal, Jefferson thought that "a natural aristocracy" of "virtues and talents" should hold high governmental positions.

Poets and poetry. Poetry became a weapon during the American Revolution, with both Loyalists and Continentals urging their forces on, stating their arguments, and celebrating their heroes in verse and songs such as "Yankee Doodle," "Nathan Hale," and "The Epilogue," mostly set to popular British melodies and in manner resembling other British poems of the period.

The most memorable American poet of the period was Philip Freneau, whose first well-known poems, Revolutionary War satires, served as effective propaganda; later he turned to various aspects of the American scene. Although he wrote much in the stilted manner of the Neoclassicists, such poems as "The Wild Honey Suckle," "To a Caty-did," and "On a Honey Bee" were romantic lyrics of real grace and feeling that were forerunners of a literary movement destined to be important in the 19th century.

Drama and the novel. In the years toward the close of the 18th century, both dramas and novels of some historical importance were produced. Though theatrical groups had long been active in America, the first American comedy presented professionally was Royall Tyler's *Contrast* (1787). This drama was full of echoes of Goldsmith and Sheridan, but it contained a Yankee character (the predecessor of many such in years to follow) who brought something native to the stage.

William Hill Brown wrote the first American novel, *The Power of Sympathy* (1789), which showed authors how to overcome ancient prejudices against this form by following the sentimental novel form invented by Samuel Richardson. A flood of sentimental novels followed to the end of the 19th century. H.H. Brackenridge followed Cervantes' *Don Quixote* and Henry Fielding with some popular success in *Modern Chivalry* (1792–1815), an amusing satire on democracy and an interesting portrayal of frontier life. Gothic thrillers were to some extent nationalized in Charles Brockden Brown's *Wieland* (1798), *Arthur Mervyn,* and *Edgar Huntly* (1799). But all such works were more interesting as beginnings than as outstanding artistic achievements.

FRENCH LITERATURE IN THE 18TH CENTURY

The years of transition. A new era in French literature can be regarded as having begun in 1685 with the revocation of the Edict of Nantes, which had led to the suppression of all forms of Protestant expression; the outbreak of the French Revolution in 1789 can conveniently be taken as the end. Molière, Corneille, La Rochefoucauld, Racine, and most other major writers had faded from the scene, but the spirit and the genius of Classicism was not yet extinct. In Jean de La Bruyère's *Caractères* this Classical attitude is tempered by the author's social conscience, which shows him to be a precursor of 18th-century philanthropy.

The shift to critical writing

Meanwhile, the main preoccupation of many writers had become critical rather than creative. Charles Perrault, remembered for his fairy tales, *Contes de ma mère l'oye* (1697; *The Tales of Mother Goose,* 1785), attacked the prestige of the ancient world in his poem *Le Siècle de Louis le Grand* (1687), in which he maintained that the age of Louis XIV was comparable with that of the Roman emperor Augustus. Perrault's views were ill received by Nicolas Boileau, La Fontaine, and La Bruyère, and reconciliation between the Anciens, Boileau's party, and the Modernes (quarreling as to whether the ancient Greek and Latin authors were the only models for literary excellence) did not take place until 1701.

Although posterity was to regard Racine's return to the stage with ***Esther*** (first performed 1689) and especially *Athalie* (1691) as by far the greatest event in the history of tragedy in Louis XIV's last 30 years, another dramatist, Prosper Jolyot de Crébillon, was at the time carried to the height of fame by blood-curdling horrific tragedies, of which the best known is ***Rhadamiste et** Zénobie* (1711), replete with murders, mistaken identities, and Oedipal conflicts.

Jean-François Regnard wrote many lighthearted but superficial comedies relying on the techniques of farce. ***Le Joueur*** (1695; ***The Gamester,*** 1705) was the best known of these. Florent Carton d'Ancourt, known as Dancourt, was another prolific writer of comedies, with a particular skill in the portrayal of social types. A higher standard was reached by Alain-René Lesage, whose first successes on the stage were ***Crispin, rival de son*** *maître* (1707; ***Crispin, Rival of His Master,*** 1915) and ***Turcaret*** (1709), both of them bold and vivacious.

The most important novel of these years was ***Les Aventures de Télémaque*** (published 1699) by François Fénelon, who was already renowned not only as an ecclesiastic but also, after his "Trait6 de l'éducation des filles" (1687; "The Education of Girls," 1805), as a man of letters and an innovator. ***Télémaque***, narrating in engaging prose the travels of Telemachus, son of Ulysses, after the siege of Troy, was written for the Duc de Bourgogne, heir to the French throne, and it was immediately extremely successful. Fénelon expressed advanced ideas of a political and social character, often involving implicit criticism of the policies of Louis XIV, and proposing a curious blend of medieval feudalism and enlightened despotism. In his hands the novel became an instrument for the spreading of specific and developed ideas.

A retreat from the categorical orthodoxy of the churchman Jacques-Bénigne Bossuet (whose work is treated in ***French literature in the 17th century,*** above) was shown by other ecclesiastics: by Esprit Fléchier, bishop of Nîmes, by Jean-Baptiste Massillon, bishop of Clermont, by the Jansenist Pasquier Quesnel, and especially by Fénelon, whose emotional and quasi-mystical ***Explications des maximes des saints sur la vie inte'rieure*** was condemned by the Roman Catholic Church. Richard Simon, whose application to biblical criticism of rigorous methods of interpretation in the ***Histoire critique du Vieux Testament*** (1678) led him into heterodoxy, was expelled from the religious order to which he belonged.

The growth of relativism

A spirit of cosmopolitan relativism—the view that good and bad are only relative concepts, that what may be good in one society may be bad in another—is revealed in the works of Charles de Saint-Denis de Saint-Évremond, who from 1661 until his death in 1703 lived in exile in England and Holland. An individualist and a freethinker, he had an acute and satirical intelligence. Relativism was encouraged also by the ***Voyages*** and ***Mémoires sur l'Amérique*** (1703) of Baron de La Hontan and by the translation (1704–17) of the ***Arabian Nights*** by Antoine Galland.

Relativism marked most of the work of Bernard Le Bovier, sieur de Fontenelle, whose style and incisive mind had won him renown before 1690. His ***Entretiens sur la pluralité des mondes*** (1686; ***A Plurality of Worlds,*** 1688) are perhaps unequalled as an example of scientific popularization as well as a vigorous attack on prejudice. Here and in the ***Histoire des oracles*** (1687) he revealed a skeptical mind, unimpressed by religious or philosophical authority.

Pierre Bayle, exiled as a Protestant, was essentially a controversialist, attacking superstition and discussing the virtues of the atheist in his ***Pensées diverses sur la comète*** (1682; "Various Thoughts on the Comet") and defending liberty of conscience in the ***Commentaire philosophique*** (1686). His greatest and most influential work was the ***Dictionnaire historique et critique*** (1697). This vast compendium of knowledge is perhaps the greatest mass of erudition ever to have come from one man. In the 18th century, Bayle was regarded as a skeptical agnostic, but some modern readers impute to him a firm religious sentiment.

Fontenelle and Bayle were the most significant precursors of the movement known as the Enlightenment, others being the anonymous writer of ***Les Soupirs de la France esclave*** (1689–90; "The Sighs of France Enslaved") and Abb6 Castel de Saint-Pierre.

Creative literature. *Drama.* The success of Crébillon tempted the young François-Marie Arouet—writing under the pen name Voltaire—to the tragic stage, and in 1718 his ***Oedipe*** was performed, the first of about 30 tragedies, of which several attained glowing success, especially *Zaïre* (first performed 1732) and ***Mahomet*** (1742). Vigorous rhetoric, melodramatic effects often inspired by Shakespeare, only intermittent respect for the decorum of the previous age—these characteristics marked a genre that became for him a weapon in philosophical controversy. Other writers enjoyed great fame in their day. Houdar de la Motte's ***Inès de Castro*** (1723) broke all records of financial success, and the plays of Buyrette de Belloy were frequently performed.

The plays of Marivaux and Beaumarchais

Philippe-Néricault Destouches wrote comedies of character with a moral aim; ***Le Glorieux*** (1732: ***The Conceited Count,*** 1923) is the best known. Alexis Piron's ***MCtromanie*** (1738) satirized poets. The two writers of comedies of greatest talent, however, were Pierre Carlet de Chamblain de Marivaux and Pierre-Augustin Caron de Beaumarchais. Marivaux wrote prose comedies of gentle intrigue, in which love is the motive force, his best known being ***Le Jeu de Pamour et du hasard*** (1730; ***The Game of Love and Chance,*** 1923). His psychological analysis is subtle, his style highly polished, and his characters sophisticated. His plays portray fashionable Parisian society of the 18th century.

Beaumarchais, a political adventurer and financial intriguer, wrote a number of plays, of which ***Le Barbier de Séville*** (1775) and ***Le Mariage de Figaro*** (1784) are the greatest. Refinement of wit, bold social criticisms of the nobility, and above all the character of Figaro, carefree, eloquent, and unscrupulous, have ensured their lasting success.

Various attempts were made to escape from the rigid classical distinction between tragedy and comedy. Nivelle de La Chaussée, whose principal plays are ***Le Préjugé à la mode*** (1735; "Fashionable Prejudice") and ***Mélanide*** (1741), wrote comedies from which the comic was banished and that represented domestic crisis and anxiety. More lasting in importance was the domestic drama, the ***tragédie bourgeoise,*** or "middle class tragedy," developed by Denis Diderot. Influenced by English examples, Diderot devised a theory of a play in which the passions are as elevated as in tragedy, but the characters are of a more humble social status. He believed that the traditional genres were exhausted. He regarded the drama, moreover. as a useful instrument for the inculcation of those social lessons that were dear to the 18th century. He wrote two plays as exemplars of the new genre, ***Le Fils nature1*** (published 1757, first performed 1771; ***Dorval; or, The Test of Virtue,*** 1767), and ***Le Père de famille*** (1758; ***The Father,*** translation published 1770). Other writers followed Diderot's example, notably Michel-Jean Sedaine with ***Le Philosophe sans le savoir*** (1765; ***The Duel;*** 1772) and Sébastien Mercier with an adaptation of ***Rorneo and Juliet*** under the title ***Les Tombeaux de Vérone*** (published 1782). The new genre was important because it prepared the way for the Romantic drama of the 19th century.

Poetry. There are several poems of merit among the ***Odes chrétiennes et philosophiques*** (published 1771) of Le Franc de Pompignan. Louis Racine, son of the great Racine, deserves mention for his poem ***La Religion*** (1742), which expresses the austere but unorthodox sentiments of Jansenist Catholicism in language that echoes that of Pascal, as does Charles-Pierre Colardeau for his popular epistles, the ***Héroïdes.*** The principal poetic figure until late in the century was, however, Voltaire, who alone in his age was able to write a successful epic, ***La Henriade*** (1723). Many of his poems are biting satires, while some, such as the ***Poème sur le désastre de Lisbonne*** (1756), show great compassion.

In the second half of the century Jacques Delille published *Les Jardins* (1782; *The Gardens,* 1798), a descriptive and didactic poem in four books, and Antoine-Léonard Thomas wrote an *Ode sur le temps* (1762; "Ode on Time"). But greater than any of these was André Chénier, who was born in Constantinople and died on the guillotine. His poems, such as *Odes, Élégies,* and *Iambes* (*Oeuvres complktes,* published 1819), show personal emotion and chastened classical beauty.

Manon Lescaut

The novel. The novel in the 18th century was in the ascendant. Alain-René Lesage, after *Le Diable boiteux* (1707; *The Devil upon Two Sticks,* 1708), a satirical tale, published *Gil Blas* in three stages at ten-year intervals (1715, 1724, 1735). In this novel, which owed much to Spanish literature, the travels of Gil Blas are described through good and ill fortune across Spain. Satirical portraits are numerous. But when Abbé (Antoine-François) Prévost (d'Exiles), in the last volume of his *Mémoires et aventures d'un homme de qualité qui s'est retiré du monde,* produced *Manon Lescaut,* the novel advanced far beyond this. The heroine Manon is beauty and treachery incarnate, while her lover des Grieux, whose conscience is always feebler than his passion, represents the heightened emotionalism that was known as *sensibilitk.* The novels of Marivaux (*La Vie de Marianne* [1731–41; *The Life of Marianne,* 1736–42] and *Le Paysan parvenu* [1735–36; *The Fortunate Peasant,* 17351) are comparatively placid. The vigour of Prévost is found also in *La Nouvelle Héloïse* (1761) by Jean-Jacques Rousseau, inspired by the English writer Samuel Richardson. This novel in letter form enjoyed great renown. The characterization is vivid, the descriptions of natural beauty, unusual in the 18th century, are frequent; the work represents the climax of *sensibilitk* in heralding Romanticism. This line of development was further marked by the publication of the exotic novel *Paul et Virginie* (1787) of Bernardin de Saint-Pierre; *Les Liaisons dangereuses* (1782; *Dangerous Acquaintances,* 1924), also in letter form, of Choderlos de Laclos; and Diderot's novels. The philosophical short stories of Voltaire, such as *Zadig* (1747) and *Candide* (1759), combine brilliant political satire with a crisp and witty style and retain their brilliance after two centuries.

Rousseau's *Confessions*

Comparable to the growth of the novel in the 18th century was that of the autobiographical memoir, which is now useful to both the historian and the literary historian. The *Mémoires* of the Duc de Saint-Simon provide a valuable source of information about court life and political activity from 1691 to 1723. The other important figure was Rousseau, whose *Confessions* (written 1765–70; published posthumously, 1781–88) are frankly and almost morbidly introspective. They give a character analysis that is of rare penetration and set forth with great literary skill. It was a new kind of autobiography, which showed Rousseau as a pre-Romantic.

The Enlightenment. The 18th century in France was remarkable less for the conventional genres of literature than for the development of its thought, the currents of which combined to form what is known now as the Enlightenment. A new orientation, skeptical and rationalist, had been given to French thought by Fontenelle and Bayle and by translations of the English philosopher John Locke. The *Lettres persanes* (1721; *Persian Letters,* 1722) of Charles-Louis de Secondat de Montesquieu contained more advanced ideas, clad in a seductively witty form: the greater importance of morality than of faith; religion viewed primarily as a social phenomenon, stripped of its revealed elements and reduced to the light of reason; the desirability of religious toleration; and the iniquity of despotism. These notions, new to the 18th-century French reader, stimulated thought and shook prejudice, and the *Lettres persanes* enjoyed great success.

Deist and atheist treatises circulated widely during the next decade. Henri de Boulainvilliers, biographer of Muhammad, argued for the rights of the nobility against the central government and produced a celebrated theory of the Germanic origin of the French state in his *Histoire de l'ancien gouvernement de la France* (1727).

Discussion of similar political themes, often deceptively innocuous, occurred at the Club de l'Entresol, where nobles and bourgeois met as equals. Established in 1724, it was organized by the abbé Alary and included the abbt de Saint-Pierre, René-Louis, marquis d'Argenson, later for a short time foreign minister, and, at least once, Montesquieu. Discussions were uninhibited and opinions unorthodox. Certainly they were displeasing to authority, for the club was dissolved in 1731 by the chief minister, the Cardinal de Fleury.

Voltaire and Montesquieu

Great developments came in the next 20 years, the first person responsible being Voltaire, in whose *Lettres philosophiques* or *Lettres sur les Anglais* (1734), criticism of the clergy, criticism of the nobility, and mockery of out-of-date philosophers were expressed in the simplest language and in a brilliant mixture of gentle raillery and mordant satire. The work's two main positive achievements were the presentation of English civilization, by inference, as a model to be imitated and the introduction on a popular level of the philosophy of Locke. After 1734, sensationalism, the theory that all ideas come from the senses, was one of the central beliefs of the French Enlightenment. The *Lettres philosophiques* was condemned by the Parlement, or court of judicature, of Paris and ordered to be burned.

In the same year appeared Montesquieu's *Considerations sur les causes de la grandeur des Romains et de leur décadence* (1734), in which the underlying causes—physical, intellectual, military, and political—of the greatness and decline of Rome are set forth. Many comments and analogies make the work one of actuality, and it contains political theory as well as historical method. The hatred of despotism shown by Montesquieu was expressed also by Fénelon in his *Examen de conscience d'un roi* (first published 1734; *Proper Heads of Self-examination for a King,* 1747).

Alexander Pope's *Essay on Man* was printed in French more than 60 times before 1789, and the doctrines of order, harmony, and optimism were welcomed by a public already prepared for them by the philosophy of Gottfried Wilhelm Leibniz. Voltaire's seven *Discours en vers sur l'homme* (1738; "Discourses in Verse on Man") were directly inspired by the *Essay on Man.*

Vigorous propaganda in favour of the ideas of Locke is the keynote of *La Philosophie du bon sens* (1737; *Philosophical Dissertations on the Uncertainty of Human Knowledge,* 1753) by the Marquis d'Argens. The same task was undertaken for Sir Isaac Newton, first by the astronomer and mathematician Pierre-Louis Moreau de Maupertuis and then by Voltaire in his *Éléments de la philosophie de Newton* (1738). The increased receptiveness of public opinion is shown by the publication in 1743 of a tiny volume of five treatises bearing the significant title *Nouvelles libertés de penser* ("New Freedom in Thought"); two of the essays have been attributed to Fontenelle and one, the most interesting, "Le Philosophe," to the grammarian and contributor to the *Encyclopédie* Ctsar Chesneau Dumarsais. "Le Philosophe" describes the philosopher as a human machine that, by means of its mechanical constitution, reflects on its movements; he is moved by reason. This notion of a Philosophe, or sage, as a member of a non-Christian elite, guided by reason, more inclined to doubt than to believe, and leaning to materialism was characteristic of the French Enlightenment. It was the 18th century's first clear definition of the term.

The *Introduction à la connaissance de l'esprit humain* (1746; "Introduction to the Understanding of the Human Spirit") of Luc de Clapiers, marquis de Vauvenargues, contains elements of Utilitarianism and traces of influence of the Earl of Shaftesbury's *Inquiry concerning Virtue,* translated into French in 1745 by Diderot, whose *Pensées philosophiques* (1746; *Philosophic Thoughts,* 1916), in which he argued clearly and succinctly for Deism, was at once burned by the public executioner. Diderot followed this by a *Lettre sur les aveugles* (1749; *An Essay on Blindness, c.* 1750), an outspoken examination of the moral consequences of sensationalism, for which he was imprisoned. *Les Moeurs* (1748; *Manners,* 1751), by François-Vincent Toussaint, was a bulkier defense of Deism and natural morality. It enjoyed great popularity and was publicly burned. The more extreme works of Julien Offroy de La Mettrie argued for complete materialism; and he openly denied the existence of the soul in his *Histoire naturelle de l'âme* (1745; "Natural History of the Soul"). In 1748 was published Montesquieu's *De E'esprit des lois* (*The Spirit of Laws,* 1750), one of the great monuments of French thought. Preferring, in relation both to religious and to political systems, to explain rathei than to judge, Montesquieu nevertheless found himself engaged in bitter controversy. His main legacy was his theory of constitutional monarchy with separation of the powers, and the theory of the influence of climate on the character of peoples. *De l'esprit des lois* gave to French 18th-century philosophy a political bias that had not previously been conspicuous.

There had also been some anti-philosophical literature. The abbé Jean-Baptiste Gaultier sought to prove the impiety of the *Essay on Man* and of the *Lettres persanes;* Diderot's *Pense'es philosophiques* was assailed by *Pense'es raisonnables* and *Pense'es anti-philosophiques;* many critiques were written of *De l'esprit des lois,* but the best anti-philosophical work of the first half of the century was the *Anti-Lucretius* (1747) of Cardinal Melchior de Polignac, who endeavoured in dignified Latin to defend the philosophy of Descartes against the growing materialism of his day.

The *Encyclopédie* of Diderot and d'Alembert

In the second half of the century the philosophical movement was marked by great achievement and by violent opposition. This is particularly evident in the case of the *Encyclope'die.* Diderot and Jean d'Alembert were put in charge of the project, which soon expanded into the production of an original compendium of knowledge with a theoretical basis of the study of knowledge and cultural history, outlined by d'Alembert in his *Discours préliminaire.* Many of the articles expressed the new ideas in relation to religion and politics. At two stages in the course of publication the *Encyclope'die* encountered grave difficulties. The first occasion was when one of the *Encyclopédie*'s contributors, the abbé Jean-Martin de Prades, was censured for heresy because of a thesis he had written; the second was a result of the coincidence of the publication of some dangerous articles (notably one by d'Alembert on Geneva) with a crisis provoked by an attempt on the King. Diderot succeeded with the aid of the chevalier Louis de Jaucourt in bringing the work to completion in 1765. This and the Comte de Buffon's monumental *Histoire naturelle* (1749–1804), which covered geology, botany, and zoology with remarkable penetration, contributed greatly to the advancement of knowledge. More significant, however, was its contribution to the theory of progress, belief in which is found in most philosophical works of this period; it was the theme of a lecture at the Sorbonne in 1750 by the future chief minister Turgot; it inspired utopian writings and received its final expression in a work that sums up all the aspirations of the 18th century, the *Esquisse d'un tableau historique des progrès de l'esprit humain* (1795; *Sketch for a Historical Picture of the Progress of the Human Mind,* 1955) by the mathematician, philosopher, and revolutionary the Marquis de Condorcet.

An apparent dissenter from this theory of progress was Jean-Jacques Rousseau, who argued for the natural goodness of man and the corrupting influence of society in his *Discours sur les sciences et les arts* (1750; "Discourses on the Sciences and Arts") and *Discours sur l'origine et le fondement de l'inégalité parmi Ies hommes* (1755; *A Discourse upon the origin and foundation of the inequality among nzankind,* 1761), in which the institution of property is denounced and the right of insurrection against a tyrant is asserted. This was followed by the boldly outspoken Socialist work *Le Code de la nature* by Morelly. He denounced property as the cause of all evil and laid down the principles of good society with a constitution and code of laws. Étienne de Condillac in his *Traite' des systèmes* (1749) surveyed previous philosophies and sought a means of avoiding their errors. He laid great stress on the linguistic element in philosophy. His *Traité des sensations* (1754) is the most rigorously sustained expression of the theory of sensationalism.

The philosopher Claude-Adrien Helvétius, in his well-written treatise *De l'esprit* (1758; "Of the Mind"), attacked all forms of morality based on religion and set forth the principles of utilitarianism. The publication of *De E'esprit,* which appeared openly and with royal privilege carelessly given, provoked a grave crisis. Helvétius was made to recant, publication of the *Encyclope'die* was stopped, and *De l'esprit* was burned, along with various other philosophical works, including Voltaire's innocuous *Poème sur la loi naturelle.*

The growth of atheism and realism

The fortunes of the Philosophes rapidly and surprisingly improved, however, because of the government's hostility to their worst enemy, the Society of Jesus, culminating in 1764 in the suppression of the order in France. The *Encyclope'die* was allowed to resume and publication was completed. Meanwhile, Rousseau published *Du contrat social* (1762; *The Social Contract,* 1902). In it he declared that primitive man is free from all constraint, that society comes into existence by man's free acceptance of the social contract, and that in any legitimate society the sovereign power rests in the people. Two years later appeared the posthumously published *Considérations sur le gouvernement ancien et présent de la France* by the Marquis d'Argenson, who argued for an extension of democracy within the framework of the French monarchy.

Almost simultaneously with *Du contrat social,* Rousseau published his *Émile, ou de I'e'ducation,* which evoked a great storm of criticism. The "Profession de foi du vicaire savoyard" ("Profession of Faith by a Savoyard Vicar"), which it contains, is perhaps the century's most eloquent plea for constructive Deism, while the book as a whole had a profound and lasting influence on European education.

An unknown parish priest, Jean Meslier, left a lengthy denunciation of all religion and an advocacy of atheism, and in 1762 Voltaire published an *Extrait des sentiments de Jean Meslier,* in which Meslier's hostility to Christianity retained its vigour of expression. Paul-Henri Dietrich, baron d'Holbach, followed up *Le Christianisme dévoilé* (1761; *Christianity Unveiled,* 1795) with a series of anti-religious treatises and in 1770 published his most famous work, the *Système de la nature.* Here the anti-religious thought of the 18th century reaches its climax. The soul is defined as a false idea, God is denied, and necessity is declared to govern all human actions.

From the finality of d'Holbach's views there could be no advance, and in relation to religion and metaphysics it is the Enlightenment's extreme position. Meanwhile, however, Diderot, in such dialogues as the *Supplément au voyage de Bougainville* (written 1772; published 1796) and *Le Rêve de d'Alembert* (written 1769; published 1830; "D'Alembert's Dream") showed materialist doctrine resting on a personal interpretation of natural science and particularly of biology. With great suppleness of mind Diderot asked more questions than he answered.

The social and political thought of the Enlightenment was taken further by Abbé Raynal in his *Histoire philosophique et politique des établissements et du commerce des Européens dans les deux Indes* (1770, revised 1774 and again in 1780 with daring additions; "Philosophical and Political History of the European Establishment and Commerce in the Two Indies") and by Abbé Mably, who expressed some Socialistic ideas in *De la législation* (1776), while d'Holbach's own *Système social* (1773) is the best French exposition of a Utilitarian political philosophy.

After the accession of Louis XVI in 1774 there was little new in French thought for many years. The Philosophes had overstated their case; they nevertheless handed on what was to be a permanent acquisition to human thought.

Provencal. In the 18th century, priests were mainly responsible for the literary output of Languedoc. The *Sermon de moussu Sistre,* of the abbé J.-B. Favre, is a masterpiece. He also wrote a mock-heroic poem *(Lou Sikge de Cadaroussa).* a novel depicting the country manners of the time (*Istoria de Jan l'an pres*), and two comedies. Another priest, Jean-Claude Peyrot de Pradinas, published *Les Quatre Saisons* in 1781. Two genuine poets are the brothers Rigaud of Montpellier: Auguste and Cyrille. Pierre Hellies of Toulouse, a poet of the people, has a certain rough charm. In Provence, Toussaint Gros, of Lyons, holds undisputed sway. His style and language are admirable, but unfortunately he wasted his gifts largely on trival occasional pieces. In Gascony the most celebrated poet is Cyrien Despourrins.

Walloon. Use of the patois broadened in the 18th century. The success of comic opera at Liège resulted in several noteworthy librettos. *Li Voyadjue di Tchaudfontaine* (1757; "The Journey to Chaudfontaine"), *Li Lîdjwès Egagî* ("The Enlisted Liégeois"), and *Les Hypocondres* ("The Hypochondriacs") resulted in the formation of the Théâtre Liégeois. In lyric poetry the cramignon (a type of song for dancing) and the *Noels* (Christmas carols and dialogue) adopted a genuine realism.

SWISS LITERATURE IN THE 18TH CENTURY

French Swiss. Although the beginnings of French Swiss literature proper go back to the Reformation, it was not until the later 18th century that French Swiss writers of international importance emerged. The chief of these was Jean-Jacques Rousseau, whose work is discussed above. Rousseau was born in Geneva and described himself in *Du contrat social* as a "citizen of a free state." He shared the political and psychological conceptions of all Switzerland, as can be seen in *Considérations sur le gouvernement de Pologne* ("Considerations on the Government of Poland"). Swiss subject matter was also dealt with by H.B. de Saussure, one of the founders of modern geology, in his *Voyages dans les Alpes* (1779–96).

German Swiss. The beginning of 18th-century literature in German-speaking Switzerland is marked by the works of a Bernese poet, Albrecht von Haller, whose "Die Alpen" gives expression to the beauty and sublimity of Alpine scenery and the moral purity of its inhabitants. Salomon Gessner is read for his prose pastoral idylls. The most important work of Johannes von Müller, *Geschichten schweizerischer Eidgenossenschaft* (1786–1808; "The History of the Swiss Confederation"), though a fragment, became the most representative piece of historical writing in German Neoclassicism and had a far-reaching influence on the style of historical writing. A Zürich school of criticism, including J.J. Bodmer and J.J. Breitinger, opposed French Rationalism and Classicism and promoted an appreciation of English literature in Germany. Their work started a revival of religious and national conceptions. J.H. Pestalozzi wrote the first village story in German but included in it the nucleus of his later meditations on the nature and destiny of man.

GERMAN LITERATURE IN THE 18TH CENTURY

The age of Enlightenment. If religion was the dominant factor in German intellectual and spiritual affairs in the 17th century, the Enlightenment of the 18th brought about a reaction. Man now claimed to be able to understand the universe by virtue of his possession of the divine gift of reason. Empirical and Idealist thinkers alike were united in rejecting traditional authority. In a rational universe, governed by the law of cause and effect, there was room neither for mystery nor for the doctrines of original sin and predestination. Evil was the result of irrational conditions of life, and man had it in his power to improve his lot by the pursuit of science and education. An optimistic belief in human perfectibility was generally held; it lay in the cultivation of reason and tireless effort in the service of human improvement. The man of the world being more highly regarded than the devout Christian, good taste and common sense came to be demanded, and literature assumed a markedly didactic character.

Rationalism. The intellectual life of Germany had recovered rapidly from the social upheavals of the Thirty Years' War. The foundations of Rationalism were laid by Gottfried Leibniz, the first of the great German philosophers. With him, the relationship of God and man to each other ceased to be considered within the limits of Christian dogma. German religious life was marked by a revival of Pietism, which left its traces in the sphere of religious poetry. The main emphasis lay not in conformity but in the individual's spiritual experience.

The influence of English translations

In literature the new ideas soon began to emerge. Translations and imitation of the English *Spectator*, *Tatler*, and *Guardian* helped to regenerate literary taste. One of the most marked features of German literature in the 18th century was the progressive influence of English literature: first, of Joseph Addison, Jonathan Swift, Daniel Defoe, and Alexander Pope; later of James Thomson, John Milton, and Edward Young. Samuel Richardson had much effect upon the growth of the moral novel, while Young's *Conjectures on Original Composition* heralded a new epoch in German literature that was to be profoundly affected by the Scots poet James Macpherson's *Ossian*, Bishop Percy's *Reliques of Ancient English Poetry*, and Shakespeare—the epoch of the *Sturm und Drang* ("Storm and Stress") movement.

The reaction against Rationalism. Between 1724 and 1740 the critic Johann Christoph Gottsched succeeded in establishing in Leipzig literary reforms in accord with French 17th-century Classicism. He purified the stage and laid down principles according to which good literature was to be produced and judged. The limitations of Gottsched soon drew resistance from two Swiss scholars, J.J. Bodmer and J.J. Breitinger. Basing their arguments on John Milton's poem *Paradise Lost*, they insisted that the imagination should not be dominated by reason. The effects of the controversy appeared toward the middle of the century in a group of Leipzig writers of Gottsched's own school, the Bremer Beitrager, or "Bremen Contributors," as they are usually called after the paper in which they published their work. In this—the *Neue Beiträge zum Vergnügen des Verstandes und* Witzes—there appeared in 1748 the first installment of an epic by F.G. Klopstock, *Der Messias* (completed 1773; *The Messiah*, first complete verse translation 1821), whose theme created a sensation when the first cantos appeared. Klopstock's genius was, however, more suited to the lyric, and his odes, in which both patriotic and sentimental themes were prominent, gained an immense reputation. Friedrich von Hagedorn in Hamburg showed to what perfection occasional verse could be brought, while Ewald Christian von Kleist wrote admirable sentimental nature poetry. Meanwhile, a rising interest in Germanic antiquity aided the growth of the "bardic" movement led by H.W. von Gerstenberg, F.F. Kretschmann, and Michael Denis, the translator of Macpherson's *Ossian*.

The influence of Lessing. As Klopstock had been the first of modern Germany's inspired poets, so Gotthold Ephraim Lessing was the first critic who brought credit to the German name throughout Europe. Like Gottsched he had unwavering faith in Neoclassicism, but classic meant for him, as for his contemporary J.J. Winckelmann, Greek art and literature rather than French pseudo-Classicism, though it is true that Lessing's own exposition of Aristotle's theory of tragedy was full of the moral preoccupations of the Enlightenment. He looked to England rather than to France for the regeneration of the German theatre. His own dramas were pioneer works in this direction. *Miss Sara Sampson* (first performed 1755) was a bourgeois tragedy on the English model; *Minna von Barnhelm* (1767), a comedy in the spirit of George Farquhar; in *Emilia Galotti* (1772) Lessing remodeled the "tragedy of common life" in a form that came to be acceptable to the *Sturm und Drang;* and finally in *Nathan der Weise* (1779; *Nathan the Wise*, 1781) he won acceptance for iambic blank verse as a medium for elevated drama.

Lessing's dramatic works

Because of Lessing, German literature made a great leap forward beyond the feeble achievements of the first half of the century. The domestic tragedy, its plot centred upon the problem of class distinction, foreshadowed plays involving marked political and social criticism in the *Sturm und Drang* period, while *Nathan* was an important forerunner of the "drama of ideas" of Weimar Neoclassicism. Lessing's theoretical work placed criticism in the forefront of affairs in literary Germany. His sharp rejection of descriptive poetry had a great effect upon the writing of the next generation, while his attack upon the literary authority of France prepared the way both for a greater attention to English examples and for the search for native originality.

C.M. Wieland

To the widening of the German imagination C.M. Wieland contributed by introducing remote and exotic literary settings, largely under French inspiration. With the exception of his verse-romance ***Oberon*** (published 1780), his work fell into neglect; he did excellent service, however, to the development of German prose fiction with his psychological novel ***Agathon*** (1766–67) and his satire ***Die Abderiten*** (1774; "The Simpletons"). Wieland had a considerable following, and the German novel owed much to the example of ***Agathon,*** but its groundwork and form were borrowed from English models.

The age of Goethe: Sturm und Drang. The period of Neoclassicism and Romanticism, the greatest epoch in German literature, fell within the lifetime of Johann Wolfgang von Goethe. The age of Goethe went beyond the Enlightenment's substitution of science for religion, inasmuch as it ascribed to science only a peripheral position in relation to the ultimate questions of life. It insisted upon the value of feeling in face of the limitations of reason. Impulse, instinct, emotion, fancy, and intuition acquired a quasi-religious significance as being the links that connect man with divine nature. The ideal of the classical age, soon to be called ***Humanität*** ("humanness"), was that of the fully developed personality in which intellect and feeling should be harmoniously balanced. Three phases may be distinguished in the evolution of this new outlook: ***Sturm und Drang,*** Classicism, and Romanticism.

Goethe belonged to and profoundly affected the movement known as ***Sturm und Drang*** ("Storm and Stress"), which aimed at overthrowing the cult of Rationalism. Seeds of the new growth were to be found in Klopstock, in the spiritual force of Pietism, and in the rising resistance to French Classical taste, while the influence of Rousseau, Edward Young, and James Macpherson and of the recently translated Shakespeare was of prime importance. Nature, genius, and originality were the slogans of the new epoch. An increasingly oppressive sense of dissatisfaction with the civilization of the day assailed the new generation. The cult of nature completely replaced orthodox religion. No law was recognized as being above the individual conscience. The standard outlook thus demanded unceasing effort, like that of Faust. Strain, protest, revolt, yearning, disillusion were obvious on all sides, and egotism became a dominating feature in literature and thought.

The influence of Herder

The critical writings of H.W. von Gerstenberg stressed personal feeling in matters of taste, but the chief impetus came from Johann Georg Hamann who emphasized the inspirational and symbolical function of language. His pupil was Johann Gottfried von Herder, who grasped, as no thinker before him had done, the idea of historical evolution and engendered the main current of the ***Sturm und Drang.*** Herder's doctrine of ***Humanitat*** was fundamental to German Neoclassicism. He stressed the value of historical continuity in literature and pointed to the folk songs, ballads, and romances of the Middle Ages as sources of inspiration to which Bishop Thomas Percy's ***Reliques of Ancient English Poetry*** had recently drawn attention. The ***Reliques*** also affected the poets who founded in 1772 the Gottinger Hain, or, literally, Gottinger Grove, and who belonged mainly to the peasant class or the lower bourgeoisie; J.H. Voss, the leader of the group and author of the famous idyll ***Luise*** (1795), was a typical North German peasant.

The ***Sturm und Drang*** was intimately associated with Goethe. Goethe had made the acquaintance of Herder, who interested him in Gothic architecture, the ***Volkslied*** ("folk song"), and Shakespeare. A pamphlet, "Von deutscher Art und Kunst" (1773; "Concerning German Nature and Art"), was a kind of manifesto of the ***Sturm und Drang.*** The new ideas seemed at once to set Goethe's genius free. His ***Gotz von Berlichingen*** (published 1773), the first important drama of the ***Sturm und Drang,*** was followed by the first novel of the movement, ***Die Leiden des jungen Werthers*** (1774; ***The Sorrows of Young Werther),*** which made the author world famous. In all forms of literature he set the fashion for his time. The Shakespearean restlessness of ***Gotz von Berlichingen*** found imitators in J.M.R. Lenz, F.M. von Klinger, J.A. Leisewitz, H.L. Wagner, and Friedrich Müller. The dramatic literature of the ***Sturm und Drang*** was its most characteristic product; it was inspired by the desire to present upon the stage figures of Shakespearean grandeur impelled by gigantic passions, all considerations of plot, construction, and form being subordinated to character, and all accepted authority—literary, social, political, or moral—being rejected.

With the production of ***Die Rauber*** (1781; ***The Robbers,*** 1792) by Friedrich Schiller, the drama of the ***Sturm und Drang*** entered upon a new phase. Schiller's tragedy was more skillfully adapted to the exigencies of the theatre than those of his predecessors had been; it and the succeeding dramas, ***Die Verschworung des Fiesko zu Genua*** and ***Kabale und Liebe,*** were fine pieces of high promise. Germany owed to the ***Sturm und Drang*** period its national theatre; permanent theatres were established in these years at Hamburg, Mannheim, and Gotha, and the Burgtheater (now the Hofburgtheater) was founded at Vienna in 1776.

Neoclassicism. Self-discipline was lacking as a ***Sturm und Drang*** theme, and the movement soon exhausted itself. A more positive form of moral idealism appeared in the poetry of Goethe and Schiller, as well as in the philosophy of Immanuel Kant. The problem of freedom was rendered acute by the impact of the French Revolution, which German literature generally regarded as a warning. Schiller believed the antagonism between duty and inclination could be resolved once morality became "second nature" and this could be achieved only through the contemplation and production of beauty. Art thus acquired an educational function, and aesthetic education was one of the major objectives of Neoclassicism.

Goethe's Weimar period

For Goethe a new phase in his development began with his departure for Weimar in 1775, while after ***Don Carlos*** (1787) Schiller turned from poetry to study history and philosophy. The first 10 years of Goethe's life in Weimar were marked by his renewed friendship with Herder, by his public service as a minister of state, and by his emotional attachment to Charlotte von Stein. He did not achieve greater clarity in his ideas until after his sojourn in Italy (1786–88). In Italy he turned his attention to three dramatic works; he gave ***Iphigenie auf Tauris*** its final form, completed ***Egmont,*** and replanned ***Torquato Tasso. Wilhelm Meisters Lehrjahre (Wilhelm Meister's Apprenticeship,*** 1824), Goethe's most important novel, had become, by the time it appeared in 1795–96, a book on the conduct of life. It is an outstanding example of the ***Bildungsroman,*** or "educational" novel, a characteristic German novel form, and profoundly affected future practitioners of the genre.

Before ***Wilhelm Meister*** appeared, however, German thought and literature had arrived at that degree of stability in form and ideas essential to a great literary period. In the year of Lessing's death (1781), Kant had published his ***Kritik der reinen Vernunft (Critique of Pure Reason).*** Under the influence of Kant, Schiller turned to the study of aesthetics, the first fruits of which were his philosophical lyrics and his treatises "Über Anmut und Wiirde" (1793), and, between 1795 and 1796, "Briefe iiber die ästhetische Erziehung des Menschen" ("Letters on the Aesthetic Education of Man") and "Über naive und sentimentalische Dichtung" ("On Naive and Sentimental Poetry").

The years 1794 to 1805, when Goethe and Schiller were united in close friendship in Jena and Weimar, mark the culmination of literary Neoclassicism. Schiller provided the theoretical basis; Goethe, as director of the ducal theatre, influenced the whole practice of dramatic production in Germany. Under his encouragement Schiller turned from philosophy to poetry and between 1798 and his death in 1805 wrote a series of Classical dramas that are Germany's greatest: the trilogy ***Wallenstein, Maria Stuart, Die Jungfrau von Orleans (The Maid of Orleans),*** Die Braut von Messina (The Bride of Messina), Wilhelm Tell, closing with the fragment *Demetrius.*

Apart from these two great poets, however, literature was in no very healthy condition; the stage was dominated by the extraordinarily popular plays of A. von Kotzebue; and there is a wide gap between K.P. Moritz's Anton Reiser or the philosophical novels which F.M. von Klinger wrote in his later years and Goethe's Wilhelm *Meister*.

Goethe's Faust

The supreme work of Goethe's latter years is Faust, Germany's greatest contribution to the literature of the world. In Part I (1808) is set out Faust's despair, his pact with Mephistopheles, and his love for Gretchen; Part II (1832) covers the magician's life at court, the winning of Helen of Troy, and Faust's purification and salvation. The doctrine of the fulfillment of life by striving and selfless activity revealed in Faust was fundamental to Goethe's mature writing. The tragic novel Die *Wahlverwandtschaften* (1809) had insisted on the theme of renunciation. Wilhelm Meisters *Wanderjahre* (1821–29; Wilhelm Meister's Travels, 1827), with its social utopianism and teaching of restraint, offered a criticism of the rise of industrialism. The autobiographical Dichtung und Wahrheit (1811–33; Poetry and Truth: From My Own Life, 1908), dramatic pieces, scientific writings, and lyrics indicate the manysidedness of Goethe's achievement.

The first phase of the Romantic movement. The Romantic movement began not so much as a protest against the Neoclassicism of Weimar but as a radical extension of some of its beliefs and interests, especially at first, its emphasis upon Greek antiquity, longed for like some lost paradise. The Romantic poet could create his own world from reality or from fancy and he could turn whatever he liked into poetry. The nature of individuality and its modes of approach being infinitely varied, there was to be no end to the innovations made in content and style by the great wealth of literary talents who now emerged all over Germany and from various strata of society. The rising generation felt free and able to revise all accepted representative values, not only in art and literature but in other spheres as well. The subconscious became a subject of serious study and the supernatural came into vogue, while in religion, Romantic subjectivism led to mysticism. In the evolution of German Romanticism no small part was played by the philosopher J.G. Fichte and the theologian F.D.E. Schleiermacher.

Holderlin and Jean Paul

It was inevitable that disappointment and disillusion should result from this new brand of egotism in thought and literature. Friedrich Holderlin, one of Germany's greatest lyrical poets and author of the novel *Hyperion* (1797–99), grew up as an admirer of Schiller; he sank into despair on realizing the impossibility of his longing for an age of heroic idealism and beauty such as that of ancient Greece. Johann Friedrich Richter, known as Jean Paul, was a disciple of Herder. His sentiment, ingenuity, whimsical style, and lavish detail gained for his shapeless novels a vast degree of popularity; his sustained attention to contemporary life was a new feature, and his work foreshadowed the Dorfgeschichte, or "village story," of later decades. His principal novels, written between 1795 and 1804, were *Hesperus*, *Quintus Fixlein*, Siebenkas, Titan, and Flegeljahre ("The Awkward Age").

The first Romantic school proper was founded at Jena in 1798. Ludwig Tieck, a leading member of the school, early developed a lifelong enthusiasm for Shakespeare and the Elizabethan and Spanish drama. The theoretical basis of Romanticism was laid down by the two brothers August Wilhelm and Friedrich von Schlegel. These Romantic critics maintained that the first duty of criticism was to understand and appreciate. The greatest imaginative achievement is to be found in the lyrics and fragmentary novels of Novalis (pseudonym of Friedrich von Hardenberg), in which Christian mysticism, romantic medievalism, and symbolism transfer the reader into the realm of the *Märchen*, or fairy tale, particularly in Hrinrich von Ofterdingen (1802). The universal sympathies of the movement were exemplified by many admirable translations, of which the greatest was A.W. Schlegel's translations of Shakespeare's plays. The critical essays and aphorisms of F. von Schlegel argued that modern (or Romantic) literature should deal with modern life in all its manifestations, and they put forward *Wilhelm* Meister as the model to be followed.

DUTCH LITERATURE IN THE 18TH CENTURY

The appearance in 1669 of the first literary society (dichtgenootschap) was an omen of a decline in Dutch literature lasting through the 18th century. Material well-being sapped the vitality of the nation. Even the talented poet Hubert Poot suffered from the delusion of his day that rococo flourish and prescribed form were the criteria of poetry. Prose, too, now consisted almost exclusively of translations and bombastic disquisitions. Significantly, Justus van Effen wrote in French before he founded De *hollandsche* spectator (1731–35). The simple style of his moralizing essays contrasts with the work of his contemporaries, and his descriptive realism links him with two popular authors on the Dutch domestic scene, Elizabeth Wolff-Bekker and Aagje Deken.

The first Dutch novel

Elizabeth Wolff-Bekker, essayist and poet, blended rationalism and romanticism in her creative genius. Her association with Agatha Deken as friend and fellow writer produced Sara Burgerhart (1782) in letter form, dedicated to "Dutch young ladies," the first Dutch novel. Remarkable for its wit and realism, it owed much to the English novelist Samuel Richardson. Elizabeth's intelligence and humour also dominated the original purpose of morally educating the young in their eight-volume *Willem Leevend* (1784–85).

By the end of the century a number of poets, including Hieronymus van Alphen, Rhijnvis Feith, Jacobus Bellamy, and Antony Staring were reacting against Neoclassicism, but only Staring achieved more than an echo of German sentimentalism, through his descriptive talent and subtlety. His contemporaries preferred Willem Bilderdijk, whose genius was almost smothered with excesses of rhetoric, but whose Protestant zeal had repercussions in the Réveil ("Revival"), a spiritual movement which gave impetus to the literary revival of the 1830s.

ITALIAN LITERATURE IN THE 18TH CENTURY

Tragedy. In 1713 Francesco Scipione Maffei, an antiquary of Verona, produced Merope—a tragedy which met with great success and pointed the way toward reform of the Italian tragic theatre. Between 1726 and 1747 Antonio Conti—an admirer of Shakespeare—wrote four Roman tragedies in blank verse. It was not until 1782 and the writing of Saul, however, that an important Italian tragedian finally emerged in the person of Vittorio Alfieri. In strong contrast with Pietro Metastasio's and Paolo Rolli's melodramtni—librettos set to music or sometimes performed as plays in their own right—Alfieri's tragedies are harsh, bitter, and unmelodious. He chose classical and biblical themes and through his hatred of tyranny and love of liberty he aspired to move his audience with magnanimous sentiments and patriotic fervour. Alfieri's influence in the Romantic period and the Risorgimento was immense, and, like Carlo Goldoni, he wrote an important autobiography, which gives a revealing account of his struggles to provide Italy with a corpus of drama comparable with that of the other European nations.

Comedy. Metastasio's reform of the operatic libretto was paralleled in the mid-18th century by Goldoni's reform of comedy. Throughout the 17th century the commedia dell'arte—a colourful pantomime of improvisation, singing, mime, and acrobatics, often performed by actors of great virtuosity—had gradually replaced regular comedy, but by the early 18th century it had degenerated into mere buffoonery and obscenity with fixed characters and mannerisms. The dialogue was mostly improvised, and the plot—a complicated series of stage directions, known as the scenario—dealt mainly with forced marriages, star-crossed lovers, and the intrigues of servants and masters. Goldoni succeeded in replacing this traditional type of improvised farce with comedies whose composition was preserved in a written script. His characters are acutely described, but, unlike Molière, he did not plunge into psychological depths. He portrayed the Venetian scene especially with wit and vigour, writing many of his best comedies in Venetian dialect. Goldoni's rival and a bitter controversialist, his fellow-Venetian Carlo Gozzi also wrote comedy and satirical verse.

The world of learning. Giambattista Vico, Ludovico Antonio Muratori, Apostolo Zeno, and Scipione Maffei were writers who reflected the awakening of historical consciousness in Italy. Muratori collected the primary sources for the study of the Italian Middle Ages; Vico, in the *Scienza nuova* (1725–44; "New Science"), investigated the laws governing the progress of the human race and from the psychological study of man endeavoured to infer the laws by which civilizations rise, flourish, and fall. Giovanni Maria Mazzuchelli and Gerolamo Tiraboschi devoted themselves to literary history. Literary criticism also attracted attention; Gian Vincezo Gravina, Vico, Maffei, Ludovico Antonio Muratori, and several others, while advocating the imitation of the classics, realized that it should be cautious and thus anticipated critical standpoints that were later to come into favour.

The Enlightenment. With the end of Spanish domination and the spread of the ideas of the Enlightenment from France, reforms were gradually introduced in various parts of Italy. The new spirit of the times led men—mainly of the upper middle class—to enquire into the mechanics of economic and social laws. The ideas and aspirations of the Enlightenment as a whole were effectively voiced in such organs of the new journalism as Pietro Verri's periodical *Il Caffè* (1764–66).

Parini and the literary revival

More than anyone else, Giuseppe Parini, the Alexander Pope or the Boileau of Italy, seems to embody the literary revival of the 18th century. In *Il giorno* (published in four parts, 1763–1801; "The Day"), a long social satire of the rights of blood, he described a day in the life of a young Milanese patrician and revealed with masterly irony, the irresponsibility and futility of a whole way of life. His *Odi* ("Odes"), which are imbued with the same spirit of moral and social reform, are among the classics of Italian poetry.

The satire in the *Sermoni* (1763) of Gasparo Gozzi (elder brother of Carlo) is less pungent, though directed at similar ends, and in his two periodicals—La *Gazzetta Veneta* and *L'Osservatore*—he presented a lively chronicle of Venetian life and indicated a practical moral with much good sense. Giuseppe Baretti—an extremely controversial figure who published a critical journal called *La Frusta Letteraria* ("The Literary Whip"), in which he castigated "bad authors"—had learned much through a lengthy sojourn in England, where his friendship with Samuel Johnson helped to give independence and vigour, if not always accuracy, to his judgments. The *Viaggi di Enrico Wanton* (1749–64), a philosophical novel in the form of an imaginary voyage by the Venetian Zaccaria Seriman, was the most all-embracing satire of the time.

SPANISH LITERATURE IN THE 18TH CENTURY

With the establishment of the Bourbon dynasty after the War of the Spanish Succession (1701–14), renewal of the country's intellectual life began. Numerous academies were created, the most influential being the Real Academia de la Lengua Española (now the Real Academia Española), founded in 1713 to maintain the purity of the language. Men of letters began again to study abroad and discovered how far Spain had deviated from the intellectual course of western Europe. A new spirit of inquiry into the national heritage caused scholars to go back and unearth forgotten medieval literature. Gregorio Mayáns y Siscar, librarian at the royal library, wrote the first biographical study of Spain's greatest novelist, Miguel de Cervantes. A church historian Enrique Flórez, embarking on a vast historical enterprise, *España sagrada*, helped resurrect the whole cultural background of medieval Christian Spain. Landmarks of even greater importance were the publication of the 12th-century epic *Poema de mio Cid*, the works of Gonzalo de Berceo, and Juan Ruiz' *Libro de buen amor*, all for the first time. From all this there resulted a debate between old and new that, waged throughout the middle decades of the century, compelled both sides to reason out their positions and marked the birth of a new critical approach to literature. Two names stand out: Ignacio de Luzán Claramunt, whose work on poetics launched the great Neoclassical polemic in Spain, and Benito Jerónimo Feijóo y Montenegro, a Benedictine monk who, in assailing error, prejudice, and superstition wherever he found them, made a monumental contribution to the intellectual emancipation of Spain. Imaginative prose produced the *Noches lúgubres* (published 1789–90; "Sad Nights") of José Cadalso Vázquez. looking forward to Romanticism, and the *Fray Gerundio* (1758) of José Francisco de Isla, a satire looking back to the picaresque novel.

First publication of the Cid

Poetry, moribund for nearly 100 years, raised a timid head in a small group at Salamanca, led by Diego González, which toward 1775 turned for inspiration to Luis de León, just as two decades later a group at Seville sought to revive the glories of Herrera. Juan Meltndez Valdés, who learned to think from the English philosopher John Locke and to feel from the English poet Edward Young, best exemplified the combination of new influences at work. A conscious artificer rather than a great poet, he helped poetry through the painful apprenticeship necessary to its rehabilitation.

In drama, the second half of the century saw a great battle over the classical "rules" (meaning chiefly the unities of place, time, and action). *La Raquel* (1778), by Vicente Antonio Garcia de la Huerta y Muñoz, showed the capabilities of the reformist school. It fell to Ramón de la Cruz to bridge the gap by his resurrection of the earlier *paso* (one-act prose skit) and longer *entremeses* (interludes) of Lope de Rueda, Cervantes, and Luis Quiñones de Benavente. Based on satirical observation of the Madrid scene, his one-act sketches did not transgress the unities or offend the purist; at the same time they delighted the public and brought the drama back to commenting on life and society. Leandro Fernández de Moratin applied the lesson to the full-length play and produced comedies imbued with deep social seriousness which were yet good theatre. His dialogue in *La conzedia nueva* (1792; "The New Comedy") and El *sí de las niñas* (1806; *The Maiden's Consent*, 1963) ranks among the best prose of the 18th century.

PORTUGUESE LITERATURE IN THE 18TH CENTURY

The 18th century, in Portugal as in Spain, was predominantly prosaic, even in poetry. Yet signs gradually appeared of a literary revolution which developed eventually into the Romantic movement. Luis António Verney poured scorn on prevailing methods of education in *Verdadeiro Método de Estudar* (1746; "True Method of Studying"). Men of liberal ideas went to France and England, and to their example were largely due the reforms that invaded every branch of letters. Of such were Alexandre de Gusmão, Xavier de Oliveira, Antdnio Ribeiro Sanches, José Correia da Serra, Avelar Brotero, and Francisco Manuel do Nascimento. New literary societies called *arcádias* cooperated in the task of reform. In 1720 King John V established the Academia Real da História Portuguesa, which counted among its members such men as Manuel Caetano de Sousa, author of a colossal *História Genealdgica de Casa Real Portuguesa* (1735–49; "Genealogical History of the Portuguese Royal House"). The Academia Real das Ciências, founded in 1779, initiated research into the study of Portuguese literary history. In its ranks were found nearly all the scholars of note at the end of the century, such as the ecclesiastical historian Frei Manuel do Cenáculo; a scientist, Antdnio Ribeiro dos Santos: João Pedro Ribeiro, perhaps his country's first modern historian; and critics Francisco Alexandre Lobo and Frei Fortunato de São Boaventura.

Establishment of the Portuguese Academy

In 1756 António Dinis da Cruz e Silva established the Arcádia Lusitana (or Ullissiponese), its first aim being the uprooting of Spanish influence. The bucolic verse of Dómingos dos Reis Quita signified a return to the native tradition of two centuries earlier. Sincerity and suffering spoke in the justly more famous *Marília de Dirceu* (1792), love lyrics in a pastoral setting, by Tomás António Gonzaga. In 1790, a Nova Arcadia came into being, its two most distinguished members being the rival poets Manuel Maria Barbosa du Bocage and José Agostinho de Macedo.

Outside the Arcadias stood the "Dissidents," among whom were at least two writers of distinction. Few Portuguese satirists have possessed such a command of ridicule as Nicolau Tolentino de Almeida, who painted the customs and follies of his day with devastating accuracy. Francisco Manuel do Nascimento (better known by his pseudonym, Filinto Elísio) addressed himself perseveringly to purifying the language and to restoring the cult of the 16th-century poets. A convert to Romanticism shortly before his death, he prepared the way for its triumph in Portugal.

The attempted revival of drama in Lisbon

Early in the century popular authors attempted a revival of the drama in Lisbon. The *Operas Portuguesas* (1733–41) of António José da Silva owe their name to the interspersing of the prose dialogue with arias, minuets, and *modinhas* (popular light songs). The Arcádia Lusitana sought to raise the tone of the stage, finding its ideals in the classics of antiquity or of the 16th century but immediate inspiration in the contemporary French theatre. Its efforts failed from lack of dramatic talent and of popular appeal. Manuel de Figueiredo was typical; setting out to write pieces "morally and dramatically correct," he produced 14 volumes of plays in prose and verse on national subjects. Utterly lacking in life, they were never acted.

SCANDINAVIAN LITERATURE IN THE 18TH CENTURY

Swedish. After the death of Charles XII (1718) and the collapse of his empire, a utilitarian attitude to life and letters gradually developed in Sweden. Olof von Dalin was the outstanding popularizer of the new ideas of the French and English Enlightenment. Educated at Lund, he later went to Stockholm and began to publish, anonymously, *Then swänska Argus* (1732–34; "The Swedish Argus"), a weekly periodical modelled on that of the Englishman Joseph Addison. One of the first serious journalistic ventures in Sweden, it marked the beginning of a new era, in which orthodoxy gave way to skepticism and enlightenment, baroque to classicism, and German influence to English and French; at this time the middle class began gradually to take over the function of chief upholder of literature. In *Argus* Dalin ridiculed the foibles of the capital and in *Sagan om hasten* (1740; "The Story of the Horse") he showed himself a master of allegorical satire. He also wrote some pseudo-classicist plays that, like many dramatic ventures of the early and mid-18th century, are academic and lifeless. An exception is *Den Svenska sprätthöken* (1740; "The Swedish Fop"), a comedy by Count Carl Gyllenborg.

With the second phase of the Enlightenment, marked by the influence of Rousseau, are associated Hedvig Charlotta Nordenflycht, the epicurean Gustav Philip Creutz, and his stoic friend Gustaf Fredrik Gyllenborg. In *Den Sörgande turturdufwan* (1743; "The Sorrowing Turtledove"), Fru Nordenflycht laments the death of her husband in highly personal lyrics. Creutz was a more sophisticated personality. He wrote little, but his few writings, of which the pastoral *Atis och Camilla* (1762) is the most important, reveal a mastery of form and versification.

The first Swedish novel

Prose—particularly the novel—developed more slowly. The first genuine novel was *Adalrik och Giöthildas äfventyr* (1742–44), by Jacob Mork and Anders Torngren, showing the influence of the Icelandic sagas. Only two 18th-century Swedish writers were of European reputation, and both were scientists: Carl von Linné (Linnaeus; 1707–78) and the mystic Emanuel Swedenborg (1688–1772).

The Gustavian period takes its name from King Gustav III (1746–92), a brilliant man and a patron of art and letters. He was especially interested in drama and opera and thanks to his patronage, a proper theatrical tradition was developed. Gustavus himself sketched out some works, the best of which was an historic opera *Gustaf Vasa*, which was finished in collaboration between Johan Henrik Kellgren and the composer J.G. Naumann. Kellgren, a great academic poet and arbiter of taste, ruled that Swedish literature should be modelled on classicist French patterns, but, beginning as a rationalist and satirist after the fashion of Voltaire, he reluctantly accepted pre-Romantic ideas later. In *Stockholmsposten*, the main organ of literary opinion in the capital, Kellgren used his polemical wit against Thomas Thorild, a truculent champion of individual genius. After Kellgren's death the controversy was carried on by Carl Gustaf af Leopold, who imposed pseudo-classical standards on the academy and applied them in his own rhetorical odes and tragedies. Johan Gabriel Oxenstierna did his most original work while a diplomat in Vienna; his *Skordarne* (1796; "Harvests") and his *Ode to Camilla* reveal pre-Romantic feeling for the beauty of nature. Bengt Lidner was the chief exponent of pre-Romanticism in poetry. His most successful work was the ode *Grefvinnan Spastaras dod* (1783).

Carl Michael Bellman stands apart from the conflicting ideals of the time. A poet and musician, he combined stylized realism with humour and the most uniquely delicate sense of language and rhythm. He was the greatest Swedish lyricist of the 18th century.

The dissertation *Om upplysning* (1793; "On Enlightenment") by Nils von Rosenstein, the first secretary of the Swedish Academy, expressed the ideals of the Gustavian epoch. Various memoirs by G.J. Adlerbeth, G.J. Ehrensvard, and others evoke the witty but artificial atmosphere of Gustav III's court. Gustav IV, who followed, did not encourage literature; nevertheless, the energetic Anna Maria Lenngren wrote some of her best verse satires between 1795 and 1800, many aimed at aristocratic foibles. The sentimental idylls of Frans Mikael Franzén are full of pre-Romantic idealism from German and English sources. As a bishop he wrote some of the most popular hymns of the Swedish Church.

The comedies of Ludvig Holberg

Danish. The 18th century was a fertile period in Danish literature. The great name in the first half of the century was that of Ludvig Holberg, a Norwegian by birth. His most important contributions, written for the Danish theatre which opened in 1722, were 32 comedies of character and manner, including some moral allegories in his old age. His aim was to create a modern Danish literature on European lines and to make people laugh at their own follies. Influenced by English and French thinking, he was a rationalist and a moderate. He also wrote satires; *e.g.*, *Peder Paars* (1719), a mock-heroic poem, and *Nicolai Klirnii iter Subterraneum* (Latin, 1741; *Journey of Niels Klim to the World Underground*, 1960). His *Moralske tanker* (1744; "Moral Thoughts") and his *Epistler* (1748–54; "Letters") are the finest examples of a Danish political essay form. He also contributed a number of valuable historical works.

Among Holberg's contemporaries the finest lyrical poets are H.A. Brorson, a mystic whose pietist hymns often have a background of personal sorrow or agony, and Ambrosius Stub, whose poems are mainly religious and moralizing verses, witty epigrams, or drinking songs. A satirist, Christian Falster, was a conservative counterpart to Holberg; Friedrich Eilschov and Jens Schelderup Sneedorff, the latter of whom edited *Den patriotiske Tilskuer* ("The Patriotic Spectator"), a Danish *Spectator*, were both rationalist disciples of Holberg.

A significant revival of Danish literature took place toward the end of the century. The Norwegian Johan Herman Wessel, one of the greatest humorists to use the Danish language, wrote *Kierlighed uden strømper* ("Love Without Stockings") in 1772, a parody of the Danish imitations of Italian operas and French tragedies that had superseded Holberg's comedies, especially those of Niels Krog Bredal and Johan Nordahl Brun. His light humorous poems and versified narratives are still very popular.

The works of Johannes Ewald

At the same time a revival of emotional poetry was taking place, influenced by German and English literature. Johannes Ewald, perhaps Denmark's greatest lyrical poet, was the first to delve into Scandinavian antiquity and discover the poetic wealth in the *Gesta Danorum* of

Saxo Grammaticus, in the myths, sagas, and ballads. He wrote verse dramas and deeply personal and descriptive poems. *Fiskerne* (1779; "The Fishermen") was the first serious Danish drama in which ordinary people were treated heroically. His memoirs, *Levnet og meninger* ("Life and Opinions"), were influenced by Laurence Sterne and Jean-Jacques Rousseau. Jens Baggesen at first imitated the satires of Holberg and Wessel but gradually developed as a poet of distinction, although the quality of his poetry varies greatly. In *Labyrinten* (1792–93; "The Labyrinth"), he described his travels in Europe in the manner of Sterne.

The end of the 18th century was marked by much literary dilettantism. P.A. Heiberg, a militant radical satirist, influenced by the ideas of the French Revolution, was expelled from Denmark and spent his last 40 years as a political refugee in France. Malte Conrad Bruun, the author of *Aristokraternes catekismus* ("The Aristocrats' Catechism"), had a similar fate.

Norwegian. Several of Denmark's leading writers of the 18th century were of Norwegian birth, pre-eminently Ludvig Holberg, and later in the century the members of Det Norske Selskab (the Norwegian Society). Established in Copenhagen in 1772 by a group of resident Norwegians, it looked to French rather than to German and English literature for models. Among its members were Johan Herman Wessel; Johan Nordahl Brun, whose tragedy *Zarine* won a literary prize in 1772; and the intellectual Claus Fasting. Within Norway itself there was little overt literary activity, though the establishment in 1760 of a Royal Norwegian Society of Learning in Trondheim was evidence that Norway was beginning to assert its cultural aspirations.

Icelandic. *House-postil* ("Sermons for the Home"), by Jón Vídalín, bishop of Skálholt, is the best example of early-18th-century prose. Among important writers of the later 18th century, Eggert Ólafsson carried out a comprehensive geographical field survey (published in Danish, 1772; partial Eng. trans., 1805) of Iceland's country and its people. In his poetry he expressed 18th-century Rationalism combined with Romantic patriotism. Jon Steingrimsson is remembered for his autobiography, in which he described a volcanic eruption of 1783 and the famine that followed. Jón Porlbksson, poet and scholar, translated Milton's *Paradise Lost* and Pope's *Essay on Man*.

Finnur Jónsson, bishop of Skalholt, wrote *Historia Ecclesiastica Islandiae* (1772–78), which covers the history of Christianity in Iceland, and Jón Espólín wrote *islands árbaekur* (1822–55; "Annals of Iceland"), a history of Iceland from 1262.

RUSSIAN LITERATURE IN THE 18TH CENTURY

Modern Russian literature dates from the first decade of the 18th century, when literary works, as distinct from religious and official books, began to be printed. Classicism was the dominant movement in 18th-century literature, but in the second half of the century it was displaced by the sentimental movement, which itself was yielding to Romanticism and Realism by 1800.

The rise of Russian classicism. The most prominent literary figure at the beginning of the century was Feofan Prokopovich, archbishop of Novgorod, one of Peter I the Great's close associates. He was the progenitor of two of the main themes in Russian literature: autocracy as a form of government and satire as a means of attack on his political adversaries.

In the early years of the century, there was some solemn panegyric verse celebrating Russian victories as well as examples of the new genre of love lyrics and elegiac poems. This period also saw the development of Russian secular drama, but, on the whole, merely prepared for later development.

The growth of Russian cultural institutions

The growth of cultural institutions—the creation of the first Academy of Sciences, in St. Petersburg (1755), the first university, in Moscow (1755), and the first theatre, in St. Petersburg (1756)—fostered the development of literary talent. The leading literary genres were the ode and tragedy, both propagating enlightened absolutism. As in western Europe, the cult of reason lay at the foundation of Neoclassical aesthetics.

The leading writers of the classical school were Prince Antiokh Dmitriyevich Kantemir. the first Russian secular poet; Vasily Kirillovich Trediakovsky, one of the most scholarly men of his time; Mikhail Vasilyevich Lomonosov, poet, grammarian. scientist, and literary critic; and Aleksandr Petrovich Sumarokov, poet and dramatist. Kantemir is known mainly for his love lyrics and his nine satires, and is regarded as the founder of Russian satire. Trediakovsky began by translating Paul Tallemant's *Voyage de l'isle d'amour* ("Voyage to the Isle of Love") the first erotic work to be published in Russia, and he helped reform Russian prosody. Lomonosov created the famous "three styles" of poetic diction characteristic of Russian Neoclassicism: the "high" (or grand) style for heroic poems, odes, etc.; the "middle" style for dramatic works demanding colloquial speech; and the "low" style for comedies, epigrams, songs, letters in prose, and precise descriptions. Lomonosov also advanced the theoretical study of the Russian language, which he maintained was no whit inferior to any European language in natural richness, beauty, and strength.

Sumarokov, who wrote the first Russian classical tragedy, *Khorev* (1747), was adept in a variety of genres, and opposed Lomonosov's "florid" style. He owed his popularity to his numerous love lyrics, elegies, eclogues, and idylls. He also wrote six tragedies, including a free adaptation of *Hamlet*, and four comedies. His tragedies, strongly influenced by Racine, were intended as a school of noblemen's moral values, and even in his comedies, he sought to improve the moral standards of the Russian aristocracy. In his periodical *Trudolyubivaya pchela* ("The Industrious Bee") he exposed corruption among officials and attacked landowners for maltreating their serfs but defended social inequality as natural and lawful. His satiric articles and fables laid the foundation of the satire of the next decade.

The first breach in classicism was made by Mikhail Matveyevich Kheraskov, famous for two epic poems modelled on Voltaire's *Henriade–Rossiyada* (written 1771–79, published 1799) on the capture of Kazan by Ivan the Terrible; and *Vladimir Vozrozhdyonny* (1785), on St. Vladimir's introduction of Christianity to Russia.

Political and sentimental themes. Influenced by a Cossack and peasant rebellion of 1773–75, writers of the period chose as their most important theme that of serfdom. The social influence of literature greatly increased, leading to a widening circle of readers and to active participation in literature of all classes. Sumarokov, in his last three tragedies, attacked the idea of a tyrannical ruler and pleaded for an enlightened emperor. In *Tilemakhida* (1766) a free rendering of Fénelon's *Télémaque*, Trediakovsky declared that while the tsar wielded power over the whole people, the laws had power over the tsar too. The empress Catherine II herself edited a journal, modelled on the English *Spectator*, in which she poured scorn on Trediakovsky.

The theme of serfdom

The two most prominent writers of the period were a playwright, Denis Fonvizin, and a writer and publisher, Nikolay Novikov. Beginning by translating fables, Fonvizin later wrote original satirical fables, such as *The Fox and the Preacher* (1762), written shortly after the death of the old empress Elizabeth (Yelizaveta Petrovna), in which he attacked her hypocritical courtiers. But it was with two prose comedies, *Brigadir* and *Nedorosl*, that Fonvizin triumphed as a playwright. Novikov, like Fonvizin, directed his satire not against serfdom itself but against the landowners' misuse of their powers. His first periodical, *Truten* (1769–70; "The Drone") had the significant epigraph: "They labour and you feed upon their labour." Catherine II suspended *Truten* for a month in 1769, after which Novikov moderated its satire. At the end of 1770 it had to suspend publication, and Novikov's second journal, *Pustomelya* ("The Windbag"), was suspended after only two issues. In his third journal, *Zhivopisets* ("The Painter") he published in 1772 two of his most famous satires: "A Fragment of a Journey" and "Letter to Fallaley." The first gives a shattering picture of the poverty and slavery of the peasants, and the second provides insight into the serf owners' mentality. The fact that he never condemned serfdom saved Novikov from Catherine's wrath but robbed his satire of efficacy as a revolutionary instrument. The most important novelist of the period was Fyodor Aleksandrovich Emin, whose best novel, *Letters of Ernest and Doravra* (1776), a free adaptation of Rousseau's *La Nouvelle Héloïse*, is interesting as the first attempt in Russian fiction at psychological analysis of the thoughts and feelings of ordinary people. In it Emin exposed the evils of serfdom but did not attack serfdom as an institution. His son, Nikolay Emin, also a popular novelist, was less interested in the social inequities of his time. His novels show the growing influence of the sentimental school. A popular novelist to emerge from the lower classes, Mikhail Dmitriyevich Chulkov, was concerned with entertaining his readers: his novels lacked the lofty moral admonitions that abound in the works of classical writers.

Comic operas

Between 1770 and 1780 the sentimental movement gained ground, becoming noticeable in adaptations of French comic operas. These were followed by native comic operas, such as *Rozana* i *Lyubim* (1778) by Nikolay Petrovich Nikolev, the main theme of which is the superiority of moral equality over social inequality. The anti-serfdom note can be detected in other comic operas of the period—such as *Neschastye ot Karety* (1779; "An Accident with a Carriage"), by Yakov Borisovich Knyazhnin; *Kofeynitsa* ("The Coffee Fortune Teller"), by Ivan Krylov, the greatest Russian fabulist; and *Melnik, koldun, obmanshchik i svat* (1779; "The Miller, the Wizard, the Quack, and the Matchmaker") by Aleksandr Onisimovich Ablesimov.

The American Revolution, fully reported by Novikov in his *Moskovskie Vedomosti* ("Moscow News"), raised again the problem of an autocratic ruler's duties to his subjects. Fonvizin maintained that the tsar should be as completely subject to the laws of the land as his lowest subjects. In his satires, and, even more, in his popular fables, Ivan Ivanovich Khemnitser touched on the same theme, while Knyazhnin, in his tragedy *Roslav* (1784), demanded that the emperor should be subject to his own laws, a sentiment that led Catherine II to order the destruction of all copies of the work. Even Nikolev declared, in his tragedy *Sorena i Zamir* (1784), that "to destroy a tyrant is a duty not a crime".

The first indication of the declining power of classicism came in the popular narrative poems of Ippolit Fyodorovich Bogdanovich, especially his *Dushenka* (1775), a free adaptation of La Fontaine's *Les Amours de Psiché et de Cupidon*. Bogdanovich was regarded by the sentimentalist writers as their predecessor.

The lyrics and odes of Gavrila Derzhavin

Gavrila Derzhavin, the greatest Russian poet of the 18th century, gave the most vigorous expression to the change from classicism to sentiment. His first works—odes as well as lyrics—showed the influence of Lomonosov and Sumarokov, but he soon took up an independent attitude toward social evils. His ode on the death of Prince Meshchersky (1779) is full of reflections on the transience of life and the inevitability of death but ended on a note of cheerful resignation characteristic of the Latin poet Horace. He reached his greatest heights during the next decade with such odes as "On the Capture of Ismail" and "To the Nobleman." His finest poems—"The Waterfall" (1794) and "The Peacock" and his poems in the tradition of the ancient Greek poet Anacreon (collected 1804) such as "Invitation to Dinner" and "Life in Zvanka" (the poet's small estate)—belong to this period. Derzhavin accepted serfdom as natural and lawful, and his poetry expressed the ideology of the ruling classes; yet his realistic representation of everyday life and his simplification of poetic diction by introducing colloquial speech tended to emphasize the democratization of poetry.

Russian literature of the 1790s, influenced by the French Revolution, raised the themes of the rights of man and the role of the nation as a whole. Aleksandr Radishchev in his "Ode to Liberty" (1781–83) hailed the American Revolution and attacked serfdom, but it was his denunciation of autocracy that made Catherine II describe the ode as "a rebellious poem," in which the tsar was threatened with execution.

Sentimentalism

Apart from violently satiric works, among them the famous comedy *Chicane* (1798) by Vasily Vasilyevich Kapnist. the main literary movement was sentimentalism, led by Nikolay Karamzin. In Russian sentimentalism subjective perception is combined with denial of classical abstractions and idealization of man's natural condition. Karamzin opposed to the classical emphasis on reason a doctrine of poetry as an expression of feeling. The sentimentalists, moreover, put forward the claims of sensibility as a precondition of aesthetic impressions, emphasizing art's emotional foundation. Unlike Derzhavin, Karamzin was interested less in reality than in feelings and moods it arouses in the poet. According to him the greatest poetic creation is "an effusion of a languorous and grieving heart." This elegiac mood, so characteristic of Karamzin's latter poetry, is expressed in his famous novel *Poor Liza* (1792), in which for the first time in Russian fiction descriptive passages, used to emphasize the hero's feelings, are an important component. In his novel *Julia* he idealized country life and withdrawal into "the embraces of nature." His other novels—The *Island* of *Bornholm* and *Sierra Morena* (both 1793)—were precursors of the Russian Romantic novel.

Georgian. Belles lettres revived in Georgia with the lexicographer Sulkhan-Saba Orbeliani (1658–1725) who wrote a book of fables, *The Book of Wisdom and Lies.* The main poets of the period were David Guramishvili and Bissarion Gabashvili.

Latvian. The loss of political independence in the 13th century prevented a natural evolution of literature out of folk poetry, and much of Latvian literature is an attempt to re-establish this connection. Written literature came late, fostered by German clergymen. Latvian secular literature began in the 18th century with G.F. Stender who, in the spirit of the Enlightenment, produced didactic tales or idyllic portrayals of country life, and vainly attempted to supplant the folk songs by ditties of his own—thus, in his own way, verifying the fact that the great wealth of folk songs (some 400,000 published, and about a million recorded but unpublished) has been in all ages a pervasive presence in Latvian literature. Already in the 17th century, Cr. Fuereccerus, a sensitive poet who introduced new metrical conventions and rhymes, at times also made use of stylistic elements from Latvian folk songs, and G. Mancelius, founder of Latvian prose, battles against folklore more in a spirit of affection than hostility.

OTHER LITERATURES OF EASTERN EUROPE IN THE 18TH CENTURY

The literary developments of the Polish Enlightenment

Polish. Close contact with western Europe, especially France and England, characterized literature of the Enlightenment period in Poland, whose writers were imbued with a desire to save the national culture from effects of partition and foreign rule. Literary developments included the rise of drama; introduction of the periodical and the novel; publication of the first Polish dictionary; and, in poetry, the introduction of *dumy* (ballads).

Drama was established late in Poland. The earliest significant event was the inauguration of a national theatre in Warsaw in 1765. There were three principal dramatists: Franciszek Bohomolec, who satirized the aristocracy in adaptations of Molikre; Wojciech Bogusławski, who wrote a popular national comic opera *Krakowiacy* i *Górale* (1794; "Cracovians and Mountaineers"); and Franciszek Zabłocki, important for *Fircyk w Zalotach* (1781; "The Dandy's Courtship") and *Sarmatyzm* (1785). Aleksander Fredro's comedies appeared when the Romantic movement was underway, and in them the influences of Molikre and Carlo Goldoni were assimilated, as *Zemsta* (1834; "The Revenge") illustrated. They were remarkable for brilliant "type" characterization, ingenious construction, and metrical facility.

Didacticism permeated most of the period's prose writing. Modern periodicals appeared at this time (*e.g.*, *Monitor,* 1765–85), and a Polish dictionary was published between 1807 and 1814. Bishop Adam Naruszewicz wrote early in the Enlightenment. His poetic works, considered chronologically, reflected the transition from Baroque to Classical, and he also wrote a history of Poland in which modern methods of scholarship were used. The most important poet, Bishop Ignacy Krasicki, of European outlook and skeptical intellect, wrote two mock-heroic poems, *Myszeis* (1775) and *Monachomachia* (1778), as well as *Satyry* (1779) and *Bajki* i *przypowieści* (1779; "Fables and Moral Tales"), and was remarkable for concise expression, formal elegance, and wit. Krasicki also introduced the novel *Mikołaja Doświadczyńskiego przypadki* (1776; "The Adventures of Mikołaj Doświadczyński"), showing the influence of Swift and Rousseau. Two other outstanding poets were Stanisław Trembecki, whose works had an important place in Polish classicism as models of stylistic fluency, and Tomasz Kajetan Węgierski, a freethinker and admirer of Voltaire, who expressed a rationalistic outlook in a mock-heroic poem *Organy* (1784).

Lyrical poetry continued to develop. Franciszek Karpiński developed features of the Baroque style in popular pastorals and religious songs, while Franciszek Dyonizy Kniaźnin, a conscious experimenter, rewrote much of his earlier work to achieve stylistic uniformity; his verse, cool and intellectual in quality, anticipated Romantic themes of folk poetry, popular superstition, and gypsy life.

Julian Ursyn Niemcewicz's writings were inspired by patriotism and concern for reform. He knew English literature thoroughly and made early translations of English Romantic ballads, his original *dumy* being the first in Poland; he introduced the historical novel to Poland with *Jan z Tęczyna* (1825), which showed Sir Walter Scott's influence. His comedy *Powrdt posła* (1790; "The Envoys' Return") was one of the best dramatic works of the period, and *Śpiewy historyczne* (1816; "Historical Songs") was widely read.

After loss of national independence, with the third partition of the country between Russia, Austria, and Prussia in 1795–96, the tradition of patriotic poetry was continued by émigré soldier-poets in the Polish legions of Napoleon's army, among them Józef Wybicki, whose popular patriotic song "Mazurek Dąbrowskiego" (1797) was adopted as the national anthem in 1918.

Hungarian. The period between 1711 and about 1770 was a time of decadence in Hungarian literature, producing notable works only in the fields of memoirs, history, and history of literature. Among historians, Mátyás Bél, Gyorgy Pray, and Istvbn Katona were most important.

The period between about 1772 and 1825, the Enlightenment, though immensely important in the development of the Hungarian spirit, produced few writers of the first rank. The Hungarian Enlightenment was more receptive to French and English ideas than it was productive of original developments.

With the publication in 1772 of the first literary work by Gyorgy Bessenyei, a translation (from the French) of Alexander Pope's *Essay on Man,* the new era began. All his works served a didactic purpose. His drama *Agis tragédiája* (1772; "The Tragedy of Agis") served as a somewhat creaking vehicle for his antidespotic ideas. His best work, *Tarimenes' Journey* (1802–04), the first real novel in Hungarian, was a bitter attack on everything that conflicts with the views of Enlightenment. With destructive irony, Bessenyei examines the shortcomings of human society. His personal influence induced several of his fellow officers—for example, Sándor Báróczi and Ábrahám Barcsay—to try to convey the ideas of the Enlightenment in Hungarian to a Hungarian public.

Spurred on by new ideas, but basically traditionalists, József Gvadanyi and András Dugonics achieved works of some literary merit and amusing enough to become popular. Gvadányi's best work, *Egy falusi nótáriusnak budai utazdsa* (1790; "The Journey to Buda of a Village Notary"), is an apology for national and traditional values, directed against encroaching foreign ideas. The novel *Etelka* (1788), by Dugonics, a sentimental love story in a historical setting, was the first Hungarian best seller. Both Gvadanyi and Dugonics used a popular language, free from foreign influence, and this was perhaps their greatest merit. Ádám Pálóczi Horváth left a collection of 450 poems, a treasure-house of popular poetry and music.

The end of the 18th century was a period of discovery, and of experiments with language. The pioneers of the use of classical metre in Hungarian verse were followed by Benedek Virág, who imbued with poetic inspiration verse forms which for his predecessors were merely formal exercises. It fell to Dániel Berzsenyi, who published a single volume of poetry, in 1813, to show what use a really great poet could make of classical metre. Berzsenyi is a giant of expression. It is the terseness and vigour of the expression, the power of conviction, that make his lines memorable. His ode "Magyarokhoz" ("To The Hungarians"), his *Fohász* ("Prayer"), and his elegy *A közelitő tél* ("On the Nearing Winter") express the transitoriness of power, friendship, of everything but God.

The ideas of the Enlightenment were not universally welcomed in Hungary. Several writers went to prison for the crime of sympathy with these ideas. The most talented among them, Jbnos Batsbnyi, secured his place in the history of Hungarian literature by his poem *A franciaorszagi változásokra* ("On the Changes in France"), a vigorous warning to all tyrants "to cast their watchful eyes on Paris."

The only work of importance by József Kármán is a sentimental novel, *Fanni hagyományai* ("The Papers of Fanny"), that marked an important step in the history of the Hungarian novel.

The first really important lyric poet since the Renaissance poet Bálint Balassi was Mihály Csokonai Vitéz, who continued the purely Hungarian poetical tradition. His many songs to "Lilla" are a happy blend of rococo grace and thoughts less superficial than they appear. The influence of Rousseau is very noticeable in some of his longer philosophical poems. Alexander Pope's *The Rape of the Lock* served as a source of inspiration for Csokonai's comic epic *Dorottya* (1804; "Dorothy") but Csokonai's poem is original and very much Hungarian. The language of the poem is vigorous, even vulgar, and the comic situations are coarse but amusing.

The place of Sándor Kisfaludy in Hungarian literature is secured by his first work, *Kesergő szerelem* ("Plaintive Love"), a song cycle depending on a very thin narrative thread. Writing in a very elaborate verse form of 12 lines, called the Himfy verse, which he devised himself, Kisfaludy displayed great ingenuity in finding new variations on the theme of unhappy love. Among his better works must also be reckoned three tales published together in *1807—Csobdnc, Tdtika,* and *Somló*—written in the mood of chivalric romances.

The contribution of Ferenc Kazinczy

Ferenc Kazinczy, a mediocre poet but an influential man of letters, was the pivot of literary life for about 40 years. His political conceptions were progressive, a luxury for which he paid with six years' imprisonment. He wanted a literature refined and limpid, neither baroque nor popular, and his interest was focussed on style. He became the head of the *neologi,* or linguistic innovators, who, opposed by the *orthologi,* or purists, tried to renew and enrich the Hungarian language to make it suitable to express the most elaborate concepts.

Romanian. Most of the 18th century presents a picture of social oppression and decadence. A rich secular and apocryphal literature circulated in manuscript (*Erotocritul, Fiziologul, Istoria lui Archir, Viaţa lui Esop, Halirnanaua);* but there was no progress in comparison with that of the past. In Moldavia a new cultural centre arose at Rădăuţi. The achievements of the century were the *Minei* ("Lives of the Saints") of 1776–80 and 1807–15 (each in 12 volumes published in Rîmnicu Vilcea and in the monastery of Neamf, respectively), whose rich and lucid language put them alongside the Bible of 1688.

Lyric poetry was cultivated toward the end of the century in love songs (1769–99), in the tradition of the ancient Greek poet Anacreon, by Alecu Văcărescu. His father Ienăchiţă, a moralist poet, also wrote the first grammar of Romanian; his son Iancu, the father of Romanian poetry, overshadowed his predecessors by his poems. The fourth Văcărescu poet was Nicolae. The lyric tradition was carried on in Walachia by B.P. Mumuleanu.

Slovene. In about 1750 there was a revival of literature in Slovene, initiated by the monk Marko Pohlin, an opponent of Germanization, who wrote educational, recreational, and religious works. On his initiative the first poetic almanac, *Pisanice* (1779–81; "Exercises in Writing"), began to be published. Anton Linhart compiled a critical history and adapted two comedies, one based on Beaumarchais's *Mariage de Figaro*. Valentin Vodnik, a rationalist didactic poet, was editor of the first periodical in Slovene *Lublanske novice* (1797–1800; "The Ljubljana News"), as well as a grammarian, lexicographer, and Francophile, since in Napoleon's province of Illyria (1809–14) the Slovene language was recognized. Literature had become secularized and was to be an important vehicle for awakening national consciousness.

CELTIC LITERATURES IN THE 18TH CENTURY

Owen's poetic reforms

Welsh. The middle of the 18th century was, after the 14th, the most fruitful period of Welsh literature. After an intervening period of stagnation, Welsh writing revived. Goronwy Owen, inspired by English Augustanism, reintroduced and improved the strict metres of the *cywydd* (verse in couplets of seven syllables) and *awdl* (ode), old verse forms which made use of *cynghanedd* ("harmony"), a complicated system of alliteration and internal rhyme. He also introduced a new and wide range of subject content, and thus founded a new classical school of Welsh poetry. The more important poets of this school were William Wynn of Llangynhafal, Edward Richard, and Evan Evans. Much of their activity was associated with the Welsh community in London and the Cymmrodorion society and led to the establishment of local eisteddfods in Wales, which perpetuated the classical forms of Welsh poetry.

Chief among Owen's successors was David Thomas (Dafydd Ddu Eryri), who, however, like other eisteddfodic bards of this period, soon departed from classical strictness.

The classicists of the 18th century stood aloof from the Methodist revival, but the intense religious fervour brought a new articulateness and greatly inspired poets in free metre, especially hymn writers like William Williams of Pant-y-celyn, almost the first poet to use the free metres for a serious purpose; the mystical Ann Griffiths, and many others. The hymn is important by virtue of its influence on later secular poetry.

For a long time after 1750, Welsh prose was narrow in range. The results of the French Revolution saw much activity in political thought. The most important political writer was John Jones (Jac Glan-y-gors). After the establishment of the periodical press, political writing became an important part of Welsh literature, extending into the 19th century.

Breton. Modern Breton is said to begin in 1659, when Julien Maunoir introduced a more phonetic orthography, but works of the Middle Breton type appeared until the 19th century. The bulk of Breton literature in this period consisted of mystery and miracle plays treating subjects from the Old and New Testaments, saints' lives, and stories of chivalry derived from French or Latin. Even plays depicting Breton saints evinced little originality. In the 18th century, many Breton dictionaries were published but little of literary significance was produced. One name survives: Claude-Marie Le Lae, who wrote satirical poems—Ar *C'hy* ("The Dog") and the sermon of *Mikel Morin*.

Scottish Gaelic. Practically no secular poetry in Gaelic was printed before 1751, and the bulk of earlier verse was recovered from oral tradition after that date. Much of the inspiration of Gaelic printing in the 18th century can be traced to Alexander Macdonald (Alasdair Mac Mhaighstir Alasdair), who published a Gaelic vocabulary in 1741 and the first Scottish Gaelic book of secular poetry, *Aish-eiridh na Sean Chánoin Albannaich* ("Resurrection of the Ancient Scottish Tongue"), in 1751. He rallied his fellow Highlanders to Prince Charles Edward's cause in the '45 rising with *Brosnachudh nam Fineachan Gaidhealach* ("Incitement to the Highland Clans"), and *Hi ri rì tha e tighinn*, a song of welcome to the Prince. His masterpiece, "The Birlinn of Clanranald," is an extravaganza that is ostensibly a description of a voyage from South Uist in the Hebrides Isles to Carrickfergus in County Antrim, Northern Ireland. He also composed nature poems, love poems, drinking songs, and bitter satires.

Duncan Macintyre (Donnchadh Bàn), who was influenced in various ways by Macdonald, published his poems in 1768, but the two men were very different in temperament and in training. Macdonald had received a formal education; Macintyre could neither read nor write. Macdonald was ambitious, independent, and restless; Macintyre was not. Macintyre fought on the Hanoverian side at the Battle of Falkirk and later praised George III in *Oran do'n Rìgh*, but he had been a forester on the Perthshire–Argyllshire borders in early manhood, and this is the setting of his greatest poems, *Moladh Beinn Dóbhrain* (The *Praise of Ben Dorain*) and *Oran Coire a' Cheathaich* ("Song to Misty Corrie"), remarkable for their emotional closeness to nature, their minute and objective detail, and their personal and lyrical quality. His most famous love song is addressed to his wife, Màiri.

Other poets of note in the 18th century include John MacCodrum, author of an elegy for Sir James Macdonald of Sleat and of much humorous and satirical poetry; also Robert (called Rob Donn) Mackay, whose gifts of observation and satire in commenting on the life of his district produce much social satire in his verse and a wealth of shrewd and humorous understanding of human nature; and William Ross, the Romantic poet of the group, several of whose best poems, such as *Feasgar Luairt* and *Oran Eile*, were occasioned by an unhappy love affair.

The greatest composer of Gaelic religious verse in the 18th century was Dugald Buchanan, who assisted Rev. James Stewart of Killin in preparing his Gaelic translation of the New Testament (1767). His *The Day of Judgement* and *The Skull* are impressive and sombre and show considerable imaginative power.

Irish Gaelic. The 18th century is a low point in Irish Gaelic literature. The last great flowering of the poetic tradition in Munster was *Cúirt an Mheánoíche* (The *Midnight Court)* by Brian Merriman, a Clare schoolmaster. After it, Irish poetry became a matter of folk songs.

During the 18th and early 19th centuries the only books in Irish prose were catechisms and devotional tracts. The manuscript tradition was carried on by a few scribes into the first half of the 19th century, when it all but died out. By the middle of the 19th century there was no literary activity in Ireland and all but a handful of the speakers were illiterate.

Manx. There is little evidence for any former existence of a rich oral literature in the Isle of Man. The first written text in the language is the translation of *The Book of Common Prayer* by a Welshman, John Phillips, bishop of Sodor and Man, which survives in a unique manuscript of about 1630.

The traditionary (or Manannan) ballad, purporting to give a digest of the history of the island, is known in manuscripts only from about 1770, in print certainly only from 1802, but was probably composed early in the 16th century. *Bause Illiam Dhone*, a ballad on the execution of William Christian in 1663, written a generation later since it refers to the fate of his accusers' families, is first known from its revival, with additional verses, in 1781. The heroic tradition survives in the *Fin as Osshin* fragment, a ballad on an Ossianic theme, recovered from oral tradition. Most Manx verse is in the form of carvals, or carols, religious songs, not all connected with Christmas (the fall and redemption of man being a favourite theme) but composed to be sung at the *Oie'l Verree*, the Christmas Eve service.

VI. The 19th century

CHARACTERISTICS OF THE AGE

The 19th century in Western literature—one of the most vital and interesting periods of all—has special interest as the formative era from which many modern literary conditions and tendencies derived. Influences that had their origins or were in development in this period—Romanticism, Symbolism, Realism—are reflected in the current of modern literature, and many social and economic characteristics of the 20th century were determined in the 19th. This is seen particularly in the economic and social influences of the increase in industrialization, beginning in England in the mid-18th century and generally termed the Industrial Revolution; but it is also discernible in many other developments of the age, particularly in the growth of what might be termed a "social conscience" toward the middle and later parts of the century.

Romanticism

The predominant literary movement of the early part of the 19th century was Romanticism, which in literature had its origins in the *Sturm und Drang* (Storm and Stress) period in Germany (see above *German literature of the 18th century).* An awareness of this first phase of Romanticism is an important correction to the usual idea of Romantic literature as something that began in English poetry with William Wordsworth and Samuel Taylor Coleridge and the publication of *Lyrical Ballads* in 1798. Moreover, although it is true that the French Revolution of 1789 and the Industrial Revolution already spoken of were two main political and social factors affecting the Romantic poets of early-19th-century England, many characteristics of Romanticism in literature sprang from literary or philosophical sources. A philosophical background was provided in the 18th century chiefly by Jean-Jacques Rousseau, whose emphasis on the individual and the power of inspiration influenced Wordsworth and also such first-phase Romantic writers in Germany as Friedrich Holderlin and Ludwig Tieck and in France Bernardin de Saint-Pierre, whose *Paul et Virginie* (1787) anticipated some of the sentimental excesses of 19th-century Romantic literature. Positive as it was, the influence of Rousseau must also be seen as a partly negative reaction against 18th-century Rationalism with its emphasis on intellect. A similar influence existed in the philosophy of Emanuel Swedenborg, who allied himself with those who sought a hidden key to life's secrets in intuitions and in sight-giving experience and thus continued a thread in German thought dating from Spinoza and emphasizing the individual and self-knowledge.

Belief in self-knowledge was, indeed, a principal article of Romantic faith. Late-18th-century French writers such as Fabre d'Olivet sought to explain the physical world by an idea of a "breath of life" similar to the "inspiration" of Wordsworth and Coleridge. The Romantics believed that the real truth of things could be explained only through examination of their own emotions in the context of nature and the primitive. Because of this emphasis on inspiration, the poet came to assume a central role—that of seer and visionary. Simultaneously, such formal conventions as imitation of the classics were rejected as binding rules. A new directness of the poet's role emphasized the language of the heart and of ordinary men, and Wordsworth even tried to invent a new simplified diction. Poetry became divorced from its 18th-century social context, and a poet was answerable only to ultimate truth and himself. Two classic poses of the Romantic poet were the mystic visionary of John Keats and the superman of Lord Byron (indeed, satirization of the Byronic hero was to become a theme of later novelists such as Fyodor Dostoyevsky, even though he himself had Romantic antecedents).

The fact that Dostoyevsky was a Russian showed how the Romantic stream flowed across Europe. In Spain and Italy, Hungary, Poland, and the Balkans, it took the form of drama, which in England failed to produce great works. The early and middle 19th century was a time of poetry and prose rather than of drama. The Romantic style in poetry was seen everywhere in Europe—in José de Espronceda in Spain; Ugo Foscolo and Giacomo Leopardi in Italy, where it became identified with nationalist sentiments; Aleksandr Pushkin and Mikhail Lermontov in Russia; Adam Mickiewicz in Poland. In America, a Romantic thread also allied with the emergence of national feeling could be seen in the adventurous stories of James Fenimore Cooper; in the supernatural and mystic element in Edgar Allan Poe; in the poetry of Walt Whitman and Henry Wadsworth Longfellow; and in the Transcendentalist theories of Ralph Waldo Emerson and Henry David Thoreau, which, as Wordsworth's pronouncements had done, affirmed the power of "insight" to transcend ordinary logic and experience.

The impetus of Romantic poetry began to slacken after about 1830 and gave way to more objective styles. although, as J.M. Cohen says, " . . . the superfluous man, the artist misunderstood, the unhappy lover, the Hamlet figure paralyzed by his own sensitiveness, continued to be the forms in which the artist saw himself." *(A History of Western Literature,* 1956)

Arguably the first post-Romantic poet was a German, Heinrich Heine, but German poetry in the mid-19th century mostly followed Wordsworth, though new tendencies were to be found in August von Platen Hallermünde and an Austrian, Nikolaus Lenau. The principal development was to be seen in France in the growth of a movement known as Parnassianism. Originating with Théophile Gautier, Parnassianism in some ways was an offshoot of Romanticism rather than a reaction against it. In concentrating on the purely formal elements of poetry, on aesthetics, and on "art for art's sake," it changed the direction of French poetry and had much influence abroad. Its most illustrious representative was Charles Baudelaire, who believed that "everything that is not art is ugly and useless." Another branch of new development was the growth of Impressionism and Symbolism, a result of "borrowing" from movements in painting, sculpture, and music. Paul Verlaine, foremost of the Impressionists, used suggestion, atmosphere, and fleeting rhythms to achieve his effects. Symbolism, a selective use of words and images to evoke tenuous moods and meanings, is conveyed in the work of Stéphane Mallarmé and Arthur Rimbaud. The advance of French poetry in the middle and latter part of the century was an achievement of individuals, based on invention of a personal idiom.

Impressionism and Symbolism

The spread of education and, in England, of circulating libraries increased a demand for novels. At the beginning of the 19th century Jane Austen had already satirized the excesses of the "Gothic" novel, a harbinger of medievalizing Romanticism in the latter part of the 18th century, in *Northanger Abbey* and the conflict of sense and Romantic sensibility in *Sense and Sensibility.* In France the conflict of intelligence and emotion appeared in the work of Benjamin Constant (*Adolphe*, 1816) and most notably in *Le Rouge et le noir* (1830) of Stendhal and later in Gustave Flaubert's *Madame Bovary* (1857). The detailed verbal scrupulousness and Realism exhibited in the work of Flaubert and of Honor6 de Balzac was carried forward by Guy de Maupassant in France and Giovanni Verga in Italy; they culminated in the extreme Naturalism of Émile Zola, who described his prose in novels such as *Thérèse Raquin* (1867) as "literary surgical autopsy."

But Realism and nationalism seem irrelevant as descriptions of the great writers of the period—for example, George Eliot, Charles Dickens, and Thomas Hardy in England and Nikolay Gogol, Ivan Turgenev, Leo Tolstoy, Fyodor Dostoyevsky, and Anton Chekhov in Russia. In such writers there was a distinct bias toward literature with a social purpose, stimulated by awakening forces of liberalism, humanism, and Socialism in many Western countries.

A decline of the Romantic theatre into melodrama was fairly general in Europe, and it was slower than the novel to take up problems of contemporary life. When revival came, through the work of a Norwegian, Henrik Ibsen, Romantic conflicts of visionary and realistic, individual and society were restated, and this was true also of the plays of August Strindberg in Sweden. In Russia a modern theatre became a vital influence that could trace its beginnings back to Gogol's *Government Inspector* (1836) but was to be felt later in the century in Turgenev's *Month in the Country* (1850) and, above all, in the work of Anton Chekhov, who was a great dramatist of the period.

ENGLISH LITERATURE IN THE 19TH CENTURY

The Romantic period. *The nature of Romanticism.* The term Romantic movement is somewhat misleading: it was hardly a movement, and the word Romantic is so vague and general as to be almost meaningless. The writers of the period differed widely in the nature of their work and in their critical approach. They did not apply the word Romantic to themselves, and recognition of their achievements by both the public and critics was slow in coming.

The features of English Romanticism

With these reservations in mind, it may be possible to isolate the distinguishing features of English Romanticism. These had obvious correspondences with political and social upheavals of the time: the American and the French revolutions, emphasizing natural rights and the importance of the individual, were reflected in attitudes of writers such as William Blake and were implicit in their statements about poetry. Where the main trend of 18th-century poetics had been to prove the general, to see the poet as a spokesman of society, addressing a cultivated and homogeneous audience and having as his end the conveyance of "Truth," the Romantics found the source of poetry in the particular, unique experience and saw the poet as an individual distinguished from his fellows by the intensity of his perceptions, taking as his basic subject matter the workings of his own mind. The implied attitude to an audience varied accordingly: Keats, for instance, declared, "I never wrote one single line of Poetry with the least Shadow of public thought." Poetry was to be regarded as conveying its own truth, sincerity the criterion by which it was to be judged.

Emphasis on feeling—seen perhaps at its finest in the poems of Burns—was in some ways a continuation of an earlier "cult of sensibility," but it now received special stress and made an appearance in most Romantic definitions of poetry. Wordsworth spoke of poetry as "the spontaneous overflow of powerful feelings." It followed that the best poetry was that in which greatest intensity of feeling was expressed. The degree of intensity was affected by the extent to which a poet's imagination had been at work; as Coleridge saw it, imagination was the supreme poetic quality, a force that made the poet a godlike being. Romantic theory thus differed from the Neoclassical in assigning the imagination a role of supreme importance. Samuel Johnson had seen the components of poetry as "invention, imagination, and judgment," but Blake wrote, "One Power alone makes a Poet: Imagination, the Divine Vision." Judgment, or conscious control, was felt to be secondary; Romantic poets emphasized the workings of the unconscious mind, dreams and reveries, the supernatural, and a childlike or primitive view of the world, valuable because it had not been overlaid by restrictions of civilized "reason." A further sign of the diminished stress placed on judgment is the Romantic attitude to form: if poetry must be spontaneous, sincere, intense, it should be fashioned primarily according to the dictates of the creative imagination. Wordsworth advised a young poet, "You feel strongly, trust to those feelings, and your poem will take its shape and proportions as a tree does from the vital principle that actuates it." This organic view of poetry is opposed to the Classical doctrine of "genres," each with its own linguistic decorum; and it led to the feeling that poetic sublimity was unattainable except in short passages.

Changes in poetic diction

With the new conception of poetry went a demand for a new diction. Wordsworth and his successors, particularly Keats, found the prevailing poetic diction of the later 18th century stale and stilted and totally unsuited to the expression of their perceptions. Wordsworth accordingly sought to bring the language of poetry back to that of common speech. When he published his views in his "Preface" to the second edition of *Lyrical Ballads* in 1800, the time was ripe for a change: the flexible diction of earlier 18th-century poetry had hardened into a conventional language, and, with the notable exception of Blake and Bums, little first-rate poetry had been produced in Britain since the 1740s.

Insistence on the importance of the individual led to an extension of literary sympathies. Most prominent and characteristic was a feeling for external nature, not unconnected with the social changes being brought about by the Industrial Revolution—especially with a shift of population from rural areas into rapidly expanding, ugly cities. In the presence of nature, man considered himself to be closer to God; in such an environment, spontaneous feeling was still possible; communion with and understanding of nature led a poet toward an understanding of his own soul. The supernatural and marvellous were a further area for an exploration of the workings of the human mind and were a prime source of sublimity; and the remote, whether in time or space, formed the subject or the setting of much Romantic writing.

Neither the suddenness nor the novelty of Romanticism should be exaggerated. The critical theories put forward had been anticipated in the 18th century and earlier, and the main difference lay in a shift of emphasis. Moreover, not all writers of this period—Jane Austen, for example, or George Crabbe or Thomas Love Peacock—could be called Romantic; and much of what seems most enduring in the novels of such a Romantic as Scott was non-Romantic. Again, the great Romantics differed among themselves: Byron was eloquent in his praise of 18th-century poetry and vehement in his attacks on Wordsworth and Keats; Shelley found Wordsworth lacking in imagination; and Blake disagreed sharply with many of Wordsworth's critical theories.

Wordsworth and Coleridge and the *Lyrical Ballads*

Poetry. The meeting in 1795 of William Wordsworth and Samuel Taylor Coleridge was the beginning of one of the most fruitful of all literary friendships. The anonymous appearance, in 1798, of their *Lyrical Ballads* marked a new departure in English poetry, and Wordsworth's "Preface" to an enlarged edition was a manifesto of a new outlook. Wordsworth's contribution to the volume was mainly a group of poems "giving the charm of novelty to things of every day," and it was these that aroused most hostility. Certainly, in the earliest versions of these poems, Wordsworth was not always sufficiently sure of himself: his extreme simplicity and matter-of-factness sometimes led him into bathos, and he did not always succeed in raising them above the level of anecdote; but at their best the poems have an extraordinary immediacy and power.

Coleridge said, "Wordsworth added two or three poems written in his own character, in the impassioned, lofty, and sustained diction which is characteristic of his genius"; foremost among these is "Tintern Abbey," a poem in which Wordsworth gave superb expression to his feeling for nature and the mystical "presence" that guided, sustained, and consoled him. He had a remarkable gift for conveying the inner significance of a scene or of an apparently trivial event. As he explains in a long autobiographical poem, *The Prelude,* there are in our existence "spots of time"—moments when, with sudden clarity, it is possible to "see into the heart of things." His ability to present these moments was his greatest strength. Poetry, he wrote, "takes its origin in emotion recollected in tranquillity"; and the quiet meditative gravity with which he is able to recall and re-create his moments of luminous perception perfectly illustrated this statement. Equally notable are the solemn dignified pathos of "Michael" (1800), the moving simplicity of the "Lucy" poems, the excitement of many passages of *The Prelude,* and the eloquence of the "Ode: Intimations of Immortality" (1807). Wordsworth is an uneven poet: when his imagination fails he can be banal or silly, or indulge in empty moral rhetoric and inflated diction; but of the greatness of his achievement there can be no doubt.

A meditative quality also informs some of Coleridge's best work; in "Frost at Midnight" (1798), he relates his state of mind to a particular time and place, moving easily and conversationally through an association of ideas. The power of a less consciously controlled association is seen in his great "magical" poems—"The Rime of the Ancient Mariner" (1798), "Kubla Khan," and "Christabel" (both published 1816). "The Ancient Mariner" is his finest achievement, in its manifestation of great narrative skill, a nightmarish atmosphere, richness of image and symbol, and a complex and all-pervading moral tone. Coleridge was a man of immense potentialities, few of which were realized; yet his achievement remains great and influential. His *Specimens of Table Talk* (1835) give an idea of his gifts and range of interests, as do his lectures and essays, especially those on Shakespeare. and collections of notes. *Biographia Literaria* (1817) is a rambling but stimulating discussion of poetry and criticism, of seminal importance: one of the most intellectually formative writers of his day, he has had a profound influence on later criticism.

Of lesser poets of this generation, Thomas Moore, whose *Irish Melodies* began to appear in 1807, was one of the most popular. Robert Southey was closely associated with Wordsworth and Coleridge and was a prominent member of the Lake poets, who sought inspiration in the Lake District of Cumberland and Westmorland; though Southey's grandiose epic poems scarcely repay the reading, some shorter poems as well as prose works, *The Life of Nelson* (1813), the *History of the Peninsular War* (1823–32), and a classic formulation of the children's tale "The Three Bears," have lived on.

A reminder of how various was the literary achievement of this period is provided by George Crabbe. He has been called the last of the Augustans, and his roots—his sensibility, his values, much of his diction, his heroic couplet verse form—are very firmly in the 18th century. He differs from the earlier Augustans, however, in his subject matter, concentrating on realistic, unsentimental accounts of the life of the poor and the middle class. In his verse tales he displays considerable narrative gifts and power of description, put to atmospheric and psychological use: *The Village* (1783), *The Borough* (1810), *Tales in Verse* (1812), and *Tales of the Hall* (1819) won him popularity, and after a long period of neglect he is widely recognized once more as a major poet.

The later Romantics: Shelley, Keats, and Byron

When *Lyrical Ballads* appeared, Byron was 10 years old, Shelley six, and Keats three; they grew up in the shadow of the Napoleonic Wars, and the overthrow of Napoleon in 1815 was an event as stirring for them as the storming of the Bastille in 1789 had been for the older poets. Shelley in particular was passionately interested in politics, coming early under the spell of the anarchistic views of William Godwin, whose *Political Justice* had appeared in 1793. This revolutionary ardour, a somewhat vague zeal for the liberation of mankind, is the basis of Shelley's critical essay "A Defence of Poetry" (written 1821; published 1840), in which poets are hailed as "the unacknowledged legislators of the world." This fervour burns throughout the early *Queen Mab* (written 1812–13), the long *Revolt of Islam* (1818), and the noble lyrical drama *Prometheus Unbound* (1820). Shelley is the poet as prophet, as the fine "Ode to the West Wind" (written 1819) makes clear; but with his prophetic spirit goes a vagueness that too often results in a blurring of poetic effect. He is an emotional poet, and critics have complained that he frequently seems simply to strike an attitude and let poetic clichés come pouring out. This is true of his weaker work: but Shelley should not be judged as if he were trying to write like Donne or Keats. His material is transcendental and cosmic, and his expression is appropriate to it; though elements of self-pity and immature posturing have caused a decline in his reputation, at his best (as in short lyrics, much of "Adonais" [1821], "Ozymandias," and *The Triumph of Life)* he is a poet of compelling excitement and power.

The reputation of John Keats, supported as it is by his superb letters, has fluctuated less. He is a richly sensuous poet whose early work, such as *Endymion* (1818), produces an overluxuriant, cloying effect. But Keats was able to discipline himself: he must have known that he had not long to live, and he devoted himself to the perfection of his art with feverish intensity. He experimented with many kinds of poem: "Isabella" (1820) is a deliberate attempt to reproduce a medieval atmosphere; his epic fragment *Hyperion* (1820) has graveness and sonority, but Keats found the style too Miltonic—"I wish to give myself up to other sensations," he wrote. Some of these are found in the poems of Keats's *annus mirabilis* —1819—"The Eve of St. Agnes" and the great odes, "To a Nightingale," "On a Grecian Urn," and "To Autumn." These, with the revised *Hyperion,* represent the summit of Keats's achievement, showing what has been called "the disciplining of sensation into symbolic meaning," the complex themes being handled with a concrete richness of detail.

Of the "great five" Romantic poets, Byron presents major differences from the others. In *Don Juan* he wrote, without irony:

> Thou shalt believe in Milton, Dryden, Pope;
> Thou shalt not set up Wordsworth, Coleridge, Southey.

There were, moreover, striking contrasts within his own work: in *Childe Harold* (1812–18) he appears as an archetypal Romantic, and an element of melodrama appears again in *Manfred* (1817) and *Cain* (1821). Very different is the detached, ironic Byron, a master of witty, almost impudent, satire. *Beppo* (1818), *Don Juan* (1819–24), a kind of gay picaresque novel in verse, and *The Vision of Judgement* (1822) represent, with a handful of lyrics, his lasting achievement.

Of minor poets of the second Romantic generation the best is undoubtedly a Northamptonshire peasant, John Clare, who led a wretched life of poverty and eventual madness, marked by a noble dedication to his poetic gifts. He has natural simplicity and lucidity of diction, intent observation, and almost classical poise, and the unassuming dignity of his attitude to life makes him one of the most quietly moving of English poets. Thomas Lovell Beddoes, whose violent imagery and obsession with death and the macabre recall Jacobean dramatists, represents a pole opposite to that of Clare. He had considerable metrical virtuosity, seen at its best in the songs and lyrical passages from his shapeless tragedy *Death's Jest-Book* (begun 1825; published 1850), but his sensationalism becomes tiring. A comic writer, Thomas Hood, once enjoyed great vogue, though he is now little read. Some of his poems of social protest, such as *The Song of the Shirt* (1843), and the graceful *Plea of the Midsummer Fairies* (1827), are by no means negligible.

Jane Austen's and Sir Walter Scott's different treatments of the novel

The novel. At the turn of the century the cult of sensibility, manifested in the novel of feeling and of terror, dominated prose fiction. The Gothic novel lingered on in such writers as Mary Shelley, whose *Frankenstein* (1818) deals, among its scientific horrors, with themes of social injustice and the "noble savage"; and Charles Maturin, whose *Melmoth the Wanderer* (1820) has, with all its absurdity, a striking intensity. But novelists dealt with the supernatural far less adequately than did poets, and their works were ripe for deflation. The spirit of parody formed a starting point for Jane Austen, who perceived the falsity of the novel based on a carefully cultivated emotional response. *Northanger Abbey* (begun *c.* 1798; published 1817, dated 1818) satirizes the Gothic novel with complex irony, and *Sense and Sensibility* (begun *c.* 1795; published 1811) examines the qualities indicated by its title with sympathy and understanding. As her novels grew in depth, so did Jane Austen's moral perceptions; her art is at its finest in *Emma* (1815). She is the spokesman for sanity and intelligence and builds up positive values against which the objects of her attack can be judged. She is by no means a simple novelist: working within chosen limits, she observes and describes closely the subtleties of relationships between people. She looks continually to discover the principles of conduct that animated them, but her analysis is sympathetic as a rule and her vision basically comic. She is a master of dialogue and writes with great economy, hardly ever wasting a word.

The same cannot be said of the sprawling talent of Sir Walter Scott. His earlier literary career was as a successful poet, with such historical romances as *The Lady of the Lake* (1810). Turning to prose with *Waverley* (1814), he wrote more than 20 novels in the next 15 years. His novels are a great achievement, to be judged by their cumulative effect. He is best when dealing with recent Scottish history, as in *Guy Mannering* (1815) or in his masterpiece, *The Heart of Midlothian* (1818). In adopting the manner and speech of "the lower orders," he was conscious of an affinity with his friend Wordsworth.

His great dramatic gifts and power of characterization often found vivid expression in dialogue but were offset by a deficient sense of form.

Thomas Love Peacock, a satirist for whom plot meant little, approached the novel differently. His novels are conversation pieces in which many pretensions of the day are laid bare. *Nightmare Abbey* (1818) humorously exposes the trappings of Romanticism, and, by including as characters, under thin disguises, Coleridge, Byron, and his friend Shelley, Peacock is able to satirize the more extravagant elements of Romantic poetry. Among lesser novelists are Maria Edgeworth, whose novels of Ireland inspired Scott; Susan Ferrier, a Scot with her own vein of racy humour; and another Scot, John Galt, whose *Annals of the Parish* (1821) is a minor classic, as was James Hogg's *Confessions of a Justified Sinner* (1824), a powerful story of Calvinism and the supernatural.

Miscellaneous prose. The Romantic emphasis on individualism is reflected in prose. William Hazlitt is a vigorous and forthright writer and an energetic and subjective critic of wide range, though lacking in sympathy with 18th-century poetry. His most characteristic work is to be seen in his *Lectures on the English Poets* (1818), *Lectures on the English Comic Writers* (1819), and in *The Spirit of the Age* (1825), especially valuable for its portraits of Hazlitt's contemporaries. Charles Lamb is an even more personal essayist, a zealous but undiscriminating critic. In the collection of essays *Elia* (1823) and *The Last Essays of Elia* (1833), he projects a carefully managed portrait of himself—charming, whimsical, witty, sentimental, warmhearted, nostalgic, sociable. Confession of a different kind is provided by Thomas De Quincey, whose addiction to opium resulted in the eloquent *Confessions of an English Opium-Eater* (1821); his *Reminiscences of the English Lake Poets* is still of interest. A critic and essayist of lighter weight is Leigh Hunt, an intimate and loyal friend of Keats. His outspoken journalism, particularly during his editorship of the *Examiner,* was influential.

Anti-Romantic criticism

Walter Savage Landor is a classicist whose detached style is best seen in brief lyrics and in a series of *Imaginary Conversations*, which began to appear in 1824. The anti-Romantic point of view received pungent expression in the quarterlies: the Whig *Edinburgh Review* (begun 1802), edited by Francis Jeffrey, was followed by its Tory rivals, *The Quarterly Review* (1809) and *Blackwood's Edinburgh Magazine* (1817). All three express trenchantly, and often savagely, adverse views of Romantic poets. These journals were not confined to literature, and in their attacks on prejudices and abuses they set a standard of fearless and independent journalism.

Drama. The great Romantic poets essayed tragedies or "lyrical dramas," and one or two of these, such as Coleridge's *Remorse* (first performed 1813) and Byron's *Marino Faliero* (published 1821), made brief appearances on the stage, but most of them were unacted and often virtually unactable; their authors had little dramatic gift or knowledge of stagecraft, and much of their work seems like a pastiche of Jacobean drama. A partial exception was Byron, who was interested in the theatre and had some talent for it, but even his dramas lack vitality. Shelley's *Cenci* (1819; first performed 1886) has some power and passion but is crudely melodramatic and has so many echoes of Shakespeare as to read almost like a burlesque.

As for the "acting theatre," performances at the theatres of Covent Garden and Drury Lane and elsewhere were of melodramas, farces, pantomimes, and extravaganzas, staged before rowdy and sensation-hungry audiences. Writers of more serious plays showed little talent; and, although some of the farces of the day are lively enough, the "regular" comedies are solemn and dreary. The early 19th century in fact failed to produce worthwhile tragedy or comedy. Against this might be set the introspective and often illuminating dramatic criticism of Coleridge, Hazlitt, and Lamb.

The aftermath of Romanticism and the Victorian age. *Tennyson.* The 19th century has usually been divided into two periods called, respectively, Romantic and Victorian, of which the first, beginning in the 18th century, was taken to have ended about 1830–32; and, as a rule of thumb, this was justifiable: Keats had died in 1821, Shelley in 1822, Byron in 1824, and Scott in 1832; Coleridge had done his best work before 1832, and the same was true of Wordsworth, although he lived until 1850. Yet there was no real break; in 1830–32 Alfred, Lord Tennyson published his first volumes of verses, and these years were not a turning point in cultural history like 1660 or 1798 or like 1914, when the 19th century came to its true end. The story was rather one of continuous development—or "evolution," a key word of 19th-century thought. The impress of the first-generation Romantics—Wordsworth, Coleridge, and Scott—was on the work of their Victorian successors; and it was, above all, Wordsworth and Coleridge who fixed the notions about poetry that were accepted for the rest of the century. Wordsworth's description of poetry as "the spontaneous overflow of powerful feelings," his avowal of the purpose of his poetry as being "to teach the young and the gracious of every age to see, to think, and feel"—these were central to later 19th-century poets and their readers. Poetry was to reveal truth, and the poet was revered as prophet and priest "Of the wonder and bloom of the world," in the words of Matthew Arnold. Never before had it been so generally accepted that nature—the outward world—should be a poet's main source of inspiration. Hence the extraordinary success attained by 19th-century writers in rendering nature with fidelity and insight—a success seen to advantage in Tennyson, Matthew Arnold, Gerard Manley Hopkins, and George Meredith but found also in such prose writers as Emily Bronte, John Ruskin, George Eliot, Charles Kingsley, Thomas Hardy, Robert Louis Stevenson, and others. From Tennyson's "Laburnums, dropping-wells of fire" to Hopkins' "Fresh-firecoal chestnut falls," from Ruskin's description of his first sight of the Alps to Hardy's Egdon Heath—19th-century literature surpassed in its command of "natural magic" any preceding century.

The great influence of Wordsworth on the 19th century

Behind this lay the "nature" religion of Wordsworth; nature was holy and morally uplifting, and the more successful a writer was in "disimprisoning the soul of fact," as Thomas Carlyle put it, the more good he was doing to mankind. Moreover, in spite of the Oxford Movement and other revivals of religious activity, the century witnessed an immense falling away from traditional religious certainties, and Wordsworth furnished many doubters with an interim faith. Increasing urbanization created a longing for escape and compensation, and readers expected poets to restore "the freshness of the early world" to them. Thus, much Victorian literature was agnostic at heart, yet nature-loving, high-minded, and morally serious. Tennyson was the representative poet of the age because he expressed most completely its misgivings as well as its underlying assumptions. *In Memoriam* (1850) represented the 19th-century mind, just as Milton's *Paradise Lost,* Pope's *Essay on Man,* and Wordsworth's *Prelude* represented their respective periods. Immortality, progress, the providential scheme, and the beneficence of nature were questioned; yet Tennyson preserved a faith in human dignity and a passionate joy in landscape and the seasons; he did not doubt that "through the ages one increasing purpose runs," and he looked forward to "one far-off divine event." Tennyson was no aesthete or escapist; he kept abreast of scientific thought and flinched from none of the implications of the nebular theory, of geology, of embryology, and of the evolutionary theory (which he had versified several years before Charles Darwin's *Origin of Species* appeared in 1859). A devout agnostic, Henry Sidgwick considered that Tennyson had given beautiful and moving expression to "the indestructible and inalienable minimum of faith which humanity cannot give up because it is necessary to life." Tennyson has emerged in reputation from a temporary eclipse, not merely as a typical Victorian but for the fidelity of his renderings of nature and for the skill with which his descriptions symbolize states of feeling.

Tennyson as the representative poet of his age

The effect of the French and Industrial revolutions. The 19th century, in one respect, was an aftermath of

The emergence of a social conscience in Victorian England

the French Revolution. English poetry of that century took its character from the Romantic age—from a period of upheaval and excitement that, originally sympathetic toward French aspirations, later swung away into conservative reaction. In England, radicals and liberals attempted to salvage from the tidal wave of reaction whatever of permanent value the French Revolution had bequeathed. Enduring social injustice and inequality lay heavily on the Victorian conscience. The French Revolution had originally fought against these, and, though England had been forced in its 20-year struggle with Napoleon to reaffirm its immemorial way of life, yet "the condition-of-England question" asserted itself with ever-increasing urgency. For the 19th century witnessed not only aftermaths of the French and the Romantic revolutions but also of the Industrial Revolution. Thus, indignation at the state of the underprivileged and a sense that something must be done before it was too late were preoccupations of Victorian writers as various as Jeremy Bentham, Carlyle, Ruskin, Dickens, Disraeli, Mrs. Gaskell, Charles Kingsley, Matthew Arnold, William Morris, and others. Naturally, they did not all diagnose the same complaint or prescribe the same cure. Bentham urged reform of the legal code to ensure the triumph of Utilitarianism—the "greatest-happiness principle"; Carlyle urged faith, sincerity, and leadership; Ruskin and Morris, love of nature and art and pride in craftsmanship; Kingsley, Christian Socialism; Arnold, the spread of culture and education of the populace; and novelists depicted or satirized abuses and brought them home unforgettably to readers' imagination.

Considering the changes going on throughout the 19th century, the tenacity with which the old social hierarchy persisted in England and the completeness with which its underlying assumptions and valuations were taken for granted by writers were remarkable. Novelists, poets, and critics presupposed a world resting on the foundation of a working class. In Dickens' *Bleak House* (serialized 1852–53), the old aristocracy represented by Dedlocks might have been on the way out, but the "new rich" Rouncewells were replacing them, and society—however little Dickens might like it—was still ascendant over the multitude of mankind. The Victorian age was uncomfortably aware that England, in the process of becoming the world's workshop, had lost much that had given an ideal character to past times and particularly to the Middle Ages. For the very reason that the age of faith and chivalry had gone forever, Scott's novels were read with avidity.

The cult of the medieval

Indeed, the cult of the medieval was significant. It appeared in Tennyson, in the Rossettis and other Pre-Raphaelites, in Carlyle, Ruskin, and Morris, and in the architecture of A.W.N. Pugin, Sir Charles Barry, Alfred Waterhouse, and Gilbert Scott. Carlyle's *Past and Present* (1843) contrasted the present—disfigured by Materialism, irreligion, Utilitarianism, dilettantism, and the "cash nexus"—with the past, typified by the harmonious life of the monks of St. Edmundsbury. Ruskin proclaimed the Gothic to an age of machinery, while Morris dreamed of a London small and white and clean and of a Nowhere where men might perhaps recapture the lost glamour of living.

Seminal influences on Victorian thought. The predominance of Thomas Carlyle in the first half of the century had wide and deep meaning. His prophetic mission was to recall a doubting and godless generation to faith in the God of the Old Testament and of Calvin. In his *French Revolution* (1837), he dramatized history as a working out of judgments upon sin and corruption; and in *Sartor Resartus* (serialized 1833–34), *Past and Present* (1843), and *On Heroes, Hero-Worship, and the Heroic in History* (1841), he summoned readers to hear "the voice of old Eternities" and to revere great men who had heard that voice more clearly than others. Carlyle's influence dwindled as people tired of his hysterical style and excessive admiration of strong leaders; but he inspired and invigorated a whole generation, and his prose at its best was moving and expressive—the embodiment of a highly individual vision.

Though Carlyle subscribed to no particular creed, he made continual use of religious language, and his message associated him with the believing side in the 19th-century struggle. John Stuart Mill once called Bentham and Coleridge the two "seminal minds" of the century. Bentham was father to philosophical radicals, Utilitarians, and freethinkers—men such as James and John Mill, Herbert Spencer, George Henry Lewes, Leslie Stephen, John Morley, Frederic Harrison—to all who believed that man could be schooled in the direction of orderly progress by education, free discussion, and the enlightened application of Utilitarian standards. Coleridge, in later prose works, *Aids to Reflection* (1825) and *On the Constitution of the Church and State* (1830), showed that old meanings in religion and politics might still be valid though outward forms had changed. The influence of Coleridge's philosophical rehabilitation of Christianity could be seen down the century in such writers as John Sterling, Julius Hare, Thomas and Matthew Arnold, and Frederick Denison Maurice: Carlyle conspicuously carried on his work, and even J.S. Mill freely acknowledged that Coleridge had deepened and enriched his understanding of the forces that hold men together in society.

J.A. Froude, historian and disciple and biographer of Carlyle, placed Thomas Carlyle and John Henry Newman where Mill had placed Bentham and Coleridge, at the head of the two currents of thought that interested him most. Though both have come to seem men of God in their own ways, to Froude they seemed to face opposite directions. Yet both belonged to the Coleridgean stream, and Newman himself counted Coleridge among influences making for spiritual reawakening. But neither the so-called Tractarian movement of High Church revival nor Newman's secession from the Church of England was much reflected in the literature of the time, except in Froude's *Nemesis of Faith* (1849) and "The Oxford Counter-Reformation" (1881), Charlotte M. Yonge's *Heir of Redclyffe* (1853), J.H. Shorthouse's *John Inglesant* (1880), a few passages in Mark Pattison's *Memoirs* (1885), and the opening of Matthew Arnold's *Discourses in America* (1885) on Emerson. But in the *Apologia pro Vita Sua* (1864), Newman produced not merely a masterpiece of prose style but one of the two or three most significant books of the century.

Darwin and the theory of evolution

Another one of the significant books of the century was Charles Darwin's *Origin of Species* (1859). Works of science and of philosophy and history did still often possess literary value, and the *Origin* was perhaps the last major scientific work written to be understood and enjoyed by a "general reader." Darwin did not invent the theory of evolution, but he familiarized his age with the concepts of natural selection and of continuous development. Powerfully aided by his disciple and champion T.H. Huxley, he compelled men to readjust many of their cherished assumptions about God, nature, and themselves. After Darwin there came a gradual ebbing of the old religious certainties and a growing suspicion that life was caught up in workings of impersonal forces moving inscrutably to an unknown end. This did not come about immediately; but, both in life and in literature, the old unthinking exuberance and confidence gave place in the latter half of the century to "doubts, disputes, distractions, fears." If Dickens' *Pickwick Papers* (serialized 1836–37) and Thackeray's *Vanity Fair* (1848) might be taken to represent the first half of the century, George Eliot's *Middlemarch* (1871–72) and Thomas Hardy's *Tess of the D'Urbervilles* (1891) or *Jude the Obscure* (1895) might stand for the second half.

The novel. Victorian literature reached its highest level of achievement in the novel. A good deal of its poetry has now worn rather thin, and some reputations—notably that of Robert Browning, so highly regarded up to the turn of the century as a subtle psychologist, deep thinker, and invigorating lay-preacher—have been somewhat eclipsed. In drama its contribution was inferior; in criticism, with a conspicuous exception in Matthew Arnold, it was uncertain of its direction. But in Charles Dickens, Emily Bronte, George Eliot, Thomas Hardy, and Henry James it produced novelists of the first order, while it brought forth also a large company of others who, though sometimes unsurpassed in their own way, were mainly of a rank below the very highest—Anthony Trollope, Charlotte Bronte, Wilkie Collins, Mark Rutherford, Robert Louis Stevenson, George Moore, and George Gissing, besides others. Thackeray and Dickens were once classed as the two best Victorian novelists. Later, Dickens was decried as melodramatic, sentimental, and chaotic, and Thackeray was preferred as the more civilized and self-conscious artist. But, seen from a modern vantage point, Dickens has once again begun to appear the supreme Victorian novelist. He began in *Pickwick* as the arch-entertainer in the comic-picaresque style of the 18th century; he ended with a group of far more subtle, intricate, and sombre novels on which his reputation came to rest: *Bleak House* (serialized 1852–53), *Little Dorrit* (1855–57), *Great Expectations* (1860–61), and *Our Mutual Friend* (1864–65). He had a unique relationship with his public, sometimes altering plots to suit what he knew it wanted or to avoid breaking its heart. Sophisticated readers have come to know about Dickens' defects: he drew caricatures rather than portraits; many of his characters were stage types who repeat catchphrases ad nauseam; his plots often seemed fantastic and contrived; and his sense of artistic form was crude. In spite of this and in spite of his manner of siding with or against his own characters, he was without an equal among English novelists for sheer vitality, inventiveness, and imaginative force combined with power of observation and a flair for significant detail; he felt his way imaginatively into scenes and situations and, by charging them with feeling, made them expressive emanations of the Dickens vision.

The great Victorian novelists—Charles Dickens, George Eliot, and Thomas Hardy

It has been said that the English novel came of age with George Eliot, and her work certainly had an intellectually adult quality that distinguished her from her predecessors and contemporaries. She was familiar with the advanced thought of her time and was the first translator of David Friedrich Strauss's *Leben Jesu* (1835–36), a work that pained the orthodox by its relentless "demythologizing" of the Gospels. Brought up in the evangelical tradition, she had been shocked out of orthodoxy by an early encounter with emancipated thinking and became an agnostic. In this she was typical of her age, and she was no less typical in retaining a firm hold upon the ethical core of religion and a lifelong sense of mission to mankind. She also retained an intense affection for the well-remembered scenes and people of her Warwickshire childhood and an imaginative sympathy with their beliefs. These qualities, combined with a powerful grasp of psychological and sociological realities, enabled her to produce a series of admirable novels, including *Adam Bede* (1859), *The Mill on the Floss* (1860), and, most notably, *Middlemarch* (1871–72), generally felt to be the greatest English novel of the period.

Both Thomas Hardy and George Eliot were at their best in dealing with rustic life, scenes, and character. But Hardy was essentially a poet; his real subject matter was states of feeling, rendered against a landscape background coloured by them. Hardy's sense of the strange disease of modern life, of the withdrawal of God and the indifference of nature, expressed itself in a vision of man caught up in a web woven by an inscrutable destiny—a vision implicit in all his work but expressly stated in a remarkable epic drama, *The Dynasts* (1903–08). Yet he was keenly alert to every aspect of nature and depicted it with imaginative faithfulness and intensity: there is nothing in English prose to equal his rendering of the night sky or the thunderstorm in *Far from the Madding Crowd* (1874), Egdon Heath in *The Return of the Native* (1878), and winter at Flintcomb Ash or early mornings at Talbothays in *Tess of the D'Urbervilles* (1891). The brooding intensity of his mind, which stamped his individuality on everything he touched, raised Hardy to a high place among novelists of the later 19th century, a position hardly threatened by his far more brilliant, witty, and sophisticated rival, George Meredith, and challenged only by Henry James, whose work fell mainly in the next century (see below *English literature in the 20th century*). The same qualities of intensity and authenticity gave Hardy as poet a rank among the later poets of the century shared perhaps by Gerard Manley Hopkins and not attained by Algernon Charles Swinburne, once considered the leading poet of the time but who, despite his sound and fury, was later felt to signify very little.

Critics. Of eminent Victorian critics Matthew Arnold has worn best. Only Leslie Stephen and Walter Pater approached a similar order of importance, with critics such as Walter Bagehot, R.H. Hutton, W.E. Henley, Robert Louis Stevenson, John Addington Symonds, Sir Edmund Gosse, and others well in the rear.

Leslie Stephen, in his *History of English Thought in the Eighteenth Century* (1876), set a high standard in what was then a new venture: the "history of ideas." Throughout numerous essays collected in *Studies of a Biographer* (1898–1902) and *Hours in a Library* (1874, 1876, 1879) and in many biographies in the *Dictionary of National Biography*, he showed unusual knowledge and skill in relating writers to their historical, social, and intellectual backgrounds and a sure instinct in discerning their moral core. Stephen was also interesting as one of a band—including T.H. Huxley, W.K. Clifford, and Henry Sidgwick—who did much to make agnosticism respectable.

Walter Pater, high priest of aestheticism, was of a quite opposite cast of mind. In him evaluation was subordinate to "appreciation." Whether he wrote of Wordsworth, Charles Lamb, or Sir Thomas Browne, his aim was to divine their uniqueness and then distill and present it "flask'd and fine." Pater did this supremely well; yet criticism that treated all subjects alike became satiating. It was Matthew Arnold who gave criticism a firm basis of principle and a sense of direction, and the best subsequent criticism took its cue from him.

The importance of Matthew Arnold

Arnold combined a poet's sensibility with a keen intellect and a flair for diagnosing the cultural and spiritual condition of the modern world. He lifted English criticism out of a rut of provinciality and brought it into line with contemporary European thought. At a time when Romanticism was in its heyday, Arnold drew attention to defects of much Romantic poetry—lack of sanity, substance, and form—and pointed to classic virtues of sobriety, weightiness, and unity. He believed that, with the advent of a new and undisciplined democracy, literature might have to play a more important part than before in maintaining civilized standards. He therefore tried to provide criteria whereby the best could be distinguished from the inferior or counterfeit.

Arnold interpreted "criticism" widely, extending it to include the whole interest of an informed intelligence. In *Culture and Anarchy* (1869), he proposed liberal education as the best remedy for the smugness and narrowness of the "Philistines," or uncultured middle class. In spite of his intellectual emancipation, Arnold recognized in religion the highest form of culture; he felt that, although men could not do without Christianity, they could not do with it as it was and set himself—in a series of remarkable works (*St. Paul and Protestantism*, 1870; *Literature and Dogma*, 1873; *God and the Bible*, 1875; and *Last Essays on Church and Religion*, 1877)—to find for religion a sure foothold in spiritual experience. Opinions have differed about the permanent value of these books but not about the sincerity of Arnold's effort to restore life and conviction to religion at a time when it was threatened, on the one side, by a too easy infidelity and, on the other, by a too easy acceptance of sacred clichés.

The end of the age. In the last two decades of the 19th century there were unmistakable signs that the original impulses that had launched the century on its course—Romanticism, evangelicalism, humanitarianism, meliorism—were losing their momentum; poets were becoming aesthetes instead of legislators of mankind, and novelists were turning more and more inward for their subject matter. In Oscar Wilde there was the iridescence of decay; in Samuel Butler the overthrow of many Victorian sanctities—the home, parental and filial affection, the church; in Rudyard Kipling the bugle notes of the new imperialism; in George Bernard Shaw the conversion of drama into an instrument of propaganda and discussion. The 19th century lingered precariously on until World War I swept away its Edwardian aftermath.

Anglo-Irish. With the 19th century there came a change in Anglo-Irish literature. Nationalism, liberalism, and revolution were felt in the Anglo-Irish colony, and Romanticism was beginning to affect the arts. Moreover, antiquarianism had begun to uncover the Gaelic past, which had never been wholly forgotten, for even Swift had translated a poem from the Irish. But Charlotte Brooke, daughter of a dramatist and novelist, was the first to make more extensive verse translations. Irish nationality was born, and thenceforth books about Ireland predominated and an enormous number were produced.

Irish Melodies and *Lalla Rookh*

Poetry. Anglo-Irish verse was established by Thomas Moore, a versatile poet and singer of his own songs. In his *Irish Melodies* (1808–34) he attempted to solve the problem of setting English words into Irish airs. Through these often vehement songs the wrongs of Ireland were warbled in English drawing rooms with great advantage to the Irish cause. Apart from his songs, his Oriental epic *Lalla Rookh* (1817) brought him wealth and fame. Moore was a national hero as no other Irish poet has been, even though he considered it unnecessary to write in dialect, or the style of folk poetry, in order to emphasize Irish nationality. George Darley, a contemporary of Moore, wrote in an exquisite style reminiscent of an Elizabethan or 17th-century poet.

In 1842 a patriotic organization named Young Ireland and led by Thomas Osborne Davis, an eloquent and sincere writer, founded a new paper, *The Nation,* to which a large number of writers attached themselves. Another journal was the *Dublin University Magazine,* which lasted about 40 years and was probably the finest magazine ever produced in Ireland. Most of the best Irish writers contributed to it, notably James Clarence Mangan, a prolific and uneven writer of every kind of verse, who, at his best, was one of the greatest poets of Ireland. He was much influenced by the German Romantics and in turn influenced Edgar Allan Poe. Other notable poets were Sir Samuel Ferguson, who wrote on Irish themes in a strikingly original way, inspired to some extent by Gaelic examples, and William Allingham, who wrote a long and interesting narrative in verse, *Laurence Bloomfield in Ireland* (1864). He settled in England and became an associate of Dante Gabriel Rossetti, adopting the Pre-Raphaelite manner, which linked him with William Butler Yeats and a succeeding period of Irish writing.

Prose. Among novelists was Charles Maturin, whose *Melmoth the Wanderer* (1820) had a wide influence in Europe. Even better known was Maria Edgeworth, who produced many novels, always with some edifying tendency. She knew Ireland well and depicted it with insight and humour in *Castle Rackrent* (1800) and in a tale, *The Absentee* (1812). A landlord's daughter, she believed firmly that the solution of Ireland's troubles lay in an improvement of the outlook of its landlords.

Joseph Sheridan Le Fanu was a successful writer of a different kind. Owner and editor of the *Dublin University Magazine* in the 1860s, he wrote many novels and short stories, chiefly dealing with mystery, crime, and the supernatural. He was the Irish master of the ghost story. He was also a poet, and there is a poetic quality about his work that raises it above the mere thriller.

Scottish literature in English. The re-emergence of vernacular prose, mainly within the novel and the short story, was a distinguishing feature of the 19th century in Scottish literature written in English. The writers fell into two groups: those using Scots for dialogue and English for the narrative, including Sir Walter Scott, James Hogg, Susan Ferrier, George MacDonald, Margaret Oliphant, William Black, and Robert Louis Stevenson; and those such as John Galt, David Moir, and (in the 20th century) Lewis Grassic Gibbon who employed a form of Scots or Scotticized English throughout. The advantage of Galt's method, in *Annals of the Parish* (1821), of casting the book as reminiscences of a Scots-speaking person, was that it secured a homogeneous tone, and the exact Scots idiom was admirable. Scott, on the other hand, had a wider range that came from contrast between English and Scots (as in *Redgauntlet,* 1824), though his English writing and the restriction of his brilliant Scots dialogue to minor characters sometimes made the result patchy and unsatisfactory.

The Kailyard school

The so-called Kailyard school of fiction, characterized by sentimental descriptions of Scottish life and the use of Scots dialect, appeared toward the end of the century and included Ian Maclaren (John Watson), S.R. Crockett, and Sir James Barrie. The sentimental viewpoint of Maclaren's *Beside the Bonnie Brier Bush* (1894) and Barrie's *Window in Thrums* (1889), the hypocrisies and idealization of village life, were pitilessly exposed in George Douglas Brown's *House with the Green Shutters* (1901). Brown's brutal tale was no more than a temporary blow against the Kailyarders.

Though much was written, the poetry of the period was disappointing. Two poems of the supernatural kept up an ancient Scottish tradition: James Hogg's "Kilmeny" (1813) and William Bell Scott's "Witch's Ballad" (1875); there were good vernacular lyrics by Scott, Allan Cunningham, and George MacDonald, but the century after Burns mostly saw a proliferation of the maudlin and jocose in well-worn verse forms. Widely read anthologies of verse called *Whistle-Binkie* (1832–53) were the poetic equivalent of Kailyard, and their popularity was a comment on the poetic taste of the age. More interesting poetry was found in James Thomson's "City of Dreadful Night" (1874), Alexander Smith's *Glasgow* (1857), and John Davidson's *Testaments* (1901–08), which were written in English and, unlike Kailyard and *Whistle-Binkie,* shpwed the impact of the contemporary world and its ideas.

CELTIC LITERATURE IN THE 19TH CENTURY

Welsh. *Poetry.* Strict metre poetry declined in the 19th century. Although the volume produced was enormous, the quality was poor. Eben Fardd was probably the last of the eisteddfodic poets to make any real contribution to Welsh literature. The influence of the hymn writers of the 18th century was seen in the development of the lyric. In fact, all the poetry of the 19th century betrays its religious origin. The influence of contemporary English songs was also seen, as in the work of John Blackwell (Alun), the father of the modern secular lyric. More originality was shown by Ieuan Glan Geirionydd (Evan Evans), who founded the Eryri school of poetry, inspired by the natural scenery of Snowdonia. The earlier lyricists were followed by a more bohemian group—Talhaiarn (John Jones), Mynyddog (Richard Davies), and Ceiriog (John Hughes), the latter the greatest lyrical writer of the century. Only one poet, Islwyn (William Thomas), made a success of the long poem: his Y *Storm* is a series of meditations on life and art.

Prose. This was the most prolific period of Welsh prose, though much of it was of poor quality, partly because it was produced by a people who had had little formal education in their own language and who had lost touch with their own literary past. Much of it was marred also by the pretentious style of the followers of William Owen Pughe, who tried to "restore" literary standards. A tremendous volume of prose was produced—periodicals, religious books and tracts, biographies, sermons, letters, and monumental works such as Y *Gwyddoniadur* ("The Encyclopaedia") and *Hanes y Brytaniaid a'r Cymry* ("History of the Britons and the Welsh"). Political writings became an important part of Welsh literature, the two great political writers of the century being Samuel Roberts and Gwilym Hiraethog (William Rees). Lewis Edwards, founder and editor of Y *Traethodydd,* tried to introduce a wider, European standard of literary criticism. There were some interesting attempts at creative writing, but the only great novelist was Daniel Owen, whose work portrays the extraordinary influence of religion on society during this period in Wales.

Breton. *The revival of Breton literature.* Interest in Breton revived at a time when France's central government was trying to impose French on Brittany and de-

Barzas-Breiz

stroy the regional language and was particularly stimulated with the publication of the celebrated ***Barzas-Breiz*** (1839; "Breton Bardic Poems"). This was a collection made in the villages by Theodore Hersart de La Villemarqué, who declared that they had survived unchanged as part of Breton folklore, though Breton-speaking scholars doubted the collection's authenticity. Attacks on the book reached their height about 1870, when R.-F. Le Men, in a reprinting of ***Catholicon,*** and François-Marie Luzel, in a paper delivered in 1872, showed that *Barzas-Breiz* was divided into three classes: (1) old poems rearranged by the editor or others, chiefly love songs and ballads; (2) modern poems made to look medieval; and (3) spurious poems on such personages as the wizard Merlin. But ***Barzas-Breiz*** was nonetheless important: the historical poems (which exalted the Breton's traditional struggle against oppression) had an especially strong influence. ***Barzas-Breiz*** led to a reawakening of Breton writers and stimulated Luzel himself to collect authentic folk songs and publish ***Gwersiou*** *Breiz-Izel* (2 vol., 1868–74; "Ballads of Lower Brittany") and, in collaboration with Anatole Le Braz, ***Soniou*** *Breiz-Izel* (2 vol., 1890; "Folksongs of Lower Brittany").

Prose. Luzel also collected folktales and legends, publishing many in Breton as well as in French translation. His collaborator, Le Braz, published stories concerning ***an*** *ankou* ("death"), as ***La Legende de la mort*** (1893; ***Dealings with the Dead,*** 1898). Traditional and literary elements combined indistinguishably in many stories. When Breton writers did not depend on folk legends for material, they fictionalized their own life stories. The many improving religious works published were not at all original; yet many Bretons who have read only one book in their own language have read ***Buez*** *ar Zent* ("Lives of the Saints").

Most playwrights were concerned to teach moral and religious lessons, such as Toussaint Le Garrec and Abbé J. Le Bayon, who revived several great mystery plays—*Nicolazig,* ***Boeh er goed*** ("The Voice of the Blood"), ***Ar*** *hent* ***en*** *Hadour* ("In the Steps of the Sower"), and ***Ar en*** *hent* ***de Vethleem*** ("On the Way to Bethlehem").

Poetry. For 200 years Bretons expressed their feelings in poems published as pamphlets--either as ***soniou*** (love songs, satires, carols, marriage lays) or ***gwersiou*** (ballads or broadsides describing recent events in Brittany and elsewhere). Their authors were people of every social class, very few of whom read ***Barzas-Breiz,*** and the poems were hawked around from fair to fair. Dozens of poets published collections after the appearance of *Barzas-Breiz;* the outstanding figure was undoubtedly Jean-Pierre Calloc'h, killed in action in 1917, whose poems were published as ***Ar en*** *Deulin* (1921; "Kneeling").

AMERICAN LITERATURE IN THE 19TH CENTURY

Early 19th century. After the American Revolution, and increasingly after the War of 1812, American writers were exhorted to produce a literature that was truly native. As if in response, four authors of very respectable stature appeared. William Cullen Bryant, Washington Irving, James Fenimore Cooper, and Edgar Allan Poe initiated a great half century of literary development.

Bryant, a New Englander by birth, attracted attention in his 23rd year when the first version of his poem "Thanatopsis" (1817) appeared. This, as well as some later poems, was written under the influence of English 18th-century poets. Still later, however, under the influence of Wordsworth and other Romantics, he wrote nature lyrics that vividly represented the New England scene. Turning to journalism, he had a long career as a fighting liberal editor of ***The Evening Post.*** He himself was overshadowed, in renown at least, by a native-born New Yorker, Washington Irving.

Irving and Cooper

Irving, youngest member of a prosperous merchant family, joined with ebullient young men of the town in producing the ***Salmagundi*** papers (1807–08), which took off the foibles of Manhattan's citizenry. This was followed by A ***History of New York*** (1809), by "Diedrich Knickerbocker," a burlesque history that mocked pedantic scholarship and sniped at the old Dutch families. Irving's models in these works were obviously Neoclassical English satirists, from whom he had learned to write in a polished, bright style. Later, having met Sir Walter Scott and having become acquainted with imaginative German literature, he introduced a new Romantic note in ***The Sketch Book*** (1819–20), ***Bracebridge Hall*** (1822), and other works. He was the first American writer to win the ungrudging (if somewhat surprised) respect of British critics.

James Fenimore Cooper won even wider fame. Following the pattern of Sir Walter Scott's "Waverley" novels, he did his best work in the "Leatherstocking" tales (1823–41), a five-volume series celebrating the career of a great frontiersman named Natty Bumppo. His skill in weaving history into inventive plots and in characterizing his compatriots brought him acclaim not only in America and England but on the continent of Europe as well.

Poe's Gothic tales

Edgar Allan Poe, reared in the South, lived and worked as an author and editor in Baltimore, Philadelphia, Richmond, and New York City. His work was shaped largely by analytical skill that showed clearly in his role as an editor: time after time he gauged the taste of readers so accurately that circulation figures of magazines under his direction soared impressively. It showed itself in his critical essays, wherein he lucidly explained and logically applied his criteria. His gothic tales of terror were written in accordance with his findings when he studied the most popular magazines of the day. His masterpieces of terror—"The Fall of the House of Usher" (1839), "The Masque of the Red Death" (1842), "The Cask of Amontillado" (1846), and others—were written according to a carefully worked out psychological method. So were his detective stories, such as "The Murders in the Rue Morgue" (1841), which historians credited as the first of the genre. As a poet, he achieved fame with "The Raven" (1845). His work, especially his critical writings and carefully crafted poems, had perhaps a greater influence in France, where they were translated by Charles Baudelaire, than in his own country.

Two Southern novelists were also outstanding in the earlier part of the century: John Pendleton Kennedy and William Gilmore Simms. In ***Swallow*** *Barn* (1832), Kennedy wrote delightfully of life on the plantations. Simms's forte was the writing of historical novels like those of Scott and Cooper, which treated the history of the frontier and his native South Carolina. ***The*** *Yemassee* (1835) and Revolutionary romances show him at his best.

American Renaissance. The authors who began to come to prominence in the 1830s and were active until about the end of the Civil War—the humorists, the classic New Englanders, Herman Melville, Walt Whitman, and others—did their work in a new spirit, and their achievements were of a new sort. In part, this was because they were in some way influenced by the broadening democratic concepts that in 1829 triumphed in Andrew Jackson's inauguration as president. In part, it was because, in this Romantic period of emphasis upon native scenes and characters in many literatures, they put much of America into their books.

Jacksonian democracy reflected in humorous writing

Particularly full of vivid touches were the writings of two groups of American humorists whose works appeared between 1830 and 1867. One group created several down-east Yankee characters who used commonsense arguments to comment upon the political and social scene. The most important of this group were Seba Smith, James Russell Lowell, and Benjamin P. Shillaber. These authors caught the talk and character of New England at that time as no one else had done. In the old Southwest, meanwhile, such writers as Davy Crockett, Augustus Baldwin Longstreet, Johnson J. Hooper, Thomas Bangs Thorpe, Joseph G. Baldwin, and George Washington Harris drew lively pictures of the people of the ebullient frontier. All showed the kind of interest in the common man that was a part of Jacksonian democracy.

New England Brahmins. Although Lowell for a time was one of these writers of rather earthy humour, his lifelong ties were to a group of New England writers associated with Harvard and Cambridge, Massachusetts—

the Brahmins, as they came to be called—at an opposite extreme. Henry Wadsworth Longfellow, Oliver Wendell Holmes, and Lowell were all aristocrats, all steeped in foreign culture, all professors at Harvard. Longfellow adapted European methods of storytelling and versifying to narrative poems dealing with American history, and a few of his less didactic lyrics perfectly married technique and subject matter. Holmes, in occasional poems and his "Breakfast Table" series (1858–91), brought touches of urbanity and jocosity to a perhaps oversober polite literature. Lowell, in poems descriptive of the out-of-doors in America, put much of his homeland into verse. His odes—particularly the "Harvard Commemoration Ode" (1865)—gave fine expression to noble sentiments.

The Transcendentalists. Concord, Massachusetts, a village not far from Cambridge, was the home of leaders of another important New England group. The way for this group had been prepared by the rise of a theological system, Unitarianism, which early in the 19th century had replaced Calvinism as the faith of a large share of the New Englanders. Ralph Waldo Emerson, most famous of the Concord philosophers, started as a Unitarian minister but found even that liberal doctrine too confining for his broad beliefs. He became a Transcendentalist who, like other ancient and modern Platonists, trusted to insights transcending logic and experience for revelations of the deepest truths. His scheme of things ranged from the lowest objects and most practical chores to soaring flights of imagination and inspired beliefs. His *Essays* (1841–44), *Representative Men* (1850), and *English Traits* (1856) were thoughtful and poetic explanations of his beliefs; and his roughhewn lyrics, packed with thought and feeling, were as close to 17th-century Metaphysical poems as any produced in his own time.

Thoreau as the great exponent of American individualism

An associate of Emerson with a salty personality of his own and an individual way of thinking, Henry David Thoreau, a sometime surveyor, labourer, and naturalist, was closer to the earthy and the practical than even Emerson was. He also was more of a humorist—a dry Yankee commentator with a flair for paradoxical phrases and sentences. Finally, he was a learned man, widely read in Western classics and books of the Orient. These qualities gave distinction to *A Week on the Concord and Merrimack Rivers* (1849) and to *Walden* (1854). The latter was a record of his experiences and ponderings during the time he lived in a hut by Walden Pond—a defense of his belief that modern man should simplify his demands if need be to "suck out all the marrow of life." In his essay "Civil Disobedience" (1849), Thoreau expounded his anarchistic views of government, insisting that if an injustice of government is "of such a nature that it requires injustice to another [you should] break the law [and] let your life be a counter friction to stop the machine."

Associated with these two major figures were such minor Transcendentalists as Bronson Alcott, George Ripley, Orestes Brownson, Margaret Fuller, and Jones Very. Fuller edited *The Dial,* the chief Transcendental magazine, and was important in the feminist movement.

New England reformers and historians. A worldwide movement for change that exploded in the revolutions of 1848 naturally attracted numerous Americans. Reform was in the air, particularly in New England. At times even Brahmins and Transcendentalists took part. William Lloyd Garrison, ascetic and fanatical, was a moving spirit in the fight against slavery; his weekly newspaper, *The Liberator* (1831–65), despite a small circulation, was its most influential organ. A contributor to the newspaper—probably the greatest writer associated with the movement—was John Greenleaf Whittier. His simple but emotional poems on behalf of abolition were collected in such volumes as *Poems Written During the Progress of the Abolition Question . . .* (1837), *Voices of Freedom* (1846), and *Songs of Labor, and Other Poems* (1850). The outstanding novelist of the movement—so far as effect was concerned—was Harriet Beecher Stowe. Her *Uncle Tom's Cabin* (1852) combined the elements of contemporary humour and sentimental fiction to dramatize the plight of the Negro.

One other group of writers—and a great novelist—contributed to the literature of New England in this period of its greatest glory. The group consisted of several historians who combined scholarly methods learned abroad with vivid and dramatic narration. These included George Bancroft, author of *History of the United States* (completed in 12 volumes in 1882), and John Lothrop Motley, who traced the history of the Dutch Republic and the United Netherlands in nine fascinating volumes (1856–74). The leading member of the group was Francis Parkman, who, in a series of books (1851–92), wrote as a historian of the fierce contests between France and England that marked the advance of the American frontier and vividly recorded his own Western travels in *The Oregon Trail* (1849).

Hawthorne, Melville, and Whitman. History also figured in tales and romances of Nathaniel Hawthorne, the leading New England fictionist of the period. Many tales and longer works—for example, his masterpiece, *The Scarlet Letter* (1850)—were set against a background of colonial America with emphasis upon its distance in time from 19th-century New England. Others, such as *The House of the Seven Gables* (1851), dealt with the past as well as the present. Still others, such as *The Marble Faun* (1860), were set in distant countries. Remote though they were at times from what Hawthorne called "the light of common day," they showed deep psychological insight and probed into complex ethical problems.

Another great American fiction writer, for a time a neighbour and associate of Hawthorne, was Herman Melville. After relatively little schooling, Melville went to sea; a whaling ship, as he put it, was his "Yale College and his Harvard." His first books were fiction in the guise of factual writing based upon experiences as a sailor—*Typee* (1846) and *Omoo* (1847); so were such later works as *Redburn* (1849) and *White-Jacket* (1850). Between 1846 and 1851, however, Melville's reading in philosophy and literary classics, as well as in Hawthorne's allegorical and symbolic writings, gave him new interests and aims. The first sign of this interest was *Mardi* (1849), an uneven and disjointed transitional book that used allegory after the model of Rabelais to comment upon ideas afloat in the period—about nations, politics, institutions, literature, and religion. The new techniques came to fruition in *Moby Dick; or, The Whale* (1851), a richly symbolical work, complex but brilliantly integrated. Only in short stories, "Benito Cereno"—a masterpiece of its genre—and others, in the psychological novel *Pierre* (1852), and in the novelette *Billy Budd* (written 1890?) was Melville later to show sporadic flashes of the genius that created *Moby Dick.*

An ardent singer of the praise of Manhattan, Walt Whitman saw less of the dark side of life than Melville did. He was a believer in Jacksonian democracy, in the splendour of the common man. Inspired by the Romantic concept of a poet as prophet and also by the Transcendental philosophy of Emerson, Whitman in 1855 published the first edition of *Leaves of Grass.* As years passed, nine revised and expanded editions of this work were published. This autobiography in verse was intended to show the ideas, beliefs, emotions, and experiences of the common man in a great period of American individualism. Whitman had a hard time winning a following because he was frank and unconventional in his Transcendental thinking, because he used free verse rather than rhymed or regularly metred verse, and because his poems were not conventionally organized. Nevertheless, he steadily gained the approval of critics and in time came to be recognized as one of the great poets of America.

The slow growth of Whitman's reputation

From the Civil War to 1914. Like the Revolution and the election of Andrew Jackson, the Civil War was a turning point in U.S. history and a beginning of new ways of living. Industry became increasingly important, factories rose and cities grew, and agrarian pre-eminence declined. The frontier, which before had always been an important factor in the economic scheme, moved steadily westward and toward the end of the 19th century vanished. The rise of modern America was accompanied, naturally, by important mutations in literature.

Literary comedians. Although they continued to employ some devices of the older American humorists, a group of comic writers that rose to prominence was different in important ways from the older group. Charles Farrar Browne, David Ross Locke, Charles Henry Smith, Henry Wheeler Shaw, and Edgar Wilson Nye wrote, respectively, as Artemus Ward, Petroleum V. (for Vesuvius) Nasby, Bill Arp, Josh Billings, and Bill Nye. Appealing to a national audience, these authors forsook the sectional characterizations of earlier humorists and assumed the roles of less individualized literary comedians. The nature of the humour thus shifted from character portrayal to verbal devices such as poor grammar, bad spelling, and slang, incongruously combined with Latinate words and learned allusions. Most that they wrote wore badly, but thousands of Americans in their time and some in later times found these authors vastly amusing.

Fiction and local colorists. The first group of fiction writers to become popular—the local colorists—took over to some extent the task of portraying sectional groups that had been abandoned by writers of the new humour. Bret Harte, first of these writers to achieve wide success, admitted an indebtedness to prewar sectional humorists, as did some others; and all showed resemblances to the earlier group. Within a brief period, books by pioneers in the movement appeared—Harriet Beecher Stowe's *Oldtown Folks* (1869) and *Sam Lawson's Oldtown Fireside Stories* (1871), delightful vignettes of New England; Bret Harte's *Luck of Roaring Camp, and Other Sketches* (1870), humorous and sentimental tales of California mining camp life; and Edward Eggleston's *Hoosier Schoolmaster* (1871), a novel of the early days of the settlement of Indiana. Down into the 20th century, short stories (and a relatively small number of novels) in patterns set by these three continued to appear. In time, practically every corner of the country had been portrayed in local-colour fiction. Additional writings were the depictions of Louisiana Creoles by George W. Cable, of Virginia Negroes by Thomas Nelson Page, of Georgia Negroes by Joel Chandler Harris, of Tennessee mountaineers by Mary Noailles Murfree (Charles Egbert Craddock), of tight-lipped folk of New England by Sarah Orne Jewett and Mary E. Wilkins Freeman, of people of New York City by Henry Cuyler Bunner and William Sydney Porter ("O. Henry"). The avowed aim of some of these writers was to portray realistically the lives of various sections and thus to promote understanding in a united nation. The stories as a rule were only partially realistic, however, since the authors tended nostalgically to revisit the past instead of portraying their own time, to winnow out less glamorous aspects of life, or to develop their stories with sentiment or humour. Touched by romance though they were, these fictional works were transitional to Realism, for they did portray common folk sympathetically; they did concern themselves with dialect and mores; and some at least avoided older sentimental or romantic formulas.

The humour of Mark Twain

Samuel Langhorne Clemens (Mark Twain) was allied with literary comedians and local colorists. As a printer's apprentice, he knew and emulated the prewar sectional humorists. He rose to prominence in days when Artemus Ward, Bret Harte, and their followers were idols of the public. His first books, *The Innocents Abroad* (1869) and *Roughing It* (1872), like several of later periods, were travel books in which affiliations with postwar professional humorists were clearest. *The Adventures of Tonz Sawyer* (1876), *Life on the Mississippi* (1883), and *The Adventures of Huckleberry Finn* (1884), his best works, which re-created the life of the Mississippi Valley in the past, were closest to the work of older humorists and local colorists. Even in his best work, however, he succumbed now and then to the temptation to play the buffoon or sink into burlesque. Despite his flaws, he was one of America's greatest writers. He was a very funny man. He had more skill than his teachers in selecting evocative details, and he had a genius for characterization.

Born and raised in Ohio, William Dean Howells was an effective advocate of a new realistic mode of fiction writing. At the start, Howells conceived of Realism as truthful portrayal of ordinary facets of life—with some limitations; he preferred comedy to tragedy, and he tended to be reticent to the point of prudishness. The formula was displayed at its best in *Their Wedding Journey* (1872), *A Modern Instance* (1882), and *The Rise of Silas Lapham* (1885). Howells preferred novels he wrote after he encountered Tolstoy's writings and was persuaded by them, as he said, to "set art forever below humanity." In such later novels as *Annie Kilburn* (1888) and *A Hazard of New Fortunes* (1890), he chose characters not only because they were commonplace but also because the stories he told about them were commentaries upon society, government, and economics.

The Naturalists. Other American writers toward the close of the 19th century moved toward Naturalism, a more advanced stage of Realism. Hamlin Garland's writings exemplified some aspects of this development when he made short stories and novels vehicles for philosophical and social preachments and was franker than Howells in stressing the harsher details of the farmer's struggles and in treating the subject of sex. His *Main-Travelled Roads* (1891) and *Rose of Dutcher's Coolly* (1895) displayed Garland's particular talents. These and a critical manifesto for the new fiction, *Crumbling Idols* (1894), were influential contributions to a developing movement.

The influence of Zola on Dreiser

Other U.S. authors of the same period or slightly later were avowed followers of French Naturalists led by Émile Zola. Theodore Dreiser, for instance, treated subjects that had seemed too daring to earlier Realists and, like other Naturalists, illustrated his own beliefs by his depictions of characters and unfolding of plots. Holding that men's deeds were "chemical compulsions," he showed characters unable to direct their actions. Holding also that "the race was to the swift and the battle to the strong," he showed characters defeated by stronger and more ruthless opponents. His major books included *Sister Carrie* (1900), *Jennie Gerhardt* (1911), *The Financier* (1912), *The Titan* (1914), and—much later—*An American Tragedy* (1925).

Dreiser did not bother with—or did not care for—niceties of style or elaborate symbolism such as were found in French Naturalistic works; but Stephen Crane and Frank Norris were attentive to such matters. In short novels, *Maggie: A Girl of the Streets* (1893) and *The Red Badge of Courage* (1895), and in some of his short stories, Crane was an impressionist who made his details and his setting forth of them embody a conception of man overwhelmed by circumstance and environment. Frank Norris, who admired Crane's "aptitude for making phrases—sparks that cast a momentary gleam upon whole phases of life," himself tried to make phrases, scenes, and whole narratives cast such gleams in *McTeague* (1899), *The Octopus* (1901), and *The Pit* (1903). Both Crane and Norris died young, their full abilities undeveloped but their experiments foreshadowing later achievements in the 20th-century novel.

Formal excellence and psychological insight in Henry James

Henry James. In the books of Henry James, born in New York but later an expatriate in England, fiction took a different pathway. Like Realists and Naturalists of his time, he thought that fiction should reproduce reality. He conceived of reality, however, as twice translated—first, through the author's peculiar experiencing of it, and, second, through his unique depicting of it. Deep insight and thorough experience were no more important, therefore, than the complicated and delicate task of the artist. *The Art of Fiction* (1884), essays on novelists, and brilliant prefaces to his collected works showed him struggling thoroughly and consciously with the problems of his craft. Together, they formed an important body of discussion of fictional artistry.

An excellent short-story writer, James nevertheless was chiefly important for novels in which his doctrines found concrete embodiment. Outstanding were *The American* (1877), *The Portrait of a Lady* (1881), *The Spoils of Poynton* (1897), *What Maisie Knew* (1897), *The Wings of the Dove* (1902), *The Ambassadors* (1903), and *The Golden Bowl* (1904). The earliest of these were international novels wherein conflicts arose from relationships between Americans and Europeans—each group with its own characteristics and morals. As time passed, he became increasingly interested in the psychological processes of his characters and in a subtle rendering of their limited insights, their perceptions, and their emotions.

Critics of the gilded age. Writers of many types of works contributed to a great body of literature that flourished between the Civil War and 1914—literature of social revolt. Novels attacked the growing power of business and the growing corruption of government, and some novelists outlined utopias. Political corruption and inefficiency figured in Henry Adams' novel *Democracy* (1880). Edward Bellamy's *Looking Backward* (1888) was both an indictment of the capitalistic system and an imaginative picturing of a utopia achieved by a collectivist society in the year 2000. Howells' *Traveler from Altruria* (1894) pleaded for an equalitarian state in which the government regimented men's lives. The year 1906 saw the publication of Upton Sinclair's *Jungle,* first of many works by him that criticized U.S. economic and political life and urged Socialism as the remedy.

Two poets embodied criticisms in songs. Edwin Markham's "Man with the Hoe" (1899) was a protest against the exploitation of labour and vaguely threatened revolution; it immediately stimulated nationwide interest. A year later William Vaughn Moody's "Ode in Time of Hesitation" denounced growing U.S. imperialism as a desertion of earlier principles; his "On a Soldier Fallen in the Philippines" (1901) developed the same theme even more effectively.

With the rise of journalistic magazines, a group of journalists became notable as critics of America—the group dubbed "the muckrakers" by Theodore Roosevelt. Ida M. Tarbell's *History of the Standard Oil Company* (1904) and Lincoln Steffens' *Shame of the Cities* (1904) were typical contributions by two members of a large group of journalistic crusaders.

Henry Adams. One of the most devastating and most literate attacks on modern life was an autobiography of a scion of an ancient New England family, the Adamses. Educated at Harvard and abroad, Henry Adams was a great teacher and historian *(History of the United States* [1889–91] and *Mont-Saint-Michel and Chartres* [1904]). *The Education of Henry Adams* (1907), however, complained that a lifelong hunt for some sort of order in the world, some sort of faith for man, left him completely baffled. The quiet, urbane style served well to underline, in an ironic way, the message of this pessimistic book.

Poetry. The latter 19th century and early years of the 20th century were a poor period for American poetry; yet (in addition to William Vaughn Moody) two poets of distinction wrote songs that survived long after scores of minor poets had been forgotten. One was Southern-born Sidney Lanier, a talented musician who utilized the rhythms of music and the thematic developments of symphonies in such fine songs as "Corn" (1875), "The Symphony" (1875), and "The Marshes of Glynn" (1878). Distressed, like many of his contemporaries, by changes in American life, he wove his doubts, fears, and suggestions into his richest poems.

The poetic techniques of Emily Dickinson

The other poet was a New Englander, Emily Dickinson. A shy, playful, odd personality, she allowed practically none of her writings to be published during her lifetime. Not until 1890, four years after her death, was the first book of her poems published, to be followed at intervals by other collections. Later poets were to be influenced by her individual techniques—use of imperfect, or eye, rhymes, avoidance of regular rhythms, and a tendency to pack brief stanzas with cryptic meanings. Like Lanier, she rediscovered the value of conceits for setting forth her thought and feeling. Such poems as "The Snake," "I Like to See It Lap the Miles," "The Chariot," "Farther in Summer than the Birds," and "There's a Certain Slant of Light" represented her unusual talent at its best.

AUSTRALIAN LITERATURE IN THE 19TH CENTURY

The literatures of Australia and New Zealand developed, from beginnings in the late 18th and early 19th centuries, their own characteristic themes, styles, and idioms. They developed at different rates and were affected by different historical and environmental circumstances, such as the fact that the first settlers in Australia included convicts; in New Zealand they were traders and organized colonizers. Differences in geography, strength of native opposition, pace of settlement, and degree of contact with Great Britain and other countries affected development of individual and indigenous literatures. Writers in 19th-century New Zealand achieved relatively little and are considered with New Zealand writing as a whole in the section below on *Australian and New Zealand literature in the 20th century.*

Three main periods may be distinguished in the history of Australian literature. In the first (1788–1880) descriptive and documentary literature predominated. The second (1880–1940) was a time of expansion and consolidation and growing responsibility, and a diversification of Australian society was reflected in literature's greater variety. In the third, from 1940 onward, an increase in literary activity matched and reflected the increase in immigration, in industrialization, and in growth of large suburban areas.

The life of the convicts, administrators, and soldiers who arrived in New South Wales in 1788 and of the free settlers who followed them was a struggle, allowing little time for the arts. Most of those with literary talent employed it in describing their new world. Memoirs and eyewitness accounts of convict life and, later, of conditions on the goldfields abounded and provided inspiration and material for later creative writers.

Novels of life in the colonies

Novelists were also primarily concerned with recording facts about life in the colonies. Often using the basis of a romantic plot, they were preoccupied with convict life, problems of pioneers and new settlers, and conditions on the goldfields, and they dwelt on details of life strikingly different from those of the Old World—landscape, drought, bush fires, bushrangers, the vastness and loneliness of the bush, and the strange flora and fauna.

Two novels by an Englishman, Henry Kingsley, reflected a change in attitude toward Australia. In *The Recollections of Geoffry Hamlyn* (1859), he wrote of an English family whose successful life as pastoralists in New South Wales enabled them to return to a life of ease in Devon. In *The Hillyars and the Burtons* (1865), his immigrant family of Chelsea blacksmiths settled in Australia and rose to a position of some eminence. Outstanding among novels of convict life was *For the Term of His Natural Life* (1874), by Marcus Clarke, an indictment of the transportation system and of conditions in colonial prisons. Rolf Boldrewood (pseudonym of Thomas Alexander Browne) in *Robbery Under Arms* (1888) and *The Miner's Right* (1890) combined excitements of the gold rushes with bushranging adventures. In most novels of this period emphasis was on colourful atmosphere and description and complicated, exciting plot.

Poetry in the first 100 years developed less confidently than did prose, with poets struggling to adapt the English Romantic tradition of landscape and reflective poetry to Australian subjects. Adam Lindsay Gordon won popularity with *Bush Ballads and Galloping Rhymes* (1870).

CANADIAN LITERATURE IN ENGLISH IN THE 19TH CENTURY

The earliest Canadian literature in English was written at about the time of the American Revolution (1775–83) and was largely a result of that event. The 40,000 American Loyalists who took refuge in the northern colonies that had not joined in the revolt were to determine the English-speaking character of Nova Scotia, New Brunswick, and Upper Canada (Ontario) and to give them the slight beginnings of a literature. Joseph Howe was a remarkable personality in the political life of Nova Scotia in his time, editor of the *Nova Scotian* (1828–41) and author of patriotic and narrative poems and several volumes of travel sketches. In the columns of the *Nova Scotian* in 1835–36 appeared the first Canadian literary work to gain international notice, *The Clockmaker; or, The Sayings and Doings of Samuel Slick of Slickville.* Its author, Thomas Chandler Haliburton, was motivated in his public life by the pro-British, anti-American, and anti-democratic prejudices of Tory Loyalist families. In ***The Clockmaker,*** sayings and doings of Sam Slick and vivid little pictures of frontier life made this book a minor classic, in spite of its formless and sometimes repetitious arrangement.

The Maritime writers

These earliest writers lived in the Maritime Provinces, which had attracted better educated Loyalists from the northern Atlantic Seaboard; those who settled in Quebec and along the St. Lawrence and Niagara frontiers were of humbler stock, mostly farmers, and literary activity began later than down east. The first name in this region was that of John Richardson, whose background and military adventures were almost as romantic as his novels. His best book was ***Wacousta*** (1832), a story of the days of the Ottawa Indian chief Pontiac's descent in 1763 on Michilimackinac and Detroit. Of early immigrants from Britain the most literary were two sisters, Catherine Parr Traill and Susanna Moodie. Although they published novels, poems, and stories for children, they have been remembered mainly for autobiographical volumes: ***The Backwoods of Canada*** (1836), by Mrs. Traill, and ***Roughing It in the Bush*** (1852), by Mrs. Moodie, accurate and spirited social records of pioneer life. About contemporary with these two gentlewomen were two immigrant poets of the working class: Alexander McLachlan, whose verses were compounded of memories of poverty and social inequality in Scotland and love for the new country, and Charles Heavysege, whose dramatic poems ***Saul*** (1857), ***Count Filippo*** (1860), and ***Jephthah's Daughter*** (1865), though often commonplace or pompous in language, were bold in design and in implied moral and religious skepticism. (For an account of literature in French in Canada, see below ***Canadian literature in the 20th century: French.)***

FRENCH LITERATURE IN THE 19TH CENTURY

The Revolutionary and Napoleonic period, 1789–1815. After the French Revolution in 1789 the Republic was forced to protect itself from threatening dangers at home and abroad, so that there was little scope for writers save as political journalists and party hacks. Of these, the republican Camille Desmoulins and the royalist Antoine de Rivarol had the force and bite that made them masters of invective. One poet, André de Chénier (see above ***French literature in the 18th century),*** was moved by the violence of the Revolution to compose ***Les Iambes,*** a savage outcry written while he awaited execution.

The other literature that survived the first Revolutionary years was the literature of such eccentrics as monarchist exiles and the work of fanatical and convinced individualists. An autobiographical novel, ***Monsieur Nicolas*** (1796–97), by Restif de La Bretonne, was a vast panorama of the social transformation of Revolutionary Paris as well as a repetitive chronicle of amorous adventure. A notorious work of the Marquis de Sade (1740–1814), ***Justine, ou les malheurs de la vertu*** (1791), with its extreme form of hedonism, has been an acknowledged influence on eroticism. Sébastien-Roch-Nicolas Chamfort left a collection of maxims and anecdotes (1795), which brought him fame. Only Joseph Joubert, a philosophical moralist, possessed the writer's traditional capacity for detachment with his subtle meditations.

The works of Chateaubriand

Napoleon, though carried to power by the Revolution, soon whittled away the liberties he had sworn to defend, and bold and original writers had to work secretly, watched by the police. The greatest of these was François-René de Chateaubriand, whose literary career spanned the first 15 years of the century. His return to belief in the Christian ideal as a principle of social cohesion, precipitated by exile and suffering, gave him a vocation as a writer, and ***Le Génie du Christianisme*** (1802) was a result. Two short novels, ***Atala*** (1801) and ***Rene',*** and his prose epic, ***Lcs Martyrs*** (1809), were illustrations to this apologetic. Chateaubriand's influence was profound; his hero René, seduced from prosaic reality by longings for the infinite that carry him into dreamworlds of the imagination, imposed the emotion of Romantic melancholy on the sensibility of the coming age. Another Romantic impulse, exoticism, was inspired by Chateaubriand's travels in America and was shown in ***Les Natchez.*** But it was in the ***Mémoires d'outre-tombe*** (written 1811–41, published 1849–50; ***The Memoirs of . . . Chateaubriand,*** 1902) that he left his most enduring work.

Influence of German Romanticism

The contemporary fashion for the introspective, self-analytical hero was stimulated by Étienne Pivert de Sénancour's influential ***Oberman*** (1804; later ***Obermann).*** Madame de Stäel asserted her tempestuous personality as a critic in essays, ***De la litte'rature*** (1800) and ***De l'Allemagne*** (1810), in which the "Romantic" genius of German civilization was first revealed to the French and played a vital role in forming the taste of the period. Gothic horror stories, Goethe's early writings, and Schiller's plays opened the way for a literary revolution that coincided with the change of political climate.

Restoration of the monarchy, 1815–48. The result of comparative peace and order under Louis-Philippe after the turmoil under Louis XVIII and Charles X was the astonishing literary achievement of the Romantic generation. Romanticism is the term used to define a new approach to human experience, a transformation of thought and taste that at this time all Europe shared. Poets, novelists, painters, sculptors, and musicians came together for discussion and the hammering out of new formulas to embody the new ideas. In 1827 Victor Hugo and Charles-Augustin de Sainte-Beuve founded the Cénacle, a group that included Alfred de Vigny, Alfred de Musset, Prosper Mérimée, Théophile Gautier, Gérard de Nerval, and Alexandre Dumas; it was here that the new forces of Romanticism worked out their strategy before a performance of Hugo's ***Hernani*** at the Théâtre Français—a clamorous evening that ended in a rout of the outdated representatives of classicism.

This literary revolution had already been prepared by vigorous manifestos. Stendhal (the pseudonym of Henri Beyle), in an essay with the significant title ***Racine and Shakespeare,*** proved that the contemporary audience was more attuned to the turbulence of Shakespearean drama than to the rigid theatrical formulas of Racine. He identified Romanticism with a new direction of taste. Hugo defined Romanticism as "liberalism in literature." However generalized this definition might seem, it summarized the spirit of the movement. The French Romantics claimed the right of an individual to place the fulfillment of his own nature above ethical imperatives. Romanticism presented diverse and even contradictory aspects: spiritual dejection and sensual delirium, the cults of the picturesque and the fantastic, humanitarianism and mysticism—these could all be discerned in major works of the time. It was a literature that recognized no law, save that of liberty.

The poetry of the French Romantics. Many Romantic poets, although free to write of their personal suffering and ecstasy, believed that the social conscience of the century would find its voice in poetry. This conviction became accentuated as the period advanced. The poems of Alphonse de Lamartine were particularly characteristic of this development. Such personal themes as an unhappy love affair ***(Les Me'ditations,*** 1820) began to be coloured by religious emotion in ***Les Harmonies poétiques et religieuses*** (1830), and in ***Les Recueillements*** (1839) he became passionately concerned with social problems, expressing remorse for his self-pitying egoism, with compassion for humanity and faith in its future.

After poetry such as ***Les Chants du crépuscule*** (1835; "Songs of Evening") and ***Les Voix inte'rieures*** (1837; "Voices from Within"), which were outlets for personal melancholy and heartbreak, Victor Hugo also gave voice to contemporary disquiet. In ***Les Rayons et les ombres*** (1840; "The Lights and the Shadows"), he had reached the conviction that he was the anointed prophet of a better age.

The poems of Alfred de Vigny appeared to be self-communings far removed from Hugo's abundant public verse. His ***Poèmes antiques et modernes*** (1826) celebrated significant moments in universal legend and history from the biblical epoch up to the 19th century. ***Les Destinies*** transcribed in symbols a long journey from agonies of doubt to a serene wisdom. The despair of the earlier poems in this collection gave way to peace and a humanist optimism.

Among great poets of this period, only Alfred de Musset retained his literary and political independence and stubbornly refused to serve a public cause. His best known work, the four "Nuits" (1835–37; "Nights"), traced the course of a heartbreaking love affair with a lyricism marred, in places, by a fatal capacity for declamation.

Of lesser poetic talents, an actress named Marceline Desbordes-Valmore wrote poems of charm and ardour, and Sainte-Beuve in *Joseph Delorme* (1829) and Auguste Brizeux in *Marie* (1831) were Romantic Intimistes—poets of emotional reticence who only suggested the deeper regions of the spirit. Such discretion contrasted with the uninhibited Petrus Borel and Philothte O'Neddy, the satiric violence of Auguste Barbier, and the wit and verve of Pierre-Jean de Beranger. Aloysius Bertrand in *Gaspard de la nuit* (1842) and Maurice de Guérin in *Le Centaure* (1840) shared a preoccupation with the effects of the prose poem.

The Romantic theatre. In the preface to *Cromwell* (1827), Hugo proposed that the unities of time and place be abandoned and that the tragic and comic be permitted to coexist in the modern poetic play, in violation of classical rules governing dramatic construction. On the stage as in poetry, Hugo, an original genius, made his own rules for his own purposes.

Alexandre Dumas *père* was the first of his generation to achieve success in the theatre in *Henri III et sa cour* (performed 1829). Hugo's triumph in *Hernani* (1830) and *Ruy Blas* (1838) enhanced the prestige of poetic drama. Vigny's *Chatterton* (1835) was closer, in its construction and dense language, to classical tragedy. Musset's *Lorenzaccio* (1834) was a victory for the Shakespearean manner.

Musset never overindulged current taste for local colour and the historical picturesque, as did Hugo and Dumas. Something of his own obsessive nature belonged to the debauched hero of *Lorenzaccio*, and it was his own wit that graced his sophisticated comedies. His plays were written only to satisfy his own needs, but they alone among the plays of the period retained their freshness and brilliance.

The Romantic novel. It was to Romanticism that the modern French novel owed its overwhelming prestige; Romanticism produced the novel of self-analysis, the historical novel. and, as in the works of Balzac and Stendhal, the novel of contemporary society.

Adolphe (1816), by Benjamin Constant, and *Volupté* (1834), by Sainte-Beuve, were disguised autobiographies in which the central character was subjected to the searchlight of a pitiless self-knowledge. Musset's *Confession d'un enfant du siècle* (1836; *The Confession of a Child of the Century,* 1892), Lamartine's *Raphael* (1849), and Eugène Fromentin's *Dominique* (1863) were further examples of self-portraiture, direct or indirect.

Both Constant in *Journal intime* and *Le Cahier rouge* (1907; *The Red Notebook,* 1948) and Stendhal in *Journal* (published posthumously) followed an impulse toward self-knowledge by writing intimate diaries in which their characters were ruthlessly analyzed. The historical novel, like the historical play, satisfied a reading public in love with the colour and movement that the past provided, and the romances of Sir Walter Scott were slavishly imitated. Vigny, in *Cinq-Mars* (1826; *The Conspirators,* 1873), related the story of a conspiracy in Richelieu's time, an apologia for his own class, the nobility. In *Notre-Dame de Paris* (1831; *The Hunchback of Notre Dame,* 1831), Hugo brought medieval Paris to life, but gave the plot certain pessimistic philosophical implications when all his characters died or were murdered.

Balzac and *La Comédie humaine*

After surrendering to the prevailing mode for historical and autobiographical novels, Honoré de Balzac gradually fitted each of his novels and novel cycles into a vast fresco of 19th-century French society, *La Comédie humaine.* Some described secret dramas of provincial life (*Eugénie Grandet*) and others the violent and complicated passions of Parisian life, a world over which modern tyrants—stock exchange, politics, and press–exercised unlimited power. Balzac was also a master of the novel of character; he understood the whole mechanism of the human heart—avarice in *Gobseck* (1830), paternal love in *Le Père Goriot* (1834–35), and envy in *La Cousine Bette* (1846). But these novels of contemporary life were, in Balzac's eyes, to be crowned by such philosophical studies as La *Recherche de l'absolu* (1834; "In Search of the Absolute"). in which man, with his limitations and ambition, confronted, in a desperate struggle, the forces of destiny.

Stendhal wrote novels whose subtlety and verve made them a major influence; his young heroes reflected his own temperament and quick, lucid intellect. But his novels, like those of Balzac, were social and political documents. *Lo Chartreuse de Parme* (1839; *The Charterhouse of Parma,* 1925) was set in the Italy to which he gave his heart and by which, in some sense, he was formed; *Le Rouge et la noir* (1830; *Scarlet and Black,* 1927) transcribed the moral climate of French society at the Restoration and mirrored all the capacity for revolt in his nonconforming nature. *De l'amour* (1822), in which he elaborated a theory of the passions, was an invaluable appendix to the novels; his travel books were a fascinating commentary on them.

George Sand displayed generosity of feeling and mastery of her craft in a love story such as *Lélia* (1833) and a regional novel such as *La Mare au diable* (1846; "The Pool of the Devil"). Dumas abandoned the theatre for inventive adventure novels, especially *Les Trois Mousquetaires* (*The Three Musketeers*) and *Le Comfe de Monte Cristo* (1845). Eugène Sue developed the peculiar techniques of the serial story in *Les Mystères de Paris* (1842; *The Mysteries of Paris,* 1843).

Shapelessness and overwriting were indigenous faults of Romanticism; yet the period was the golden age of the short novel and short story. Charles Nodier with a seeming artlessness that concealed art conducted readers into regions of dream, as in *Trilby* (1822); Prosper Mérimée, finest technician of them all, gave body to Romantic mystery and an almost unbearable verisimilitude to the violence of Romantic passion (*Colomba* [1840; *Carmen.* 1845]).

Historical and critical writers of the Romantic movement. Romanticism affected the disciplines of historical and critical writing to the point at which they became accepted literary genres in their own right. Augustin Thierry, for example, transformed the early chronicles by his narrative skill. Other historians such as François-Pierre-Guillaume Guizot in *Histoire de la révolution d'Angleterre* (1826–27) and Alexis de Tocqueville in *De la de'mocratie en Ame'rique* (1835–40) set out to verify their theories by an examination of the facts. Jules Michelet, greatest of the Romantic historians, regarded historical writing as an integral re-creation of life; his monumental *Histoire de France* evoked successive advances to the attainment of liberty.

The only literary criticism to have survived from this period was the work of Sainte-Beuve. Sainte-Beuve possessed so lucid a comprehension of the human character behind the written word that *Portraits litte'raires* and his weekly *Causeries de lundi* (1851–62; "Monday Conversations") were often models of perspicacity.

Comte and Positivism

The movement of ideas. Continuity with 18th-century thought was sustained by the Idéologues, a group who carried the propositions of the Encyclopédistes to their logical conclusion and who were opposed by the Transcendentalists, whose philosophy emphasized the intuitive and spiritual over the empirical. The Positivism of Auguste Comte, a mystique of science and humanity, was more attuned to the century of progress. Positivism excluded speculation about the metaphysical, concerning itself only with facts. A disciple of Comte, Émile Littrt, compiled a monumental *Dictionnaire de la langue française.* His precise definitions and historical grasp of the growth of the French language made this dictionary an indispensable instrument in French literature.

The social structures of the period were scrutinized by

Henri de Saint-Simon, Charles Fourier, Pierre-Joseph Proudhon, and Félicité de Lamennais, and their criticism foreshadowed modern Socialism.

The Second Republic and Second Empire, 1848–70. In February 1848 the people of Paris rose against the intolerant bourgeois regime of Louis-Philippe, and from this democratic revolution the Second Republic was born. But from 1852 under Napoleon III, censorship was reintroduced.

Positivist thought, more reassuring to the bourgeois determination to consolidate itself, grew in power; scientific method dominated intellectual life. Ernest Renan and Hippolyte Taine analyzed historical fact and literature with the sole purpose of arriving at an objective truth. Romantic idealism survived in the work of a few poets, but many writers reacted against the previous generation. This reaction was known in poetry as the Parnassian movement and in the novel and drama as Realism.

The aftermath of the Romantic movement. Three poets, Hugo in his later work, Gérard de Nerval, and Charles Baudelaire, formed a bridge between the first and second phases of Romanticism.

Gérard's life was a search for the paradise of which his childhood had been an intimation—a paradise from which he was constantly expelled by the pains and deceptions of existence. The essential Gérard was distilled in two works of prose, *Sylvie* (1854) and *Aurélia* (1855), and in the 12 sonnets called *Les Chimdres* (1854). GCrard was a pure incarnation of the Romantic spirit; the bold innovations of his art looked forward to the experiments of modern poetry.

Hugo (who lived in exile from the Second Empire on Jersey and Guernsey in the Channel Islands), in his finest lyrical verse *(Les Contemplations,* 1856), wrote the memoirs of his spirit—the life of a man who had endured, without defeat, 25 years of struggle and suffering. He composed an epic of humanity, *La Légende des siècles* (1859), which he set within the vaster epic of creation; he wrote the novel of human society *Les Misérables* (1862), in which he gave free rein to all the longings and hopes of his generous heart.

Influence of Baudelaire

Charles Baudelaire profoundly influenced the course poetry was to take in the years after his death. In controlled and rigorous verse, he combined melancholia and the pursuit of the ideal. His poems were collected under the title of *Les Fleurs du mal* (1857; *The Flowers of Evil,* 1926). Hugo spoke of a new excitement in French verse, and certainly these emotionally tense poems explored forbidden regions of human existence as never before. *Petits Poèmes en prose,* experimenting with a language in which metrical freedom was demanded by inner necessity, discovered the emotional tone of a life beginning to be known as "modern." Baudelaire's notebook, *Mon Coeur mis à nu* ("My Heart Laid Bare"), survived as an intimate and pathetic reflection of a tortured spirit.

The Parnassian reaction. The Parnassian poets were determined to rescue French verse from the emotional and verbal imprecision of the Romantics. The movement was heralded by Théophile Gautier, whose poems proposed a belief in formal perfection without which, he asserted, no work of art, however sincerely felt, could hope for immortality.

Charles-Marie-RenC Leconte de Lisle was the first leader of the Parnassians. His reserve and deliberate discipline inhibited movement of life within his verse; his poems were magnificent but sometimes curiously dead. *Poèmes antiques* (1852), and *Podmes barbares* (1862) voiced a longing to escape into the savage grandeur of the past or untouched regions of the natural world, but the rigorously controlled language of his verse paralyzed his inspiration.

Gautier, Baudelaire, and Leconte de Lisle figured in the magazine of the Parnassian movement, as did Theodore de Banville, Sully Prudhomme, François Coppée, and José Maria de Heredia—all proclaiming a belief in formal perfection and flattering themselves that they had brought poetry back to the sublime purity of its nature.

The Realist novel and the drama of social manners. A development of scientific studies, after 1850, impelled novelists to base their work on accurate observation of facts. Champfleury (Jules Husson), who wrote a Realist manifesto in 1857, demanded that novelists choose subjects from the most ordinary aspects of the human condition.

The impersonal, exact writing of Flaubert

Gustave Flaubert, master of the Realists, was a "lost" Romantic who devoted himself to an elaboration of an impersonal, classical art. Whether, as in the historical novels *La Tentation de Saint Antoine* (1874) and *Salammbô* (1862), he chose a Romantic subject or whether, as in novels of modern life, *Madame Bovary* (1857) and *L'Éducation sentimentale* (1869), a subject more in accord with Realist theory, his work was always based on microscopic research and observation. He subjected each novel to an exact articulation of plot and demands of an impersonal style. This made *Madame Bovary* and *L'Éducation sentimentale* two of the great novels of the 19th century. Flaubert's clarity and Realism did much to exorcise the century's sentimental bourgeois complacencies and emotional self-indulgence. Flaubert's *Correspondence* was a moving witness to the devotion with which he served his art.

Jules de Goncourt and his brother Edmond collected documents, took notes in the suburbs and slums of Paris, made clinical studies of pathological cases, and translated this into their novels. Like Flaubert, they were conscious prose stylists, and the same gifts of observation and style were applied to their famous *Journal.*

In the theatre, the failure of Hugo's *Burgraves* in 1843 marked the end of the Romantic dream. Émile Augier took the rich bourgeoisie for his subject, and Alexandre Dumas *fils* portrayed the social and moral problems of his time in beautifully constructed plays, such as La *Dame aux came'lias* (performed 1852). Eugène Labiche used his gift for farce to explore the deeper strata of human character.

The Third Republic, from 1870. The Franco–German War of 1870 marked a new point in French history, but the literary world seemed unaffected. Novelists, with Émile Zola at their head, carried Realism a stage further and formed a Naturalist school. By 1885, however, sustained attacks on the reigning ideology by a critic, Ferdinand Brunetikre, and a philosopher, Henri Bergson, incited a new generation to recover idealism for the novel and the poem.

The Naturalist school. Influenced by Taine, a literary historian, and Claude Bernard, a scientist, Zola maintained that the novel should study "the human temperament and the profound modifications of the human organism under the pressures of environment and events."

Zola's cycle of 20 novels

He set out to compose a cycle of 20 novels, *Les Rougon-Macquart,* that would become a "natural and social history of a single family during the Second Empire." Two of these novels stand out for their savage depiction of working class life: *L'Assomoir* (1877; *The Drunkard,* 1958) and *Germinal* (1885).

A collection of short stories about the Franco–German War, *Les Soirées de Médan,* signalled in 1880 the birth of the Naturalist school; Zola was among the contributors, and Guy de Maupassant made his first public appearance with a masterpiece, "Boule de suif." In addition to 300 short stories, he completed six novels, the best known being *Bel-Ami* (1885), in which he painted the Parisian newspaper world with ruthless realism.

Several writers adhered to Naturalism: Edmond de Goncourt, writing after his brother's death; Alphonse Daudet, a novelist *(Les Aventures prodigieuses de Tartarin de Tarascon,* 1872) and a writer of lively and touching short stories *(Lettres de mon moulin* [1869; "Letters from My Mill"], *Les Contes du lundi* [1873; "Monday Tales"]); and, though on the fringe of the school, Jules Vallès, with a bitter semi-autobiographical trilogy, *Jacques Vingtras* (1879–86). In the theatre, Henry Becque displayed in *Les Corbeaux* (published 1882; *The Vultures,* 1913) many Naturalist traits. André Antoine founded the widely influential Théâtre-Libre for Naturalist drama. Jules Renard had all the Naturalist observation, but used it only to serve a narration and prose style reduced to barest essentials.

The reaction against Naturalism. Some writers scorned this literature of exact description and opposed a materialist philosophy of the democratic age. Barbey d'Aurevilly's best known work, *Les Diaboliques* (1874; *Weird Women*, 1900), was a collection of six short stories written by one "who believes in the devil and his power over the world." In *Les Pleiades* (1874; *The Pleiads*, 1928), the Comte de Gobineau related the travels of three young men who thought of themselves as exceptional beings, with an exclusive scornful pride of race. Villiers de l'Isle-Adam *(Contes Cruels* [1883; *Cruel Tales,* 19631) arraigned the pride of scientists, the tyranny of finance, and insolent triumphs of mediocrity.

Joris-Karl Huysmans was a disciple of Zola in the first stages of his literary career, but despair at the limitations of Naturalist "reality" drove him toward religious faith. Between Naturalism and religion, Huysmans paused to explore the ivory tower of aestheticism. This provoked a most extraordinary and best known novel, *A rebours* (1884: *Against the Grain,* 1922). His last novels described the stages of this journey toward a full, mystical experience of the Christian religion. Lion Bloy, a fanatical Catholic, wrote two autobiographical novels, *Le Désespéré* (1836) and *La Femme pauvre* (1897), in which he thundered imprecations against the impious century.

Novelists who reacted profoundly but less violently were Paul Bourget and Maurice Barrès. Barrès' early work was complex—a series of novels published under the collective title of *Le Culte de moi* (1888–91); here he transcribed his philosophy of egoism—a systematic cult of rare, even violent, sense experiences. Pierre Loti, a naval officer, wrote exotic novels analyzing the sensibility of those who sought relief from the perplexities of their nature in travel and yet could not disperse innate melancholy and despair.

The Dreyfus affair

Anatole France began his writer's career as an escapist into a dream of pagan beauty and wisdom, but the impact of the Dreyfus affair (a bitter dispute over a Jewish army officer wrongly condemned for spying that divided even the most detached men of letters) drew him into the struggle against the right wing. The hero of his novel series *Histoire contemporaine* (1897–1900), first encountered as a convinced skeptic, was also drawn into the political struggle and compelled to fight for justice and humane government.

The Symbolist movement. The Parnassian doctrines of impassiveness and objectivity were opposed by those who still took it for granted that the writing of poetry was a metamorphosis of inner experience. Baudelaire himself had shown the path that Symbolism was to follow. Natural forms, this new movement asserted, are the multiple symbols of a single reality, and beneath their diversity the poet rediscovers a profound unity, mysterious and intangible.

Arthur Rimbaud "studied to make himself a visionary," and, as a way of access to obscure regions of consciousness, he cultivated delirium in all its forms. This experimental initiation into the sources of poetic consciousness was recounted and condemned in *Une Saison en enfer* (1873; "A Season in Hell"); his prose poems *(Illuminations,* published 1886) witnessed the urgency with which his whole being strained toward the hidden mystery of existence and a new language with which to utter it.

Others foreshadowed this new poetic consciousness. The prose poems of Lautréamont (pseudonym of Isidore Ducasse), *Les Chants de Muldoror,* which appeared to celebrate the apotheosis of evil, were composed with a haphazard juxtaposition of images that made them a source book of Surrealism. Tristan Corbière's poems seethed with explosive, savage images, while those of Charles Cros were composed with a metropolitan bitterness that reduced everything to nothing.

Paul Verlaine liberated himself rapidly from the precepts of the Parnassians, which only constrained his voluptuous sensibility. A dangerous friendship with Rimbaud culminated in an emotional crisis echoed in *Romances sans paroles* (1874; "Songs Without Words") and above all in *Sagesse* (published 1880, dated 1881; "Wisdom"), in which the verse was infused with Christian repentance. Verlaine had great sincerity and charm; his verbal music prefigured the more elaborate orchestrations of Symbolism.

The Decadents

To emancipate the self from tradition, to live within the claustrophobia of ennui, as did the poets between 1880 and 1885 known as the Decadents, produced either vague or contorted verse. But decadence did produce one true poet, Jules Laforgue, whose dislocated verse rhythms perfectly transmitted the diffidence and the impudence of his paradoxical nature.

Symbolism was a manifestation of the climate of Hegelianism, the philosophy of Idealism. Symbolist doctrine was based on a belief that the world transmitted by the senses was a reflection of the spiritual universe. The poets thought of a poem as an instrument of metaphysical knowledge. Several of them experimented with a free verse liberated from rhyme and measure.

Stéphane Mallarmé was a central figure of this movement: the simplicity of his character, his poverty, and selfless devotion to his aesthetic made him, in the eyes of disciples, a saint of literature. With *Hérodiade* and *L'Après-midi d'un faune* ("The Afternoon of a Faun") he achieved his particular and profound originality: words brought into relationship not only with other words but with the pauses between them. Each of his poems was an outcome of protracted meditation and required an equal patience of the reader to allow the poetic meaning to be disclosed.

Mallarmé transcended the Symbolist movement. Symbolism failed as a poetic doctrine because it asked more of poetry than it could give. Jean Moréas, one of the movement's founders, broke with it in 1891 and called his countermovement the *école romane* ("Roman school"). This reaction bore little but dead fruit, though Paul Valéry, Mallarmé's close friend but not his disciple as a poet, might have been touched by it, as was Charles Maurras.

The year 1900 did not, on the surface, mark any change in literary taste, but literature, like other arts, was secretly preparing, in a few minds in advance of their time, a revolution as radical as were coming changes that transformed the old Europe and undermined accepted conventions of society (see below *Belgian literature in the 19th century: French).*

Provençal. After the French Revolution, scholars of Provensal literature such as F.-J. Raynouard occupied themselves with the brilliant literary traditions of the Middle Ages; newspapers sprang up, and poets banded together and collected their pieces in volume form; *e.g.*, the nine *troubaire* ("troubadours") who published *Lou Bouquet Prouvençaou* ("The Provençal Garland") in 1823. The "forerunners of the Félibrige" were all more or less academic and appealed to the cultured few; addressing a far wider public were four poets of the people: Jean-Antoine Verdié of Bordeaux, who wrote comic and satirical pieces; Jean Reboul, the baker of Nîmes, who never surpassed his first effort, *L'Ange et l'enfant* (1828; "The Angel and the Child"); Victor Gelu of Marseille, relentless but undeniably powerful of his kind; and, greatest of them all, the true and acknowledged forerunner of the Félibres, Jacques Jasmin. In Languedoc the Marquis de la Fare-Alais is noteworthy.

Mistral and the Félibrige

Joseph Roumanille and Frédéric Mistral, by writing in native dialect, showed a desire to stir the Provensal nation to renewed awareness of its glory. Roumanille's secular verse was moving in the sincerity of its poetry, and his prose works sparkled with delightful humour, but there was no doubt that Mistral's finest work (the long narrative poems *Mirèio* [1859] and *Calendau* [1867], among others, towel-ed above that of his fellows. It was he who in 1852 collected and published an anthology in which all the names yet to become well-known and most of those known already (such as Jasmin) were represented. In 1854 Roumanille was one of seven poets who, on May 21, foregathered at the castle of Font-Ségugne, near Avignon, and founded the Félibrige, a society of writers desirous of promoting Provensal literature. The other six were Mistral, Théodore Aubanel, Anselm Mathieu, Jean Brunet, Alphonse Tavan, and Paul Giéra. Of these, Théodore Aubanel of Avignon, a printer, alone proved himself worthy to rank with Mistral and Roumanille. He was, without a doubt, the deepest nature and temperament among the Félibres, and his lyrics were the most poignant. The first Félibres were followed by Louis Roumieux, Felix Gras (a vapid imitator of Mistral), the Irish-born W.C. Bonaparte-Wyse, and the popular song-writer Charles Rieu (called Charloun). Mistral's doctrine spread from province to province and gave impetus to poetic works throughout the entire oc-speaking region, such as those of Joseph Roux of Limoges, the Auvergnat Arsène Vermenouze, and the Rouergue J. Bessou. Mistral's poetic doctrine, however, did not receive universal acceptance among Provençal writers. Among those who most strongly dissented was A. Foures of Languedoc (Les *Grilhs*, 1887). Many of these poets whose careers were so brilliant at their outset continued to develop in the first decades of the 20th century (Mistral, for example, received the Nobel Prize for Literature in 1904).

GERMAN LITERATURE IN THE 19TH CENTURY

The second Romantic school. The first Romantic school had dispersed by 1804. Two years later, however, another phase of Romanticism was initiated in Heidelberg by Clemens Brentano, Achim von Amim, and J.J. von Gorres, the most characteristic production of which was the collection of folk songs published under the title Des *Knaben Wunderhorn* in 1805–08. Brentano's and Joseph von Eichendorff's lyrical poetry was also striking on account of its melodious quality. Unlike the earlier school, the Heidelberg writers wrote historical works and also collected folk songs and popular prose romances. Their immediate influence on German intellectual life was consequently greater; from that period dates the study of German philology and medieval literature and the nationalism that prepared for the rising against Napoleon.

The plays of Kleist

Around 1809, when the Heidelberg school broke up, the Romantic movement followed two lines of development, one north German, the other associated with Wiirttemberg. In the north, Heinrich von Kleist, Prussia's greatest dramatic poet, was closely allied to the Romantic movement. In his plays, such as *Amphitryon* (published 1807), *Penthesilea* (published 1808), and Prinz Friedrich *von* Homburg (posthumously published 1821), he depicted the confusion of mind resulting from the conflict between the intellect and feeling. Zacharias Werner was popular because of his fate tragedies and martyr dramas. During the same period E.T.A. Hoffmann, a novelist of genius, created a world of fantasy, often of a morbid character, and gave European currency to German Romanticism. Wilhelm Muller, following Byron's example, stirred up German sympathy for oppressed Greeks and Poles.

The Swabian school represented a last phase of Romanticism. The ballads of its leader, Ludwig Uhland, were held second only to Schiller's in popular esteem. Uhland, Justinus Kerner, Gustav Schwab, and others carried on the Romantic tradition but without the impetus of its original force.

The death of Goethe in 1832 marked the end of cosmopolitan humanism, typical of the 18th century. Although some leading Romantics outlived Goethe by a decade, the movement had lost its impact. In conservative, nonliterary fields, Romanticism was more tenacious, and in politics its alliance with rising nationalism coloured German thinking for more than a century. In literature, Goethe himself was of more lasting significance; few were unaffected by his work, though the effect sometimes took the form of rebellion against his olympian predominance.

Rule by conservative governments repressed liberty of thought, and writers' efforts to prescribe solutions for social ills were foiled by severe censorship. Literature was dominated by disillusionment with man's capacity to achieve lofty ends, and pessimistic appraisal of man's role replaced once optimistic or solidly constructive attitudes. In keeping with this change in attitude, writers sought to free themselves from the bondage of Neoclassical and Romantic thought, not always by rejection but often by adaptation.

Grill-parzer's tragedies

The greatest dramatist of this post-Napoleonic era, Franz Grillparzer consciously cast his dramas in the tradition of Neoclassicism, yet they lacked self-confident vigour and assurance. His tragedies had weakness of will as their theme; an individual's effort was of no avail against the overwhelming power of circumstance and the world could not give contentment to man, for only cultivation of inner resources and self-abnegation were able to grant inner peace. In his first play, Die Ahnfrau (performed 1817; "The Ancestress"), fate was an external force; but in a trilogy, Das goldene *Vliess* (1820; The Golden Fleece, 1942), it was the combination of character and circumstance. In Medea both partners in a marriage stand in isolation, incapable of reconciliation because of their burden of guilt.

This element pervaded Grillparzer's plays: *Sappho* (1818), modelled on Goethe's *Torquato Tasso*, depicted the tragedy of a poetess seeking in vain to hold her younger lover; Der Traum *ein* Leben (1834; *Dream* Is Life, 1946) shows the folly of ambition. Konig Ottokars *Glück* und Ende (1823; King Ottokar, His Rise *and* Fall, 1938) treated a ruler's downfall through overweening ambition, while in Ein Bruderzwist in Habsburg (c. 1848; Family Strife in Hapsburg, 1940) untimely action was the source of evil.

The sombre mood of Grillparzer prevailed in the period, notably in the work of C.D. Grabbe and Georg Biichner; it invaded even the popular Viennese comedies by F. Raimund and J.N. Nestroy. The heritage of Neoclassicism and Romanticism appeared discredited, but no new faith had yet grown out of the resulting disillusionment. In the new drama, therefore, established values were questioned and disparaged. Grillparzer's doubts of his creative capacity and his conflicting aspirations gave rise to contradictions in his work, held in check by an impulse to create an almost classical form. Grabbe did not aim at formal perfection. The tragic error of his heroes sprang from a mistaken belief in their ability to control the course of history; the chaos of a meaningless world was stronger than the efforts of individuals. For Biichner, too, the world appeared meaningless; but he was a much more incisive dramatist than Grabbe, for his was a more radical vision. He had been disappointed by the political ineffectiveness of would-be revolutionaries, and his writings waged war against unrealizable ideals and sham rhetoric. His comedy Leortce und Lena (published 1850) satirized Romantic drama and the nebulousness of Romantic ideas; Danrons Tod (published 1835; *Danton's* Death, 1939) was a tragedy of heroic pessimism; and in Woyzeck (posthumously published 1879) the individual was caught in a net of hostile social forces. Buchner's work was thus an indictment of the social and cosmic order.

Lyric poetry. Melancholy and morbid subjectivity characterized the lyric verse of Nikolaus Lenau. The great tradition of Neoclassical and Romantic poetry made it possible to achieve formal excellence but difficult to be original. Friedrich Riickert and August von Platen-Hallermiinde, two most consummate formal lyricists, illustrated this amply. Ruckert was a scholar and poet of no mean importance whose real poetic achievement was a portrayal of the conflict between reason and will. In Platen-Hallermunde's *Sonnette aus* Venedig (1925; "Venetian Sonnets") the German language was skillfully adapted to what was essentially an alien verse form.

For Eduard Morike, too, classical poetry was a model; Morike's formal excellence was matched by an idyllic and melodious portrayal of nature and country life. In Mozart *auf* der Reise nach Prag (1856; "Mozart's Journey to Prague"), he humorously examined the problems of artists in a world uncongenial to art. Nature was also a source of inspiration of Annette von Droste-Hulshoff, whose powerful rhythm and sombre language expressed apprehension of the irrational forces of life, as in Das geistliche Juhr (1851; "The Spiritual Year"). Religious feeling helped balance her vision and gave her poetry greater maturity.

No one attacked Romanticism more ruthlessly than Heinrich Heine. In Germany, Heine was, and still is, a con-

Heine's attack on Romanticism

troversial figure, mainly because he subjected German national susceptibilities and Romantic nationalism to scathing criticism, but his *Buch der Lieder* (1827; "Book of Songs") became one of the best known anthologies of love poetry. Heine described dreams and yearnings, and his realism showed that they were only make-believe. In *Romanzero* (1851) and the posthumously published poems, his poetry conveyed the hopes and anguish that were so real during his last long, drawn-out illness, and his early Saint-Simonian belief in the "rehabilitation of the senses" had given way to a belief in God. His less well-known prose, *Reisebilder* (1826–31; "Travel Sketches"); *Lutezia* (1854), a collection of reports on life in France; his analysis of German intellectual life and history in *Zur Geschichte der Philosophie und Religion in Deutschland* (1834; "On the History of Religion and Philosophy in Germany"), reveal him as a master of ironic prose, dissatisfied with solemnity and pretension. His most effective political satire is a verse epic, *Deutschland: Ein Wintermärchen* (1844; "Germany: A Winter's Tale"), a savage attack upon personal enemies and upon political conditions of his time.

In 1835 an edict of the federal Diet banned Heine's writings, together with those of Ludolf Wienbarg, Karl Gutzkow, Heinrich Laube, and Theodor Mundt. Wienbarg's *Ästhetische Feldzüge* (1834; "Aesthetic Campaigns") had been dedicated to "Young Germany," and this name was then given to the writers who were in tune with the radicalism of the Young Hegelians and political liberals, frustrated by severe censorship and authoritarian government. Wienbarg proclaimed a need for literature to deal with political and social problems. Karl Gutzkow criticized conventional morality and orthodoxy in a novel, *Wally, die Zweiflerin* (1835; "Wally the Doubter"), and a drama, *Uriel Acosta* (performed 1846). Exile was often the fate of those who dared to criticize the established political and social order. Ludwig Borne, a highly talented prose writer, went into self-imposed exile in Paris, and his *Briefe aus Paris* (1830–33; "Letters from Paris") were valuable social documents. Two important lyric poets, Georg Henvegh and Ferdinand Freiligrath, had to flee, one to Switzerland, the other to London.

More conservative was Emanuel Geibel, whose collections *Zeitstimmen* (1841; "Voices of the Age"), *Juniuslieder* (1848), and *Heroldsrufe* (1871; "Herald Calls") contain patriotic verse. He was the leader of the popular Munich school of poetry, whose patron was Maximilian II of Bavaria. Of these poets only Heinrich Leuthold struck a deeper note. Paul von Heyse's novellas rarely roused the reader, and Victor von Scheffel's verse tale *Der Trompeter von Sackingen* (1854) and a novel *Ekkehard* (1857), although popular at the time, came to sound unconvincing.

Realism and regionalism. Realism, often of regional inspiration, was a source of real originality at this time. Poetic Realism, a term coined by Otto Ludwig, aimed at portraying life but only insofar as life was artistically significant and appeared to possess intrinsic value. Attention focussed on social reality but not, as in later Naturalism, on its ugly, pathological side. The main object of a realist writer was to discover positive values in everyday life without reference to transcendental ideas. Changes in the social order had made a host of social critics to question developments that accompanied the beginning of urbanization. Of them Karl Marx is now the best known, but Arthur Schopenhauer's pessimistic philosophy and Ludwig Feuerbach's materialistic thought also amounted to a more sober appraisal of man's capacity. In the same period there developed a positivism that, by way of analogy, sought to apply to the study of literature and society methods that they mistakenly believed to be those of natural science. This, in turn, led to the study of sources and texts, formalized by the first important organized school of literary history in Germany, that of Wilhelm Scherer. Together with his successors, Erich Schmidt and Jakob Minor, Scherer established criticism of modern literature as an academic discipline; Jacob and Wilhelm Grimm and Karl Lachmann had already given academic respectability to German medieval studies.

In fiction, writers concentrated on subjects that they could, through familiarity, accurately describe. Karl Immermann, whose work was still greatly influenced by classicism, in *Der Oberhof* (1839; "The Manor") portrayed peasants rooted in their work and their countryside. The Low German novels *Ut de Franzosentid* (1859) and *Ut mine Stromtid* (1862–64), by Fritz Reuter, contained a wealth of individual character made more convincing by a lively dialect style. With *Quickborn* (1853), a collection of lyrical poetry, Klaus Groth became a prototype of the regional poet; his dialect clearly linked his work to colloquial speech and recalled the folk song.

The gentle art of Stifter

Adalbert Stifter, too, drew much strength from his native Bohemian forest; some of his tales, collected in *Studien* (1844–50) and *Bunte Steine* (1853; "Bright Stones"), were set there. For Stifter, the world of everyday events was a symbol of emotional significance; he therefore portrayed it carefully in his novel *Der Nach Sommer* (1857; "Indian Summer"). Influenced by Goethe's *Wilhelm Meister*, Stifter stressed the power of art to educate; he sought to show how the "law of gentleness"—that is, of humane action, based on justice, simplicity, self-control, restricted activity, and admiration of the beautiful—was effective in bringing about an exemplary life true to nature.

The Realism of Otto Ludwig had a psychological flavour; in *Zwischen Himmel und Erde* (1856; "Between Heaven and Earth"), a conflict within an artisan family was explored with striking objectivity and careful characterization. Another important Realist was Theodor Storm, whose work was stamped by the atmosphere of his native Schleswig-Holstein. Romantic elements in his work were subordinated to realistic description, though both his prose tales and lyric poetry were permeated by a sense of the ephemerality of life. Romantic preoccupation with the past had stimulated historical thought and writing. Some of Storm's novellas had dealt with the past, but other writers, such as Willibald Alexis (W. Häring), Wilhelm Hauff, Gustav Freytag, and W.H. Riehl, made history a main theme. The stories of Alexis—of which *Die Hosen des Herrn von Bredow* (1846; "The Trousers of Squire Bredow") was the best known—were imbued by a delicate sense of humour and a feeling for the landscape of Brandenburg. Freytag's best work is his portrayal of social and economic change in *Soll und Haben* (1855; *Debit and Credit*, 1855) and his comedy *Die Journalisten* (performed 1852; *The Journalists*, 1888).

Wilhelm Raabe's analysis of social life was more profound. He anticipated 20th-century methods of storytelling in focussing attention both on the story and the way in which it was told. He attacked the narrowness of the bourgeois philistinism and the nationalism of Bismarck's empire. In his later work—*Stopfkuchen* (1891) and *Die Akten des Vogelsangs* (1895; "The Vogelsang Documents")—he depicted eccentric characters with a rich inner life who achieved spiritual freedom. His pessimism and humour were paralleled by Wilhelm Busch, a satirical poet most widely known for his comic-strip verse-story *Max und Moritz* (1865). His down-to-earth laughter over human imperfection savagely exposed hypocrisy and illusion.

The plays of Friedrich Hebbel revealed poetic Realism at its most powerful. His work was a synthesis of psychological analysis and metaphysical beliefs. Deep inner impulses drove his characters to doom: his dramas, *Judith*, *Herodes und Mariamne*, and *Gyges und sein Ring* (all between 1840 and 1856), depicted the tragedy of those who suffered defeat because their outraged individuality did not allow them to compromise. In his last play, *Die Nibelungen* (performed 1861), he interpreted an old legend in terms of his own psychological and metaphysical ideas. His lyrical poetry depended on close identification with the growth and decay of nature. Theodor Fontane's Realism was subtle and impressive for its humour and irony. His novels about life in Brandenburg were characterized by psychological insight and an understanding of social problems in *L'adultera* (1882), *Irrungen Wirrungen* (1888), *Frau Jenny Triebel* (1892), and *Effi Briest* (1895), human relationships clashed with society and survived or broke down according to their innate strength.

Nietzsche's theory of art

Friedrich Nietzsche was a harbinger of 20th-century literature. His distinction, in *Die Geburt der Tragodie aus dem Geiste der Musik* (1872; *The Birth of Tragedy,* 1909), between the Apollonian and Dionysian elements of art was of considerable consequence. The view spread that classical art could not only be serene but also be ecstatic, and that the origins of Greek drama sprang from the orgiastic intoxication of Dionysian religious mysteries. Nietzsche's emphasis on a need to liberate personality from shackles of conventional Christian morality—he denounced Richard Wagner after the composer had turned to Christianity—skepticism as to the validity of the artist's statements and his place in society, and prophecy of nihilism to come provided an arsenal of ideas and intellectual ferment for the next generation of writers.

The keynote of Naturalism in Germany was scientific objectivity, and the principal model was the work of Émile Zola in France. An anthology of lyric verse, *Dichtercharaktere,* appeared in 1884, in which urban life was the theme, but the real revolution was made by Arno Holz, who, in *Buch der Zeit* (1886; "Book of the Times"), revealed himself as the first important poet of Naturalism. Together with Johannes Schlaf, he wrote three tales published as *Papa Hamlet* (1889), in which they attempted faithfully to depict the minutiae of life, even its pathological and sordid aspects. Many of his poems dealt with the life of the poor. Gerhart Hauptmann was the chief naturalist playwright. His play *Vor Sonnenaufgang* (1889; "Before Dawn," 1912) was memorable for its novel technique, for it was a drama without hero or proper plot. *Die Weber* (1892; *The Weavers,* 1962), also without a plot, was an indictment of dire poverty caused by industrialization, while *Der Biberpelz* (1893; *The Beaver Coat,* 1912), one of the few successful German comedies, was a satire on Prussian officialdom outwitted by the common sense of a clever washerwoman. In *Hanneles Himmelfahrt* (1893; "The Assumption of Hannele"), Hauptmann began to experiment with symbolist drama, and *Der Nurr in Christo, Emanuel Quint* (1910; *The Fool in Christ,* 1911) revealed his force as a prose narrator.

In Hauptmann's wake several writers wrote in a straightforward Naturalist manner. The best known among them were Herman Sudermann, notable for *Ehre* (1889; translated as *What Money Cannot Buy)* and *Heimat* (1893; translated as *Magda,* 1896), plays criticizing middle class morality; and Max Halbe, whose *Jugend* (1893; "Adolescence") was a drama of adolescent love.

In the course of the 19th century, German literature had increasingly abandoned an idealistic conception of man and turned to a more down-to-earth and deprecating appraisal of reality, reflecting the rise of positivist and materialist thought in science. This proved too narrow; and in consonance with the new relativist scientific cosmology of the 20th century, artistic imagination began to portray a more complex vision of the world.

The revival of drama in Vienna

Austrian literature. The reforms of Maria Theresa and her son Joseph II encouraged a revival of Austria's German language and culture and in particular of German theatre. Vienna became once more the centre of drama, with dramatists such as Josef Schreyvogel, Heinrich Laube, Eduard von Bauernfeld, and Franz Grillparzer (see above *The second Romantic school)* writing for the court theatre, or *Burgtheater.* The popular state theatre, or *Volksbühnen,* produced the fairy-tale plays of Ferdinand Raimund, the peasant drama of Ludwig Anzengruber (see below), and the popular pieces of Johann Nestroy. Dramatists such as Joseph von Hammer-Purgstall and Jakob Fallmerayer dealt with far-off places, while Franz Stelzhamer, one of several regional dramatists, wrote in an Upper Austrian dialect. Other playwrights active at this time included Charles Sealsfield, Friedrich von Schwarzenberg, Adalbert Stifter (see above *Realism and regionalism),* and Franz Kranewitter. Toward the end of the century Ferdinand von Saar and Marie von Ebner-Eschenback provided realistic accounts of both bourgeois and peasant society. A novelist, Peter Rosegger, and a dramatist, Ludwig Anzengruber, wrote about peasant life. Anzengruber's humour, as seen in his plays *Der Pfarrer von Kirchfeld* (1870; "The Minister from Kirchfeld") and *Das vierte Gebot* (1877; "The Fourth Offer"), softened his polemical didacticism.

SWISS LITERATURE IN THE 19TH CENTURY

German Swiss. In the 19th century, international fame and critical estimate at home did not always occur in German Swiss literature. Thus, *Der schweizerische Robinson* (1812–27; *Swiss Family Robinson,* 1818), by Johann David Wyss, and *Heidi* (1881), by Johanna Spyri, were examples of worldwide success by works little considered at home. Meanwhile, Bern produced its greatest novelist in Albert Bitzius, who, as Jeremias Gotthelf, combined epic power with a profound knowledge of human nature in a series of great novels accurately depicting the life of Swiss peasant communities. It was in fiction and poetry that the Ziirich writers Gottfried Keller and Conrad Ferdinand Meyer achieved fame. Keller first reached the height of his creative powers with *Der grune Heinrich* (1854–55; revised 1879–80), and Meyer with his verse epic *Huttens letzte Tage* ("Hutten's Last Days") in 1871. Meyer drew upon the mountains and upon the princely courts of the past and, as a writer of fiction, immersed himself in the life and culture of the Renaissance.

Keller's prose and poetry were full of genial humour and exuberance, matching his fundamentally human attitude, and his love for his country took different forms, from hymns of praise to pungent social criticism. The importance of this period was further emphasized by the work of a Basel historian, Jacob Burckhardt, and his disciple Heinrich Wölfflin, who continued and completed that part of his master's work that was interrelated with the history of fine arts.

French Swiss. At the beginning of the 19th century, French Swiss literature was largely connected with the influential circle of Madame de Staël, who, in her house on Lake Geneva, often entertained leading intellectuals of western Europe. She and Benjamin Constant were both of Swiss extraction, and much of their work was inspired by their background. Another member of Madame de Stael's circle was a Genevan, Simonde de Sismondi, who wrote the *Histoire des rhpubliques italiennes du moyen âge* (1807–18).

Rodolphe Toepffer depicted a more homely humorous world in *La Bibliothèque de mon oncle* (1832; "My Uncle's Bookshop"), while the poems of Juste Olivier testified to a pronounced patriotism. A powerful international influence was exercised by the fragments of H.F. Amiel's *Journal intime* (1883–84).

Italian Swiss. About 9 percent of Switzerland's population use Italian as their native language and for their literature. The Swiss Italians included Francesco Soave of Lugano, who translated John Locke and Salomon Gessner into Italian and, following the example of the latter, wrote his own *Novelle morali,* which appeared in many editions and made the story of William Tell popular in Italy; and Stefano Franscini, who helped to bring about the unity of the Ticino and connect it with the spirit of the Swiss Confederation. His main work, *La Svizzera italiana* (1837–38), bore as a Latin motto: "The Alps looking toward Italy are my greatest hope." His *Nuova statistics della Svizzera* (1847) was a particularly effective description of Switzerland.

DUTCH LITERATURE IN THE 19TH CENTURY

The growth of nationalistic Romanticism

Although Jacob Geel's essays in *Onderzoek en phantasie* (1838; "Inquiry and Fantasy") set a new standard in philological and philosophical criticism in Dutch literature, he, like the poet Antony Staring, was almost swept aside by the growing wave of Romanticism. Simultaneously, the reactionary freethinking of the Enlightenment roused the militancy of the Calvinists, who realized that their entrenched position was being threatened. Willem Bilderdijk and his disciple Isaac da Costa reminded the nation of its divine mission, and foreign historical novels (particularly the work of Chateaubriand

and Scott) provided models for historical national Romanticism. In 1826 David van Lennep published a paper calling for novels modelled on Sir Walter Scott; his son Jacob responded with *De pleegzoon* (1833; *The Adopted Son,* 1847). Although now little read, Jacob was a prolific writer of poetry, novels, drama, librettos, translations (of Byron, Scott, Shakespeare), history, and literary criticism. More important was *Hermingard van de Eikenterpen* (1832; "Hermingard of the Oak Burial Mounds"), by Aernout Drost, with the makings of a modern novel. Drost also started a new literary journal, *De muzen* (1834), which, like his novel, was true to the spirit of the "Réveil." Two men on its staff, a historian, R.C. Bakhuizen van den Brink, and a future leader of the literary revival, Everardus Johannes Potgieter, continued the campaign to improve critical standards in *De gids* ("The Guide"), known as the "Blue Butcher" because of its merciless treatment of complacency. Potgieter defined the historical novel, and Anna Bosboom-Toussaint put his ideas into effect, transposing the universal Christian idealism of Drost to the national Protestant faith of the Golden Age. Her best known book, *Majoor Frans* (1874; "Major Francis"), was not historical, belonging rather to an era of liberal politics and female emancipationists. One historical novel had an effect far beyond its literary value. Hendrik Conscience's *Leeuw van Vlaanderen* (1838; *The Lion of Flanders,* 1853–57) aroused a national consciousness in Flanders that matured in the Flemish revival (see below *Belgian literature in the 19th century: Flemish).*

Nicolaas Beets was the most popular poet of his period, but, although he was feted as a national Protestant poet, his claim to fame rests on his sketches in *Camera obscura* (1839), which is still overrated for its so-called Dickensian observation of detail. The more subtle humour of Potgieter never had the appeal of Beets or of Johannes Kneppelhout, who wrote *Studententypen* (1839–41), and his quest for originality tended to deprive his style of simplicity and clarity. The sober, personal, and merciless standards of the critic Conrad Busken Huet, a progressive who left the church, placed Dutch writing in a truer perspective with western European writing. His essays were collected in *Litterarische fantasien en kritieken* (1868–88, "Literary Fantasies and Criticisms"), and his later work was best represented by *Het land van Rembrandt* (1882–84). Meanwhile, a furor had been caused by an entirely unknown writer, Multatuli, pseudonym of Eduard Douwes Dekker, whose *Maw Havelaar* (1860), *a* masterly satire of Dutch exploitation of the Dutch East Indies, suddenly revealed the full poignancy of the language. Dekker's writing, in *Wouterje Pieterse* (1865–77) and *Minnebrieven* (1861; "Love Letters"), vibrated between extremes of sentimentality and anarchy, iconoclasm and utopianism. Although poetry as a convention was anathema to him, Dekker was nevertheless rather greatly admired by the young men of the new generation; and the first of the 1880s poets, Jacques Perk, wrote sketches in Dekker's humorous style before writing a sonnet cycle, *Mathilde* (published posthumously in 1882), which opened a new epoch in Dutch literature.

The appearance of the periodical *De nieuwe gids* ("The New Guide") in 1885 marked the beginning of an important renaissance of literature in the northern Netherlands. It differed from the earlier periodical *De gids* in that it pursued an exclusively aesthetic ideal. Leaders of the movement were the poets Willem Kloos and Albert Verwey and the violent and lyrical critic Lodewijk van Deyssel. Among others involved in the movement was the dramatist, poet, and prose writer Frederik van Eeden, who occupied a special position. Herman Gorter, who became the foremost poet after *Mei* appeared in 1899, and the poets Hélène Lapidoth-Swarth, Pieter Cornelis Boutens, and Jan Hendrik Leopold also stood close to the movement. Also belonging to this generation were the writers Jacobus van Looy, who wrote pictorial prose, and Louis Couperus, who described the social activities of The Hague. (For literature in Flemish, see the following.)

BELGIAN LITERATURE IN THE 19TH CENTURY

Flemish. Before the end of the 18th century, Willem Verhoeven and Jan Baptist Verlooy started a reaction in Flanders against French influence, returning to the work of 16th-century Humanists but neglecting medieval masterpieces. Revival was helped by the *rederijkers,* who continued to use Flemish, not French. Karel Broeckaert wrote dialogues in a spirit of rational liberalism modelled on Joseph Addison's *Spectator* essays and also wrote the first Flemish prose story, *Jellen en Mietje* (1811). A poet, Peter Joost de Borchgrave, embodied the transition from classicism to Romanticism, and Jan Baptist Hofman wrote "bourgeois" dramas.

Romanticism made its influence felt in the 19th century and was linked to a revival of nationalist consciousness. The older generation—Jan Frans Willems, Jan Baptist David, Philip Blommaert, and Ferdinand Snellaert—while remaining rationalists, rediscovered the rich medieval inheritance. To their group belonged two important poets of the new age, Karel Lodewijk Ledeganck and Prudens van Duyse. The younger generation was more spontaneously Romantic, as was illustrated by the work of Hendrik Conscience, creator of the Flemish novel. Theodoor van Rijswijck and Johan Alfried de Laet freed poetry from classical concepts and forms, and the ultra-Romantic stories of Eugeen Zetternam and Petrus Frans van Kerckhoven denounced social evils.

Led by a Realist, Jan Lambrecht Domien Sleeckx, a reaction against Romanticism set in about 1860. Writing became characterized by acute observation, description of local scenery, humour, and, not infrequently, a basic pessimism, as could be seen in novels such as Anton Bergmann's *Ernest Staes* (1874) and Virginie Loveling's *Een dure eed* (1892; "*A* Dear Oath"). The poets Johan Michiel Dautzenberg, Jan van Beers, and Rosalie Loveling, together with the first important Flemish critic, Max Rooses, also reflected a new Realistic approach.

The reaction of Flemish Realism

Parallel to this Realism was a remarkable revival in West Flanders, headed by a lyrical poet, Guido Gezelle, and a nationalist agitator, Hugo Verriest, who revitalized prose. Albrecht Rodenbach, who died before his talents had matured, wrote militant songs, thoughtful lyrics. monumental epics, and a verse tragedy, *Gudrun* (1882).

The review *Van nu en straks* ("Of Today and Tomorrow"), which was to make Flemish literature of European importance, was influenced more by Gezelle, Verriest, and Rodenbach than by the generation of the 1880s. Led by Pol de Mont, an already complex modern poet, the writers of the 1880s had, however, widened horizons and, by emphasis on individualism and "art for art's sake," prepared the ground for their successors.

French. Impetus for the long-awaited Belgian literary renaissance came from Max Waller, founder in 1881 of an influential review, *La Jeune Belgique* ("Young Belgium"), which gave its name to a movement aiming for literary expression of a genuinely Belgian "consciousness." It was a group of writers of highly individual talent pursuing similar aims in different ways.

Of novelists early associated with the movement, Camille Lemonnier, Eugène Demolder, and Georges Eekhoud were the most influential. Later *Jeune Belgique* novelists were Georges Virrès, who lyrically depicted peasant life in the Kempenland (Campine), and Maurice des Ombiaux, who portrayed his compatriots with sly humour.

Stimulated by the *Jeune Belgique* movement was a group of poets much concerned with style and language; Grégoire Le Roy, a gifted lyrical Symbolist poet, and a poet and critic, Albert Mockel, founder of an influential Symbolist review, La *Wallonie,* were in this group.

They were overshadowed, however, by three poets of international stature: Émile Verhaeren, Maurice Maeterlinck, and Max Elskamp. Verhaeren extolled humanity's struggle toward social justice; Maeterlinck, creator of the Symbolist poetic drama, was concerned to illuminate life's inner meaning; and Elskamp treated themes from folklore and legend.

Outstanding dramatists were Maeterlinck (*L'Oiseau bleu* [performed 1908; *The Blue Bird,* 1909]), Verhaeren, and Georges Rodenbach. A playwright, novelist, and critic, Edmond Picard propounded Socialism, Symbolism, and Impressionism.

The period also saw a beginning, in the work of Godefroid Kurth, of modern historiography and of modern criticism. An outstanding historian was Henri Pirenne. Art and literary criticism flourished, and an atmosphere for a flowering of scholarship culminated in the foundation (1920) of the Belgian Académie Royale de Langue et de Littérature Françaises.

Walloon. The number of Walloon poets and other dialect writers increased during the 19th century. Charles-Nicolas Simonon wrote the moving stanzas of "Li Côpareye" (the name of the clock of the cathedral of Saint-Lambert), François Bailleux his charming "Mareye," and the first great Walloon lyric poet, Nicolas Defrêcheux, his famous "Leyiz-m'plorer" (1854; "Let Me Weep"). The establishment at Liège, in 1856, of the Société Litgeoise de Littérature Wallonne had considerable influence on both language and literature. The number of poems, songs, plays, and even translations into Walloon of such authors as La Fontaine, Ovid, and Horace increased.

Other parts of Belgium, apart from prolific Liège, still remained active centres of dialect writing. In the 19th century, Namur could boast especially of Charles Wérotte and Nicolas Bosret, poet of the touching song "Bia Bouquet." The works of Jean-Baptiste Descamps and others originated in Hainaut. Walloon Brabant was the home of a truculent Abbe Michel Renard.

By the end of the 19th century many writers working in Walloon dialects chose a rather doctrinaire Realism to depict workaday existence and remained somewhat hidebound by social conventions. Poets included Joseph Vrindts and, above all, Henri Simon, who sang of working peasantry. Successful playwrights included André Delchef and Édouard Remouchamps, whose vaudeville comedy in verse, *Tâtî l'pèriquî* (performed 1885; "Tati the Hairdresser"), married observation and technical dexterity.

ITALIAN LITERATURE IN THE 19TH CENTURY

The 19th century was a period of political ferment in Italy, and many outstanding writers were involved in public affairs. Much of the literature written with a political aim, even when not of intrinsic value, became part of Italy's national heritage and inspired not only those for whom it was written but all who valued freedom.

Early Romanticism. Foremost among writers in early struggles for his country's unity and freedom from foreign domination was Ugo Foscolo, who reconciled passionate feeling with a formal perfection inspired by classical models. His *Le Ultime lettere di Jacopo Ortis* (1802; "Jacopo Ortis' Last Letters") was an epistolary story of a young man forced to suicide by frustrated love for both a woman and his fatherland. It was extremely moving and popular, as was a poem, "Dei sepolcri" (1807; "Of the Sepulchres"), where, in fewer than 300 lines, he wrote lyrically on the theme of inspiration to be found at the tombs of the great, exhorting Italians to be worthy of their heritage. This poem influenced the Italian Risorgimento, or national revival, and a passage in which Florence was praised because it kept in Santa Croce the ashes of Michelangelo, Machiavelli, and Galileo is still very popular in Italy. Two odes celebrating the divine quality of beauty, 12 sonnets ranking with the best of Petrarch's and Tasso's, and an unfinished poem, "Le Grazie" ("The Graces"), also testified to Foscolo's outstanding poetic merit. As an exile in England from 1816 until his death (1827), he wrote for English readers remarkable critical essays on Italian literature.

In Foscolo patriotism and classicism united to form almost one passion, but Vincenzo Monti was outstanding for mobility of feeling. He saw danger to his country in the French Revolution and wrote "Il pellegrino apostolic~'(1782; "The Apostolic Pilgrim") and *La Bassvilliana* (1793; *The Penance of Hugon,* 1805); Napoleon's victories aroused his praise in "Prometeo" ("Prometheus"), "Il bardo della selva nera" ("The Bard of the Dark Wood"), and "La spada di Federico 11" (1806; "The Sword of Frederick II"); in "Il fanatismo" and "La superstizione" (1797) he attacked the papacy; later he extolled the Austrians. Thus every great event made him change his mind, through lack of political conviction, yet he achieved greatness in *La bellezza dell'universo* (1781; "The Beauty of the Universe"), lyrics inspired by domestic affections, and in a translation of the *Iliad,* a masterpiece of Neoclassical beauty.

Classicism, purism, and the reaction of Romanticism

Melchiorre Cesarotti occupied a prominent position in the world of learning at the end of the 18th century, and his *Poesie di Ossian* (1763–72) influenced Foscolo, Leopardi, and others by their mysterious and gloomy fantasy, so alien from classical inspiration; *Saggio sulla filosofia delle lingue* (1785) was an important essay in the dispute on the Italian language. The trend was toward pedantic classicism as a reaction against an excessive Gallicism favoured by some 18th-century writers. Among the purists was Antonio Cesari, who brought out a new enlarged edition of the *Vocabolario della Crusca* (the first Italian dictionary, published by the Accademia della Crusca in 1612). He wrote *Sopra lo stato presente della lingua italiana* ("On the Present State of the Italian Language") and endeavoured to establish the supremacy of Tuscan and of Dante, Petrarch, and Boccaccio as models. But a Lombard school opposed this Tuscan supremacy. Vincenzo Monti, its leader. issued *Proposta di alcune correzioni ed aggiunte al vocabolario della Crusca* ("Proposal for Some Corrections and Additions to the Crusca Dictionary"), which attacked the Tuscanism of the Crusca. All Italy took part in the disputes about language, literature, and politics.

A patriot, a classicist, and a purist at one and the same time was Pietro Giordani; he was almost a compendium of literary movements of the time; his works were few and comparatively unimportant, but his position among contemporaries singled him out as the last great exponent of *purismo.* French domination in Italy made people realize that traditional aspirations toward political unity and independence were possible, for the separate states and rulers had been obliterated by Napoleon; and Italians were convinced that they could not rely on foreign help, for the French, who had entered Italy as liberators, had proved oppressive masters. An artificial form of classicism was associated with the Napoleonic domination, so that when Napoleon fell, forces antagonistic to classicism arose. Literary Romanticism had already won favour with the French, who erroneously thought themselves akin to German Romantics. Between 1816 and 1818 a battle was fought for Romanticism, particularly in Milan, where a Romantic periodical, *Il Conciliatore.* was published. G. Berchet (patriotic poet and author of *La lettera semiseria di Grisostomo,* a manifesto of Romanticism), Silvio Pellico, L. di Breme, Giovita Scalvini, and E. Visconti were among its contributors. Their efforts were silenced in 1820 when several were arrested by the Austrian police because of their liberal opinions; among them was Pellico, who later wrote a famous account of his experiences, *Le mie prigioni* (1832; *My Prisons,* 1853). It appears that the new school, though advocating the study of modern works in different languages and a certain independence from classical imitation and conventions, was on the whole consistent with Italian tradition. The chief exponent of Italian Romanticism was Alessandro Manzoni, whose major work, *I promessi sposi* (1825–27; *The Betrothed,* 1889), was a novel about two peasant lovers whose marriage is opposed by a wicked nobleman and who are finally reunited after many adventures through war, revolution, and plague. This was a great masterpiece of modern Italian fiction and has ranked with the greatest novels of any literature. Manzoni's fluency, elegance, and humour pervaded the work, which had a vivid 17th-century background integrated with a complicated plot. Manzoni's genius as a poet showed in the odes "Il cinque maggio" ("The Napoleonic Code"), written on the death of Napoleon, and "Marzo 1821" and in passages of *Inni sacri* ("Sacred Hymns"; five poems in celebration of church festivals) describing human affections. His tragedies, *Il conte di Carmagnola* (performed

The publication of *Il Conciliatore*

1820) and *Adelchi* (1822), marked a victory of Romanticism over classicism; they contained passages of great lyrical beauty but lacked strong dramatic power.

Giovan Battista Niccolini was another notable dramatist of the Romantic period who gradually evolved from classical to Romantic in style but never achieved more than topical significance. although his eloquence stiired his compatriots to love of liberty.

Leopardi's pessimism

The foremost Italian poet of the age was Giacomo Leopardi, an outstanding scholar and thinker whose philological works together with his philosophical writings, *Operette morali*, would alone place him among the great writers of the 19th century. Embittered by solitude, sickness, and near penury, from the age of 20 he realized the vanity of his hopes. Though he developed a doctrine of universal pessimism, seeing life as evil and death as the only comfort, the poetry based on these bitter, despairing premises was far from depressing. Leopardi's poems were all contained in one book, *I canti* ("Songs"), first published in 1831. Some were patriotic and were once very popular; but the best came from deeper lyrical inspiration. Among them were a meditation on infinity; "A Silvia," on the memory of a girl who died when he was 20; "Le ricordanze," an evocation of his childhood; "Il passero solitario," comparing the lonely poet and a sparrow that sings by itself; and "La quiete dopo la tempesta" and "Il sabato del villaggio," two pictures of village life. They balance depth of meaning and formal beauty, simplicity of diction, intensity, and verbal music.

The Risorgimento and after. After 1821 the literature of Italy was mainly concerned with political events tending toward the Risorgimento. Romantics and classicists were both actuated by patriotism. Manzoni's novel created so deep an impression that several writers followed his example. Massimo Taparelli d'Azeglio, painter, statesman, and author, had a political object in writing *Ettore Fieramosca*, a heroic episode of the Renaissance. Though overcrowded with incidents, it was full of zest. D'Azeglio, a Piedmontese monarchist, was several times prime minister and during his leisure composed excellent memoirs, *I miei ricordi.* Francesco Guerrazzi, Republican follower of Mazzini, wrote several historical novels, notably *L'assedio di Firenze* ("The Siege of Florence"). Writing was, to him, a means of fighting oppression. He led the Tuscan revolution of 1848 and wrote in the intervals of his political activity. His vigorous rhetoric impressed his readers, but he was not always historically accurate. Garibaldi knew Foscolo's "Dei Sepolcri" by heart, and his followers called each other by the names of Guerrazzi's characters. It was among Garibaldians that the last historical novelists were to be found. G.C. Abba wrote *Da Quarto al Volturno*, a moving account of Garibaldi's campaign in 1860; and Ippolito Nievo wrote *Le confessioni di un italiano* (1858), a cyclic novel that just fell short of being a masterpiece through lack of revision because of his premature death. Giuseppe Mazzini, the greatest political leader of the Risorgimento and advocate of an Italian republic, wrote profusely on literature and politics; at first championing Romanticism, he later opposed it as laying too great stress on the individual, but at all periods his style was vigorous and his convictions sincere. His political works, mainly written during his exile in London, included *Fede ed avvenire* (1835; "Faith and the Future"), *Dei doveri degli uomini* (1844; "The Duties of Men") and *Ai giovani d'Italia* (1860; "To the Youth of Italy"). Vincenzo Gioberti had to leave his country because of his love of liberty. In 1843 he published *Del prirnato morale e civile degli italiani* ("The Moral and Civil Superiority of the Italians"), in which he supported the moral and historical right of Italians to independence and advocated an Italian confederation under the presidency of the pope. This book formed a basis of the political creed of those Italians, other than followers of Mazzini, who took part in the revolution of 184849; the failure of that revolution caused Gioberti to revise his opinions and to advocate the institution of a monarchy under the House of Savoy in *Rinnovamento civile d'Italia* (1851; "The Civil Renovation of Italy"). Patriotic literature was common throughout the period. Luigi Settembrini's *Ricordanze della mia vita* ("Memoirs of My Life") was one of the best examples. Others were written by Manzoni, Leopardi, Berchet, and Gabriele Rossetti; by a satirist, Carlo Porta, who made French and Austrians a target of his invective; and by Giuseppe Gioacchino Belli, who used extraordinary gifts of observation to depict, in 2,000 sonnets, the life of the people in papal Rome. Giuseppe Giusti's satires were also avidly read. Francesco Dall'Ongaro, Aleardo Aleardi, and Giovanni Prati were sentimental Romantics and patriots. Prati was standard-bearer of the "second Romanticism," a movement that thrived between 1850 and 1870, characterized by sentimentality and manneiism. Prati's *Ermengarda* (1841) was acclaimed as a poem of originality, and Aleardi was highly praised; but they and the mediocre imitators of Manzoni brought Romanticism into disrepute and caused Carducci's reaction. Niccolò Tommaseo was also a patriot and a Romantic and, after Manzoni, the greatest Roman Catholic poet and essayist of his age; his personality and work had many conflicting characteristics, but some of his poems are outstanding.

Guerrazzi's novels

The works of Carducci

Giosuè Carducci was an outstanding figure whose enthusiastic support for the national cause during the struggle of 1859–61 was changed to disillusionment by the difficulties in which the new kingdom was involved. The bitterness of some of his poetry revealed frustration and rebelliousness. *Rime nuove* ("New Rhymes") and *Odi barbare* ("Barbarous Odes"), both of which appeared in the 1880s, contained the best of his poetry: memories of childhood, evocations of landscape, laments for domestic sorrows, an inspired representation of historical events, an ambitious effort to resuscitate the glory of Roman history, and an anachronistic but sincere cult of pagan civilization. He tried to adapt Latin prosody to Italian verse, which sometimes produced good poems, but his opposition to Romanticism and his rhetorical tirades provoked a strong reaction, and his metrical reform was short-lived. He was also a scholarly historian of literature, and his literary essays had permanent value, although philosophical criticism such as that of De Sanctis was uncongenial to him. Francesco De Sanctis, exiled by the Bourbons of Naples, considered literature to be an expression of social and moral conditions; and on the basis of a sound philosophical training, he inaugurated aesthetic criticism. *Saggi critici* ("Critical Essays") and particularly his *Storia della letteratura italiana* ("History of Italian Literature") were not fully appreciated in his day but later received recognition and started the philosophical revival of the later 19th century.

While Carducci was still alive, Giovanni Pascoli acquired a reputation and succeeded him in the chair of Italian literature at the University of Bologna. His art was often impressionistic and fragmentary, his language occasionally laborious: but his lyricism. at first timid in inspiration (*Myricae,* 1891), rose to fuller tones when he attempted the loftier themes of antiquity: Roman heritage and greater Italy. His original vein still found expression in *Canti di Castelvecchio* (1903; Songs of Castelvecchio") and in the classicism of *Poemi conviviali* (1904; "Convivial Poems"). In later years he produced—both in humanistic Latin and in self-consciously elaborate Italian—heroic hymns in honour of two sacred cities of Italy, Rome and Turin. But what Pascoli wrote in imitation of Carducci and D'Annunzio was inferior to poems derived from humbler inspirations.

The realist novel

Giovanni Verga stands out as by far the greatest Italian novelist of his time; he was the master of Italian Realism, and his two best novels, *I malavoglia* (1881; *The House by the Medlar Tree,* 1953) and *Mastro-don Gesualdo* (1889; "Mister-Sir Gesualdo"), which describes the dismal conditions in Sicily during the first part of the century, achieved a poignant intensity. Verga was also an exceptionally gifted writer of short stories (such as those in *Novelle rusticane*). Although his success was not immediate, it was lasting, and his influence on the development of Italian literature was paramount. Followers of his were Luigi Capuana and Federico De Roberto, who wrote a masterly novel, *I vicerè* (1844; "The Viceroys"). In order to secure a realistic reproduction of life, these writers drew their subjects from the life of their district, thus creating a tradition of "regionalist" novels. The same was true of Matilde Serao and of Renato Fucini and to some extent of Antonio Fogazzaro, too, although his characters moved in a wider circle. Fogazzaro was a deeply religious man, much influenced by Manzoni and the German Romantics. His later novels—*Il santo* (1906; "The Saint"), well known because of the criticism it provoked from the ecclesiastical authorities, and ***Leila*** (1911)—did not equal the mellowness and sentimental seductiveness of such previous works as ***Piccolo** mondo **antico*** (1895; ***The Patriot,*** 1906), generally acknowledged as his masterpiece, and ***Piccolo rnondo** moderno* (1900; ***The Man of the World,*** 1907). Fogazzaro tried to contrast the Realism of his times with a personal idealism and was more interested than the Realists in psychological problems, yet he never rose to the power of Verga. A lasting though less resounding fame was won by Emilio De Marchi, whose novels (particularly *Demetrio **Pianelli*** [1890]) revealed a thoughtful and kindhearted nature. A very different writer was Edmondo De Amicis, whose *Cuore* (1886; first English translation, 1887; ***The Heart of a Boy,*** 1960) became a classic of children's literature and whose short stories and travel books were truly representative of the mood of the masses in his time.

Finally, writers of the scapigliatura milanese (Milanese libertinism) created a movement that was strongly imbued with real Romanticism—not the "second Romanticism" (against which it rebelled)—and that also disliked bourgeois orderly life, taking its cue from the French Bohémiens. The novelist Giuseppe Rovani (*I cento **anni*** [1859–60; "The Hundred Years"]), Emilio Praga, and a poet and musician, Arrigo Boito (Giuseppe Verdi's librettist), were foremost among the scapigliati.

In literary criticism the historical method prevailed. Under the inspiration of Alessandro D'Ancona, A. Bartoli, C. de Leva, Pasquale Villari, D. Comparetti, and G.I. Ascoli much research was carried out that aimed at clarifying political and literary history.

SPANISH LITERATURE IN THE 19TH CENTURY

Spanish Romanticism

Literature in Spain in the first third of the 19th century was still affected by the Napoleonic Wars and their long aftermath. Many liberals driven into exile by Ferdinand VII after 1823 sought refuge in France; and, when they returned to Spain after his death in 1833, they had been so influenced by French Romanticism that this date has been taken as the beginning of the Romantic movement in Spain. The ground had been prepared in Cádiz from 1814 onward in a debate on literary values initiated by a German, Johann Niklaus Bohl von Faber; in Barcelona with the founding in 1823 of a review, ***El europeo;*** and in Madrid with Agustin Durán's essay on the drama of the Siglo de Oro (the Golden Age of the Spanish Renaissance) in 1828 and his ***Coleccidn de romances** antiguos* (1828–32; "Collection of Ancient Ballads"). Romanticism in Spain was, in many respects, a return to the spirit of its own earlier classics. The formal characteristics of Spanish Romantic drama—mingling of genres, rejection of the unities, metrical variety—had characterized the drama of Lope de Vega and his contemporaries, who had, moreover, treated many of its themes.

The movement arrived in Spain a generation later than elsewhere and had a short life. It never became a school or had a particular leader. José de Espronceda was the one Romantic who lived his Romanticism. His ***Estudiante de** Salamanca* ("Student of Salamanca"), ***Canciones*** ("Songs"), and an unfinished work, ***El diablo** mundo* ("The Devilish World"), were the only subjective lyrics of value that the period produced, and they marked a milestone in the development of poetic form. A play, ***Don Álvaro o la** fuerza **del sino*** ("Don Alvaro; or, The Force of Destiny"), by Angel de Saavedra, duque de Rivas, and the preface, by a critic, Antonio Alcalá Galiano, to Saavedra's narrative poem ***El** moro expósito* (1834; "The Foundling Moor") came nearest to expressing a philosophy of Romanticism.

Three poets revealed how one of Romanticism's concerns was liberation of the individual personality. Gustavo Adolfo Bécquer, in 76 *Rimas,* expressed his own tortured emotions; Rambn de Campoamor y Campoosorio wrote *Doloras, Pequeños poemas,* and ***Humoradas*** (new poetic forms of his invention), attempting to bring poetry back into the realm of ideas; and Gaspar Núñez de Arce wrote ***Gritos del** combate* (1875; "Combat Cries"), patriotic, declamatory exhortations defending democracy. A later poet, Salvador Rueda, anticipated Modernism with his innovations and influenced Juan Ramón Jiménez. He was an exuberant local colorist.

Costumbrismo was a movement that started earlier than Romanticism and flourished at the same time. It concerned realistic prose writing, often within a narrative framework, and resembled the Coverley papers of Joseph Addison and Richard Steele. Both the *cuadro **de** costumbres* and the ***articulo de costumbres*** were short literary sketches on customs, manners, or character, but the ***cuadro*** inclined to description for its own sake, whereas the ***articulo*** was more critical and satirical. ***Cartas de un pobrecito** holgazán* (1820; "Letters from a Poor Idler"), by Sebastián de Miiiano, was probably the first work of this kind, but the most important were by José de Mariano Larra, outstanding prose writer and most critical mind of his age, who dissected society pitilessly in ***Articulos*** (1835–37); by Ramón de Mesonero Romanos, whose ***Escenas** matritenses* (1836–42; "Scenes of Madrid") gave a vivid picture of contemporary life; and by Serafín Estebanez Calderón, who portrayed the manners, folklore, and history of Andalusia in ***Escenas andaluzas*** (1847; "Andalusian Sketches"). These writings, and others similar, with their realistic and regional elements helped to prepare for a revival of the novel.

Revival of the Spanish novel

For more than two centuries the novel, with which Spain had made its greatest contribution to literature, had been almost extinct. The first novels of the revival were more interesting for powers of observation and description (a continuation of *costumbrismo*) than for imaginative or narrative quality. A woman novelist, Fernán Caballero, determined the technique of observation that was to rehabilitate the novel in ***La gaviota*** (1849; ***The Seagull,*** 1867). In 1874 the great period of the regional novel began with the masterpiece ***El sombrero de** tres picos **(The Three-Cornered Hat,*** 1918), a sparkling tale of peasant malice by Pedro Antonio de Alarcbn. Although local Realism prevailed too in Juan Valera's novels, they were also remarkable for their psychological insight and notes of cultured detachment. The greatest regional writer was José Maria de Pereda, whose Realism was a powerful re-creation of nature as the abiding reality, more important than the individual. ***Sotileza*** (1884; "Subtlety") and ***Peñas arriba*** (1893; "Up the Mountains") revealed his support for traditional values of religion, family, and country life. An attempt by Emilia, condesa de Pardo Bazán in her novels of Galicia, ***Los** pazos **de** Ulloa* (1886; ***The Son of a Bondwoman,*** 1908) and ***La** madre **naturaleza*** (1887; "Mother Nature"), to combine a Spanish Christian outlook with French Naturalism made clear their incompatibility, and in later novels she returned to a genuinely Spanish Realism. Armando Palacio Valdés was the novelist of Asturias.

Besides this cult of regionalism, the fecundity in the novel of Benito Pérez Galdbs created a world in itself. After living through one of the most turbulent chapters in Spanish history, he imaginatively re-created its antecedents, beginning with the war against Napoleon. The 46 novels of his ***Episodios nacionales*** (1873–79 and 1898–1912) embraced 70 years of the country's story and taught generations of Spaniards much of what history they knew.

Post-Romantic drama

The drama, disoriented after a brief Romantic efflorescence, produced few outstanding works. Manuel Tamayo y Baus achieved fame by an isolated work, *Un **drama nuevo*** (1867; ***A New Drama,*** 1915), in which the characters are members of Shakespeare's company of actors (including Shakespeare himself). Adelardo Ldpez de Ayala pilloried bourgeois vices in ***El tejado de vidrio*** (1857; "The Glass Roof") and *Consuelo* (1870). The 1904 Nobel Prize for Literature set a seal on José Echegaray y Eizaguirre, whose 63 plays fall into two main groups. In the first, drama became melodrama with a constant striving after effect and a basic falsity of character, passion, and situation alike. The second comprised a serious, often tragic, drama of social problems. Joaquín Dicenta brought the drama nearer to earth with an unpleasantly realistic study of working class conditions in *Juan José* (performed 1895). Failure to master the problems involved in the change from novel to drama robbed Pérez Galdós of full success in the theatre, to which he turned increasingly in his desire to stimulate the social conscience: *Realidad* (1892; "Reality"), *La loca de la casa* (1893; "The Madwoman in the House"), *Electra* (1901).

Catalan. In 1814 appeared the *Gramdtica y apologia de la llengua cathalana* ("Grammar and Apology of the Catalan Language") of Josep Pau Ballot i Torres, a forerunner of the literary and linguistic renaissance (Renaixensa) that marked the Romantic period in Catalonia. The pioneers, however. saw the inadequacies of the ancient language for the expression of spiritual and intellectual ideas. The Institut d'Estudis Catalans, founded in Barcelona in 1907, has played a notable part in the deliberate enrichment and purifying of Catalan as a vehicle for contemporary thought.

Bonaventura Carles Aribau's patriotic *Oda a la pátria* (1832; "Ode to the Fatherland") and the poems of Joaquim Rubió i Ors and Victor Balaguer prepared the way for the mysticism of Jacintó Verdaguer Santaló, a great epic poet (*L'Atlántida* [1877], *Canigó* [1886]). Miguel Costa i Llobera cultivated a classical perfection of form. In Joan Maragall i Gorina, Catalonia found its first great modern poet who, in spiritual quality (as also, if less happily, in his theory that inspiration absolved the artist from the need for formal perfection), exerted a powerful influence on later poets.

The foundations of modern Catalan prose were laid by the critical writings of Rubió i Ors, Francisco Pi i Margall, one of the four presidents of the Spanish Republic of 1873, and Josep Torras i Bages (*La tradicid catalana,* 1892). One of the best and most influential writers in prose was the essayist Eugeni d'Ors (pseudonym "Xenius"), whose philosophical novel *La ben plantada* (1911; "Firmly Rooted") was one of the most notable works in modern Catalan literature.

Catalan dramatists have produced plays of considerable originality. Angel Guimerb achieved a European reputation with *Terra baixa* (1896; *Martha of the Lowlands,* 1914). The many social dramas of Ignasi Iglésias, inspired by Gerhart Hauptmann, a 19th-century German dramatist, included one near-masterpiece, *Els vells* (performed 1903; "The Old Men").

The suppression of languages other than Castilian by the dictator (1923–29) Miguel Primo de Rivera cast responsibility for the literary survival of Catalan largely onto expatriate intellectuals in Mexico City and Santiago, Chile.

The Galician revival

Galician. The Romantic movement, like the Peninsular War, revived local feeling and interest in things Galician, but not in the language. The *Xogos Froraes* (Floral Games, or poetry congresses; an equivalent of Catalan and Provençal *Jocs Florals*) of 1861, with the first dictionary (1863) and first grammar (1864) of Galician, marked a change. Francisco Añón y Paz was the first notable poet in the resurrected idiom, his most stirring notes being love of country and of freedom. Rosalia de Castro, the greatest name in Galician literature, identified herself with the spirit and people of the Galician countryside in *Cantares gallegos* (1863; "Galician Songs"); her *Follas novas* (1880; "New Leaves"), introspective to the verge of despair, reflected deep personal sorrows. Eduardo Pondal y Abente, a bard of a dimly sensed heroic past, was concerned with nature and Celtic mythology. Valentin Lamas Carvajal has been remembered as the voice of the peasant.

Prose showed no comparable achievement. Aurelio Ribalta, Manuel Lugris Freire, and Heraclio Pérez Placer wrote short stories but were overshadowed by novelists of stature—Emilia Pardo Brazbn and Rosalia de Castro—who chose to write for a larger public in Castilian.

PORTUGUESE LITERATURE IN THE 19TH CENTURY

The 19th century witnessed a general revival of Portuguese letters. The chief exponents of Romanticism were, in poetry and drama, João Baptista de Almeida Garrett and, in prose, Alexandre Herculano; both lived for some years in exile, the price of their political liberalism. Almeida Garrett read literature in English and French and introduced Portugal to nationalistic Romanticism through two epics, *Camões* (1825) and *Dona Branca* (1826). Antonio Feliciano de Castilho, half-Romantic, half-Classicist, exercised much influence over a younger generation of poets; João de Lemos, Soares de Passos, Tomás Ribeiro (who won fame with an ardently patriotic *Dom Jaime,* 1862). In 1865 Antero de Quental, a student of German philosophy and poetry, and Teófilo Braga, disciple of Auguste Comte, led a revolt against the primacy of Castilho, much to the benefit of literature, and of poetry in particular. The *Campo de Flores* (1893; "Field of Flowers") of João de Deus contained some of the finest short poems in the language, of a spontaneous simplicity. Abilio Manuel Guerra Junqueiro, heir to Victor Hugo, was a would-be social revolutionary overprone to grandiloquence. In *Os Simples* (1892) he turned to the portrayal of peasant life, and this work constituted his finest poetry. Akin to him was António Duarte Gomes Leal (*Claridadesdo Sul* [1875] and *O Anti-Cristo* [1886]), who could likewise achieve quiet sincerity on humble themes. Antonio Cândido Gonçalves Crespo stood out as the first of his country's Parnassians. By contrast Cesbrio Verde addressed himself to the poetic essence of common realities. The *Só* (1892; "Alone") of António Nobre was intensely Portuguese in themes, mood. and rhythms; he and Teixeira de Pascoais developed a cult of *saudosismo* ("yearning," "nostalgia") that inspired a whole school of poets. French Symbolism found an enthusiastic adept in Eugénio de Castro.

Portuguese drama

Almeida Garrett, seeking to reinvigorate drama. found he had to create alike theatre, plays, actors, and audience. In *Urn Auto de Gil Vicente* (1838), *O Alfageme de Santarkm* (1841; "The Swordsmith of Santarém), and especially in *Frei Luis de Sousa* (1843; *Brother Luiz de Sousa,* 1909), he proved himself, after Vicente, his country's most notable dramatist. António José Enes dramatized topical questions in a spirit of combative liberalism; Ernesto Biester wrote social drama; and Fernando Caldeira produced comedy. João da Câmara was the outstanding dramatist of his day, and his works included *Alfonso VI* (1890), *Rosa Enjeitada,* and *Os Velhos* (1893). Other historical playwrights included Henrique Lopes de Mendonça, Marcelino Mesquita, and the gifted Júlio Dantas (*A Severa* [performed 1901], *Santa Inquisição* [1910], *A Ceia dos Cardeais* [1902], *Rosas de Todo o Ano*). Herculano, returning from exile with an enthusiasm for Sir Walter Scott, launched the historical romance with *O Monasticon* (1844) and *Lendas e Narrativas* (1851). Many followed suit, including Oliveira Marreca, Arnaldo da Gama, and Pinheiro Chagas, popular successes being *A Mocidade de D. João V* (1852), by Rebelo da Silva, and João de Andrade Corvo's *Urn Ano na Côrte.* This was the great age of the novel: Camilo Castelo Branco, J.G. Gomes Coelho (better known as Júlio Dinis), and especially José Maria Eça de Queirós were names that would stand high in any country. The first was a master of the language and of dramatic, or melodramatic, plot; Julio Dinis depicted country life, as in *As Pupilas do Senhor Reitor* (1867); while Eça de Queirós introduced Realism with a powerful novel, *O Crime do Padre Amaro* (1876). Naturalism claimed Júlio Lourenço Pinto, Luis de Magalhães, and the much superior Francisco Teixeira de Queirós, who sought to expose contemporary society in his two series, *Comédia do Campo* and *Comédia Burguesa.*

With his magnum opus the *Histdria de Portugal* (1846–53) and the *Da Origem e Estabelecimento da Inquisição em Portugal* (1854–59), Herculano established himself as a leader of modern Peninsular historians. Historiography flourished with the Visconde de Santarém, historian of the Cortes, José Simão da Luz Soriano of constitutionalism, Luis António Rebelo da Silva of the period of Spanish rule under the Philips, and José Maria Latino Coelho of the dictatorship of Pombal. Henrique da Gama Barros and António da Costa Lôbo followed in the footstep of Herculano. The works of Joaquim Pedro de Oliveira Martins gave proof of psychological imagination, a notable capacity for general ideas, and a gift of picturesque narration. He left in his numerous writings a vast portrait gallery of great figures of his country. In the 20th century the trend toward scientific documentation and objectivity was accentuated. The monumental ***Histdria de Portugal*** edited by Damião Peres provides an exhaustive treatment of the entire field.

LATIN-AMERICAN LITERATURE IN THE 19TH CENTURY

Literature of rebellion. Political unrest spread in Latin America as the 18th century advanced, and French ideas broke through the relaxed controls held by Spain and Portugal over New World thought. Early stirrings became more purposeful when French Revolutionary doctrines were tested by the action of Britain's colonies to the north. Printing presses and periodicals sprang up everywhere, and ideals quickened in literary societies founded by Latin America's young liberals returning from abroad.

One of the earliest fruits of these contacts with foreign thought was an abortive Brazilian Inconfidência Mineira of 1789, a "conspiracy of poets" headed by Joaquim José da Silva Xavier and supported by a number of exceptional writers who were members of the Minas school of epic and Neoclassical poets, which had no equal in the Spanish colonies of that day. José Basilio da Gama's *Uraguai* (1769) and José de Santa Rita Durão's ***Caramuru*** (1781) were two great native epics of Brazil. Love of country and an appearance of the Indian as a literary character stamped both works as forerunners of Brazilian intellectual independence. Tomás António Gonzaga, author of love songs to ***Marília de Dirceu*** (1792), has remained the most popular love poet in the Portuguese language. When, after a bloodless victory in 1822, Brazil finally emerged as an empire, only one literary figure, José Bonifácio de Andrada e Silva, stood out as a patriarch of Brazil's struggle for independence. As author of such vigorous, passionate verse as ***Poesias*** (1825), he revealed his affinity to Byron and has been acclaimed Brazil's first Romantic poet.

A Venezuelan, Francisco de Miranda, left a remarkable journal that revealed how he profited from contacts in the United States with the "great American experiment." His compatriot Simón Bolivar was later christened "the thinker of the Revolution." Through his prophetic analyses of the sociopolitical scene filtered some of the best of French thought, particularly that of Montesquieu and Rousseau. Bolivar's dream of a confederation of Latin-American states received support from a fiery Argentine pamphleteer, Bernardo Monteagudo. But it was the Mexican José Joaquin Fernández de Lizardi, referred to as ***el pensador mexicano*** ("the Mexican thinker"), after a revolutionary journal founded by him in 1812, who was the only liberator remembered as a man of letters. His fame has rested largely on a picaresque tale, ***El periquillo sarniento*** (1816; ***The Itching Parrot,*** 1942), conceded to be the first Latin-American novel.

Patriotic poetry

The Revolution found popular expression in balladry and heroic verse. More enduring patriotic poetry came with victory and represented but a phase in the work of three outstanding poets whose significance transcended the struggle for independence. Of the three, only an Ecuadorean, José Joaquin Olmedo, limited his work almost exclusively to the themes and spirit of Revolutionary years; his best known poem, "La victoria de Junín, canto a Bolivar" (1825), has been generally considered to be the finest example of heroic poetry in the classical style written in Latin America. In his Virgilian ***Silva a la agricultura de la zona tórrida*** (1826), Andrés Bello of Venezuela exhorted his fellow Latin Americans to turn their swords into plowshares and live the life that rich, tropical America offered freemen; in later years Bello distinguished himself as a grammarian, critic, translator, lawgiver, and educator and as the father of Chile's first intellectual generation.

José Maria de Heredia of Cuba, Latin America's first great lyric poet and Romanticist, lived most of his life in exile (Venezuela, New York, Mexico), where nature's grander moods inspired his widely acclaimed poems ***En el teocalli de Cholula*** (1820 and ***Oda al Niágara*** (1824). He participated in the new literary movement of Romanticism with which the young Latin-American nations were born to political independence.

The Romanticists. Political independence from Spain and Portugal did not bring freedom from the despotism of political anarchy in the early formative years. Economic and political stability for most new nations came in the closing years of the century. American themes had fired the imagination of the "liberators," but Neoclassic forms were still held sacred and untouchable. European Romanticism pointed the way to cultural independence also, even though that way lay largely along a route marked out by French, English, and Spanish writers. Neoclassic form and restraint yielded to Romantic freedom, individualism, and emotional intensity, tenets suited to Latin-American temperaments and conditions. The most illustrious early Romanticists were Argentine political refugees who fled from Juan Manuel de Rosas. Their leader was Esteban Echeverria, who, after a stay in France (1826–30) when Romanticism was at its height, indicated the way by writing "La cautiva" (1837), his best poem and the earliest fusion of native themes and scenes with new, more appropriate verse forms. It remained for Domingo Faustino Sarmiento to present, in *Facundo* (1845), written in passionate denunciation of Rosas, a first serious study of the great plain, or Pampas, and of gaucho (cowboy) lore. This emphasis on the national scene gave birth to an indigenous literary genre without European prototype, the gaucho literature of Argentina and Uruguay. The gaucho had long been the subject of folktale and ballad and became a subject of some of the best verse, as in Rafael Obligado's poem (1887) on a legendary minstrel, Santos Vega, and a humorous gem, *Fausto* (1866; ***Faust,*** 1943), by Estanislao del Campo, until finally he received epic treatment in ***Martin Fierro*** (1872–79; ***The Gaucho, Martin Fierro,*** 1936), by José Hernández.

The search for national roots in Latin-American writing

These "primitive" poems in imitation of gaucho minstrels' ballads and speech were one of two of Latin America's original contributions to Western letters; the other sprang from Romanticists' interest in a search for national roots. This romantic evocation reached its poetic climax in the elegiac *Tabaré* (1886; Eng. trans., 1956), by an Uruguayan, Juan Zorrilla de San Martin, which related the fate of the aboriginal Charrúas, vanquished by Spanish invaders. In prose, this same past produced the *tradición*, the other unique Latin-American genre patterned after the historical anecdote and perfected by a Peruvian master of humorous prose, Ricardo Palma, whose ***Tradiciones peruanas*** appeared between 1872 and 1910 ***(The Knights of the Cape,*** 1945). The Brazilian Romantics had no great past and therefore extolled the beauties of their tropical homeland and the simple living of idealized Indians with languid nostalgia. Domingos José Gonçalves de Magalhães' poems, *Suspiros* ***Pokticos e Saudades*** (1836; "Poetic Sighs and Longings"), have been held to have launched the Romantic movement, earlier phases of which found their best and most representative expression, however, in inspired and richly variegated writings by Antônio Gonçalvez Dias, one of Latin America's truly outstanding poets of the Romantic period. Later phases were best exemplified in a poetry of doubt and despair by Manuel Antônio Alvares de Azevedo of the Satanic school and author of ***A Noite na Taverna*** (c. 1851) and in the sociopolitical verse of Antônio de Castro Alves, author of ***Os Escravos*** (1876).

The Realists and the Naturalists. The Romanticist's interest in the picturesque and unusual helped him discover evidence of a budding national way of life. The sketch of contemporary manners developed into a realistic novel of manners shortly after midcentury. From that date, the novel assumed a more commanding role in Latin-American letters but appeared almost concurrently in the several types representative of successive literary trends in

Europe whose masters provided the molds into which Latin-American themes were poured. José Marmol gave the outside world a dramatic depiction of life in Argentina under the tyranny of Rosas in a Romantic novel, *Amalia* (1851–55; Eng. trans., 1919). Across the Andes, a Realist, Alberto Blest Gana, began producing a series of ***costumbrista*** novels on Chilean life, of which ***Martin Rivas*** (1862; Eng. trans., 1918) was considered his best, while the Colombian Jorge Isaacs wrote the tragic idyll ***Maria*** (1867) and Juan León Mera contributed *Cumandá* (1879) to a growing number of highly Romantic fictional portrayals of idealized Indian life.

The true novel appeared first in Brazil in 1844 when ***A Moreninha*** ("The Little Brunette"), by Joaquim Manuel de Macedo, was published. After that, the form enjoyed a more orderly and continuously progressive growth in Brazil. Still one of the most widely read of his country's novelists, José Martiniano de Alencar initiated a vogue of the Brazilian *Indianista* novel with ***O Guarani*** (1857; "The Guarani Indian") and ***lracema*** (1865; Eng. trans., 1886). But these romantic tales of love between Indian and white represented only one aspect of Alencar's varied literary activity; he also turned to the hinterland to depict life and customs of Brazil's backlands, and *O Gaúcho* (1870) and **O *Sertanejo*** (1876; "The Man of the Backlands"), although still markedly Romantic in spirit, were among the forerunners of a flourishing regional genre that foreshadowed the truly national novel. Two other contributors to this transitional genre were Alfredo d'Escragnolle Taunay, whose *Inocência* (1872; Eng. trans., 1945) became a universal favourite, and Bernardo Guimarães, whose abolitionist ***Escrava Isaura*** (1875) was a decisive step forward in the direction of the novel of social protest.

True Realism

True Realism—with a definite leaning toward Naturalism—was initiated by ***Memdrias de um Sargento de Milícias*** (1854), by Manuel Antônio de Almeida. But it was not until the mid-1870s that the novel began to expose cankers of social and psychological maladjustments to a rapidly changing economic scene. Aluizio Azevedo, an early example of social protest in the manner of 20th-century novelists, wrote such favourites as O *Mulato* (1881) and ***O Cortiço*** (1890; ***A Brazilian Tenement,*** 1926). Less occupied with external aspects of Brazilian life, Joaquim Maria Machado de Assís pried into the psychological complex of the Brazilian and distinguished himself as his country's most original and gifted writer. His trilogy, ***Bras Cubas*** (1881; ***Epitaph of a Small Winner,*** 1952), ***Quincas Borba*** (1891; ***Philosopher or Dog?,*** 1954), and ***Dom Casmurro*** (1900; Eng. trans., 1953) was a landmark in Latin-American letters. Two other novelists, Júlio Ribeiro and Raúl de Avila Pompéia, also helped set the Brazilian novel on a course not followed elsewhere until almost a generation later.

After midcentury, several late Romantic political writers distinguished themselves in essay form. Among these were Juan Montalvo, Eugenio Maria de Hostos, Joaquim Nabuco, and Rui Barbosa. An ironic experimental poet and essayist, Manuel González Prada was the chief figure and his verse paved the way for the new poetry of a coming generation of rebels.

SCANDINAVIAN LITERATURE IN THE 19TH CENTURY

Swedish. Political changes in Sweden up to 1804 meant that ardent nationalism emerged as a characteristic of Swedish Romanticism. The idealism at the core of this movement was laid by the Kantian teaching of Benjamin Hoijer and the impact of Schiller, Goethe, and the German Romantics on Swedish literature. Student societies and their periodicals, such as ***Polyfem*** (1809–12) and ***Phosphorus*** (1810–13), led the attack on the traditional school. Most gifted of the Forforister, or Phosphorists, Per Daniel Atterbom wrote a verse "Prolog" (1810) to ***Phosphorus*** revealing both talent and commitment to Romanticism. Vilhelm Palmblad, publisher and prose writer, led the polemics at Uppsala; Lorenzo Hammarskold, with Johan Askelöf and Clas Livijn, advanced the same views at Stockholm; while Johan Valerius, P.A. Wallmark, and Johan Wallin championed the traditionalists.

Meanwhile, another society, Gotiska Forbundet (Gothic Society), advocated, from its start in 1811, that study of the "Gothic" past could morally improve society. One of its members, Esaias Tegnér, wrote a most popular poem, ***Frithiofs saga*** (1825), based on an Old Norse theme.

The return to the literary past

Tegnér, like E.G. Geijer, valued old Northern mythology for the patterns he discerned in it—patterns also found in Greek mythology and Romantic metaphysics, in which religion, philosophy, and poetry appeared to be one and the same. Nevertheless, Tegnér's ideals of clarity of thought and formal perfection led him sometimes to side with traditionalists in their struggle against obscurities and formal innovations.

Several leading Romantics were learned men whose poetry strove to embody a philosophical system or an interpretation of history. The most ambitious attempt of this kind was Atterbom's ***Lycksalighetens*** o (1824–27: "The Isle of Bliss"), an allegory dealing with adventures of a legendary king, Astolf, and a history of poetry as an illustration of man's alienation from the divine. The greatest poet was perhaps Erik Johan Stagnelius, who held aloof from schools and coteries. The recurrent theme in his *Liljor i Saron* (1821; "Lilies of Sharon") was the lament of the human soul, imprisoned in a world of darkness and sin. Among noteworthy minor Romantic writers were Adolph Torneros, Per Elgstrom, Samuel Hedborn, and Arvid August Afzelius, collector of folk songs.

In prose the most complex personality among the later Romantics was a novelist, Carl Jonas Love Almqvist, who combined an extravagant imagination with realism. A master of prose style, he was at his best in the long short story, in which he foreshadowed Strindberg's method of raising problems for debate. The novel was really established by Fredrika Bremer, author of *Grannarna* (1837; "The Neighbours"), whose "sketches from ordinary life" appeared from 1828. Sophie von Knorring wrote principally about aristocratic families, and popular Emilie Flygare-Carlén produced many stories dealing with west-coast life, including ***Rosen på*** *Tistelön* (1842; ***The Rose of Tistelon,*** 1844–50).

Realism made only slow headway, in spite of the example of a great Finnish–Swedish poet, Johan Ludvig Runeberg. Literature of the 1840s and 1850s was mainly an aftermath of Romanticism: though a critic, Bernhard Elis Malmstrom, was ostensibly an opponent of the movement, even his work showed Romantic influence. His elegy "Angelika" (1840) had affinities with the poetry of a contemporary, Carl Vilhelm Bottiger, and older Karl August Nicander. A movement known as Scandinavism produced a good deal of verse: Carl Vilhelm August Strandberg (pseudonym "Talis Qualis"), fieriest poet of this type, later made excellent translations from Byron. Popular reading was provided by August Blanche in ***Bilder ur verkligheten*** ("Pictures of Real Life"), successful short stories depicting Stockholm life with humour and vivacity, while Frans Hedberg wrote pompous historical plays.

Poetic Realism became an official program of the "pseudonym poets" of the 1860s, including Carl David of Wirsén, Edvard Backstrom, Pontus Wikner, and Carl Snoilsky. Only Snoilsky had the temperament and poetic gift needed to carry out the program. Wirsén, on the other hand, as secretary of the Swedish Academy, launched formidable opposition against innovators; and Viktor Rydberg fell between idealism and Naturalism, for, as novelist, poet, and critic, he began as a radical journalist and ended as a professor and author of philosophical poems. His important early work consisted of an ideological novel, ***Den siste athenaren*** (1859; ***The Last Athenian,*** 1869), and a treatise, ***Bibelns lära om Kristus*** (1862; "The Teaching of the Bible About Christ"), which prepared the way for scientific Rationalism.

The sources of modern Swedish literature

Four influences combined to free Swedish literature from petrifying conventions: the English writings of Charles Darwin, Herbert Spencer, and J.S. Mill; the French Naturalism of Émile Zola; the drama of the Norwegians Henrik Ibsen and Bjørnstjerne Bjemson; and the criticism of Georg Brandes. The modern literature growing out of this was first and best represented in the work of August Strindberg, Sweden's greatest writer. Modern drama has dated from his play Muster Olof (published 1872), and the modem novel from Roda *rummet* (1879; The Red Room, 1913). Strindberg overshadowed all the writers of the 1880s, including Gustaf af Geijerstam, author of Erik *Grane* (1885), Anne Charlotte Edgren-Leffler, and the gifted Victoria Benedictsson, who both wrote about the adverse position of women in society. Benedictsson's stories, such as *Från Skåne* (1884; "From Skåne"), revealed the regional character of new prose literature. Regional poetry was written by Albert Bååth and Ola Hansson, both of Skåne.

In 1888 Verner von Heidenstam began the reaction against Utilitarianism and Naturalism with a volume of verse, *Vallfart* och *vandringsår* ("Pilgrimage and Wander Years"). His later poetry and historical tales won him the Nobel Prize for Literature in 1916. Oscar Levertin, stimulated by Heidenstam's example, wrote poetry full of colour and lore of the past and as a critic was influential in molding contemporary taste. Gustaf Fröding was also influenced by Heidenstam, and his verse constantly mingles the melancholy and gay. Despite the threat of madness, he published five collections of poems before his reason gave way in 1898, by which time he had established himself as one of the greatest Swedish lyric poets. Regionalism entered Neoromantic poetry with Fröding, who was from Varmland, and with the work of Erik Axel Karlfeldt the province of Dalarna came into its own. Karlfeldt's mature poetry won him the Nobel Prize in 1931.

Meanwhile, Selma Lagerlof, the first Swede to win a Nobel Prize for Literature (1909), had developed the prose tale; her long series of novels and short stories, beginning with Gosta Berlings saga (1891), reached an international public through translation. Per Hallstrom was a more skillful writer of short stories than of novels. Romantic, too, in his love for the skerries (rocky isles) was Albert Engstrom, a great humorist and one of the most beloved of Swedish writers.

Finnish literature in Swedish. Only in the 19th century did a strong literature in Swedish develop in Finland. Previously it had been somewhat sporadic, although there had been several excellent poets, notably Jacob Frese, Gustaf Philip Creutz, and Frans Mikael Franzén. In the work of Johan Ludvig Runeberg, the Finnish people and landscape first came to life in literature, and he became the national poet. Zacharias Topelius wrote historical novels in Scott's manner, as *Fältskärns* berattelser (1853–67; The Surgeon's Stories, 1872), poems, and children's stories. Johan Vilhelm Snellman, a disciple of the German philosopher Hegel, developed ideas that emphasized the importance of a local and national culture: he decisively influenced the status of Finnish, which was given parity with Swedish as an official national language in 1863.

Toward the end of the 19th century, Swedo-Finnish literature began to draw away from literature in Finnish. Important writers of the transition were Josef Julius Wecksell and Karl August Tavaststjerna. Early in the 20th century a school of prose writers known as Dagdrivarna (Idlers or Saunterers) emerged with a crisp, skeptical, analytical tone. Important among them was Runar Schildt, whose stories and plays dealt with ethical and artistic problems. Among poets, Arvid Morne and Bertel Gripenberg may be mentioned.

Norwegian. After 1814 a new, exciting, and difficult age began for Norway: an opportunity seemed to be offered to develop an independent Norwegian culture and way of life, but there were deep differences of opinion as to how this could best be achieved. A poet and critic, Johan Sebastien Welhaven was chief representative of those who insisted that the existing Danish element in the culture should not be neglected. Henrik Wergeland was a spokesman for those whose nationalistic pride led them, on the other hand, to demand a complete break with Denmark. Welhaven stood for a coolly intellectual approach, for restraint and control, and for conscious artistry, as his own sonnet cycle Norges daemring (1834; "The Dawn of Norway") exemplifies. Wergeland was more passionate and revolutionary, and his enormous epic, *Skabelsen*, mennesket og messias (1830; "Creation, Humanity and Messiah"), was a youthful but typical example of the spirit he admired.

Wergeland dominated the age as poet, orator, and social reformer, and the clash between him and Welhaven and between the two factions associated with them—the "patriots" and the "intelligentsia"—began an ideological conflict that has continued to persist in modified forms.

The ideological conflict between Wergeland and Welhaven

The literature of the mid-19th century, known as Norway's "national Romanticism," continued to reflect the country's larger aspirations. The compilation and publication, between 1841 and 1844, of Norske folkeeventyr ("Norwegian Folk Tales"), by Peter Christen Asbjørnsen and Jørgen Engebretsen Moe, and the 1853 collection by Magnus Brostrup Landstad, Norske folkeviser ("Norwegian Folk Ballads"), indicated a lively interest in the past, as did Peter Andreas Munch's eight-volume history of the Norwegian people (1857–63). Ivar Aasen was the creative spirit behind the Landsmål movement to establish a literary language based on rural dialects linked with Old Norse. Many publications of these years—including earlier works of Ibsen and Bjørnson—turned consciously to Norway's heroic past and its peasants. To these years belonged also the lyric poetry of Aasmund Olafsson Vinje, founder of a periodical *Døien*, who adopted Nynorsk (New Norwegian) as his literary language.

In 1855 Camilla Collett, Wergeland's sister, published Amtmandens *døttre* ("The Governor's Daughters"), which, by considering the place of women in society, marked a beginning of a trend that, encouraged by the immensely influential Danish critic, Georg Brandes, culminated in the 1870s and the 1880s in the realistic "problem" literature of Ibsen, Bjnmson, and their contemporaries. *Samfundets støtter* (Pillars of Society), in 1877, was first of a succession of problem dramas by Ibsen to win him worldwide fame. By then he had already written two verse dramas, Brand (published 1866) and Peer Gynt (1867), and his long "double drama" *Kejser* og Galilaeer (1873; The Emperor and the Galilean, 1876). The first substantial drama of this type, by Bjørnstjerne Bjørnson, En *fallit* (The Bankrupt, 1914), appeared in 1875. Although never the world figure that Ibsen became, Bjørnson was a leading personality of his age in Norway, as novelist, dramatist, and lyric poet and in public affairs.

The novelists Jonas Lie and Alexander Kielland, together with Ibsen and Bjørnson, were the major figures of modern Norwegian literature and were responsible for a remarkably large body of important work between 1870 and 1884, as the following translated titles illustrate: Ibsen's works Et *dukkehjem* (A Doll's House), Gengangere (Ghosts), En *folkefiende* (An Enemy of the People), and Vildanden (The Wild Duck); Bjørnson's dramas Det ny system (The New System), En handske (*A* Gauntlet) and Over aeve (Beyond Our Power) and his novel Det *flager* i *byen* og *på* havnen (The Heritage of the *Kurts*); Lie's novels Gaa Paa! (Go Ahead!), Livsslaven (One of Life's Slaves), and *Familjen* paa Gilje (The Family at Gilje); and Kielland's Skipper Worse, Gift ("Poison"), and *Fortuna.* The foremost stylist of his age, Kielland was an elegant, witty novelist with a strong social conscience and an active reforming zeal stemming from an admiration for John Stuart Mill.

The literature of the 1870s emphasized individual development and expression in keeping with the optimistic attitude of the times to social change and improvement; in the following decade, growing skepticism and disillusionment made writers more bitter in their attacks on "established social institutions. The publication of Fra *Kristiania-Bohêmen* ("From the Christiania Bohemia") in 1885 by Hans Henrik Jaeger created, by its seeming advocacy of sexual license, a public scandal. The most extreme exponent of Naturalism was Amalie Skram, especially in a four-volume novel, *Hellemyrsfolket* (1887–98; "The People of Hellemyr"). Arne Garborg, poet, novelist, dramatist, and critic, was a much superior writer whose work reflected successive movements of Romanticism, Realism, Naturalism, and Neoromanticism. His wider reputation was first established with a novel, *Bondestudentar* ("Peasant Students"), in 1883, but perhaps his greatest achievement was the poem cycle *Haugtussa* (1895).

Danish. *The Romantic period.* The Romantic movement came to Denmark from Germany, inspired partly by the German Jena Romantics, partly by the Neoclassicism of Goethe and Schiller. F.W. Schelling's philosophy was interpreted in Denmark by the Norwegian Henrik Steffens, but the leading Danish Romantics gave it a form very different from the original. A.W. Schack von Staffeldt was hardly recognized in his lifetime, and his metaphysical poetry lacked popular appeal. The leader of the Romantic movement in Denmark was Adam Oehlenschlager, whose unparalleled versatility in poetry, drama, and prose showed the influence of German Romanticism, especially of the poets Goethe and Schiller. His plays *Sanct Hansaften-Spil* ("Play for Midsummer Eve") and *Aladdin; Hakon Jarl*, one of his many Northern tragedies; and a cycle of dramatic poems, *Helge* (1814), were outstanding. The popular and historical songs and hymns of the poet N.F.S. Grundtvig, as well as his personal poetry, have given him a lasting place in Danish literature. Sharing the Romantic enthusiasm for antiquities of Scandinavia, he translated the 13th-century historians Saxo Grammaticus and Snorri Sturluson and translated *Beowulf* even before it had appeared in English. His *Håndbog i verdens-historien* ("Handbook of World History") was a mixture of scholarship, prophetic visions, and insanity. Bernhard Severin Ingemann wrote historical novels and a poetic cycle, *Holger Danske*, around the themes of chivalry and nationalism as well as his unsophisticated *Morgen og aftensange* ("Morning and Evening Songs"). Johannes Carsten Hauch wrote tragic and philosophic dramas, novels, and contemplative poetry.

Romantic realism. New elements of reason and realism appeared after the first quarter of the century in the works of Poul Møller, who wrote the first Danish novel on contemporary events, *En dansk students eventyr* (1824; "The Adventures of a Danish Student"), and dramatic poems and fables, sometimes showing personal disillusionment, and Steen Steensen Blicher, who, in *Traekfuglene* ("The Birds of Passage"), interpreted human nature with sad resignation. Some of his best poems were in the Jutland dialect. His many *noveller*, or short stories. beginning in 1824 with the masterly "En landsbydegns dagbog" ("The Journal of a Parish Clerk"), struck notes varying from sorrow and resignation to humour and irony.

Minor writers of the same period were Thomasine Gyllembourg-Ehrensvard, whose novel *En hverdagshistorie* (1828; "A Story of Everyday Life") was much admired; Andreas de Saint-Aubin, who wrote novels under the nom de plume of "Carl Bernhard"; and Carl Bagger, whose novel *Min broders levned* (1835; "My Brother's Life") shocked the literary world by its bold realism.

Poetic realism. About 1830, early Romanticism gave way to a less naive poetic realism, more contemplative and more concerned with form than with content. Johan Ludvig Heiberg, who led this movement, attempted to revivify Danish drama by importing French vaudeville, and in his serious romantic plays *Elverhøj* ("The Elfinhill") and *Syvsoverdag* ("Day of the Seven Sleepers") he juxtaposed poetic and pedestrian reality. His finest achievement was a verse comedy, *En sjael efter døden* (1841; "A Soul After Death"). He was the leading literary critic of his time, profoundly influenced by the philosophy of Hegel. Henrik Hertz also regarded the perfection of poetic form as more important than its content, as was clearly expressed in *Gjenganger-breve* (1830; "Letters of a Ghost"). He also wrote comedies and serious Romantic plays, including *Kong Renés datter* (1845; *King Rene's Daughter*, 1850).

The revival of lyrical poetry

A revival of lyrical poetry took place in the 1830s and 1840s, led by poets concerned with the aesthetic treatment of love and nature. Christian Winther, most famous for a long verse novel, *Hjortens flugt* (1885; "The Flight of the Stag"), sang the praises of his native island, Zealand, and of woman. Ludvig Bødtcher wrote delicate and sensitive poetry, some of which was inspired by the Italian scene. Emil Aarestrup is Denmark's erotic poet par excellence. Frederik Paludan-Miiller became an uncompromising moralist; *Adarn Homo* (1841–48), a poetic epic, was a bitter contemporary satire. Hans Christian Andersen was most important for his fairy tales, most of whose plots were his own invention, though he also wrote novels, plays, travel books, and poems. His own story is told in *Mit livs eventyr* (1855; *The Fairy Tale of My Life*, 1954).

Søren Kierkegaard holds a position entirely isolated in Danish literature. His highly personal religious philosophy was expressed in such works as *Enten Eller* (1843; *Either/Or: A Fragment of Life*, 1944) and *Stadier paa livets vei* (1845; *Stages on Life's Way*, 1940). He spent his last years in a violent and passionate attack on "official Christianity."

Meïr Aron Goldschmidt edited a rebellious, anti-royalist weekly, *Corsaren* ("The Corsair"), while many of his novels and short stories were concerned with Jewish life in the Danish community. The 1850s and 1860s produced hardly any new Danish writers of importance: the most original was Hans Egede Schack, whose novel *Phantasterne* (1857; "The Daydreamers") revealed great psychological gifts.

Icelandic. The literary and linguistic renaissance in Iceland at the start of the 19th century was fostered by three men in particular: a philologist, Hallgrímur Scheving; a poet and lexicographer, Sveinbjörn Egilsson; and a philosopher and mathematician, Bjorn Gunnlaugsson. The principal movement in this renaissance was Romanticism, introduced to Scandinavia by a German philosopher, Henrik Steffens, in his Copenhagen lectures of 1802–04. Inspired by Steffens, Bjarni Thorarensen produced nationalistic poetry that became a model for 19th-century lyrical poetry. Jónas Hallgrímsson, however. surpassed Thorarensen as a metrist. He was one of four involved in the periodical *Fjölnir* ("The Many-Sided"), which aimed to revolutionize literary theory and practice. The Fjolnismenn were anti-traditional and rejected the use of rhymes. Among those they criticized were Sigurdur Breidfjord and a satirist, Bólu-Hjálmar, both working class poets.

The group was replaced in the 1840s by another group of poets, of whom the most outstanding were Benedikt Grondal, Steingrimur Thorsteinsson, and Matthías Jochumsson. Grondal wrote powerful lyric poetry, two prose fantasies, and an autobiography, *Daegradvol* (1923; "Day-Spending"). Thorsteinsson wrote nature poetry and satirical epigrams but is best remembered as translator of *King Lear* (1878) and *A Thousand* and *One Nights* (1857–64). Jochumsson's *Hallgrímur Pétursson* (1874) and hymn *Fadir andanna* ("Father of Spirits") established him as the greatest lyric poet of the three. He, too, translated Shakespeare in addition to Ibsen's *Brand*. A poet, Grimur Thomsen, was contemporary with but distinct from this group; his poetry was less lyrical, more austere and rugged, as *Hemings flokkur Áslákssonar* (1885; "The Story of Heming Aslakssonar") exemplified.

The latter part of the century produced three talented poets: Thorsteinn Erlingsson, author of *Aldaslagur* (1911; "Sound of the Ages"); Einar Benediktsson, a Neoromantic; and Stephan G. Stephansson, an embittered expatriate whose irony passed in Iceland for Realism.

The 19th century also saw a renaissance in imaginative prose. Jón Thoroddsen wrote two novels that have acquired a position not incommensurate with that of the medieval sagas: *Piltur og Stúlka* (1850; *Lad and Lass*, 1890) and the incomplete *Madur og kona* (1876; "Man and Woman"), distinguished in prose style, narrative skill, and witty and perceptive observation of peasant and small-town life.

RUSSIAN LITERATURE IN THE 19TH CENTURY

Poetry. The greatest Russian poet of the early 19th century was Aleksandr Pushkin. Although his early poems displayed surpassing mastery of diction and verse, The simplicity and profundity of Pushkin's language

it was not until his great narrative poem, ***Yevgeny Onegin*** (written 1823–31), that he achieved what no other Russian poet had achieved: integration of idea and character; warm, humane humour; really deep understanding of the human heart; and mastery of language astonishing in its simplicity and profundity. His other narrative poems written at this time, ***Count Nulin*** (1825) and ***Domik v Kolomne*** (1830; "The Small House in Kolomna"), revealed a playful satiric vein, a blend of realism and irony with good-humoured sympathy for human foibles. The ***Tsygane (The Gypsies)*** and ***Poltava*** were masterpieces of two quite different types, the first a tragedy of frustrated passion and the second the apotheosis of a hero-king in the person of Peter I the Great. During the same period Pushkin wrote his greatest lyrics and the often-quoted rhetorical ***Prorok*** ("Prophet"). At the beginning of the 1830s he wrote fairy tales in verse and the magnificent "King Saltan." In 1833 he wrote ***Medny Vsadnik*** ("The Bronze Horseman"), in which Peter the Great personified the elemental forces that sweep puny man out of their path. Pushkin never surpassed this poem, with its onomatopoetic undertow of horror. His poetic dramas included a historical play, ***Boris Godunov*** (1831).

The period's other major poet, Mikhail Lermontov, sprang to fame in 1837 with a poem on the death of Pushkin that resulted in his exile. Lermontov's most Romantic poems, such as ***Mtsyri*** (1840; "The Novice") and the celebrated ***Demon*** (1839), dealt with Caucasian themes.

Less important poets of the Romantic period were Ivan Ivanovich Dmitriev, a follower of the sentimentalist Karamzin, and Krylov, whose simple and profound fables are among the great treasures of Russian literature. Vasily Zhukovsky, most important as a translator, created a new poetic diction on Karamzinian principles. Many poets of this golden age, such as Yevgeny Baratynsky, wrote elegaic poetry, while many others, such as Kondraty Fyodorovich Ryleyev, gave expression to their abhorrence of tsarist oppression and glorified death in the struggle against it.

Prose. With Pushkin and Lermontov the golden age of poetry ended, and the great age of creative prose began. Pushkin turned increasingly to prose in the last years of his life. Lermontov's novel ***Geroy nashevo vremeni*** (1839–40; ***A Hero of Our Time,*** 1886) was immediately successful. Its hero was one of the typical misfits of Lermontov's generation, a man of great gifts but with an insatiable and self-destroying desire for novelty.

The first widely popular 19th-century novelist was Mikhail Nikolayevich Zagoskin, whose ***Yury Miloslavsky*** (1829), on the expulsion of the Poles from Russia in 1612, appealed to the public by its crude nationalistic spirit, while a most popular Romantic novelist was Aleksandr Aleksandrovich Bestuzhev (pseudonym Marlinsky). An outstanding writer was Nikolay Gogol, who was a great influence on Russian literature and on the Russian revolutionary movement; partly due to the fact that he was a Ukrainian by birth and saw the Russians from the standpoint of an outsider, he revealed the Russians to themselves and raised turbulent ideas in their minds. The revolutionary impact of Gogol's writings became evident in 1836 with the performance of his comedy ***Revizor (The Government Inspector,*** 1968). It dealt with the officials of an obscure provincial town, but it presented a microcosm of the Russian state. It was saved from being banned by its laughter, which Gogol considered the only positive character in it. Protests at the play forced Gogol to go to Italy, where he finished the first part of his famous novel ***Myortvye dushi*** (1842–52; ***Dead Souls,*** 1877). Gogol was a conservative and an upholder of autocracy and serfdom who never thought that his play and his novel were destroying the things he stood for.

The "natural" school

One of Gogol's last stories, "Shinel" (1842; "The Overcoat," 1949), led to the creation of a whole school of writers, the "natural" school, of which Fyodor Dostoyevsky was the most brilliant follower. The first part of Dostoyevsky's literary career was notable mainly for his stories of St. Petersburg slums and the life of poor degraded civil servants, such as ***Bednye lyudi*** (1845; ***Poor Folk,*** 1894). In 1849 he was imprisoned in Siberia for his membership in a revolutionary group, and on his return—faced with the impending liberation of the serfs and the new relationship that would inevitably ensue between landowners and peasants—published a magazine to help solve this problem. In ***Epokha*** he took up an attitude midway between the Slavophiles and conservatives, on the one hand, and, on the other, the westerners and revolutionaries grouped around the magazine ***Sovremennik*** ("The Contemporary"). Dostoyevsky's attempt to bridge the gulf between the mass of illiterate peasants and the educated classes, by advocating that the latter should "return to the soil," ended in failure. His four greatest novels, ***Prestupleniye i nakazaniye*** (1866; best translation, ***Crime and Punishment,*** 1955), ***Idiot*** (1868–69; ***The Idiot,*** 1955), ***Besy*** (1871–72; ***The Devils,*** 1953), and ***Bratya Karamazovy*** (1879–80; ***The Brothers Karamazov,*** 1958), all showed his intuitive understanding of the human heart and of the tragic aspects of life, together with unsurpassed moments of illumination, that makes them immortal.

Although Ivan Turgenev lacked Dostoyevsky's moments of illumination, his stories and novels, too, transcended topical problems and laid bare the eternal. ***Zapiski okhotnika*** (1852; ***A Sportsman's Notebook,*** 1950), in which serfs were shown to be superior to their masters in intelligence, decency, and humanity, foreshadowed the abolition of serfdom. With his first novels he enhanced his reputation as a liberal-minded thinker and as one of the finest prose writers of the day. His greatest novel, ***Otsy i deti*** (1862; ***Fathers and Sons),*** dealing with the generation gap, set left-wing writers (who took Bazarov, its nihilist hero, for a caricature of themselves) against him, and he never regained popularity with young progressives. *Dym* (1867; ***Smoke,*** 1949), in which he had bitter truths to say about all classes of Russian society, and ***Nov*** (1877; ***Virgin Soil,*** 1877), in which he analyzed the Russian revolutionary movement and forecast the rising power of the managerial class, deepened the enmity of the radicals. It was only on his last visit to Russia that he was again popularly acclaimed.

Another great creative writer of the century, Leo Tolstoy, mostly stood aloof from the literary scene. Two great novels, ***Voyna i mir*** (1865–69; ***War and Peace,*** 1925) and ***Anna Karenina*** (1875–76), raised him to a pinnacle; and even his denial of his art, following his conversion to Christianity in 1876, failed to affect his supremacy. Tolstoy's distinguishing mark as a creative artist was his genius for analytical dissection of character. In "Kreytserova sonata" (1891; "The Kreutzer Sonata"), "Smert Ivana Ilyicha" (1886; "The Death of Ivan Ilich"). and ***Voskreseniye*** (1899; ***Resurrection,*** 1900), his creative genius asserted itself over the philosophic and religious convictions that marred many of his last stories.

Ivan Goncharov achieved shattering results in his finest novel, *Oblomov* (1859), by an accumulation of seemingly unimportant details. Aleksey Pisemsky was another major novelist, but his play ***Gorkaya sudbina*** (1859; "A Hard Lot"), though popular, showed the artistic dangers of too realistic a representation of life.

Late-19th-century drama, satire, and short stories. The most prolific professional playwright of the century, Aleksandr Ostrovsky, in such plays as ***Bankrut*** (1850; "The Bankrupt") and ***Groza*** (1860; "The Storm"), exposed and protested against the low business morality, ignorance, and brutality of the Moscow merchant class.

The great satirist of the century, Mikhail Saltykov, was a pillar of the radical movement as editor of ***Otechestvennye zapiski*** ("Fatherland Annals"). In ***Gospoda Golovlyovy*** (1876; ***The Golovlyov Family,*** 1955), Saltykov wrote one of the century's most brilliant novels; its hero, Iudushka (Little Judas), surpassed even Gogol's most villainous characters in idleness, uselessness, and drunkenness.

Chekhov, master of the Russian short story

Although the short story was developed by many writers, such as Nikolay Leskov, author of ***Ocharovanny strannik*** (1873; ***The Enchanted Wanderer,*** 1924) and "Zayachy remiz" (1894; "The March Hare," 1949), it was

Anton Chekhov who became its acknowledged master, as well as a brilliant innovator in drama. Chekhov was the first great Russian writer to adopt a neutral attitude to politics. Most of his short stories were written while he was a medical student and later as a young doctor. In his greatest short stories, "Step" ("The Steppe") and "Skuchnaya istoriya" ("A Dreary Story"), as well as in four famous plays, *Chayka* (1896; *The Seagull*), *Dyadya Vanya* (1897; *Uncle Vanya), Tri sestry* (1901; *The Three Sisters),* and *Vishnyovy sad* (1903–04; *The Cherry Orchard),* his training can be seen both in the medical men he introduces and in his attitude toward his characters.

Gorky's social themes

Maksim Gorky, in contrast, overflowed with pity and compassion, though only for the working classes. Self-educated, he early threw himself into revolutionary work. A romantic poem in prose, "Pesnya o burevestnike" (1901; "The Song of the Stormy Petrel")—with its refrain, "The Storm! The Storm is about to break!"—became a powerful rallying cry of the revolutionary movement. In his first and most popular play, *Na dne* (1902; *The Lower Depths),* Gorky introduced philosophizing down-and-outs onto the Russian stage. Apart from his early novels—especially *Mat* (1906; *Mother)—his* most impressive work is an autobiographical trilogy: *Detstvo* (1913; *Childhood,* 1950), *V lyudyakh* (1915–16; *In the World,* 1917), and *Moi universitety* (1923; *My Universities,* 1952).

Ukrainian. Ukrainian literature in the 19th century reflects the rapid development of Ukrainian national consciousness. Ukrainian writers at first were primarily concerned with reawakening that consciousness and the sense of historical continuity and establishing the vernacular as a literary language. Virtually the entire Ukrainian literary process in the 19th century occurred under official and unofficial Russian disfavour, and in 1863 and 1871 all Ukrainian publications were prohibited. Not until 1905 did the Russian Academy of Sciences concede that Ukrainian was indeed a separate language.

Ivan Kotlyarevsky, classicist poet and playwright, inaugurated modern Ukrainian literature with his *Eneida* (1798), a travesty of Virgil's *Aeneid.* Kotlyarevsky's work had enormous (at times undue) influence on later Ukrainian writers. Classicist prose appeared only with Hryhoriy Kvitka Osnovyanenko's novel *Marusya* (1834). His *Konotopska vidma* (1837; "The Witch of Konotop") is a first-rate novella that is reminiscent of Gogol.

Around 1830 Kharkov became the centre of Ukrainian Romanticism (Izmail Sreznevsky, Levko Borovykovsky, A. Metlynsky, and Mykola Kostomarov). In the western Ukraine Romanticism was represented by the "Ruthenian Triad": Markiian Shashkevych, Yakov Holovatsky, and Ivan Vahylevych. The Romantic movement reached its peak in the work of the Kiev Romantics and found its highest expression in the Brotherhood of SS. Cyril and Methodius (1846). Its ideology was reflected in Kostomarov's biblical *Knyhy bytiia ukrainskoho narodu* ("Books of Genesis of the Ukrainian People").

The early poetry of Taras Shevchenko, the outstanding Ukrainian poet of the 19th century, expressed the interests of the Romantics, but it soon moved to a more sombre portrayal of Ukrainian history, especially in the long poem *Haidamaky* (1841; "The Haidamaks"), and to works satirizing Russia's oppression of the Ukraine—*e.g.*, *Son* ("The Dream"), *Kavkaz* ("The Caucasus"), and *Poslaniie* ("The Epistle"). His later poetry, written after his release (1857) from exile, treats broader themes.

After Shevchenko, the most important Romantic was Panteleymon Kulish, poet, prose writer *(Chorna rada;* "The Black Council"), translator, and historian.

Ukrainian Realism, which begins with Marko Vovchok *(Narodni opovidannia,* 1857; "Tales of the People"), was long confined to populist themes and the portrayal of village life. Anatoli Svydnytsky's novel *Lyuboratski* was exceptional for its artistry but had little influence on the development of Ukrainian Realism since, though written in 1862, it was not published till 1898. Ivan Nechuy-Levytsky's literary oeuvre was larger and thematically broader, ranging from the portrayal of village life in *Kaydasheva simya* (1879; "The Kaydash Family") to that of the Ukrainian intelligentsia in *Khmary* (1908; "The Clouds"). Panas Myrny (pseudonym of Panas Rudchenko) was the major representative of Ukrainian Realism. His depiction of social injustice and the birth of social protest—in *Khiba revut voly, yak yasla povni?* (1880; "Do the Oxen Low When the Manger Is Full?"), written with his brother Ivan Rudchenko (pseudonym Ivan Bilyk), and in *Poviya* ("The Whore")—has a new psychological dimension. The novels of the influential Ivan Franko describing the Boryslav oil industries of Galicia were written in the naturalistic manner of Zola. The long narrative poems *Moysey* ("Moses"), *Panski zharty* ("Nobleman's Jests"), and *Ivan Vyshensky* are the heights of Franko's poetic achievement.

Writing at the turn of the century were such outstanding figures as the Neoromantic poet Lesya Ukrainka, Mikhaylo Kotsiubynsky, and Vasyl Stefanyk.

POLISH LITERATURE IN THE 19TH CENTURY

A second distinct period of Polish classicism, called pseudoclassicism, occurred in the early 19th century. In general, pseudoclassicist writing lacked freshness. During the Enlightenment period, literature had re-established contact with the West and became the voice of national consciousness. Although it was the literature of a community undergoing a severe political crisis, it exercised an influence on neighbouring countries.

The Romantic period began later in Poland than in England or Germany, and it lasted longer. It has been regarded as the greatest period in Polish literature. The rise of Romanticism coincided with the loss of independence, and great writers found in it an expression of their own mood. A need to interpret their country's destiny gave the work of the three great Romantic poets—Adam Mickiewicz, Juliusz Siowacki, and Zygmunt Krasiński—visionary power and moral authority. Writing in exile, they kept faith in the restoration of Polish independence alive, and their concern gave the literature of the Polish Romantic movement its strength and passion. Mickiewicz was the greatest Polish poet and leader of the Romantic period. His *Poezye* (1822–23; "Poetry") was the first major literary event of the period. In its second volume were included parts ii and iv of *Dziady,* in which he combined folklore and mystic atmosphere to create a new kind of Romantic drama. Mickiewicz' greatest works were written after 1824, when he was deported to Russia for revolutionary activities as a student; they included *Sonety Krymskie* (1826; "Crimean Sonnets"); a visionary third part of *Dziady* (1833); a messianic interpretation of Poland's past and future destiny, *Księgi narodu polskiego i pielgrzymstwa polskiego* (1832; "Books of the Polish Nation and Its Pilgrimage"), written in biblical prose; and a great epic, *Pan Tadeusz . . .* (1834; "Master Thaddeus . . .").

The poets in exile

The suppression of the insurrection of 1830–31 drove the cultural elite into exile in France; among poets whom Mickiewicz joined there were Słowacki, Krasinski, and Cyprian Norwid. Słowacki, a Romantic in the fullest sense, who in all his work showed genius, wrote verse narratives in the style of Byron and accomplished lyric poetry. He was inspired by patriotic themes: *Kordian* (1834) was a drama of conspiracy and problems of commitment. His subtle W *Szwajcarii* (1839; "In Switzerland") is probably the finest lyrical work in Polish. Much of Slowacki's work was in dramatic form, and although it was not intended for production, it laid the foundations of Polish tragic drama. His plays showed the influence of French Romantic drama, Shakespeare, classical tragedy, and Calderon. The last years of Słowacki's life were devoted to writing *Król-Duch* (1847; "The Spirit King"), an unfinished lyrical and symbolic epic describing the history of a people as a series of incarnations of the essential spirit of the nation.

Zygmunt Krasinski, when 23, published (anonymously, as with all his works) *Nie-Boskq komediq* (1835; *The Undivine Comedy,* 1875), which presented, for the first time in Europe, a struggle between opposed worlds of aristocracy and disinherited masses. *Irydion* (1836), his second play, was an allegory of Poland's fate. In *Przed-Swir* (1843; "The Moment Before Dawn") he developed a messianic interpretation of Polish history, and this conception of Poland as "the Christ among the nations" was also expounded in ***Psalmy przyszłości*** (1845; "Psalms of the Future"). The introduction of fantastic or supernatural elements into a realistic setting was characteristic of many Polish Romantic works.

The genius of Cyprian Kamil Norwid was not fully recognized until the 20th century. During his lifetime he was misjudged and remained obscure, partly because he accepted some ideas of Romanticism while criticizing others but even more because of his ironic intellectual reserve. Among the most important works published in his lifetime was a verse dialogue on aesthetics, "Promethidion" (1851), which expounded a theory of the social and moral function of art anticipating that of John Ruskin. An authentic text of his most important lyrical collection, the ***Vade-mecum,*** was first published in 1947. Norwid experimented with free verse and with the rhythms of speech, and, furthermore, he foreshadowed the French Symbolists in his analogical method of presenting the poetic concept.

The lesser talents of early Romanticism formed the Polish Ukrainian school, of whom Antoni Malczewski was outstanding as the author of a single poem, the Romantic verse narrative ***Maria . . .*** (1825), a tale of love and treachery remarkable for original diction, dramatic tension, and unity of mood.

There were fewer prose writers than poets among the exiles. Zygmunt Miłkowski (pseudonym Teodor Tomasz Jeż) wrote on a wide range of subjects, including folklore and the history of the Balkan countries. The literary criticism of Maurycy Mochnacki, a passionate advocate of Romanticism and the first Polish critic to link literature with Poland's political progress, exercised a strong and not wholly beneficial influence on literary theory. The historical works of Joachim Lelewel, a great, many-sided scholar, were an impressive example of the prose of the period.

As a result of partition, Romantic poetry in Poland was limited to closed provincial circles. In Warsaw a group of young poets was formed, but its activities were restricted by political pressure. Its most fully developed talent was that of Teofil Lenartowicz. Ryszard Wincenty Berwiński, a poet of social radicalism, wrote *Poezje* (1844) and *Studia o literaturze ludowej* (1854; "Studies on Folk Literature"), which marked a step away from Romantic nationalist interpretations and stressed the international community of folk tradition.

Prose was more popular with writers in Poland than with those in exile. Henryk Rzewuski belonged spiritually to the 18th century: *Pamiątki Seweryna Soplicy* ("Memoirs of Seweryn Soplica") evoked the atmosphere of the Baroque tradition. As the century progressed, signs of a Realistic tendency were discernible in Józef Korzeniowski's novels *Spekulant* (1846) and ***Kollokacja*** (1847). A woman novelist, Narcyza Zmichowska (pseudonym Gabryella), produced ***Poganka*** (1846; "The Pagan"), a psychological allegory anticipating 20th-century sensibility in its subtle analysis of feeling. The dominant figure among prose writers was Józef Ignacy Kraszewski, whose output ran into hundreds of volumes of fiction, history, ethnography, criticism, and so on. His imaginative writings reflected a change of literary styles during his long career. Banished in 1863, he continued to influence Polish writers at home and in exile, maintaining the Polish cause by his manifold activities.

The influence of Polish Romanticism

Polish Romanticism, conscious of its role as the torch of national spirit, retained its force as a mode of thinking beyond the period of the political circumstances that fostered it. It produced works of highest artistic value, which excited the interest of foreign writers. Mickiewicz influenced Slavonic literatures and was compared by George Sand with Goethe and Byron. Slowacki's poetic technique proved of fundamental importance to writers at the end of the 19th century, whereas Norwid's poetic technique was increasingly felt in the mid-20th century. The political implications of the Romantic movement led to the insurrection of 1863, which ended in Poland's becoming a province of Russia.

OTHER LITERATURES OF EASTERN EUROPE AND THE EASTERN MEDITERRANEAN IN THE 19TH CENTURY

Lithuanian. After the death of creative secular writings in Lithuania in the 18th century—only the poem *Metai* ("The Four Seasons," 1818), by Kristijonas Donelaitis, depicting village life, stood out from the grammars and folk song collections—the first half of the 19th century produced a new movement attempting to create a Lithuanian literary language and fostering a new Romantic interest in the early history of the country. In the literature of the period, notably in the poetry of Simanas Stanevičia and Dionyzas Poška, there appeared a surge of Western influence, in the wake of the French Revolution. Despite a Russian prohibition of the printing of Lithuanian writings in Latin letters, this renaissance was continued by Bishop M. Valančius, noted for religious and educational works, and by Bishop A. Baranauskas, a poet whose greatest work was *Anykščiu šilelis* (1858–59; ***The Forest of Anykščiai***, 1956). The literature of this era sought to rally Lithuanians against the political control of Russia and the cultural influence of Poland.

Influence of *Aušra* and of Maironis

The first modern Lithuanian periodical, *Aušra* ("Dawn"), founded in 1883 by Jonas Basanavičius and printed in Lithuanian Minor, gave its name to the literature of the ensuing generation. One of the poems of Vincas Kudirka, a leading publicist and short-story writer, became the national anthem of independent Lithuania. The most famous Lithuanian poet, J. Mačiulis (pseudonym Maironis), was noted for both dramatic and lyric poetry and has been called "the poet-prophet of the Lithuanian renaissance." Other distinguished names were V. Storasta (pseudonym Vydūnas), philosopher, poet, and dramatist; J. Biliūnas, a sensitive short-story writer; and J. Tumas (called Vaižgantas), a literary critic.

Latvian. During the "national awakening" of the mid-19th century, the Latvians established their literary independence. J. Alunāns' book of verse *Dziesminas* (1856; "Little Songs") founded the modern Latvian lyric. Folk poetry became a source of literary inspiration, as in the lyrics of Auseklis (M. Krogzems) and in A. Pumpurs' epic poem *Lāčplēsis* (1888; "Bearslayer"). The first major novel, ***Mērnieku laiki*** (1879; "The Times of the Land-Surveyors"), by R. and M. Kaudzites, portrayed Latvian peasant life realistically. Modern Latvian plays and short stories began with R. Blaumanis.

In the 1890s the "new movement" demanded Realism, but the major poet of that time, Rainis (J. Pliekšāns), wrote in a Symbolic manner, using the imagery of folk poetry in his depictions of contemporary problems. His wife, Aspazija (E. Pliekšāna, *née* Rozenberga), took up the struggle for women's rights, but displayed rather Romantic tendencies in her later work. J. Poruks introduced New Romanticism, whereas in the following decade "Decadents" or "Symbolists" propounded art for art's sake.

Estonian. Written literature began in the Estophile period (*c.* 1750–1840) with moral tales and manuals written by Balto-German enthusiasts for the native language and culture. The philological journal ***Beiträge zur Genauern Kenntniss der Ehstnischen Sprache*** ("Contributions to a Better Understanding of the Estonian Language") contained examples of folk poetry and essays, including work by the first native Estonian poet, Kristjan Jaak Peterson. More significant for literature was an epic, ***Kalevipoeg*** (1857–61, part translated in ***The Hero of Estonia,*** 1895), part authentic tradition and part a creation of Fr.R. (F. Reinhold) Kreutzwald, for this inspired the Romantic nationalistic movement soon to emerge. Popular patriotic Romantics were the poets Lydia Koidula and Anna Haava, and the first novelist was Juhan Sommer, whose book ***Luige Laus*** appeared in 1843. The historical novel first appeared with Eduard Bornhohe's *Tasuja* (1880; "The Avenger"), while Jakob Pam's ***Oma tuba, oma luba*** ("Own House, Own Master") approached a realistic style fully developed by the later work of Juhan Liiv.

Czechoslovak. ***Czech.*** The historicism and antiquarianism of the 18th century led many Czech scholars to investigate the earlier literature and history of their country; a revival of Bohemian patriotism was, moreover, a natural reaction to the centralizing tendencies of the Habsburg government under Maria Theresa and was encouraged by the local nobility in the interests of their own power. Josef **Dobrovský** became a codifier of a revived Czech literary language, and in his history of older Czech literature he brought to light neglected models that Czech writers might follow. He himself wrote in Latin and German; but several of his contemporaries restored Czech to literary currency, notably **Antonín** Jaroslav Puchmajer with his poetical almanacs. At the same time, social and political developments of the late 18th century created a public for Czech literature. Vernacular plays acted in Prague from 1786 and the beginnings of modern Czech journalism represented by the newspapers of Václav Mattj Kramerius were directed at a new middle class of growing Czech towns.

The establishment of a modern Czech literary language

Dobrovsky had contented himself with codifying the Czech language: Josef Jungmann set out to extend its possibilities so that it could become a modern literary language and achieved this in translations (including that of Milton's ***Paradise Lost,*** 1811) and above all in a monumental Czech–German dictionary (1835–39). The revival of Czech culture was also greatly furthered by **František Palacký,** whose history of Bohemia and Moravia was the last masterpiece of classical Czech.

Some Slovaks were being drawn into the Czech literary revival; foremost was **Ján Kollár,** whose allegorical sonnet cycle *Slávy **dcera*** ("The Daughter of Slava") was the first significant work in the revived language.

A special position in this period belonged to the ***Rukopis*** *Královédvorský* ("Manuscript of Králové Dvůr") and ***Rukopis*** *Zelenohorský* ("Manuscript of Zelená Hora"), ostensibly poems from the early Middle Ages but in fact forgeries executed by a brilliant and unscrupulous poet, Václav Hanka, and by others. Though their spurious character was finally proved toward the end of the century, they have held their own as examples of Romantic poetry, with echoes of popular poetry of southern Slavs.

The greatest poet of Czech Romanticism, perhaps of all Czech poets, was Karel Hynek **Mácha,** whose lyrics, prose fragments, and a lyrical epic, ***Maj*** (1836; ***May,*** 1932) showed the influence of Byron, Sir Walter Scott, and the Polish Romantics, but it was an influence transcended by its intensity of poetic vision and perfection of language.

In the 1840s a reaction against Romanticism became evident in a number of authors, notably **Karel Havlíček Borovský** and **Božena Němcová.** Both gave attention to practical issues and were masters of prose style who did much to bring Czech prose nearer to the everyday language. **Havlíček** was a political journalist whose achievement lay in critical articles in which he stood out for Czech rights against Habsburg absolutism, and in brilliant satirical poems, especially those written in exile just before his death: *Král Lávra* ("King Lavra"), ***Tyrolskd elegie*** ("Tyrolese Elegies"), and *Křest **svatdho** Vladimíra* ("The Baptism of St. Vladimir"). **Božena Němcová,** the first woman to make a significant contribution to Czech literature, became best known for ***Babička*** ("The Grandmother"), a portrayal of country life in a series of sketches held together by the central figure. Descriptions are precise and realistic; and the language has a firmness and individuality new in Czech prose.

The work of **Havlíček,** Erben, and **Němcová,** all at their peak in the 1850s, formed a bridge from a first, Romantic phase of revived Czech literature to the second half of the century, in which authors, writing for an expanding public, tried to establish a literature that might occupy a worthy place among other literatures of Europe. In 1858 an almanac, *Máj,* appeared, dedicated to the memory of **Mácha** and containing the work of young authors, among whom Jan Neruda and **Vítězslav Hálek** were pre-eminent. A significant novelist of the *Máj* group was **Karolina Světlá** (Johanna **Mužáková**); her novels of early-19th-century Prague society and of the north Bohemian countryside contained Romantic elements of theme and plot but showed considerable skill in psychological analysis. Moral and social problems and, in particular, the position of women in 19th-century society were dominant themes of her work. It was under the auspices of the *Máj* group that Adolf Heyduk began his literary career. His patriotic poetry best stood the test of time, notably the poems in which he was inspired by the kinship between Czechs and Slovaks.

Slovak. Until the 18th century there was no systematic attempt to establish a literary language from the Slovak dialects, which, though closely related to Czech, developed a separate identity from the early Middle Ages. The decline of literary Czech in the early 18th century caused an increase of local colouring in devotional texts produced by both Catholics and Calvinists. With Anton **Bernolák,** these tendencies were consolidated in a grammar (1790) and a dictionary (6 vol., 1825–27), which codified a literary language based mainly on western Slovak usage. In an era of reviving national consciousness, this language was taken up by a number of writers. Preeminent was **Ján Hollý,** who wrote lyrics, idylls, and national epics. The didactic novel ***Rend*** (1783–85) was written in a strongly Slovakized Czech by Jozef Ignác Bajza.

Attempts to establish a literary Slovak language

Bernolák's language was not accepted by the Protestants of Slovakia. who used Czech as their liturgical language. By the 1840s, however, Slovak national feeling among the Catholic majority was so strong that there was little possibility of their accepting Czech as their written language. The western Slovak of Bernolhk was inadequate; but L'udovit **Štúr** codified a new form of literary Slovak based on the central dialects that was immediately taken up by a group of talented poets. Andrej **Sládkovič** (Andrej Braxatoris) produced, in ***Marina*** (1846), a national epic that had individuality and freshness. The most significant figure was Janko **Král',** whose exploits in the revolution and war of 1848–49 made him a legend. His ballads, epics, and lyrics are among the most original products of Slavonic Romanticism.

The first Slovak dramatist of note was **Ján** Chalupka, whose works, including a lively satire, *Kocúrkovo* (1830), were written in Czech. **Ján** Palhrik used the new literary language: his comedies, produced in the 1850s and 1860s, portrayed Slovak society with sympathetic irony.

Yugoslavian. ***Serbian.*** The most important Serbian representative of the Enlightenment period was Dositej **Obradović,** whose writings (autobiography 1783, and others) greatly influenced Serbian literary development. Many characteristics of European Romanticism could be observed in the literature of the period 1820–70, especially the cult of folklore and national self-assertion. A central figure was Vuk **Stefanović Karadžić,** a reformer of literary language who did much to forward the national movement. Petar II **Petrović Njegoš,** a Montenegrin ruler, was a poet whose gifts are best seen in his dramatic poem ***Gorski vijenac*** (1847; "The Mountain Wreath"). The lyrical verses of Branko **Radičević** contributed to the break with earlier didactic-objective poetry. Among younger Romantic writers were the lyrical poets **Jovan Jovanović** (known as **Zmaj**), **Đura Jakšić,** and Laza **Kostić.**

From 1870 to 1900 there was a tendency toward Realism, and its outstanding propagator was Svetozar **Marković,** founder of Socialism in Serbia. The most significant narrative writing was by Laza **Lazarević,** a master of the short story; Simo Matavulj, a keenly observant novelist and short-story writer with a strong sense of humour; and Stevan **Sremac,** satirist and humorist. Vojislav **Ilić** was a fine descriptive poet, resembling the French Parnassians.

Toward the end of the 19th century Serbian literature was influenced by European currents, particularly the French Symbolistr: outstanding figures were Jovan **Dučić** and Milan **Rakić,** a patriotic poet of a melancholy nature. Vladislav **Petković** (called Dis), a subjective poet, chose themes halfway between reality and the irrational. Prose changed from objective description to psychological narrative, often with a lyrical quality, as in the work of Petar **Kočić,** Ivo Cipiko, and especially Borisav **Stan-**

ković. A literary historian and critic, Jovan Skerlić asserted concepts of aesthetics based on belief in the social and humanitarian functions of art.

The popularity of patriotic verse

Croatian. Romanticism in Croatian literature originated in a cultural and national reformation brought about by the Illyrian movement (1835–48), which aimed at a union of all south Slavs within the Habsburg federation. Ljudevit Gaj introduced the *Btokavski* dialect and *ijekavski* speech as the literary language of Croatia and also a unified orthography. Personal, patriotic, and reflective lyrics were popular and were best represented by the sensitive, moving poems of Stanko Vraz, the patriotic songs and poetic drama of Petar Preradović, and the dramatic works of Dimitrije Demeter. Some of the best poetry includes *Smrt Smail–Age Čengića* (1846; *The Death of Smail–Aga Čengić*, 1925) by Ivan Mažuranić.

Many writers collected Croatian folk poetry, which continued to inspire writers long after the Illyrian movement was over. In the development of national characteristics of Croatian Romanticism an outstanding figure was August Senoa, poet, dramatist, critic, journalist, and creator of the Croat historical novel. Šenoa was a subtle stylist and a champion of Realism, which was at its height between 1881 and 1895. Writers were concerned with contemporary problems, particularly with conditions among the lower classes. Representatives of this movement were Evgenij Kumičić, a naturalist and poet; Ante Kovačić, satirist and author of a lively novel, *U registraturi* (1888; "In the Registry Office"); and Ksaver Šandor Gjalski, a short-story writer and novelist concerned with political and cultural problems. A poet, Silvije Strahimir Kranjčević, was a notable social rebel whose work was rich in imagery and rhythmic variety.

Slovene. The Romantic movement purified both language and literature in Slovenia. Its initiator was Jernej Kopitar, who established standards for the literary language. Matija Cop supported the poetic miscellany *Kranjska Cbelica* (1830–34, 1848; "The Carniola Bee"), which included work by France Preieren, an outstanding poet of the Slovene Romantics. He created some of the greatest poems in Slovene and gave new life to his country's literature. Preieren published a tale in verse, *Krst pri Savici* (1836; "The Baptism at the Savica"), and only one collection of poems, *Poezije* (1847; "Poems"), in which he expressed his love for a woman, for his people, and for mankind. With his works Slovene literature reached an unexpected summit in European poetry. The Romantics produced the first popular narrative and were enthusiastic about folk poetry and the newly established literary language. A battle for the national alphabet was won by the progressive Cop-Preieren group, which defeated attempts to replace the literary language by Illyrian, an amalgam of Serbo-Croatian dialects. The conservative national leader J. Bleiweis established the newspaper *Novice* (1843–1902; "News"), of which the leading poet was the bombastic, emotional J. Vesel (called Koseski), a champion of Slovene nationalism at the time of the March revolution of 1848. The revolution and the constitutional period after it induced the young intelligentsia to begin to organize politically and culturally the awakening peasant masses and native bourgeoisie, in spite of the opposition of the German administration. Poets and novelists who shared in this national revival were Janez Trdina, Fran Erjavec, and Janez Mencinger, best known for short stories and novellas, and Simon Jenko, a significant poet. The most dynamic writer was a critic, Fran Levstik, founder of a program for popular literary prose. His program was completed by Josip Jurčič, author of the first Slovene novel, *Deseti brat* (1866; "The Tenth Son"). Josip Stritar, a poet, novelist, playwright, essayist, and critic, directed Slovene writers toward Western literatures; and Fran Levec achieved distinction as a literary historian. Simon Gregorčič, a melodious lyrical poet, was celebrated for his patriotic verses, and Anton Aškerc wrote some of the best Slovene ballads. Neoclassical Realism was represented by Fran Detela, Janko Kersnik, and the novelist Ivan Tavčar. Drama developed in a parallel manner and is represented by Fran Levstik, Josip Jurčič, Jože Ogrinec, and Josip Voinjak. In the second half of the 19th century several literary reviews were established in Slovenia: *Slovenski glasnik* (1858–68; "The Slovene Herald"), *Zvon* (1870,1876–80; "The Bell"), *Zora* (1872–78; "The Dawn"), *Kres* (1881–86; "The Bonfire"), *Ljubljanski zvon* (1881–1941; "The Ljubljana Bell"), and *Dom in svet* (1888–1945; "The Home and the World"). Later, two associations which were also publishing houses were established, Mohorjeva Družba (1852; "St. Mohor's Society") and Slovenska Matica (1864; "The Slovene Literary Society"), as well as Dramatično Društvo (1872; "The Dramatists' Association"), and Pisateljsko Društvo (1872; "The Writers' Association"). Celovec and Maribor joined Ljubljana as Slovene cultural centres.

Macedonian. Early-19th-century literature in Macedonia echoed the medieval religious enlightenment of the school of Kliment, and the only original contributions were from folk literature. Later, original lyrical poetry was written by Konstantin Miladinov, who, with his brother Dimitrije, compiled a notable collection of legends and folk songs contributing to the development of a national Macedonian literature.

Greek. After the capture of Constantinople by the Turks in 1453, the destruction of Greek national life and Byzantine civilization naturally involved a cessation of Greek literary production in the subjected regions. Learned Greeks in exile sometimes retained their own language when they wrote, but they often used foreign languages. Modern Greek literature can nevertheless be traced to distant native origins, such as the 10th-century epic cycle of Digenis Akritas; verse romances of the 13th and 14th centuries; a chronicle of 15th-century Cyprus by Leontios Makhairás, who used the Greek vernacular; and writings inspired by the Franks, who introduced rhyming verse when they ruled Crete, Rhodes, and Cyprus. Crete was captured by the Turks in 1669 and until then had a vigorous and remarkable literature, influenced by Italian during its Venetian occupation. It showed development in versification and in drama, including comedies and a mystery play that have some extraordinarily modern aspects. There were also many interesting popular songs. Above all, the language of Cretan literature, though incorporating some Venetian words, was vigorously native. Those parts of Greece under Turkish domination only produced folk songs and klephtic ballads (a klepht was an armed Greek living as a freeman and outlaw). The roots of these songs, some reaching far into antiquity, were a source of inspiration to modern Greek poetry.

The literary styles of modern Greek

Modern Greek has inherited two literary styles—the consciously archaic, the Katharevusa, and the living popular, the Demotiki. Both were promoted by the invigorating influence of Italian and French literature. The Demotic movement of the 1880s ended with a decisive victory of the popular language for all writings of an imaginative nature. The purist language for its part was unsuited for production of a really living drama, but with the use of Demotic and a choice of contemporary themes, plays showed a fresh vitality, and a national drama began to emerge, inaugurated mainly by G. Xenopoulos.

The Greek Romantics. The liberation of Greece from the Turks in 1828 made the capital of the new kingdom a centre of Greek political and intellectual life. The Phanariotes (Greeks who had gathered around the ecumenical patriarch in Constantinople and who endeavoured to continue the purist Byzantine tradition in the 18th century and link it with the great literatures of the West) moved finally to Athens. In this mixed and unsettled society the Romantic school of Athens flourished, founded and led by Aléxandros Soútsos, who had studied in Paris and was influenced by French Romanticism but never captured its spirit. The influence he exerted upon Greek poetry was considerable. All poets in the early days of the Romantic school were painstakingly patriotic and often excessively Romantic; they mainly used the Katharevusa. The most striking figure was Aléxandros Rizós Rankavis, a man of extraordinary fertility of mind; his works covered many genres and were considerable achievements in the archaic style.

Akhilléfs Paráskhos was a leading poet in the last period of the school. His contemporaries Angelos Vláchos, Aléxandros Vyzántios, Dhimitrios Paparrigópoulos, Geórgios Viziïnós, and Geórgios Paráskhos were overshadowed by his rhetorical profuseness, but often showed greater sincerity and finer technique.

The Ionian school of poetry. Another school of poetry, parallel to the Romantics of Athens, flourished in the Ionian islands, under British rule until 1864. Its founder and greatest representative was Dhionísios Solomós, who wrote at first in Italian, following his education in Italy, but soon developed a preference for Greek. His work showed a gradually deepening approach to life, and, although temperamental instability prevented his completing any major works, his verses possessed a delicacy and balance unsurpassed in modern Greek. Solomós marked a turning point in the struggle between Katharevusa and Demotiki; his choice of the latter foreshadowed the course of Greek poetry after the Demotic movement of the 1880s. He also introduced some Western metrical forms, thus freeing Greek poetry from the monotony of the 15-syllable political verse mainly used previously. Two other major poets of the Ionian school were Andréas Kálvos, who was inspired by Greek classics and used a classicized form of language; and Aristotelís Valaorites, a deeply Romantic poet, greatly admired by his contemporaries, who influenced Kostís Palamás in adopting the spoken tongue for poetry and thus linked the Ionian school with the new school of Athens.

The new school of Athens. About 1880 some young poets formed a group in reaction to the stilted Katharevusa and the weaknesses of Romanticism. They aspired to a restrained and objective art, drawing inspiration from contemporary Greece and using the living idiom. A climax in the so-called Demotic movement came with Ioinnis Psicharis' *Tó taxídi mou* (1888), a prose work intended to awaken Greek linguistic consciousness.

Prose. Before the 19th century, prose works were mainly theological or educational and were fostered by the clergy, the great Phanariote families of Constantinople, and the patronage of wealthy merchants. The greatest single forerunner of the revival of prose was Adamántios Koraïs, who exercised an enormous influence on his Greek contemporaries by issuing many editions of classical authors, accompanied by stirring patriotic introductions and by efforts to reform the written language to bring it closer to the living idiom.

In the first decades that followed the liberation of Greece, prose was largely confined to journalism and scholarship. The few writers of creative literature were attracted by the works of Sir Walter Scott and Alexandre Dumas *père*, and the historical novel was cultivated between 1821 and 1888, notably *Authéntis tou Móreos* ("Lord of the Morea") by Aléxandros Rizos Rangavis; Emmanouil Roi'dis' vivid historical satirical novel about Pope Joan (1865); and *Loukis Láras* (1879) by Dhimitrios Vikelás. The short story was also developed in this period, particularly by Georgiós Viziïnos and Ioánnis Kondylákis. There also occurred the development of a remarkable classical style by certain distinguished scholars. Two works of major importance were a history of the Greek revolution by Spyridon Trikoupis and a history of the Greek nation by Konstantinos Paparrigópoulos, both written in a balanced and flowing style.

Bulgarian. Modern Bulgarian literature dates from the mid-19th-century awakening of national consciousness. Consonant with this was the formation of *novobulgarski*, the new (or modem) literary Bulgarian language based on the vernacular of its eastern dialects, as opposed to the medieval Church Slavonic, which until then had always been used for literary purposes. Pioneers in this were Bishop Sophrony, whose *Nedelnik* (1806; “Sunday-Book”) is the first modern Bulgarian printed book; Neophyt Rilski, grammarian and founder of the first modern Bulgarian school in 1835; V. Aprilov; N. Gerov, compiler of the first major dictionary of Bulgarian; I. Bogorov; and the Russian antiquary Y. Venelin. Harbinger of this whole awakening of Bulgarian national consciousness (known as the *Vuzrazhdane)* was Father Paisy of Chilandari, whose single work, *Zstoria slavyano-bulgarska* (1762; "Slavo-Bulgarian History"), by its romantic evocation of Bulgaria's past and appeal to national self-respect inspired Bulgaria's renascence, including its first able modern writers. These, who often combined capacities of poet, scholar, publicist, and revolutionary, shaped through works of unequal literary merit an effective image of the resurgent nation. Conditions of the time—lack of freedom, the strength of Greek cultural domination, and strong Russian utilitarian influences—taught these writers, many of whom were educated in Odessa or Moscow, that literature should serve social and national needs. Thus inspired, D. Voynikov, I. Bluskov, and especially L. Karavelov and V. Drumev founded modern Bulgarian Realism with their narrative prose and drama taken from rural and small-town life; H. Botev, in his single-minded devotion to ideals of liberty and fatherland, wrote impassioned revolutionary poetry, dying (like his fellow martyr-poet Bacho Kiro) in 1876 a victim of the Turks; Petko Slaveykov, an irrepressible journalist, Bible translator, and agitator for the Bulgarian independent exarchate, worked all his life in Bulgarian and Macedonian lands and in Stambul itself (never in emigration like Karavelov and Botev), drawing for his verse on folklore and Greek popular songs: and G. Rakovski, a typical *vuzrozhdenets* ("Renaissance figure") in his versatility and vitality, exploited often with mole zeal than discretion the two main indigenous resources for Bulgarian writers then and since, an illustrious medieval past and a richly surviving folklore.

The liberation of Bulgaria in 1878 created a climate far more favourable to literary development than that of the preceding five centuries of Turkish rule. Ivan Vazov almost alone links as a writer the epochs before and after liberation. His immense output from the early 1870s to 1921, reflected in all main genres every facet of his people's life, past and present, has earned him the title of "national poet." An epic cycle, *Epopeya na zabravenite* (1881–84: "Epic of the Forgotten"), evoked with visionary power his pantheon of heroes from the struggle for independence; a novella, *Chichovtsi* (1895; "Uncles"), was a realistic portrait gallery of Bulgarian provincial "notables" in Turkish times. Vazov's narrative gifts were at their highest in Bulgaria's "national novel" *Pod igoto* (1893; translated as *Under the Yoke,* 1894), which vividly described the Bulgarian struggle against the Turks; his short stories entertained and his travelogues of rural and historical Bulgaria informed his readers; his most popular play, *Hushové* (1894), depicting these pre-liberation expatriates' ordeals in Romania, together with his dramas of medieval Bulgaria, dominated the repertoire of the Sofia National Theatre (founded 1907).

Without equalling Vazov's powers of imagination and synthesis, Konstantin Velichkov shared his ideals. His unusual poetic temperament was best expressed in sonnets inspired by travels to Constantinople and Italy. An exponent of Italianate influence, he contributed to the then fashionable literature of memoirs. Most notable here was Z. Stoyanov, whose *Zapiski po bulgarskite vuzstaniya* (1883–85; translated as *Notes on the Bulgarian Uprisings)* recorded eyewitness experiences of then recent history with a directness rarely equalled since in Bulgarian prose.

Writers of the new independent state, not preoccupied with celebrating the recent or distant past, eyed critically contemporary society's more negative aspects. In satire, fable, and epigram, S. Mihaylovski with unrelenting bitterness castigated corruption in public life. His most ambitious satire, *Kniga za bulgarskia narod* (1897; "Book on the Bulgarian People"), took the form of a moral-philosophical allegory. In a lighter vein, Aleko Konstantinov created in *Bay Ganyu* (1895; subtitled "Incredible Tales of a Contemporary Bulgarian [on his European Travels and at Home]") a tragi-comic prototype of the Bulgarian peasant turned *parvenu* and demagogue. In his travelogue *Do Chicago* i *nazad* (1894; "To Chicago and Back”) he measured Bulgaria against cultures of Europe and the United States, not always to the latters' advantage.

Romanian. The first landmark of the 19th century in Romania was a national popular rising of Tudor Vladimirescu (1821) in Walachia and the return of the national rulers. Romanticism carried forward the falling wave of the Latinist movement. In the second half of the 19th century a serious literary criticism, which originated in German philosophy and French culture, inaugurated modern literature.

Transylvanian Latinism crossed the Carpathians and had beneficial effects on the Greek-inspired culture of Walachia. Eliade Rădulescu, who came under this influence, founded the first Romanian newspaper in Walachia and the Societatea Filarmonică (1833), which later treated a national theatre in Bucharest. He was a pioneer of Italian influence, which was taken up in Moldavia by Gheorghe Asachi, who created the historical short story, wrote verses in Romanian and Italian, and founded a periodical, Albina *Românească.* The outstanding literary personality among a galaxy of minor poets and translators who enriched the Romantic heritage was Grigore Alexandrescu. Alexandrescu wrote Poezii (1832, 1838, 1839) and *Meditații* (1863), fables and satires influenced mostly by French writers. A literary magazine, Dacia *Litterară*, edited by Mihail Kogălniceanu, a leading statesman and father of modern Romanian historiography (1840), marked a beginning of the traditionalist trend in literature. Alecu Russo, another leader of 1848, enriched literature with a biblical prose poem, *Cântarea României.*

Leading writers of the second half of the century were Vasile Alecsandri and Mihail Eminescu. Alecsandri's rich output comprised poetry (Doine *şi lăcrimīoare* [1853], *Suvenire si Mărgăritărele* [1856]), prose (*Buchetiera* din *Florența* ["A Bouquet from Florence"], *Călătorii* in Africa), and plays (*Fîntîna* blanduziei, Ovidiu, Despot *Vodă*). He also revealed treasures of Romanian folklore in *Balade* (1852–53) and Poezii populare (1866). Eminescu, a philosophical lyric poet, created modern Romanian poetry. He was influenced by Hindu thought and German philosophy but remained rooted in tradition. He raised Romanian poetry to new heights and was the guiding star in every aspect of cultural life. His writings include short stories and political and philosophical essays.

Hungarian. Literary leadership of Hungary at the beginning of the 19th century was assumed by Károly Kisfaludy when, in 1822, he founded a literary magazine, Aurora, to which all the important writers of the period contributed. He was also the first representative of Romanticism and the first playwright to achieve popular success.

While Kisfaludy's tragedies were applauded all over the country, Bdnk *bán* (the bdn was a high Hungarian dignitary), one of Hungary's best tragedies, by József Katona, was published in 1821 but was overlooked. Set in the 13th century and written in vigorous prose, the play was a masterful combination of national and individual conflicts, and one of its characters, Tiborc, a poor peasant, has remained ever since a symbol of the oppressed.

Ferenc Kolcsey was a politician of standing and a brilliant orator; his literary criticism was of a high standard, though unduly severe. His later poems, grave, vigorous in thought and expression, often dealt with problems of his country's situation; the impressive "Hymnus" (1823) became the Hungarian national anthem. After Kisfaludy's death, Mihály Vorosmarty became a central figure in literary life, producing works of value in every literary genre. In particular he succeeded with a long epic poem *Zalán futása* (1825; "The Flight of Zalan"), written in a Romantic vein but with an eye to contemporary problems, which appeared also in many of his best lyric poems and in his symbolic fairy play Csongor *és* Tiinde (1830).

Hungarian novels

In Hungarian literature, poetry was far ahead of drama, and the novel seemed slow in taking root. Miklós Jósika, a disciple of Sir Walter Scott, was the first successful novelist. His first and best work, a historical novel, *Abafi* (1836), marked a turning point in the history of the Hungarian novel. József Eotvos, who after the 1848 revolution became a political theorist, produced two of the best novels in Hungarian literature—A *falu jegyzője* (1845; "The Village Notary"), a portrait of feudal life in his own time, and *Magyarország 1514-ben* (1847; "Hungary in 1514"), about Gyorgy Dózsa's peasants' revolt. They possessed exceptional qualities of characterization, both of individuals and of periods, and were political manifestos in support of the oppressed and against the appalling injustices that led to revolutions—of which Eotvos disapproved as a political means.

The folk song and ballad collections of János Erdélyi and Janos Kriza exerted an influence on the further development of Hungarian poetry. "Popular poetry is the only real poetry" was the opinion of Sándor Petőfi, one of the greatest Hungarian poets, whose best poems ranked among masterpieces of world literature. He was an innovator and made a break with conventional subjects and poetic language. His poems are striking in immediacy of perception and directness of language and could treat almost any subject. The revolutionary violence of his patriotic poems was an inspiration during the revolution of 1848. Petofi's many songs were enchanting in their simplicity, and in this genre he remained unsurpassed. His death on the battlefield in 1849 at 26 was a tragic loss to Hungarian literature.

János Arany shared Petőfi's conviction of the value of popular poetry, but his approach was different, for his subjects were often taken from history and showed deep understanding of the human mind. He had the assurance of one who knew that what he wrote was the language of the people, lifted to a degree never surpassed in Hungarian. His ballads, often Romantic, had vigour, conciseness, and uncommon evocative power. His great narrative poems, the Toldi trilogy (1847–79) and Buda *halála* (1864; "Death of Buda"), reflected eternal human problems; Arany's philosophy appeared through his characters and not in lengthy digressions and was accompanied by subtle humour.

The peaks of poetry reached by Petofi and Arany remained inaccessible to other poets during the rest of the 19th century. Hungary, prostrated by the unsuccessful revolution of 1848–49, remained under Austrian rule until 1867. Defeat was followed by half-hearted acceptance of the compromise and by a self-deceiving optimism that produced complacency and uninspired traditionalism. The strongest talent in this period of decadence, János Vajda, beset with gloomy visions, was little recognized by contemporaries but influenced coming generations.

The national theatre

In 1837 a national theatre was formed to produce works of merit, but the level of plays was low. Ede Szigligeti's comedies were amusing, and he created a special genre, the *népszinmü*, plays giving an idealized picture of village life. Szigligeti catered for a public delighted to watch plays presenting no problems; but problems—social, political, and religious—constituted a principal preoccupation of Imre Madách, whose Az ember *tragédiája* (published 1861; "The Tragedy of Man") was one of the glories of Hungarian literature. This dramatic poem followed man's destiny from creation through stages of history into a future of a phalanstery and the ultimate extinction of life; the play not only displayed original thought incisively expressed but was also effective on the stage. It remains a favourite with the Hungarian public.

The first really good novelist, Zsigmond Kemtny displayed, in such novels as Zord id6 (1862; "Grim Times"), A *rajongók* (1858–59; "The Fanatics"), and *Férj és nő* (1852; "Husband and Wife"), a masterly skill in psychological analysis. His characters were pushed toward their gloomy fate by a host of circumstances determined by their own deeds. Analysis often took the place of action in Kemény's novels, which were therefore difficult to read and not popular. On the other hand, Mór Jókai was undoubtedly a popular Hungarian novelist, an exceptional storyteller able to evoke any epoch, any milieu; characters were idealized and descriptions inaccurate, but the dream world so created was real. In his works (more than 100 volumes), historical novels alternated with others centred on problems of contemporary society, of which almost all strata interested him. Kálmán Mikszáth was also popular; he recorded with keen observation and sly humour the shortcomings of society but, although a politician, was little concerned with improvement. Though his principal works were published early in the 20th century, Géza Gárdonyi belonged to the 19th century. ***Egri csillagok*** (1901; "The Stars of Eger") and ***A láthatatlan ember*** (1902; "The Invisible Man") were well constructed.

During the 19th century, literary life in Hungary became organized to some extent: Kisfaludy and Petofi societies, founded in 1836 and 1876, respectively, were particularly influential, though the authority of the Hungarian Academy, founded in 1825, remained unchallenged. The principal critic of the second half of the century was Pál Gyulai, himself a poet and writer. Literary criticism, the history of Hungarian literature, attracted some of the best minds, including Jeno Peterffy and Frigyes Riedl.

Finnish. Although a number of talented poets wrote in Finnish in the 17th and 18th centuries, it was only with Aleksis Kivi that a genuine Finnish literature came into being. His *Seitsemän veljestä* (1870; ***Seven Brothers,*** 1929) was the first Finnish novel. It endowed Finnish with a style (realism spiced with humour) and a theme (the Finnish people) that were to become traditional. Kivi was also a poet, and among his fellow poets were August Ahlqvist-Oksanen: Suonio (Julius Krohn); an austere and powerful poet, Kaarlo Kramsu; and J.H. Erkko, whose style was based on folk song.

In 1872 Kaarlo Bergbom founded the Finnish National Theatre. The 1880s saw the formation of a group of liberal writers known as Nuori Suomi (Young Finland), who founded a paper, ***Yaivalehti*** (from 1904, ***Helsingin Sanomat).*** Influenced by Norwegian and French writers, they introduced Realism and social criticism to Finland. Similar views were already being put forward by a formidable dramatist, Minna Canth. Among the group's members were Juhani Aho, a master of the lyrical nature novel; and Arvid Jarneflet, who attracted attention with *Isänmaa* (1893; "The Fatherland"), a novel of student life, and who tried to apply Tolstoy's teachings to his novels and his life. In *Vanhempieni romaani* (1928–30; "The Novel of My Parents"), he produced a classic portrait of his parents, who had played a part in Finland's cultural life. A younger writer was Eino Leino, one of the major Finnish poets, the scope of whose talent ranged from the visionary and mystical ***Helkavirsia*** (1903–16; "Sacred Hymns") to topical novels, pamphlets, and critical journalism. (For literature in Swedish, see above ***Scandinavian literature in the 19th century: Finnish literature in Swedish.)***

Hebrew. In the 18th century the conservative mystical movement of Ḥasidism spread rapidly over all eastern Europe except Lithuania. There, Elijah ben Solomon of Vilna, a writer of unusually wide scope, advocated a better graded course of Talmudic training. Shneur Zalman of Ladi created the highly systematized Ḥabad Hasidism, which was widely accepted in Lithuania. The Musar movement of Israel Salanter encouraged the study of medieval ethical writers.

The Haskala movement

In the Berlin of Frederick II the Great, young intellectuals from Poland and elsewhere, brought in as teachers by the wealthy, met representatives of the European Enlightenment and came under the influence of Moses Mendelssohn and also met some representatives of Italian–Dutch Hebrew culture. One was a Dane, N.H. Wessely, who wrote several works on the Hebrew language, and another, an Italian, S.A. Romanelli, who wrote and translated plays. Out of these contacts grew Haskala ("Enlightenment"), a moderate tendency toward Westernization that venerated Hebrew and medieval western Jewish literature. Among German Jews, then already in rapid process of Germanization, this Hebrew movement had no place. The Enlightenment was somewhat brusquely introduced in Galicia (Austrian Poland), a centre of Hasidism, by the Edict of Toleration (1781) of the emperor Joseph II. By supporting some of its aims, Hebrew writers incurred hatred and persecution. Their chief weapon was satire, and the imitation by J. Perl of the ***Epistolae obscurorum virorum*** of 1515 of Crotus Rubianus and the essays of I. Erter were classics of the genre. One poet, M. Letteris, and one dramatist, N.I. Fischman, wrote interesting biblical plays.

Galicia's chief contribution was to the Jiidische Wissenschaft, a school of historical research with Romanticist leanings. The impact of Haskala ideas upon the humanistic Italo-Hebrew tradition produced a short literary renaissance. Its main connections were with the Jiidische Wissenschaft, to which I.S. Reggio contributed. S.D. Luzzatto, a prolific essayist, philologist, poet, and letter writer, became prominent by his philosophy of Judaism, while a poet, Rachel Morpurgo, struck some remarkably modern chords. For the Jews of the Russian Empire. the Enlightenment proper began with I.B. Levinsohn in the Ukraine and with M.A. Ginzberg (Günzburg), in Lithuania. In the 1820s an orthodox reaction set in, coinciding with the rise of a Romanticist Hebrew school of writers. A.D. Lebensohn wrote fervent love songs to the Hebrew language, and his son Micah Joseph (Mikhal). the most gifted poet of the Haskala period, wrote biblical romances and pantheistic nature lyrics. The first Hebrew novel, ***Ahavat Ziyyon*** (1853; "The Love of Zion"), by A. Mapu, was a Romantic idyll, in which Mapu, like all Haskala writers, employed phrases culled from the Bible, more or less ingeniously adapted to the thought the writer wished to express.

Mapu's third novel, ***Ayit Tsavu^{c}a*** (1857–69; "The Hypocrite"), marked a departure. It dealt with contemporary life and attacked its social evils and portrayed a new type, the ***maskil*** (possessor of Haskala), in a fight against orthodox obscurantism. The new, aggressive Haskala soon came under the influence of Russian left-wing writers, such as Nikolay Gavrilovich Chernyshevsky and Dmitry Pisarev. Judah Leib Gordon, like Mapu, had started as a Romantic writer on biblical subjects. From 1871 onward he produced a series of ballads exposing the injustices of traditional Jewish life. M.L. Lilienblum began as a moderate religious reformer but later became absorbed by social problems, and in "Mishnat Elisha ben Abuyah" (1878) he preached Jewish socialism. P. Smolenskin created in six novels a kaleidoscope of Jewish life in which he rejected the Westernized Jew as much as orthodox reactionaries did.

Armenian. The 18th century witnessed an Armenian cultural and intellectual renaissance, and by the middle of the 19th century, the time was ripe for the development of a modern Armenian literature. The Armenian language, however, was in a chaotic state, and the question of which form should serve as the vehicle for new ideas led to controversies, both in Turkish and Russian Armenia, between champions of the old classical language and those of the modern spoken languages. Eventually the latter prevailed, with the result that the eastern literature became written in a modified form of the Erevan dialect ***(rusahayeren)*** and that of the west in a similarly modified form of the dialect of Istanbul (***dachgahayeren).*** For their models, and for many of their ideals, Armenian writers looked to Europe. Among western authors, Hakob Paronian and Ervand Otian were outstanding satirical novelists, and Grigor Zohrab wrote realistic short stories; the theatre was best represented by Paronian, whose comedies (as ***The Dowry, Master Balthazar, The Oriental Dentist)*** still remain popular.

Russian Armenian novels

The novel, weak in western Armenian literature, was strongly represented in Russian Armenia, where it became a vehicle for Armenian moral, social, and political aspirations. Khachatur Abovean, "father of modern Armenian literature," wrote ***Wounds of Armenia*** in ***1841.*** The most celebrated Armenian novelist was Hakob Meliq-Hakobian, or Raffi. Among eastern poets, Hovhannes Thumanian wrote lyric and narrative poems; and his masterpiece, a short epic, ***Anush,*** full of songs that have become traditional, was early adapted as an opera. The most outstanding Armenian dramatist was Gabriel Sundukiantz, whose comedies ***(Hullabaloo, Pepo, The Broken Hearth)*** portrayed the contemporary Armenian society of Tbilisi, in whose dialect most of them were written.

The rapid decline of Istanbul as the principal western Armenian literary centre (after massacres in the first quarter of the 20th century) brought about a new period of decadence in Armenian literature, although Armenians scattered abroad continued to write in Paris, Beirut, and Boston. Some Turkish Armenians fled to the east, where they enjoyed a certain degree of autonomy and where, after the foundation of the Armenian Soviet Socialist Republic in 1936, national literature was encouraged, and controlled, by the state.

VII. The 20th century

CHARACTERISTICS OF THE PERIOD

When the 20th century began, social and cultural conditions that prevailed in Europe and America were not too different from those of the middle and late 19th century. Continuity could be seen, for example, in the work of four novelists writing in English at the turn of the century and after. Joseph Conrad, Thomas Hardy, Henry James, and D.H. Lawrence all demonstrated in the progress of their work the transition from a relatively stable world at the end of the 19th century to a new age that began with World War I (1914–18). The awakening of a new consciousness in literature was also to be traced in such works of fiction as the first volume of Marcel Proust's ***Remembrance of Things Past*** (*Swann's Way*, 1913), André Gide's ***Vatican Cellars*** (1914), James Joyce's *Ulysses* (1922), Franz Kafka's ***Trial*** (published posthumously in 1925), and Thomas Mann's ***Magic Mountain*** (1924).

The influence of Nietzsche and Freud

Various influences that characterized much of 20th-century writing from the 1920s were at work in these writers. An interest in the unconscious and the irrational was reflected in their work and that of others of about this time. Two important sources of this influence were Friedrich Nietzsche, a German philosopher to whom both Gide and Mann, for example, were much indebted, and Freud, whose psychoanalytical works, by the 1920s, had had a telling influence on Western intellectuals. A shift away from 19th-century assumptions and styles was not limited to writers of fiction. André Breton's first *Manifeste du surréalisme* (1924; "Manifesto of Surrealism") was the first formal statement of a movement that called for spontaneity and a complete rupture with tradition. Surrealism showed the influence of Freud in its emphasis on dreams, automatic writing, and other antilogical methods and, although short-lived as a formal movement, had a lasting effect on much 20th-century art and poetry. The uncertainty of the new age and the variety of attempts to deal with it and give it some artistic coherence can be seen also in Rainer Maria Rilke's ***Duino Elegies*** and ***Sonnets to Orpheus*** (1923); in T.S. Eliot's ***Waste Land*** (1922); and in Luigi Pirandello's play about the instability of identity, *Henry IV* (1922).

The international and experimental period of Western literature in the 1910s and 1920s was important not only for the great works it produced but also because it set a pattern for the future. What was clearly revealed in the major works of the period was an increasing sense of crisis and urgency, doubts as to the 19th century's faith in the psychological stability of the individual personality, and a deep questioning of all philosophical or religious solutions to human problems. In the 1930s these qualities of 20th-century thought were not abandoned but, rather, were expanded into a political context, as writers divided into those supporting political commitment in their writing and those reacting conservatively against such a domination of art by politics. Nor did World War II resolve the debate concerning political commitment—issues similar to those that exercised major creative imaginations of the 1930s were still very much alive during the 1960s and the early 1970s.

The scarcity of great writers after World War II

It would be tempting to explain what seemed to be a relative scarcity of great writers in the period after World War II as an inevitable result of the cumulative pressure of disturbing social and technological developments accelerated by that war. Under such fluctuating and doubtful circumstances, it would not seem altogether strange if writing and reading, as traditionally understood, should cease. Indeed, in certain technologically highly developed countries, such as the United States, the printed word itself seemed to some critics to have lost its central position, having been displaced in the popular mind by a visual and aural electronic culture that did not need the active intellectual participation of its audience. Thus the communications media that helped to create something resembling an international popular culture in many Western countries did nothing to make the question of literary value easier to answer. Given the extraordinary conditions in which a modern writer works, it was not surprising, as the last quarter of the century was about to open, that reputations were difficult to judge, that radical experimentation characterized many fields of literature, and that traditional forms of writing were losing their definition and were tending to dissolve into one another. Novels might acquire many features of poetry or be transformed into a kind of heightened nonfictional reportage, while experimentation with typography gave poems an appearance of verbal paintings, and dramatic works, shorn of anything resembling a traditional plot, became a series of carefully orchestrated gestures or events. But formal experimentation was only part of the picture, and to say that modern writing since World War II has been primarily experimental would be to ignore other characteristics that writing acquired earlier in the century and that still continued to be issues. Most good critics felt that there was no lack of good literature being written, despite the lack of major reputations and despite the possibly transitional nature of much of the period's work in its variety of styles and subjects.

ENGLISH LITERATURE IN THE 20TH CENTURY

The turn of the century was a period of complex change, and combinations of social forces not at first apparent left their mark on literature. Rudyard Kipling's verse, for example, expressed pride in Britain's overseas expansion; but it also looked back in admiration to mid-Victorian industriousness and technical skill, upon which that expansion had been based, and registered with recurrent disquiet a sense that they were being replaced by an age in which a financier stood for more than an industrialist or pioneer. In his short stories, too, Kipling was often critical not of the imperial idea but of a failure to live up to its ideals. He was no vulgar imperialist, and his writings reflected the anxieties as well as the buoyancy of his period. Yet as the 20th century opened it was a very different poet, Thomas Hardy, who, in pieces about the South African War, brought out the realities of empire building and war in their effects on ordinary people, so anticipating and providing an example for many later 20th-century "war poets."

Early-20th-century novelists. Joseph Conrad also showed a sense of the less attractive side of overseas expansion. His novel ***Lord Jim*** (1900) closed with a memorable picture of the lowest kind of overseas adventurer, petty yet ruthless, half trader and half bandit. The story ***Heart of Darkness*** (1902) struck a balance between what was dark and bestial in primitive black Africa and the greed and aimless vindictiveness of its white exploiters. Conrad's most outstanding novel, ***Nostromo*** (1904), recorded how finance imperialism operated in a Central American country to replace colourful barbarism with nothing better than a soulless, sterile materialism. The deepest note running through this and his other works was that of disillusionment with a society that, for all its range and power, was losing its values and often preferred elaborate sham to genuine human worth. In the end, Conrad's conclusion seemed to be that only at sea could human integrity both maintain itself and achieve anything meaningful. This was made explicit in ***Chance*** (1913), the only one of Conrad's major novels to be set in roughly equal proportions in normal English society and at sea. The extraordinary atmosphere of constriction and gloom so memorable in ***The Secret Agent*** (1907) and ***Under Western Eyes*** (1911) came partly from their subject matter—the ruthless and far-reaching power of the Russian police state and, in the latter, the terrorism this created in its opponents. But the taut, dark atmosphere of the two novels was due also to Conrad's tight-lipped recognition that the claims of devotion and integ-

Conrad's disillusionment with society

rity, though as real as ever, were, in a modem society, either pathetic and ineffective or ruthlessly overriden.

If this note of frustration, almost of despair, at the fate of human ideals in a modern world was Polish-born Conrad's response to the society of his time, it was significant that American expatriate Henry James showed much the same feeling. The world James depicted—in a series of novels that were an outstanding literary product of the turn of the century—was the world of Edwardian plutocracy (the great country manor, the great town house); and repeatedly his novels revealed the sordid reality beneath the elegant surface of contemporary life: greedy connoisseurship in *The Spoils of Poynton* (1897); calculating promiscuity in *What Maisie Knew* (1897) and *The Awkward Age* (1899); materialism and selfishness in *The Wings of the Dove* (1902) and *The Golden Bowl* (1904). One value not shown explicitly in Conrad's work was conspicuous in James's: a conviction of the importance in life of art in its highest sense. This was a reminder of something lying behind his whole effort as a writer. In his deep and discriminating study of the great 19th-century French novelists, in his attention to a range of other European writers, and in his intent, sustained reflection on the task and art of the novelist, James marked a departure from the general character of English literature in the mid-19th century.

The influence of the continental novel in England

At that period, English literature had proceeded mainly along its own lines, independent of trends elsewhere; but by 1900 a steady tide of influence was beginning to flow from the mainland of Europe. The desire of James and Conrad to excel as artists was a sign of the influence of a self-conscious and sophisticated writer such as Gustave Flaubert. All in all, the influence of the Continent was pervasive at this time. Ford Madox Ford, especially in *The Good Soldier* (1915), displayed an economy of narrative and intricacy of structure that must be related to French originals. Somerset Maugham, especially in *Of Human Bondage* (1915), showed an affinity with Émile Zola and Flaubert; and John Galsworthy, in *Villa Rubein* (1900), for example, was influenced by Ivan Turgenev. Arnold Bennett also had an outstanding affinity with Zola, as his posthumously published *Journals* (1932–33) made explicit.

Other strands were interwoven with these continental affiliations. The influence of Charles Dickens was marked in Conrad, both in idiosyncracy of characterization and in the sinister and macabre quality of some of his most remarkable scenes. More surprisingly, in the novels of George Moore, Ford, and Bennett there was often a warmth of character or inclination to sentimentalize that undoubtedly recalled Dickens. Interest in continental writing, moreover, did not lead to mere slavish imitation: James studied Flaubert and learned from him but on balance condemned him; Arnold Bennett studied Zola while resolving to avoid what he saw as Zola's lack of sympathy and humanity. Nor, of course, was there a break with earlier English tradition and continuity. James saluted George Eliot as the great master of English fiction, and Bennett acknowledged how much he owed to Herbert Spencer, the Victorian sociologist. Bennett's achievement, though minor compared with that of James, was a genuine one. More than any other novelist of the early 1900s, he revealed the other side of Edwardian opulence: the drab and deprived life of the lower classes, whose emergence into political and social power during the next 50 years was to have such an effect on writing itself. In this context it has been convincingly argued that it was to Bennett—and to H.G. Wells, who, in his realistic fiction, much resembled him—that many novelists writing of everyday lower middle class life from the 1950s onward would look back.

Continental influence affected the manner of two other novelists whose own work had international repercussions. James Joyce (see below *World War I's impact and the years to World War II*) in his early writings—the sketches and stories in *Dubliners* (1914)—showed a link with the aesthetic movement of the 1890s in his occasional melancholy, spare preciosity, and sense of form, while these stories and *A Portrait of an Artist as a Young Man* (1916) also show a link with French Naturalism. Another early and lasting continental influence was that of Ibsen's realism. All these, combined with the influence of French Symbolism and the experimental poetic techniques developed along with Ezra Pound and T.S. Eliot, were among the strands from which *Ulysses* (1922) and *Finnegans Wake* (1939) were woven. Like that of Joyce, the main work of D.H. Lawrence belonged to a somewhat later period, and his intense individuality was always conspicuous in it; but admiration for late-19th-century Russian fiction, then becoming available in English for the first time, was visible in the humane sensitiveness and fluid organization of his work.

The early development of modem poetry. How "modern" poetry emerged in the early years of the century is not wholly clear; and it was certainly more than a swift and drastic substitution, in the first decades of the century, by two American-born poets, Ezra Pound and T.S. Eliot, of realism and intelligence for escapist romanticism. Throughout these years there was a clear tradition of verse that dealt with everyday life and used everyday language—characteristics of much modern verse—and there was no lack of vitality in poems written about this time by Hardy, Edward Thomas, and D.H. Lawrence. The poetry that appeared in Georgian anthologies, however, does offer a useful contrast to what was to become a dominant 20th-century poetic idiom. These anthologies arose out of a feeling among poets who about 1910–11 were beginning to complain that their poetry was neglected in comparison with that of established poets—Robert Bridges, John Masefield, and Lascelles Abercrombie, for example. Rupert Brooke and Edward Marsh decided there was a need to make new poetry more accessible to the public and, with Harold Munro, editor of *The Poetry Review*, John Drinkwater, and Wilfrid Wilson Gibson, they planned the anthologies. They aimed to make people aware that poetry was being written and could be bought, read, and enjoyed. Brooke and Marsh applied the name "Georgian" to suggest that a new poetic age had opened with the accession of George V (1910).

The Georgian poets

Though their work remains readable, the poets represented in the anthologies did not wholly bear out the promise of a new age. The real gifts of Rupert Brooke, W.H. Davies, Walter de la Mare, Edmund Blunden, and Ralph Hodgson should not be overlooked; but, taken as a whole, too much of the Georgians' work was lifeless or repetitive. As a result, "Georgian" came to be used in a sense not intended by its progenitors: rooted in its period and looking backward rather than forward. Even the work of Hardy and Edward Thomas hinted at weaknesses that later proved decisive: both were inspired by nature and the countryside, but it was already to some extent a nostalgic inspiration, for to most young English poets nature could no longer be at this time a source of deep inspiration. For this reason, the manner and themes of these poets easily became, in the hands of others less gifted, diluted and middlebrow; and it was against this that the avant-garde authors of the period reacted.

The work of D.H. Lawrence was not in total contrast with that of the Georgian group, but from the start his insistent originality and responsiveness to experimentation outside England showed in his verse as in his prose. His earliest poems were written in traditional forms; but from about 1915 his original and cosmopolitan outlook led him to reject conventional lyric forms for the sake of a blunt yet sometimes delicate spontaneity, and it was this that made him one of the first major English poets to use free verse. As developed in France, *vers libre* (free verse) meant a defiance of the rigid exigencies of French metre, which was determined by the number of syllables. According to *vers libre* theory, what mattered was musical cadence rather than the number of syllables, and consequently the prosodic unit became the phrase rather than the foot or the line. Lawrence responded to this continental influence but in his own distinctive way: for him free verse was a means of embodying the unconstrained and vital immediacy of experience.

Although embracing more than versification, the movement called Imagism also had as one of its principles "to

The Imagist movement

compose in sequence of the musical phrase, not in sequence of the metronome." The most important writers associated with the movement were Pound and Eliot; T.E. Hulme, a philosopher, aesthetician, and poet whose emphasis in his essays on the precise use of words and the psychological effect of juxtaposing concrete images provided the movement's metaphysical background; F.S. Flint, a poet; Richard Aldington, a poet and novelist who contributed to all four Imagist anthologies and who was on the staff of *The Egoist,* unofficial organ of the group; and his wife, the poet Hilda Doolittle ("H.D."). The first "Imagist manifesto" (in *Poetry,* Chicago, 1913) was based on ideas of brevity, rejection of rhetoric, and truth to perception. The movement itself was short-lived, though its effects were considerable. It showed one more aspect of a varied and self-conscious reaction, about 1910, against rhetorical diffuseness and dilution.

William Butler Yeats had, before 1900, recognized something of this in 19th-century poetry and had declared that true poetry could be written for scholars or for peasants but for no one in between. In so saying he was responding to the indigenous peasant life of his own country, Ireland, on the one hand, and to France (as a source of literary ideals), on the other. He turned from English Victorian poetry to the uncompromising work of the French Symbolists. During the 1900s Yeats, both in poetry and in plays written for the Dublin Abbey Theatre, was breaking with literary mannerisms such as he had owed to the Pre-Raphaelites. He was turning, however, not to the "ordinary life" or "ordinary speech of the middle class but to a diction he found among the peasantry or the landed Ascendancy of Ireland and to an austere ideal of poetic drama remote from the everyday.

From about 1910 onward, Pound and Eliot constituted a sophisticated minority movement that looked to the Symbolists and to the French poet Jules Laforgue as models; and their poetry must be seen as a deliberate reaction against more conventional work accepted by a large middlebrow audience. Moreover, it turned emphatically away from nature and the countryside to represent attitude and interests of a metropolitan intellectual elite.

World War I's impact and the years to World War II. Even as early as in Eliot's *Waste Land* (1922) something was to be observed in literature that was fundamentally different in kind from what had gone before. What this poem attempted lay quite outside the interests and range of the Symbolists. It sought to depict the plight, indeed, the breakdown, of a whole civilization. This fundamental change of direction could be seen in the sharp break between early war poets such as Julian Grenfell and Rupert Brooke and later ones such as Siegfried Sassoon, Wilfred Owen, and Isaac Rosenberg. The early poets of World War I inherited their attitude from literature about wars in the past; what dominated the work of the later poets was a tragic sense of the sufferings that this war had brought, or angry satire directed against those who endorsed and prolonged it. In their poems written after 1920, Pound and Eliot shared the concern that mankind had reached a time of crisis and that the urgent call on the poet was to confront it. The poet's utterance rose from social satire almost to prophecy. Yeats, who had a long personal association with Pound at this time, was moving in the same direction, as was plain in poems such as "The Second Coming."

World War I had a decisive effect on fiction as well as on verse, though its impact showed itself in rather different ways. The early years of the century saw a resurgence of political liberalism, which enjoyed a great victory in the election of 1906 and which in the following years was responsible for the Britain of social welfare and material advance that was to become a pattern for the future. This liberalism was reflected in the intellectual life in the great houses of leading aristocratic families, and it had a certain affinity with more radical groups such as those of the Fabian Society. In general, the writers and intellectuals who formed this circle—Beatrice and Sidney Webb, George Bernard Shaw, Virginia Woolf, and E.M. Forster—were characterized by the term Bloomsbury group, in reference to the university area of London in which many of them lived and worked. In culture the word liberal during these years pointed to a whole movement of emancipation, humanity, and concern for the higher and finer values of life.

This intellectual liberalism was reflected in the novels of E.M. Forster and Virginia Woolf, whose values were those of the true liberal intellectual. Nonetheless, their work leaves an impression—clear even in a novel such as Forster's *Where Angels Fear to Tread* (1905) from before World War I—that there was a weakness in their position that might make it untenable in the postwar world. In Forster, as in Virginia Woolf's first postwar novel, *Jacob's Room* (1922), a concern for freedom, individuality, and the development of the personality was combined with a fear that these qualities lacked strength to survive in the violent modern world. Even in these authors' best work some readers discerned a lack of vitality or conviction.

It would be doctrinaire, however, to see Virginia Woolf as primarily a representative of liberalism. *Mrs. Dalloway* (1925) might show a consciousness of tyranny over the individual; *To the Lighthouse* (1927) might show the self-doubts and tensions of the intellectual; *Between the Acts* (1941) might hint at how the life of a nation is affected by its environment and history. But, in each case, what gave these books their distinction and power was their compassionate awareness and a rare delicacy of response to the world of intuitive consciousness and spontaneous feelings.

The novels of Aldous Huxley signalled a new stage in the anxieties of a liberal intellectual confronted by the modern world. Huxley seemed to doubt not only the ability of liberal culture to survive but the reality of that culture itself: his portrayals seemed like the work of one who had lost faith in what he depicted and who was preoccupied with its destruction at the hands of the violent and irrational, especially in his early novel *Crome Yellow* (1921). His brilliance and lively, witty display of ironic intelligence only emphasized the emptiness that he found at the centre of a culture to which he still belonged. In comparison with those of Forster and Virginia Woolf, his novels seemed to many to lack the sense of values, the humane responsiveness to life, and (in the case of Woolf) the sensuous poetry that made their work positive in its ultimate quality.

The sometimes almost hysterical satire of Huxley's later novels, such as *Eyeless in Gaza* (1936), had a counterpart in the cynical attitude to liberal culture of Evelyn Waugh, particularly in such satirical novels as *Decline and Fall* (1928), *A Handful of Dust* (1934), and *Brideshead Revisited* (1945). It was only in the context of this developing attitude that feelings conveyed by the novels of Graham Greene (*Brighton Rock* [1938], *The Heart of the Matter* [1948], *The End of the Affair* [1951], for example)—that man's life, if lacking in a sense of religion and, more particularly, of sin and a need for grace, was essentially drab, futile, and horrifying—could be rightly understood.

A very different writer, T.E. Lawrence, the famous "Lawrence of Arabia," is remembered in the literature of the age for *Seven Pillars of Wisdom* (1926), a dryly mannered but powerful and authentic autobiographical account of the Arabian struggle for independence from the Turks, in which he played a leading role.

Some of the earlier novels of Wyndham Lewis, such as *The Apes of God* (1930), with their disgusted sense of the triviality underlying much cultivated or "intellectual" life, were reactions to social change in something of Huxley's manner. But Lewis' work registered a more deeply felt and pondered awareness of the realities of a modern, machine-age society than did Huxley's, as may be seen in *Tarr* (1918) and more fully developed in *The Childermass* (1928), *The Human Age* (1955–56), and *Self-Condemned* (1954). Lewis saw the machine and city as menacing individual personality because something in man himself accepts and responds to mechanical control. His insight into fundamentals of modern life is at once intellectualized and humane, and it constituted a notable positive achievement.

Influences on Joyce

James Joyce. Joyce's first affiliations were with Dublin and with the 1890s; and something of the conviction that art was and should be a self-contained world of its own remained through all his work. Some of his later realism he might have owed to Ibsen or to such French Realist fiction as that of Zola. But his unique style seemed to have emerged mainly from contact with Pound and Eliot and the French writers who had influenced them. In his later prose ***(Ulysses*** [1922], ***Finnegans Wake*** [1939]) Joyce crystallized the reaction from middlebrow writing toward a style in which art made no compromise with the reader. Ideas come one after another in apparently unpremeditated but highly organized juxtaposition. The effect was to present what seemed like the disorganized thinking of a character who was bringing his thought only partially into conscious focus. In developing this style Joyce created a dense texture of associations that purported to represent the mind at work. Yet ***Ulysses*** was not only a work of stylistic virtuosity: Joyce's constant experimentation with language created a total impression of daily life in a large city (Dublin) that had few if any parallels in literature and that was also outstanding for its interpenetrating humour and humanity.

D.H. Lawrence. Coming from England's north Midlands, where rural life was fast giving place to industrialism, D.H. Lawrence was influenced by Thomas Hardy, who shared something of his experience of the passing of an old way of life. His earlier novels frequently rehearsed his own development: the central character outgrows provincial origins and acquires a modern, urbanized, or cosmopolitan consciousness. This was clearest in ***Sons and Lovers*** (1913) and ***The Rainbow*** (1915).

Lawrence's hostility toward modernity

As his acquaintance with modernity extended, Lawrence seemed to become more and more aware of its dangers and to set himself more and more urgently to find some solution for them. At the core of his novels lay a conviction that men were being deprived of traditional sources of strength, balance, genuineness, and dignity and that they could replace these only by rediscovering the source of life-giving spontaneity that they all possessed in their own innermost selves. Lawrence's treatment of sex was subordinate to this; for him, it was in their sexual lives that men could be most genuinely themselves and least the puppets of a machine age. This was revealed most clearly in short stories such as "The Captain's Doll" and "The Fox" and a short novel such as ***St. Mawr.*** From this conviction—that the truest, most unassailable source of vitality resides in the psyche—Lawrence was naturally drawn toward technical experiments not altogether unlike those practiced by Pound, Eliot, and Joyce. In the fluid, episodic construction of his later novels (notably ***Women in Love*** [1920]), with their seemingly disjointed characters who yet have an intense reality and individuality, he too displays something of the trends of the time.

But Lawrence, as he himself recognized, sought the deepest values of life, though he did not always master them. His successes came out of the intensity of his effort, but his work often displayed a bullying insistence that, when truest to himself, he rejected. Yet it is easy to forget his faults and to remember instead his astonishing vitality, often sharply perceptive intelligence, immensely real and tender sense of life's potentialities for good, and passionate stand for the real against the sham.

W.B. Yeats. Like Lawrence, Yeats too was forced to a reappraisal of a local and traditional culture by the threat of modernity. His earliest work was a product of Celtic legend, Irish peasant life, and the aesthetic movement of the 1890s. But more and more he was drawn into the modern world: first, by the part he played in the Irish dramatic movement (see below ***Anglo-Irish);*** then by the "troubles" in Ireland before and after 1920; and, more generally, by an awareness that the values that he most cared for were being threatened by "new commonness" on every hand.

Yeats's interest in the occult

Troubled more and more deeply after 1920 by a sense of living in a hostile world, Yeats turned again, but with a disciplined intensity, to the occultism and esoteric philosophies that had interested him as a young man. It was out of this material that he developed his later views, expressed in ***A Vision*** (1925), asserting that human vitality was in itself an answer to what was dehumanizing in modern life and looking back to the ordered society of the 18th-century Irish Ascendancy, based on a way of life rooted in the land but essentially aristocratic, rhetorical, and stylized.

Other writers. Other authors between 1920 and 1950 responded to what they sensed as a period of crisis in human affairs. L.H. Myers, especially in ***The Root and the Flower*** (1935), showed Lawrence's concern for what was real and what was sham in life, though he lacked Lawrence's gift for embodying a philosophy in narrative form. Joyce Cary, in such novels as ***The Horse's Mouth*** (1944) and ***A Prisoner of Grace*** (1952), attempted to deal with the disintegrating impact of war and modernity but reasserted the power of the simple human emotions. and especially of social and family affection, to overcome it. George Orwell, despite a conventional background of Eton and service in Burma in the Anglo-Indian police, showed perceptiveness and vigour in semi-autobiographical writings about the life of the underdog ***(Down and Out in Paris and London*** [1933], ***The Road to Wigan Pier*** [1937]), works remarkable for simple directness of style, free from all "literariness" and formality. By a repudiation of the upper class, empire-building ethos of the "English gentleman," he prepared the way for much of interest in later English writing. His postwar novels, ***Animal Farm*** (1945) and ***Nineteen Eighty-four*** (1949), rendered unforgettable by a deep general pessimism about the future of mankind, reflected more particularly a disillusionment with Stalinist Russia as the embodiment of Socialism. The Hungarian-born Arthur Koestler, who migrated to England in 1940, was another notable writer of fiction and essays with marked political interests. ***Darkness at Noon*** (1940) is a powerful novel of the Stalinist purges, and in his autobiography ***(Scum of the Earth*** [1941]) and essays (*e.g.*, ***Reflections on Hanging*** [1956], ***The Act of Creation*** [1964]) Koestler shows greater affinities with continental traditions than most English writing of the period does.

The poets of the 1930s

Although the best known poets of the 1930s followed Pound and Eliot in their satire and terse colloquialism, for the most part they did not continue the Symbolist side of their work. This partial lack of continuity was an indication of the new social and political conditions that were being reflected in the writing of the period before World War II. From the ***New Signatures*** anthology (1932) onward, W.H. Auden, Stephen Spender, Louis MacNeice, and C. Day Lewis were writing verse with a markedly left-wing political involvement—a reflection of the period of the Spanish Civil War, idealism about the Soviet Union before the "purges," and the political Popular Front in France. More diffuse than T.S. Eliot, and more traditional than his work up to that time, their verse revealed the poetry of Gerard Manley Hopkins (written 1884–89; published 1918) as an important influence in technique and that of Thomas Hardy as influential for its plainness of diction and lyricism. But Hopkins' obscurity, compression, and bold linguistic experimentation were not followed, so that his influence was seen mainly in a rhythmic fluency derived from a use of irregular metrical patterns called "sprung rhythm." While these poets of the 1930s used plain language to reach a wider audience rather than as a sophisticated revolt against poetic orthodoxy, much of their verse sacrificed depth of vision and particularity of language for a too journalistic or too doctrinaire engagement with politics. MacNeice, after World War II, finally found a new medium in radio drama. Spender, in some ways a more personal poet than the others, increasingly turned his gaze from the external, topical situation to subjective experience and, after 1939, was better known for his perceptive criticism and editorial association with the influential reviews ***Horizon*** and ***Encounter*** than as a poet. Day Lewis, the least adventurous of the poets of the 1930s, reverted in later work to a quiet, meditative, and discursive style, more "Georgian" than contemporary. Auden—who, like Christopher Isherwood (see below ***Drama in the first half of the 20th century***), later became a United States citizen—continued to develop significantly as a poet (see below *Poetry after* 1945).

The decline in technical experimentation

The changing scene after 1930. Broadly speaking, the avant-garde movement represented by Joyce, Pound, the early Eliot, Virginia Woolf, and D.H. Lawrence was only briefly dominant. L.H. Myers and Orwell made no original contribution to technique, and Cary was, in the main, a traditional novelist, writing in the chronicle form and owing much to Dickens. But since the political poets of the 1930s were more interested in content and wide readership than in form, after 1930 both fiction and verse began to make a significant break with the immediate past and to regain some continuity with 19th-century and earlier trends in English writing. The period of the Continent's strongest influence was over. Almost imperceptibly, new writing came to be coloured by something more popular in quality and to read as if written for an audience both larger and at the same time more limited and insular in outlook.

T.S. Eliot's conversion to the Church of England in 1927, and his profound study during this period of Dante, marked something of a turning point in both his prose and his poetry. Emphasis on literature as art and on the poet's need to break with the past (expressed in an essay, "Tradition and the Individual Talent" [1917], and in *The Sacred Wood* [1920]) was replaced by concern for the interrelation of literature, society, religion, and culture (*The Idea of a Christian Society* [1939], *Notes Towards the Definition of Culture* [1948]). And in the poetry from "Ash Wednesday" (1930) to *Four Quartets* (1943), Eliot's reappraisal of the crisis of contemporary civilization in terms of his own religious experience led to a widening of philosophic outlook as well as to a new feeling for English writing of the 17th and 18th centuries. Many people who had acknowledged Eliot's earlier greatness, but who did not accept his later religious and ecclesiastical opinions, all the same found in his later poetry a fundamental spiritual experience that had not been present in the earlier, more experimental work.

Pound and the earlier Eliot had been part of a small elite group writing in reaction to an effete literary "Establishment." The same ends were being served by such critical reviews of the time as *The New Age* (1908–21), which published much of Pound's early criticism. The literary movement of the first 30 years of the century comprised a comprehensive set of new principles about writing expressed by such philosophic critics as T.E. Hulme and a new evaluation of English literature based on these principles.

In time, these minority views, in becoming those of the majority, tended to be changed, partly because of an expansion of higher education and, in consequence, the growth of a sizable reading public of a new kind. This educational change had been brought about by social change: the advance of the lower middle and working classes to positions of much greater prominence and influence in the life of the country. Literature after 1930 was coming to be written and read by those with a new outlook and new interests that were more earnest and more straightforward than what had gone before. The reaction, led by Pound and Eliot, against the "ivory tower" of the aesthetic movement became more taken for granted. That literature should be intelligent and "common-sensical" also became an accepted view, as did the concept of fiction as mainly interesting for the light it shed on moral values and on how men ought to live. These changes were powerfully advanced by the critical writings of F.R. Leavis, an acute, uncompromising, passionately engaged critic whose influence was diffused, from the 1930s onward, through his teaching at Cambridge and through *New Bearings in English Poetry* (1932), the magazine *Scrutiny*, and *The Great Tradition* (1948). A somewhat distinct critical movement, begun in the 1920s by I.A. Richards with *The Principles of Literary Criticism* (1924) and *Practical Criticism* (1929), was equally directed toward the view that literature and its study were social disciplines, capable of close, quasi-scientific analysis.

The new criticism

Partly because the changes were already underway, World War II had no effect on writing comparable to that of World War I. Only three war poets (all of whom died in the war) achieved even minor distinction: Alun Lewis, with a few poems that showed promise of an assurance he did not live to develop; Sidney Keyes, whose *Collected Poems* (1945) made an impression greater than their achievement warranted; and Keith Douglas, who, of the three, was the only one to show an original poetic response to the fact of war and a potential greatness.

But none of these poets made more than a passing impact. The pervasive literary transformation was already taking place as a reflection of a social transformation. The political poets of the 1930s were lucid, even diluted, by comparison with Pound and Eliot. William Empson, George Barker, and Dylan Thomas were the most important poets of that decade to write verse that was in any way "difficult." Of the three, Barker and Thomas were the most visionary and apocalyptical and also the most uninfluenced by the mainstream of English poetry in the 1930s.

Drama in the first half of the 20th century. Developments in English drama in the early years of the 20th century showed a complex interaction of indigenous forces and continental influence. A vogue for Henrik Ibsen's drama of social comment came to dominate the English theatre in the 1890s and strongly influenced the early plays of George Bernard Shaw. As he went on, Shaw's material increasingly diverged from that of Ibsen: the incisive, if hilarious, political involvement and topicality of the early plays (such as *Man and Superman* [first performed 1905] and *Major Barbara* [1905]) changing into the much more timeless and speculative interests of the later ones (such as *Back to Methuselah* [1922], *Saint Joan* [1923], and *Geneva* [1938]). James Joyce's *Exiles* (1918) was another play to show Ibsen's influence. John Galsworthy, in *The Silver Box* (1906), *Strife* (1909), and *Justice* (1910), at least succeeded in showing that a didactic aim and a serious theme—that of social justice—could be good theatre. Harley Granville-Barker, with some notable works, especially *The Voysey Inheritance* (1905), treated the conflict between public and personal moral responsibility with intellectual subtlety and feeling. The "Manchester school" of playwrights wrote plays that at their best were both realistic and good theatre—Stanley Houghton's *Hindle Wakes* (1912), for example. But the most important development of the period was in poetic drama; and the greatest contribution came from Ireland (see below *Anglo-Irish*).

The revival of verse drama

The revival of verse drama was continued in the 1930s by T.S. Eliot (*Murder in the Cathedral* [1935], *The Family Reunion* [1939]). His lead was followed by other writers, including Auden and Isherwood, who collaborated in political verse plays, notably *The Dog Beneath the Skin* (1935) and *The Ascent of F 6* (1936). Eliot inspired a group of religious poets and verse dramatists, including Ronald Duncan (*This Way Is the Tomb* [1945]), Norman Nicholson (*The Old Man of the Mountains* [1946]), and Christopher Fry. Fry, influenced by Charles Williams, a Christian apologist with unusual power to present theological ideas symbolically in an everyday setting (such as *The Place of the Lion* [1931]), developed a colourful if somewhat precious style of his own in his best plays, notably *The Lady's Not for Burning* (1948).

Directions of writing after World War II. *Poetry after 1945.* Empson and Thomas represented a recession rather than a continuance of the early Pound–Eliot line. Later poets were influenced by Empson's criticism (*Seven Types of Ambiguity* [1930], *The Structure of Complex Words* [1951]) as well as by his poetry, but they adopted his laconic wit and nonchalant tone rather than his cramped intense feeling and intellectual ingenuity. Many who seemed to have learned something from Empson, such as John Wain, Donald Davie, and D.J. Enright, drew as much or more from the poet, novelist, and critic Robert Graves. Graves, especially in poetry, preserved a continuity with an earlier tradition of English writing. In his novels the creative use of myth and history sprang from idiosyncratic erudite interests (as he outlined in *The White Goddess* [1948]); in criticism he was a pioneer of close analysis. Like W.H. Auden, however, he to some extent withdrew from the battle.

Auden's later poetry, like that of T.S. Eliot, was partly religious in inspiration and showed a highly individual interest in historical learning and linguistic virtuosity (*About the House* [1965], *City Without Walls* [1969]). It maintained a continuity with his earlier poetry by its development of a certain lyrical quality and by use of a vocabulary and syntax reminiscent of Thomas Hardy rather than Pound or Eliot. Philip Larkin stated his belief that Hardy was the great and authentic poet of the 20th century; and Larkin's own work, especially in *The Less Deceived* (1955) and *The Whitsun Weddings* (1964), showed the influence of Hardy and Edward Thomas as well as a restrained but often lyrical quality of "Englishness." Roy Fuller (*Collected Poems* [1962], *New Poems* [1968]) illustrated, especially in his later verse, the terse and fluid elegance of this trend at its best.

Characteristics of the "new poetry"

By the 1960s shrewd common sense, carefully calculated tone, disciplined craftsmanship, and sharp intelligence had become dominant in new English verse, particularly that included by A. Alvarez in a collection, *The New Poetry* (1957), which contained work of important British poets of the period after World War II. Three of these, Thom Gunn (*Fighting Terms* [1954], *My Sad Captains* [1961], *Touch* [1967], among others), Ted Hughes (*The Hawk in the Rain* [1957], *Wodwo* [1967], *Crow* [1971]), and Davie (*Essex Poems* [1969]), were generally ranked with Larkin as being outstanding poets of this period, although writing in very different styles. Hughes showed the influence of D.H. Lawrence's early verse in much of his poetry, while Gunn showed the influence of continental and American writing. Later judgments rated American-born Sylvia Plath no less highly as an outstanding poet writing in England during this period. The wife of Ted Hughes, she achieved in her early delicate work, *The Colossus* (1960), distinction that foreshadowed the achievements of the posthumously published *Ariel* (1965) and *Crossing the Water* (1971), in which she confronted the most painful and terrifying possibilities of death and oblivion with skill and intelligence.

The work of Charles Tomlinson (*A Peopled Landscape* [1963], *The Way of a World* [1969]) showed the influence of the U.S. poet Wallace Stevens, but without the latter's flamboyance and philosophizing. Tomlinson's best work was notable for its scrupulous precision of workmanship and delicate integrity of vision. Elizabeth Jennings shared with other poets of the period a quality of plainness, clarity, and directness in such collections as *A Way of Looking* (1955) and *Collected Poems* (1967). Another widely admired poet of the 1950s and 1960s was a Welsh clergyman, R.S. Thomas (*Song at the Year's Turning* [1955], *Pietà* [1966]), who resembled Wordsworth in technique but whose abrupt and tight-lipped utterance placed him in the mainstream of the poetry of the time.

Postwar fiction. It became clear, after 1945, that a second literary revolution was in progress, though one less spectacular and on the whole less interesting than that of the 1910s. The change may perhaps be seen by a comparison of Virginia Woolf with Elizabeth Bowen (*The Death of the Heart* [1938], *The Heat of the Day* [1949], *Look at All Those Roses* [1941]). The latter's feeling and subject matter were not unlike Virginia Woolf's, but her framework of ideas was slighter and her technique much more conventional. Only one earlier novelist of high distinction continued into the 1950s and 1960s something of the originality and the technical innovation of an earlier generation. This was Ivy Compton-Burnett (*Pastors and Masters* [1925], *A House and Its Head* [1935], *Parents and Children* [1941], *The Present and the Past* [1953]). Ostensibly her narrow but deeply worked achievement is to deal closely with upper middle class family life in the countryside of late Victorian England. But character and incident are intensely stylized, and the dry wit of a ruthlessly sustained dialogue suggests an elaborate hidden world of aggression, vanity, deprivation, and tangled family affection and hatred. As a result, her work is both timeless and intensely dramatic.

Ivy Compton-Burnett has a limited link with the stylishness and wit of such early-20th-century novelists as Frederick Rolfe (*Hadrian the Seventh* [1904] or Ronald Firbank (*Concerning the Eccentricities of Cardinal Pirelli* [1926]) and ultimately with a 19th-century novelist such as Meredith. But this, in suggesting certain continuities for her work, also shows her isolation. Postwar, the movement away from avant-garde and also from upper middle class fiction became clear. Angus Wilson (*The Wrong Set* [1949], *Hemlock and After* [1952], *Anglo-Saxon Attitudes* [1956], *No Laughing Matter* [1967]) dealt with current middle class conditions and predicaments with mingled irony and compassion, but through loosely episodic plots, characters reminiscent of Dickens, and a general adherence to traditional modes. C.P. Snow (*Strangers and Brothers* [1940], *The Masters* [1951], *Corridors of Power* [1964]) and William Cooper (*Scenes front Married Life* [1961]) also wrote fiction that was traditional in form, dealing with the postwar middle class world of bureaucracy and social life, though Snow did not achieve Cooper's lightness and colour. The latter's *Scenes from Provincial Life* (1950) broke new ground in its relaxed but witty and frank portrayal of the life of young and obscure people in the provinces, though it cannot be seen merely as a forerunner of others. Cooper's relaxed satire came out of an attitude of general acceptance, but several later novelists, such as Kingsley Amis (*Lucky Jim* [1954], *Take a Girl Like You* [1960], *The Anti-Death League* [1966]) and John Wain (*Hurry on Down* [1953], *Living in the Present* [1955], *The Smaller Sky* [1967]), who emerged in the 1950s seemed to be in deliberate revolt against earlier avant-garde ambitions, hostile to anything that smacked of the "highbrow" or "phoney." As *a* result, they and others came collectively to be known as the "Angry Young Men." Forerunners of these attitudes can be found in the lower class provincial setting and in the rejection of "highbrow" values found in such early-20th-century writers as Bennett and Wells.

The "Angry Young Men"

A striking development of the later 1950s and early 1960s was the emergence of an authentic working class or near-working class fiction. It was remarkable that this took so long to happen, but with the arrival of John Braine (*Room at the Top* [1957], *Life at the Top* [1962]), Alan Sillitoe (*Saturday Night and Sunday Morning* [1958], *The Loneliness of the Long Distance Runner* [1959]), Keith Waterhouse (*Billy Liar* [1959]), Sid Chaplin (*The Day of the Sardine* [1961]), Stan Barstow (*A Kind of Loving* [1960]), and above all David Storey (*This Sporting Life* [1960]) a vigorous fiction dealing with working class life from something like a working class point of view came into being. There may have been little or no innovation in technique, but this far from negligible development should be seen alongside the assertion of a north-of-England cultural distinctiveness, the increased opportunities for working class children to take part in higher education at all levels, and the impact of a book like Richard Hoggart's picture of present-day English working class culture, *The Uses of Literacy* (1957). These novelists may have developed from Wain, Amis, and the rest in what concerns their attitudes and technique, but whereas the earlier group largely dealt with "opting out" of middle class situations, the later one concerned itself with the process of moving upward in working class life or with moving over into the middle class.

Certain other novelists may be loosely linked—sometimes even against their own professions—with this traditional and realist mode of writing. Thus Lawrence Durrell, author of *The Alexandria Quartet* (four novels, 1957–60), claimed that his fiction was based on the scientific principle of relativity, though little in his superficially scintillating work seemed to bear this out. Durrell's scintillations result chiefly from setting his fiction in an exotic foreign city, partial relic of the imperialist phase of British history. In almost the same years Anthony Burgess was writing his best work, *A Malayan Trilogy* (1956–59), in which satire, wit, vivid portraiture, and marked descriptive power embody a deep sense of the passing of the British "raj." These works, however (like Waugh's African novels or Graham Greene's *The Heart of the Matter* [1948] or *The Quiet American* [1955], set in Southeast Asia), also belong to a realist tradition in fiction. Malcolm Lowry's opulent and complex Mexican novel *Under the Volcano* (1947) is a notable work that also indirectly reflects the decline of the "raj." Burgess' varied output includes more experimental work, like the savagely violent near-farce *A Clockwork Orange* (1962).

The link with Iris Murdoch's work is more tenuous. An academic philosopher by early profession, her first novel (*Under the Net* [1954]) showed the influence of French Existentialist thought, while later works, such as *The Flight from the Enchanter* (1956), *The Bell* (1958), and *An Unofficial Rose* (1962), seemed less cosmopolitan. Her attempts to deal with the movements of deep human feeling (*e.g.*, *Bruno's Dream* [1969]), her poetic suggestiveness, idiosyncratic fertility of character and incident, and sometimes memorable use of archaic myth (*A Severed Head* [1961], *The Unicorn* [1963]) make her probably the most interesting and complex of English novelists to appear since the war.

With Iris Murdoch, however, the trend moved toward another continuity in fiction, one more romantic, philosophical, or indeed fantastic, which was less conspicuous because in the nature of the case it was more diverse—yet one that came to stand recognizably over against the "realist" continuity. Here, the Powys brothers, early in the century, were undoubtedly key figures. The most important novels of J.C. Powys (*Wolf Solent* [1929], *A Glastonbury Romance* [1932], *Weymouth Sands* [1934]) were set largely in "Wessex" and to some extent recalled the work of Thomas Hardy; but they showed originality in their variety and profusion, eloquent if sometimes artless power and depth, and intricate development of an almost transcendental awareness. Powys also had a certain kinship with Lawrence (who was in fact younger than he, though so much less long-lived). T.F. Powys, his brother, was part of the same development (*Mr. Weston's Good Wine* [1927]), as was L.H. Myers, with his Indian trilogy *The Root and the Flower.*

As the avant-garde movement receded, J.C. Powys came to seem less isolated and more important. Mervyn Peake's *Titus Groan* (1946), *Gormenghast* (1950), and the sketchier but no less remarkable *Titus Alone* (1959), written during his last illness, all displayed powers of rich description and fantastic but intense characterization in what can now be seen as an ambitious allegory of decline and modernity. Andrew Sinclair's erudite, bawdy rhetorical fantasy *Gog* (1967) seems to stem from Powys, though other novels (*My Friend Judas* [1959], *The Breaking of Bumbo* [1970]) are more conventional; a similar distinction shows in the fiction of John Fowles, as between his ambitious myth-novel *The Magus* (1966) and *The Collector* (1965) or *The French Lieutenant's Woman* (1969).

In this general context J.R.R. Tolkien's elaborately topographical "fairy tale," *The Lord of the Rings* (1954–56), may also be mentioned. Much of William Golding's work (*Free Fall* [1959], *The Spire* [1964], *The Pyramid* [1967]) has been notable for its imaginative and poetic qualities, though his most assured success is *The Lord of the Flies* (1954), which converts the children's desert-island adventure story into a horrifying moral fable of human evil nature and degeneration.

Certain writers stood outside the general picture. The fiction of the Irish-born, French-writing Samuel Beckett (see below *French literature in the 20th century*), in its compassion, anxiety, and daring stylistic experiments, looked back to the work of his friend James Joyce and to the innovators of the earlier part of the century. Another distinctive postwar novelist was Anthony Powell, whose saga *The Music of Time* (1951–) and earlier groups, including *Afternoon Men* (1931), showed the two aspects of his development. *Afternoon Men* had affinities with the satire of Evelyn Waugh and was concerned with bohemian London in the 1930s, while *The Music of Time* sequence recalled Marcel Proust and *The Alexandria Quartet* of Durrell. Powell's characteristically uppercrust, eccentric, or seedy world was unique in the fiction of the period.

Postwar drama. After World War II a significant change in English drama set in. It became Eliot's ambition to reintroduce a kind of verse drama into the commercial theatre. His position was thus already remote from that which lay behind his earlier verse; and later plays (*The Cocktail Party* [1949], *The Confidential Clerk* [1953]) were to be regarded as a distinctive kind of work for the commercial theatre rather than as verse drama in a more ambitious or poetic sense. Thus, even he, in his later drama, illustrated a movement away from avant-garde ideas of the 1920s toward a style more traditional in kind and general in appeal.

T.S. Eliot's later plays

From the late 1950s on, however, the most interesting theatrical developments were in prose, and to some extent went parallel with those in the novel. A number of young dramatists, for example, encouraged by the English Stage Company at the Royal Court in West London, Joan Littlewood's Theatre Workshop in East London, and several provincial theatres, rejected the middle-class interests and values of drawing-room comedy and depicted everyday attitudes and values in the postwar life of the common people. John Osborne's *Look Back in Anger* (1956)—followed by *The Entertainer* and other plays down to *West of Suez* (1971)—was a landmark at the beginning of this change. Other playwrights were Arnold Wesker (*Chicken Soup with Barley* [1958], *Roots* [1959]), whose loosely organized, panoramic plays often depended on explicitly left-wing political attitudes; John Arden (*Live Like Pigs* [1958], *Serjeant Musgrave's Dance* [1959]); Shelagh Delaney (*A Taste of Honey* [1958]); Charles Wood (*Cockade* [1965]); and Harold Pinter (*The Caretaker* [1960], *The Homecoming* [1965]), whose often highly stylized "absurdist" dialogue and action suggested a connection with Eugene Ionesco. Something similar may be said of Anne Jellicoe's richly inventive *The Sport of My Mad Mother* (1956) and *The Knack* (1961), where singing, games, and nonsense are woven into a complex of psychological perception. Freer methods of stage production, such as multiple simultaneous scenes and other innovatory techniques (*e.g.*, Wesker's *Their Very Own and Golden City* [1966], Donald Howarth's *A Lily in Little India* [1966], or Henry Living's *Eh?* [1965] and *The Little Mrs. Foster Show* [1969]) were also cause, as well as effect, of freer dramatic writing.

By the middle 1960s, direct social reporting and protest were in part giving way to plays that (though mostly in prose) showed more poetry, fantasy, and even "absurdist" stylization. Giles Cooper (*Happy Family* [1967]) and Peter Terson (*The Mighty Reservoy* [1970]) may be noted among many such writers, and others of interest were Tom Stoppard (*Rosencrantz and Guildenstern are Dead* [1968]), Peter Shaffer (*Five Finger Exercise* [1958], *The Royal Hunt of the Sun* [1964]), David Storey (*Home* [1970], see also above *Postwar fiction*), Peter Nichols (*A Day in the Death of Joe Egg* [1968], *The National Health* [1969]), and Joe Orton (*Entertaining Mr. Sloane* [1965], *Loot* [1968]). In this respect the influence of writing for radio drama was important. One of the earlier outstanding radio plays was Dylan Thomas' *Under Milk Wood* (1953). In later years one may note Barry Bermange's crisp but nightmarish *No Quarter* (1968) and Pinter's intensely dramatic interior-monologue piece, *Landscape* (1968). Robert Bolt's popular stage success *A Man for All Seasons* (1960) was one of many plays to succeed as films. Television also produced dramas of interest in the work of Dennis Potter (*Stand Up Nigel Barton* [1968], *Vote Vote Vote for Nigel Barton* [1968]) and John Hopkins (*Talking to a Stranger* [1967]), as well as popular successes like Jeremy Sandford's *Cathy Come Home* (1971). In all these interrelations between drama and the mass media, and with the increasing dominance of the visual dimension in much dramatic art, lie possibilities of change that may profoundly affect future literature—and perhaps increase its transient appeal while reducing its lasting value.

Television drama and film

Conclusion. A general view of the literary and cultural scene in Great Britain in the early 1970s revealed, in general, limited achievement and reason for disquiet. In all kinds of creative writing in the 1950s and 1960s it had seemed easy enough to find young writers of interest; but hopes of their maturing and consolidating their work were often disappointed, and an emergence of any genuinely major talent was lacking. This situation was perhaps associated in part with absence of literary continuity; older writers of outstanding achievement had fallen silent or, like Auden and Graves, had for many years been expatriates. Related perhaps to this limited literary achievement was the comparative poverty of nonliterary writing of a kind that provides an atmosphere in which creative writing itself may flourish. In the first half of the century there had been many prose writers whose work in one way or another bore strongly and directly on society and culture as a whole. The philosopher G.E. Moore, the psychologist Havelock Ellis, the critic and aesthetician John Middleton Murry, the philosopher and political theorist Bertrand Russell, the economist J.M. Keynes, the economic historian R.H. Tawney, the historian G.M. Trevelyan were all outstanding in a broad spectrum of intellectual and cultural life. T.S. Eliot was a major critical and cultural as well as poetic influence; F.R. Leavis' writings had marked significance in the field of social thought. But, in the second half of the century, philosophy, economics, and criticism all became more academic or technical; sociology seemed to replace social philosophy; and good political writing had long been rare. Thus, despite much good writing, the cultural climate as a whole was fragmented, specialized, lacking in the authority of profound nd widely meaningful achievement and tending to be indifferent or inimical t [illegible] ti life. The relative lack of major literary talent or achievement has to be seen in this perturbing context.

Scottish literature in English and Scots. Scottish prose fiction continued its traditional emphasis on physical description and reliance on action rather than on psychological subtleties, as was seen in the stories of John Buchan. More original were the witty and often fantastic narratives of Eric Linklater; the passionately written ***A Scots Quair*** (1932–34) of Lewis Grassic Gibbon; the Celtic visionary novels of Neil Gunn; and the sharp character studies of Robin Jenkins. The drama, revived to some extent by James Barrie (as in ***Dear Brutus*** [1917]), discovered a distinctively native voice in James Bridie, whose entertaining plays combined the humorous, ethical, and supernatural in a very Scottish manner.

Hugh MacDiarmid's poetry

It was in poetry that the greatest changes occurred. In the second decade of the 20th century the explosive personality of Hugh MacDiarmid became a centre and motive force of what has been called the Scottish renaissance. This movement, by re-expanding the Scots vocabulary after its stagnation following the great lyric poet Robert Burns, aimed to restore the intellectual prestige of Scots and break through the barrier created by the domination of the English language. MacDiarmid, in his encyclopaedic, polemical, metaphysical verse, himself wrote the best lyrics since Burns. He was followed by others, including Sydney Goodsir Smith and Douglas Young.

Of poets using English, Edwin Muir, with his meditative, myth-haunted verse, was the most distinguished, while Norman MacCaig, W.S. Graham, Edwin Morgan, George Mackay Brown, and Iain Crichton Smith wrote with craftsmanship and force. In a class by itself was the delicate experimental poetry of Ian Hamilton Finlay.

Although the Scottish sense of separateness was too real to be swamped, English was dominant in prose and educated speech, and the modern Scots poet needed to try many different forms in order to find his characteristic mode.

Anglo-Irish. At the end of the 19th century a new movement became apparent. In the 1890s two Irish dramatists appeared: Oscar Wilde and George Bernard Shaw. Although they were outstanding dramatists in the Anglo-Irish tradition, neither wrote of or for Ireland. A new movement, closely allied with a strong political nationalism, was growing quickly. This movement brought about a literary renaissance in both Irish and English. The latter, led by the poet William Butler Yeats, was deeply influenced by the work of Douglas Hyde, whose translations from the Irish were the model for the stylized peasant dialect used by Lady Gregory. Yeats and Lady Gregory were the founders of the first national Irish theatre, the Abbey Theatre, which soon made its name through the magnificent dramas of John Millington Synge, notably ***Riders to the Sea*** (1904) and ***The Playboy of the Western World*** (1907), which used the dialect evolved by Hyde and Lady Gregory with great beauty and power. Later, the theatre became absorbed in realism, mostly rural, apart from Sean O'Casey's three great dramas of the Dublin slums: ***The Shadow of a Gunman*** (1923), ***Juno and the Paycock*** (1924), and ***The Plough and the Stars*** (1926).

The plays of J.M. Synge

In poetry, besides Yeats there was George Russell, a mystic and patriot. Notable among their younger contemporaries were Padraic Colum, author of the beautiful "Drover," one of the finest poems in the language, and Seamus O'Sullivan, founder and editor of ***The Dublin Magazine***. The Republican movement also had its poets: P.H. Pearse, Thomas MacDonagh, and Joseph Mary Plunkett, all executed in 1916 after the Easter Rising.

In the novel there were historical tales by Emily Lawless and Standish James O'Grady and imaginative stories by James Stephens. By contrast there was also George Moore, who attached himself for a time to the Irish cultural movement; his irreverent account of its leading figures in ***Hail and Farewell*** (1911–14) was on the whole more readable than his novels.

By far the greatest of modern Irish writers was James Joyce, whose work, although it was complex and cosmopolitan in inspiration (see above ***English literature in the 20th century***), was firmly rooted in the daily life of Dublin. Like his friend Samuel Beckett (see below ***French literature in the 20th century***), who wrote mainly in French, Joyce left Ireland but never lost the essential "Irishness" of his character.

Writers such as Sean O'Faolain, who remained in Ireland after the achievement of independence in 1921, were most successful when treating the waning of the nationalist struggle or the failings of Irish Catholicism, as in O'Faolain's ***Bird Alone*** (1936) and ***Come Back to Erin*** (1940). In general, the short story, as written by Frank O'Connor and Liam O'Flaherty, was more successful. One exception, a writer who showed the direct influence of James Joyce, was Flann O'Brien (pseudonym of Brian Ó'Nualláin), who in a novel, ***At Swim-Two-Birds*** (1939), created a brilliant and under-rated experimental work.

Dramatists, while continuing to write on Irish themes, became more cosmopolitan in outlook and technique. At the Dublin Gate Theatre, founded in 1928, Micheál MacLiammdir appeared in and designed nearly 300 productions, including his own ***Ill Met by Moonlight*** (1946). Donagh MacDonagh broke new ground with ballad comedies in the style of the German dramatist Bertolt Brecht, but the most outstanding playwright was Brendan Behan, with a tragicomedy, ***The Quare Fellow*** (1954), and with ***The Hostage*** (1957).

Among poets, though no successor to Yeats emerged, Patrick Kavanagh showed originality and realism in ***A Soul for Sale*** (1947) and ***Come Dance with Kitty Stobling*** (1960); and Thomas Kinsella, in ***Another September*** (1958) and ***Nightwalker*** (1967), showed unusual mastery of subject and style. (See above ***English literature of the 20th century*** for separate mentions of Yeats and Joyce and other Irish-born writers.)

CELTIC LITERATURE IN THE 20TH CENTURY

Welsh. The establishment of the University of Wales (1872–93) widened literary horizons. Men like Sir Owen M. Edwards, Sir John Morris-Jones, and Emrys ap Iwan (Robert Ambrose Jones) made the Welsh conscious of their literary identity and set new standards for correctness of language and sincerity of thought. The great literary revival that followed was marked by T. Gwynn Jones's masterly use of the old strict metres to express modern thought and W.J. Gruffydd's lyrical use of the free metres to express his rebellion against society and his love for the countryside of his youth. R. Williams Parry showed a superb gift of poetic observation, while Sir Thomas Parry-Williams combined a mystical love for his native Gwynedd with an almost scientific analysis of his own instincts and emotions in relation to man's place in the universe. Older poets, such as Cynan (A. Evans-Jones), William Morris, and Wil Ifan (William Evans), clung to earlier lyrical models, although many others, like D. Gwenallt Jones and Saunders Lewis, drew increasingly on the rhythms and vocabulary of colloquial speech. Waldo Williams, Gwilym R. Jones, the younger Bobi Jones, and particularly Euros Bowen experimented with form and subject. Their work has been followed and developed by writers of the younger generation, the most distinctive and prolific being Gwyn Thomas. Most poets, old and young, reflect a varying involvement in contemporary Welsh political activity.

The high standard of the periodical Y *Llerzor* (1922–51) indicated the advances made in prose. Contributors were generally involved in a wide range of activities: its editor, W.J. Gruffydd, was both poet and essayist; Saunders Lewis was a poet, dramatist, and politician; Sir Thomas Parry-Williams a poet and essayist; and R.T. Jenkins an essayist and historian. Together with novelists and short-story writers like Tegla Davies, T. Rowland Hughes, Kate Roberts, and D.J. Williams, they effectively mirrored the contemporary Wales of which they were part. John Gwilym Jones and Islwyn Ffowc Ellis were innovators in form and subject matter, and they are being followed by an enthusiastic younger generation of writers of which none is, as yet, outstanding. As in other literatures, there has been an increasing preoccupation with, and an improvement in, literary criticism.

Drama in Wales began in the 20th century. At first realistic, it developed into poetic, symbolic drama, often based on historical and mythological themes but dealing with moral, social, and psychological contemporary problems. The outstanding names are Saunders Lewis, John Gwilym Jones, and, of the younger generation, Gwenlyn Parry.

Scottish Gaelic. A considerable amount of prose appeared in 19th- and 20th-century periodicals in the form of short stories and essays. Alongside these were numerous religious translations from the 17th century onward, including Calvin's catechism of 1631, Gaelic translations of the Old and New Testaments, Kirk's Psalter and his Irish version of the Bible, and the 1807 Gaelic Bible. Other translations included works by John Bunyan, Richard Baxter, Thomas Boston, and Philip Doddridge. Among original prose writers in the earlier part of the century were the Rev. Donald Lamont, Donald Mackechnie, and Angus Robertson. The most notable modern short stories have appeared from Colin Mackenzie, John Murray, and Iain Crichton Smith.

Little vital poetry appeared in the 19th century, and a 20th-century movement to free Gaelic poetry from its traditional shackles began with the work of Sorley Maclean, George Campbell Hay, and Derick Thomson. Later work in this vein came from Donald MacAulay and Iain Crichton Smith.

Valuable collections of poetry date from the mid-18th century. One of the earliest was that of the Rev. Alexander Pope. Jerome Stone collected Ossianic ballads, while James Macpherson's ***Fragments of Antient Poetry*** (1760) also stemmed from this collecting zeal. The work was continued in the 19th century by later scholars and clerics—for example, the Rev. Alexander Cameron's ***Reliquiae Celticae*** (1892–94), which printed the Fernaig manuscript, parts of the ***Books of*** *Clanranald*, and some of the Edinburgh manuscripts for the first time. Gaelic proverbs were published with an English translation by Donald Macintosh (1785), a new edition by M. MacInnes appearing in 1951. Frances Tolmie's collection of folk songs (1911), Margaret Shaw's *Folksongs and* ***Folklore of South*** *Uist* (1955), and J.L. Campbell and F. Collinson's *Hebridean Folksongs* (1969) continued the work in the 20th century, while the ***Transactions of the Gaelic Society of*** *Inverness* and ***Scottish Gaelic Studies*** have published much material from manuscript and oral sources.

Irish Gaelic. An antiquarian interest in the Irish language had begun early in the 19th century, especially among the comparatively small educated classes who spoke English. The movement suffered because speakers of Irish were a minority, and the language of Irish nationalist agitation had been English. As a result, the revival was led by antiquarians and nationalists rather than by the linguistic minority itself. Irish today is a minority language in the republic, although it has a firm place in the educational system.

The revivalists

The revivalists were faced with the problem of writing a language that had split up into dialects. The result was considerable linguistic diversity that continued until the mid-20th century, when a governmental standard was introduced. Patrick Pearse, a leader of the 1916 Easter Rising, and Pádraic Ó Conaire, who introduced the short story into Irish. were the first writers who took their craft seriously. Tomás Ó Criomhthain's *tOileánach* (1929; *Islandman*, 1934), the life story of a Munster fisherman, was a unique document. The short story had been handled in masterly fashion by Liam O'Flaherty, who also wrote in English (see above ***Anglo-Irish),*** and Máirtín Ó Cadhain. Generally, however, both the novel and the drama were undistinguished, excepting Brendan Behan's *Giall* (1957; *The* ***Hostage,*** 1958). While some lyric poetry was produced. no poet of Yeats's stature emerged.

Breton. The 20th century saw the publication of Jabez Riou's short stories in ***Geotenrz ar Werc'hez*** (1934; "The Virgin's Herb"), which were remarkably concise and unexpectedly combined sensitivity with profound skepticism. Youenn Drezen wrote one of the best modern Breton novels, *Itron* ***Varia Garmez*** (1942; "Our Lady of the Carmelites"). In drama the work of Tanguy Malmanche (d.1953) dominated the first half of the century, and it was compared with that of Paul Claudel and J.M. Synge. His work was carried on by Pierre Hélias, a talented younger writer who had written hundreds of radio plays and had had stage plays produced and whose writing was both popular and extremely polished. In poetry an outstanding figure was Jean-Pierre Calloc'h, killed in action in 1917, whose poems were published with a French translation in 1921 as *Ar en Deulin* ("Kneeling").

The struggle to resist the encroachment of French words and to make Breton self-sufficient continued, with the encouragement of many literary reviews and the work of scholars such as François Vallée, a lexicographer. In 1941 a new standardized spelling was introduced, not wholly successfully, and in 1951 Breton was given a place in state schools.

AMERICAN LITERATURE IN THE 20TH CENTURY

Writing from 1914 to 1945. Important movements in drama, poetry, fiction, and criticism took form in the years before, during, and after World War I. The eventful period that followed that war left its imprint upon books of all kinds, for it was a time when writers were much involved with interpreting life about them. Literary forms of the period were extraordinarily varied, and in drama, poetry, and fiction leading authors tended to experiment with radical innovations in technique.

Experiments in drama. Although drama in the 19th century had not been a pre-eminent form, no type of writing embodied wider experimentation than a new drama that arose as a result of rebellion against the glib commercial stage. In the early years of the 20th century, Americans travelling abroad found a vital theatre flourishing in Europe; returning home, some of them became active in founding a Little Theatre movement in every corner of their country. Freed from commercial limitations, playwrights experimented with dramatic forms and methods of production and in time producers, actors, and dramatists appeared who had been trained in college classrooms and community playhouses. Some Little Theatre groups became commercial producers; for example, the Washington Square Players, founded in 1915, which became the Theatre Guild (first production in 1919). The drama that resulted was marked by a spirit of innovation and by a new seriousness and maturity.

The plays of Eugene O'Neill

Eugene O'Neill, the most admired dramatist of the period, was a product of this movement. He worked with the Provincetown Players before his plays were commercially produced. His dramas are remarkable for their range. *Beyond the Horizon* (first performed 1920), *Anna Christie* (1921), *Desire Under the Elms* (1924), and *The Iceman Cometh* (1946) were Naturalistic works, while *The Emperor Jones* (1920) and *The Hairy Ape* (1922) made use of the Expressionistic techniques developed in German drama in the period 1914–24. He also employed a stream-of-consciousness form in *Strange Interlude* (1928) and produced a work of subtle psychological analysis in *Mourning Becomes Electra* (1931).

No other dramatist was as generally praised as O'Neill, but many others wrote plays of a high order that reflected the growth of a serious and varied drama. Marc Connelly wrote touching fantasy in a Negro folk biblical play, *The Green Pastures* (1930). Like O'Neill, Elmer Rice made use of both Expressionistic techniques *(The Adding Machine* [1923]) and Naturalism *(Street Scene* [1929]). Beginning as a Realist, Maxwell Anderson turned to verse drama in plays such as *Elizabeth the Queen* (1930) and *Winterset* (1935) and then to musical comedy satire in *Knickerbocker Holiday* (1938). Robert Sherwood produced a distinguished body of work, writing comedy *(Reunion in Vienna* [1931]) and tragedy *(There Shall Be No Night* [1940]). Clifford Odets, in *Waiting for Lefty* (1935), a plea for labour unionism, utilized auditorium as well as stage for action and in *Awake and Sing* (1935) wrote in the vein of Naturalism. Thornton Wilder used stylized settings and poetic dialogue in *Our Town* (1938) and turned to fantasy in *The Skin of Our Teeth* (1942).

Robinson and Frost

The new poetry. Poetry ranged between traditional types of verse and experimental writing that departed radically from the established forms of the 19th century. Two New England poets, Edwin Arlington Robinson and Robert Frost, who were not noted for technical experimentation, both won critical as well as popular acclaim in this period. Robinson, whose first book appeared in 1896, found sonnets, ballad stanzas, and blank verse satisfactory to his thought. In the 1920s he won three Pulitzer Prizes—for his *Collected Poems* (published 1921), *The Man Who Died Twice* (1925), and *Tristram* (1927). Like Robinson, Frost used traditional stanzas and blank verse in volumes such as *A Boy's Will* (1913), his first book, and *North of Boston* (1914), *New Hampshire* (1923), *A Further Range* (1936), and *A Masque of Reason* (1945). The best known poet of his generation, Frost, like Robinson, saw and commented upon the tragic aspects of life and the complexities of human existence and was skeptical of pat solutions.

Just as modern U.S. drama had its beginnings in little theatres, modern U.S. poetry took form in little magazines. Particularly important was *Poetry: A Magazine of Verse,* founded by Harriet Monroe in Chicago in 1912. The surrounding region soon became prominent as the home of three poets: Vachel Lindsay, Carl Sandburg, and Edgar Lee Masters. Lindsay's blend of legendary lore and native oratory in irregular odelike forms was well adapted to oral presentation, and his lively readings from his works contributed to the success of such books as *General William Booth Enters into Heaven, and Other Poems* (1913) and *The Congo, and Other Poems* (1914). Sandburg wrote of life on the prairies and in Middle Western cities in Whitmanesque free verse in such volumes as *Chicago Poems* (1916) and *The People, Yes* (1936). Masters' very popular *Spoon River Anthology* (1915) consisted of free-verse monologues by village men and women, most of whom spoke bitterly of their frustrated lives.

Writing traditional sonnets and brief, personal lyrics, Edna St. Vincent Millay and Sara Teasdale were innovative in being unusually frank (according to old standards) for women poets. Three fine Negro poets—James Weldon Johnson, Langston Hughes, and Countee Cullen—also found old molds satisfactory for dealing with new subjects, specifically the problems of their race. In general, however, the range of experimentation was great. While Conrad Aiken experimented with poetical imitations of symphonic forms often mingled with stream-of-consciousness techniques, e.e. cummings used typographical novelties to produce poems that had surprisingly fresh impact. Marianne Moore invented and brilliantly employed a kind of free verse that was to make her one of the most distinguished voices in modern U.S. poetry. Stephen Vincent Benét, in *John Brown's Body* (1928), produced a stirring novel in verse. Robinson Jeffers used violent imagery and modified free or blank verse to express perhaps the most bitter views voiced by a major poet in this period.

Pound, Eliot, and the influence of the 17th-century Metaphysical poets

Except for a period after World War II when he was confined in St. Elizabeth's Hospital, Washington, D.C., Ezra Pound had lived outside the United States since 1908. He had had, nevertheless, a profound influence on 20th-century writing in English, both as a practitioner of verse and as a patron and impresario of other writers (see above *English literature in the 20th century).* His most controversial work, which he was said to be revising in the early 1970s, remained *The Cantos,* the first installment of which appeared in 1925 and the latest in 1959 *(Thrones: 96–109 de los cantares).*

Like Pound, to whom he was much indebted, T.S. Eliot lived abroad most of his life, becoming a British citizen in 1927. His first volume, *Prufrock and Other Observations,* was published in 1917. In 1922 appeared *The Waste Land,* the poem by which he first became famous. As a poet and critic, Eliot exercised a strong influence, especially in the period between World Wars I and II. In what some critics regard as his finest work, *Four Quartets* (1943), Eliot explored through images of great beauty and haunting power his own past, the past of the human race, and the meaning of human history.

Eliot was the acknowledged master of many members of a varied group of poets whose work was indebted to 17th-century English Metaphysical poets, especially to John Donne. Eliot's influence was clear in the writings of Archibald MacLeish, whose earlier poems were similar both in manner and thought to *The Waste Land.* In later poems MacLeish voiced a positive belief in social advance that contrasted with the religious attitude advocated by Eliot. A number of Southern poets showed a loose affiliation and metaphysical influence; among them were John Crowe Ransom, Donald Davidson, and Allen Tate. Hart Crane had a similar metaphysical manner but a subject matter of his own. Other American metaphysicals having individual qualities of thought and method were Louise Bogan, Léonie Adams, Muriel Rukeyser, Delmore Schwartz, and Karl Shapiro.

The influence of the little magazines

Fiction. The little magazines that helped the growth of the poetry also contributed to a development of the fiction of the era by printing material that diverged from older patterns and by attacking conventional writing and established writers. The most powerful of the little magazines were two edited by the journalist-critic H.L. Mencken—The *Smart Set* (editorship 1914–23) and *American Mercury* (which he co-edited between 1924 and 1933). One of Mencken's favourites, James Branch Cabell, sprang to fame with *Jurgen* (1919), a novel that attacked America's orthodoxies and institutions. Other authors whom Mencken favoured launched "a revolt against the village," an attack on the narrow, provincial quality of much rural American life. The most distinguished of these was Sherwood Anderson, who dealt with the subject in two collections of short stories, *Winesburg, Ohio* (1919) and *Triumph of the Egg* (1921).

In 1920 a new school of fiction rose to prominence with the success of books such as F. Scott Fitzgerald's *This Side of Paradise* and Sinclair Lewis' *Main Street.* Thereafter, writers tended toward Realism and Naturalism—frank portrayals of contemporary life. Portrayal of character and motive was much influenced by the psychology of Sigmund Freud and others. In the decades that followed, fiction voiced reactions to changing times:

novels of the 1920s expressed disillusionment and protest against established institutions and ideologies. Some of those of the 1930s protested against the economic and political system. World War II led many novelists to see qualities of excellence in American life not before realized.

Critics of society. F. Scott Fitzgerald's ***This Side of Paradise*** (1920) showed the disillusionment and moral disintegration of post-World War I America. His masterpiece, ***The Great Gatsby*** (1925), was a more poignant and unified development of the same theme. Like Fitzgerald, Sinclair Lewis was best as social critic. His onslaught against the "village virus" (***Main Street*** [1920]); against average businessmen (***Babbitt*** [1922]); against materialistic scientists (***Arrowsmith*** [1925]); and against the racially prejudiced (***Kingsblood Royal*** [1947]) were satirically sharp and thoroughly documented. Similar careful documentation characterized James T. Farrell's Naturalistic "Studs Lonigan" trilogy (1932–35). Richard Wright protested against the plight of the Negro in ***Uncle Tom's Children*** (1938) and ***Native Son*** (1940). Particularly admired as a novelist of protest was John Dos Passos, who first attracted attention with an anti-World War I novel, ***Three Soldiers*** (1921). His ***Manhattan Transfer*** (1925) and the "U.S.A." trilogy (***The 42nd Parallel, Nineteen Nineteen,*** and ***The Big Money*** [1930–36]) employed various narrative innovations to attack society from the left. His later books, attacks on leftists, had less merit. A bitter vision of an inhuman and brutal world and a black-comedy style distinguish Nathanael West's novels, ***The Dream Life of Balso Snell*** (1931), ***Miss Lonelyhearts*** (1933), and ***The Day of the Locust*** (1939).

Hemingway, Faulkner, Steinbeck. Three Nobel laureates whose writings showed a shift from disillusionment were Ernest Hemingway, William Faulkner, and John Steinbeck. Hemingway's early short stories and first novels, ***The Sun Also Rises*** (1926) and ***A Farewell to Arms*** (1929), were full of the disillusionments of the postwar years, but the Spanish Civil War led him to believe in the possibility of collective action to solve social problems, and his novels ***To Have and Have Not*** (1937), ***For Whom the Bell Tolls*** (1940), and ***The Old Man and the Sea*** (1952) embodied the new belief.

William Faulkner brilliantly handled point-of-view, stream-of-consciousness techniques, and complex plots in a body of work concerned chiefly with a mythical Mississippi county, Yoknapatawpha. Such novels as ***The Sound and The Fury*** (1929), ***As I Lay Dying*** (1930), ***Light in August*** (1932), and ***The Hamlet*** (1940) presented a generally grim picture of the human condition, though more optimistic convictions emerged in ***Intruder in the Dust*** (1948) and ***The Reivers*** (1962).

Steinbeck's career, marked by uneven achievements, began with a historical novel, ***Cup of Gold*** (1929), wherein he voiced the distrust of society, the glorification of the anarchistic individualist typical of the rebellious 1920s. Later, however, a belief in the possibilities of collectivist action was implied in ***In Dubious Battle*** (1936) and ***The Grapes of Wrath*** (1939).

Lyric fictionists. An interesting development in fiction was a movement toward poetry. The tendency to select details and endow them with symbolic meaning, to stress thoughts and emotions, and to use rhythmical prose gave fiction more of a lyrical quality. In varied ways, Hemingway, Steinbeck, and Faulkner all showed the trend. Lyricism was prominent in such writings of Willa Cather as ***My Ántonia*** (1918) and ***Death Comes for the Archbishop*** (1927). Katherine Anne Porter, writing more in the style of the Metaphysical poets, used the stream-of-consciousness method in ***Flowering Judas*** (1930) and ***Pale Horse, Pale Rider*** (1939) with the complexity and the symbolic sophistication of poetry. Another leading poetic fictionist was Thomas Wolfe, the author of four large novels that in effect were one long lyrical recording of the author's life: ***Look Homeward, Angel*** (1929), ***Of Time and the River*** (1935), ***The Web and the Rock*** (1939), and You ***Can't Go Home Again*** (1940).

Criticism. The first half of the 20th century was noteworthy for its achievements in literary criticism. The period began with a battle between the New Humanists—a group supporting the older values in judging literature—and a group who urged that new standards be adopted. The Humanists were led by Irving Babbitt, who in books and vigorous essays attacked the modern tendency toward Naturalism as a vicious influence. The leader of the opposition was the pugnacious H.L. Mencken, who, notably in ***Prejudices*** (1919–27), claimed that the duty of writers was to present "the unvarnished truth" about life. In the end, the progress of Naturalism was assured, and literature was liberated from a number of ancient restrictions.

The New
Humanists
and their
opponents

In this period of social change, it was natural for a number of critics to consider literature's relationships to society and politics. They consistently judged books as reflections of society or as expressions of social truth. Van Wyck Brooks and Vernon Parrington wrote greatly admired studies that embodied such an approach. Two outstanding critics, Kenneth Burke and Edmund Wilson, stressed social influences less and illuminated literary works by analyzing them and relating them to their creators' sensibility.

After World War II. The United States emerged from World War II as a world power, and its literature correspondingly took on a new worldwide interest. Three of the nation's leading novelists received Nobel Prizes—Faulkner in 1949, Hemingway in 1954, and Steinbeck in 1962—but in each instance the award was presumed to have been made on the strength of work done in earlier decades. Other distinguished writers were also still on the scene, notably the poets Robert Frost, Wallace Stevens, Marianne Moore, e.e. cummings, William Carlos Williams, and Ezra Pound. But the most interesting work of these writers seemed to lie behind them. The period 1955–72 was marked by the deaths of Frost, Stevens, Faulkner, Hemingway, Cummings, Williams, and Pound. Impending for postwar writing, then, was a changing of the guard, but there was little agreement as to who were the most worthy successors. The literature produced between 1945 and the 1970s was characterized by a great diversity of style, goals, and accomplishments.

Fiction. The multiplicity and variety of postwar writing was nowhere more evident than in the novel and short story. Many writers who became prominent in the United States after World War 11—Henry Miller, Vladimir Nabokov, Eudora Welty, Mary McCarthy, John Cheever, Bernard Malamud, Ralph Ellison, William Burroughs, and Saul Bellow, for example—had very little in common except their ages: they all had been born before 1920. What was striking about these authors was the variety of their achievements. Miller came into his own as something of an American classic in the 1960s after a Supreme Court decision allowing the publication of his works in the United States. His ***Tropic of Cancer*** (1934) and ***Tropic of Capricorn*** (1939) contained uninhibited portrayals of sexual experiences that could, in the 1960s, be seen as characteristic of the general vitality and candour of Miller's prose. Nabokov, who became a U.S. citizen in 1945, had long been a master of high style and intricate literary effects before he won recognition with the publication of ***Lolita*** (1955) and ***Pale Fire*** (1962). His other novels written in English—The ***Real Life of Sebastian Knight*** (1941), ***Bend Sinister*** (1947), ***Ada or Ardor: A Family Chronicle*** (1969), and ***Pnin*** (1957)—together with those originally written in Russian—The ***Defense*** (1930), ***Camera Obscura*** (1933; Eng. trans., ***Laughter in the Dark,*** 1938), ***Invitation to a Beheading*** (1935), and ***The Gift*** (1937), for example—secured Nabokov's reputation as a writer of first importance. Eudora Welty, usually grouped with Southern writers such as Faulkner, Flannery O'Connor, and Carson McCullers, achieved immediate recognition with her first collection of stories, ***A Curtain of Green*** (1941), and went on to produce a distinguished body of work, both short stories and novels, chiefly in the postwar period. For all her brilliance as a stylist, her work was neither mannered nor precious. Mary McCarthy was perhaps better known as a critic, but ***The Groves of Academe*** (1952), ***The Group*** (1963), and ***Birds of America*** (1971) were novels

Nabokov's
novels
in
English

that often reflected an acute satirical intelligence. John Cheever, a writer long associated with *The New Yorker* magazine, was the author of five books of short stories and three novels, the last of which, *Bullet Park* (1969), showed a distinct turning to a highly satirical point of view. Like Saul Bellow, Bruce Jay Friedman, Philip Roth, and Herbert Gold, Bernard Malamud wrote of specifically Jewish subject matter in such novels as *The Assistant* (1957) and *The Fixer* (1966) and in his collections of short stories, such as *Pictures of Fidelman* (1969). Ellison's subject was the travail of the Negro in America, searching for "his place in history." He had thus far written only one novel, *Invisible Man* (1952), but it attained the standing of a contemporary American classic. Burroughs' first book, *The Naked Lunch* (1959), was made up of fragments, many of them considered obscene, from the life of a drug addict. His experimental techniques in prose as well as his unique vision had a great influence on younger writers. Bellow's subject—most notably in *The Adventures of Augie March* (1953), *Herzog* (1964), and *Mr. Sarnrnler's Planet* (1970)—is modern man in urban America. His novels are marked by wit, intelligence, a high comic sense, and extraordinary craftsmanship.

There had been a general awareness among these and other American novelists that the categories, styles, and modes of thought traditional to American literature no longer applied with the old force. Novelists whose work had not been touched by this awareness had little or no effect on U.S. readers in the postwar period. That most American novelists appeared to be perceiving the world in radical flux was illustrated well in the work of Norman Mailer. Although regarded as one of the nation's most talented novelists, Mailer had been called a genius who had not yet found his subject. There would seem to be some truth in this view since he had written four novels —*The Naked and the Dead* (1948), *Barbary Shore* (1951), *The Deer Park* (1955), and *An American Dream* (1965)—on four different subjects in four utterly different styles. Two other important writers of Mailer's generation were James Jones and William Styron. Jones's *From Here to Eternity* (1951) and Styron's *Lie Down in Darkness* (1951) were remarkable first novels, but their succeeding efforts did not maintain such a high mark. In *The Confessions of Nat Turner* (1967), a controversial novel about an antebellum slave revolt, Styron reminded readers of his unusual imaginative powers.

Irony, satire, and "black humour" in fiction

If the postwar period in American fiction had a dominating style, that style was irony. Joseph Heller, in *Catch-22* (1961), attacked military bureaucracy and the inhumanity of war with irony as perhaps sole weapon. In *Goodbye, Columbus* (1959) Philip Roth, a writer of admirable comic talent, used irony in criticizing those of his fellow American Jews who had grown indifferent to the suffering of others. Later, in *Portnoy's Complaint* (1969), he created a brilliantly sustained comic performance. In short stories and a novel, *Snow White* (1967), Donald Barthelme demolished cliché in all aspects of modern life. J.P. Donleavy, in *The Ginger Man* (1955), presented a hero who, from all appearances, was totally irrational; but then the novel gradually revealed an entire world that was itself more than a little mad.

There arose among American novelists a group of writers—prominent among them were Terry Southern, Kurt Vonnegut, Jr., and Thomas Pynchon—whom critics labelled "black humorists." Essentially satirists, they were able to wring comedy out of the bleakest of situations.

John Updike, William Gass, John Barth, and Flannery O'Connor were novelists who fitted conveniently into no group or category. In *Rabbit, Run* (1960), *The Centaur* (1963), *Of the Farm* (1965), and *Rabbit Redux* (1971), Updike dealt with such traditional themes as adolescence, marriage, and the family, but his lyrical treatment of these themes set him apart from his contemporaries. In a single novel, *Omensetter's Luck* (1966), and a collection of short stories, *In the Heart of the Heart of the Country* (1968), Gass showed himself to be a unique stylist. Barth, in *The Floating Opera* (1956), *The End of the Road* (1958), *Giles Goat-Boy* (1966), and *Chimera* (1972), established a reputation as something of an American Nabokov. His *Sot-Weed Factor* (1960) was an outstandingly great imaginative tour de force. Flannery O'Connor (died 1964) demonstrated in her last book. a collection of stories, *Everything That Rises Must Converge* (1965), her continued growth as a wholly original talent. A Roman Catholic, she combined immense comic gifts with an essentially religious concern with corruption in the human heart.

In the midst of this profusion of talent, an insistent literary question of the period was whether the novel was a dying form. Those who believed that it was argued that sociologists had taken over much of what had once been an exclusive domain of the novelist—namely, the description of manners and milieu and the topical investigation of values. There was no doubt that nonfiction such as Truman Capote's *In Cold Blood* (1966), a chronicle of the murder of a Kansas family, and later works by James Baldwin and Norman Mailer aroused a great deal of interest (see below *Literary and social criticism).* Though there might be a paucity of fictional masterpieces during the period—the work of Saul Bellow comprised perhaps the only major fictional *oeuvre* of these years—novelists continued to produce work of high quality, and, so long as they did, the question concerning the vitality of the novel as a form came to seem increasingly academic.

Poetry. Poets who reached literary maturity after World War II moved in different directions. One group tended to forsake the obscurantism, .the Surrealism, the free forms of previous American poetry for dramatic or dialectical organization. Stanley Kunitz, little appreciated heretofore, won a Pulitzer Prize in 1959 for *Selected Poems 1928–1958*. Karl Shapiro, who was later to change his style decisively, was a Pulitzer Prize winner for *V-Letter and Other Poems* in 1945; Robert Lowell for *Lord Weary's Castle* in 1947; Richard Wilbur for *Things of This World* in 1957; W.D. Snodgrass for *Heart's Needle* in 1960; Alan Dugan for *Poems* in 1962; Louis Simpson for *At the End of the Open Road* in 1964; John Berryman for *77 Dream Songs* in 1965; and Anne Sexton for *Live or Die* in 1967. Two other distinguished poets in this group were Randall Jarrell and Theodore Roethke.

All the above-named poets could be said to represent the poetry establishment in postwar American literature. But there were other forces at work in American poetry. There was, for example, something of a Midwest renaissance, evidence for which was supplied by the appearance in 1967 of an anthology entitled *Heartland: Poets of the Midwest.* Allied with this trend was a rather loosely organized movement, led by the poet Robert Bly, that disdained English verse models and called for an image-driven poetry written in a wholly American idiom. The late William Carlos Williams proved an influence on a group known as the Black Mountain school (from a short-lived college of that name in North Carolina); its reigning doyen was Charles Olson and its leading younger light was Robert Creeley. On the West Coast, mostly in and around San Francisco, a group known as the Beats began to be heard from in the mid-1950s. Among a number of poets more or less associated with the Beats, the best known were Allen Ginsberg, Lawrence Ferlinghetti, and Gregory Corso.

Drama. Perhaps because of ever-increasing costs of production, which permitted only "smash hits" to prosper, most dramatic writing in the postwar period was unexperimental and unexceptional. The best young literary minds showed a decided lack of interest in the theatre, many preferring to write for films or television. So marked was this apathy that the Ford Foundation, in the early 1960s, began to offergrants to novelists and poets to study with theatre companies. The only significant theatre to come out of the project was Robert Lowell's anthology piece *The Old Glory* (performed 1964). In the main, these were not rich years for the American theatre.

Over the postwar period two playwrights predominated: Tennessee Williams and Arthur Miller. Williams' experimentation in *The Glass Menagerie* (1944), which used a narrator, unorthodox settings, and poetic dialogue, continued in *A Streetcar Named Desire* (1947), *Cat on a*

Hot Tin Roof (1955), *Orpheus Descending* (1957), *Suddenly Last Summer* (1958), and *The Night of the Iguana* (1961). Miller was experimental in a variety of ways in *All My Sons* (1947), *Death of a Salesman* (1949), *The Crucible* (1953), and *A View from the Bridge* (1955). By the 1960s, however, there were signs that the talents of both playwrights were wearing rather thin.

The situation among younger American playwrights was more fluid. One of the most promising, Lorraine Hansberry, whose *Raisin in the Sun* (1959) centred on Negro family life in Chicago, died young. Several younger dramatists seemed likely to have noteworthy careers —among them, Jack Gelber and Frank Gilroy—but then appeared to fade out rather quickly. Edward Albee, after early critical successes off-Broadway, established his reputation firmly with *Who's Afraid of Virginia Woolf?* (1962).

The rise of off-Broadway theatre

From a literary point of view, the greatest share of the excitement of American dramatic writing after the war shifted from Broadway to off-Broadway and, indeed, in many cases, from off-Broadway to "off-off-Broadway." Much avant-garde theatre in the 1960s was produced in lofts and cabarets. Toward the end of the 1960s, nudity became a theatrical issue, and everywhere improvisation, shock, and the Absurd were features of a good deal of theatrical writing. The inspiration for it came less from American than from older European playwrights—Samuel Beckett, Eugène Ionesco, and Bertolt Brecht—whose plays were much in vogue in the U.S after 1945.

Literary and social criticism. The poet Randall Jarrell once described the postwar era in American literature as the "age of criticism." At no other time in American literature, indeed, had so much criticism, both literary and social, been written. Yet not much changed in criticism itself. Yvor Winters, John Crowe Ransom, Cleanth Brooks, and others who had first made their marks in an earlier era still predominated.

In literary scholarship three American critics turned out biographies that were acknowledged masterworks. In his five-volume study of Henry James, *Henry James: The Untried Years, 1843–1870, The Conquest of London 1870–1881, The Middle Years, 1882–1895, The Treacherous Years, 1895–1901,* and *The Master, 1901–1916* (1953–72), Leon Edel produced a work that was at once subtle, masterly, and provocative. Ernest Samuels' three-volume biography of Henry Adams, *The Young Henry Adams*, *The Middle Years,* and *The Major Phase* (1948–64), demonstrated a cultivated and refined literary craftsmanship rare in any era. Richard Ellmann's biography of James Joyce (1959) was properly regarded as definitive.

The broadening concerns of the critics

Alongside more traditional critics in the United States after World War II arose another kind who tended to eschew the highly theoretical and preferred to deal with each new work or movement in a direct, topical, and often personal way. They did not restrict themselves to the exclusively literary. Lionel Trilling completed a major body of work whose concern was the relation between life and literature. History, social theory, and politics, as many critics saw it, also fell within their purview. Irving Howe established a reputation with *Politics and the Novel* (1957); Alfred Kazin, after producing a history of American literature, *On Native Grounds* (1942), turned to autobiography; Norman Podhoretz, who became editor of an intellectual journal, *Commentary*, in 1960, flatly announced in *Doings and Undoings* (1964) that issues brought to the surface by authors sometimes interested him more than their books.

Bridging the two critical traditions—the scholarly and the intellectually committed—was Edmund Wilson, critic, historian, playwright, novelist, social reporter, travel writer, chronicler. Wilson was the complete man of letters, and his death (in 1972) left a gap that did not seem soon likely to be filled.

Among others who turned away from an exclusive interest in literature was James Baldwin, a Negro novelist, whose *Notes of a Native Son* (1955) and *The Fire Next Time* (1963) combined autobiography and radical social criticism in elegant and powerful essays. Originally a novelist and poet, Paul Goodman became best known for a study of the young in America, *Growing Up Absurd* (1960), and for other polemical and utopian writings. Norman Mailer, in *The Presidential Papers* (1963), *The Armies of the Night* (1968), and *Miami and the Siege of Chicago* (1968), wrote some of the most stunning personal-political reportage in that genre's history. A novelist and poet, Robert Penn Warren took time out from literary work to compose *Who Speaks for the Negro?* (1965), a particularly forceful and intelligent book on the civil rights movement. Indeed, many problems and issues to arise in postwar America seemed so urgent that literary men often found themselves dealing with them directly in social criticism.

CANADIAN LITERATURE IN THE 20TH CENTURY

French-Canadian. The cultural history of Canada has been conditioned by the country's dual origin, so that a certain historical tension is a constant of Canadian civilization. Founded as New France in 1608, conquered by Britain in 1763, and exposed to the overflowing vitality of the United States since 1850, Canada, in its search for a national identity, has been variously the beneficiary and victim of these three cultural forces.

Within French Canada itself, two other internal pressures complicate the situation: the claustrophobic and the agoraphobic complexes. The former dominated the 17th, 18th, and 20th centuries, the latter holding sway in the 19th, when a strong urge to hold to the past animated official thought and literature. In terms of literary themes, these opposing tendencies are seen as the patriotic pride in the feats of the early adventurers and of 20th-century separatists on the one hand, and the regionalist praise of an ordered and settled rural society on the other.

History and related fields. François-Xavier Garneau laid the foundation for a later blossoming of both historical studies and creative literature; his *Histoire du Canada* (1845–52) reawakened a dormant pride of nation and inspired poets, novelists, and other historians of the day. Among his successors, Robert Rumilly wrote an ambitious chronicle, *Histoire de la province de Québec* (41 vol., 1940–69).

Various aspects of French-Canadian society were examined in J.-C. Falardeau's *Essais sur le Québec contemporain* (1953), Father P. Angers' *Problèmes du culture au Canada français* (1960), and M. Rioux and Y. Martin's *French-Canadian Society* (1964). Among many theologians and philosophers who contributed to religious thought was Louis-Adolphe Paquet, who developed the theme of the providential mission of the French in America. This theory cast strong reflections on regional literature earlier in the 20th century. Two comprehensive views of French Canada were to be found in I.F. Fraser's *Spirit of French Canada* (1939) and Mason Wade's *French Canadians: 1760–1967* (1968).

Poetry. The first literary movement of French Canada, the Patriotic school of 1860, had been imbued with a strong nationalism to which religious fervour was a natural adjunct. Literary inspiration was found principally in Victor Hugo and Alphonse de Lamartine. Later in the 19th century, Charles Baudelaire and the Symbolists gradually replaced Romantic models, and literary activity was oriented more around Montreal. In the 1930s, Surrealist influences began to predominate; after World War I1 the impact of Existentialism was felt, and poets gave an impression that they sensed and wrote with more independence than they had formerly.

Joseph-Octave Crémazie was the first important poet of French Canada. His patriotic poems were still popular early in the 20th century; Louis Fréchette won success with a patriotic epic entitled *La Légende d'un peuple* (1887), while other followers of Crémazie celebrated Canadian life, its picturesque customs, and religious faith. These poets were precursors of Le Terroir, or regionalist school of Quebec, who recorded humble aspects of French-Canadian life. Blanche Lamontagne-Beauregard, best known of the group, published *Par nos champs et nos rives* (1917) and *La Vieille Maison* (1920).

In 1895 Jean Charbonneau and Paul de Martigny founded the Montreal literary school. Canadian life, Canadian landscape, and patriotism were no longer the principal inspiration for these poets. Symbolists and aesthetes, they wished to live in the presence of beauty and liked to picture their souls as beautiful things, Émile Nelligan picturing his as a transparent ship sculptured in gold and laden with treasures, "Le Vaisseau d'or."

Paul Morin reflected European experience in the exotic poems of *Le Paon d'émail* (1911; "The Enamel Peacock"). Louis Dantin (real name Eugène Seers), the critical conscience of the group, contributed *Le Coffret de Crusoé* (1932; "Crusoe's Casket"), and Alfred DesRochers, author of *L'Offrande aux vierges folles* (1928; "The Offering to the Foolish Virgins"), expressed the gusty spirit of the north in his "Hymne au vent du nord" ("Song to the North Wind").

In 1934 Robert Charbonneau, François Hertel, and Saint-Denys Garneau founded a literary journal, *La Relève*. Charbonneau subsequently made a name for himself as a novelist and editor. For Garneau, author of *Regards et jeux dans d'espace* (1937), poetry was a means of probing despair; the importance of his *Journal* was only realized in the 1960s, when his intense feeling and suggestive technique began to exert a strong influence on many poets who followed him. Hertel, as seen in such titles as *Strophes et catastrophes* (1943), was a humorous man with a double vision, torn between sensuous delights and a consciousness of sin; the full spectrum of his poetry is found in *Poèmes d'hier et d'aujourd'hui* (1967; "Poems of Yesterday and Today").

Along with Saint-Denys Garneau, the respected poets of the older generations are Alain Grandbois, Anne Hébert, and Rina Lasnier. Alain Grandbois had a powerful grasp of cosmic images; his mastery was clearly maintained from his first collection in 1944 to his collected works, *Poèmes* (1963). Anne Hébert shows conscious poetic progress from *Les Songes en équilibre* (1942) to the surer and more positive *Poèmes* (1960) and *The Tomb of the Kings* (1967); the same Symbolist and Surrealist elements are to be found in her poetic prose volumes *Le Torrent* (1950) and *Kamouraska* (1970). One of the best poems of Rina Lasnier, a Christian poet looking for perfection, was "Malemer," in *Présence de l'absence* (1956); a later collection is *L'Arbre blanc* (1966; "The White Tree").

The Hexagone

The most important contemporary poetic movement, the Hexagone, was founded in Montreal in 1954, principally by Gaston Miron and Jean-Guy Pilon. Perhaps the best way to describe this dynamic group of poets and publishers is to say, at the risk of oversimplification, that they rediscovered the value of eroticism and combined it with a fierce love of Quebec; sometimes the loved one is indistinguishably the woman-country; tenderness and violence are inextricably bound together. A few of the principal names follow with titles of their representative works: Jacques Brault, *Mémoire* (1965); Gaston Miron, *La Vie agonique* (1964) and *L'Homme rapaillé* (1970); Paul-Marie Lapointe, *Choix de poèmes: Arbres* (1960) and *Le Réel absolu* (1971); Gatien Lapointe, *Ode au Saint-Laurent*, *J'appartiens à la terre*, and *Le Chevalier de neige* (1963); Roland Giguère, *L'Âge de la parole* (1965); Yves Prefontaine, *Pays sans parole* (1967); and Fernand Ouellette, *Poésie* (*poèmes 1953–1971*), (1972).

A comprehensive view of French-Canadian poetry can be found in a number of anthologies: *The Oxford Book of Canadian Verse in English and French* (1960), Guy Sylvestre's *Anthologie de la poésie canadienne française* (4th ed., 1963), J. Cotnam's *Poètes du Québec* (1969), and J. Glassco's anthology, *The Poetry of French Canada in Translation* (1970).

The novel. The history of fiction followed a course parallel to that of poetry. Romantic in mid-19th century, it lapsed into a long period of regionalism before the social novel developed later under Naturalist influences, providing an authentic picture of urban French-Canadian society.

According to its author, Philippe-Ignace Aubert de Gaspé, *Le Chercheur de trésor* (1837; "The Treasure Seeker") was the first Canadian novel, telling a story of adventure and black magic. The best-known 19th-century novel, by Philippe-Joseph Aubert de Gaspé, was *Les Anciens Canadiens* (1863; *Canadians of Old*, 1864), describing seignorial life and the tragedy of divided loyalties caused by the country's conquest. *Jean Rivard* (1862), by Antoine Gérin-Lajoie, depicted pioneer life, forerunning *Maria Chapdelaine* (1915), by Louis Hémon, and novels of pioneering. This theme was the mainspring of "deserter" novels of writers such as Claude-Henri Grignon, author of *Le Déserteur* (1934), in which a farmer who moves to the city is called a deserter.

Germaine Guèvremont dramatized the progressive decadence of a pioneer family in novels of peasant life (*Le Survenant* [1945] and *Marie Didace* [1947]; published together in English as *The Outlander*, 1950). Ringuet (Philippe Panneton) brought a clinical talent into fictional art and in *Trente arpents* (1938; *Thirty Acres*, 1940) viewed migration from the farm with disarming frankness, while in *Fausse monnaie* (1947) he photographed the frustration of an insignificant life in the city.

Works of historical fiction exploited the natural interest in past glories: *Les Habits rouges* (1923; "The Redcoats"), by Robert de Roquebrune, revived memories of the rebellion of 1837; Léo-Paul Desrosiers went back to the early days of the colony for *Les Opiniâtres* (1941) and to the fur trade of 1800 for *Les Engagés du Grand Portage* (1938), works on historical themes that contrasted with his later style as found in *Vous qui passez* (3 vol., 1958–60).

Gabrielle Roy and Roger Lemelin

Gabrielle Roy recorded her observations of life in a working class district of Montreal in a widely acclaimed novel, *Bonheur d'occasion* (1945; *The Tin Flute*, 1947), and continued her skillful revelation of the frustrating insignificance of modern living in *Alexandre Chenevert* (1954; *The Cashier*, 1955). While Gabrielle Roy was portraying Montreal, Roger Lemelin was depicting life in Quebec: the hero of his trilogy, including *Au pied de la pente douce* (1944; *The Town Below*, 1948), *Les Plouffe* (1948; *The Plouffe Family*, 1950), and *Pierre le magnifique* (1952; *In Quest of Splendour*, 1955), was torn between a desire for grandeur and a paralyzing sense of impotence.

Yves Thériault, a prolific novelist, reaches beyond his own milieu for subjects: *Aaron* (1954) was concerned with the strain on a Jewish family in a gentile world; *Agaguk* (1958) powerfully depicted an Eskimo family faced with the white man's code of law; and *Ashini* (1960) gave a lyrical picture of the Indian way of life. Gérard Bessette is another keen observer of French-Canadian society—*La Bagarre* (1958; "The Brawl") dealt with antagonistic social forces in a big city, *Le Libraire* (1960; *Not for Every Eye*, 1962) treated small-town hypocrisy, and *L'Incubation* (1965) and *Le Cycle* (1971) examined the tragedy of human solitude.

Certain novelists, feeling the influence of Mauriac, Gide, Camus, and others, probed the depths of the individual soul. Robert Charbonneau located his psychological studies in a small fictional town that also provided the name of his novel *Fontile* (1945). André Giroux chose two unique situations for his penetrating studies, a crime of passion in *Au delà des visages* (1948; "Behind the Faces") and a man near death from cancer in *Le Gouffre a toujours soif* (1953; "The Abyss Is Always Thirsty"). Solitude and despair were the main themes of André Langevin in *Évadé de la nuit* (1951; "Night Fugitive") and in *Poussière sur la ville* (1953; *Dust over the City*, 1955); primitivism and ennui are delicately orchestrated in *L'Élan d'Amérique* (1972).

The novels of the 1960s demonstrate a new anguish, stimulated by political and spiritual turmoil. Representative names include Réal Benoît, *Quelqu'un pour m'écouter* (1964; "Someone to Listen to Me"); Hubert Aquin with his intriguing baroque novels, *Prochain Épisode* (1965), *Trou de mémoire* (1968), *L'Antiphonaire* (1969); Jacques Godbout and his novels of revolt, *Le Couteau sur la table* (1965) and *Salut Galarneau* (1967); V.-L. Beaulieu presenting the saga of the Beauchemin family in *Jos Connaissant* (1970) and *Les Grands-Pères* (1971); the prolific and whimsical Jacques Ferron, among whose best titles are *Cotnoir* (1962) and *Le Salut de l'Irelande* (1970). The adolescent occupies centre stage in the work of Marie-Claire Blais, beginning with *La Belle Bête* (1959; *Mad Shadows,* 1960) and continuing through to *Une Saison dans la vie d'Émmanuel* (1965) and *Le Loup* (1972); the same subject matter occupies Réjean Ducharme in *L'Avalée des avalés* (1966) and *Le Nez qui voque* (1967).

Drama. The theatre is the poor relation of French-Canadian literature. Excellent theatrical companies were nonetheless founded, such as Les Compagnons du Saint-Laurent (1937–52), Le Théâtre du Nouveau Monde (1951), and Théâtre de la Comédie Canadienne (1956). Gratien Gélinas won great popular acclaim as author and principal actor of *Tit-coq* (published 1950) and *Bousille et les justes* (1960; *Bousille and the Just,* 1961). Robert Elie dealt in Symbolist form with an anguish for lost purity (*L'Etrangère*, 1954), and his contemporary Jacques Languirand confirmed the atmosphere in *Les Insolites* (1956) and *Les Grands Départs* (1957). Marcel Dubé won a wider audience with social themes in *Un Simple Soldat* (1958), *Florence* (1960), and *Pauvre Amour* (1969). His post has since been occupied by Michel Tremblay, best known for *Les Belles-Soeurs* (1971). Robert Gurik introduced science fiction to the stage with such plays as *Api 2967* and *La Palissade* (1971). Radio and television, indeed, offered much scope to dramatists, and a number were engaged in writing specifically for that medium. Examples of their art were to be found in the series *Écrits du Canada français*, published from 1954 onward.

Criticism. Literary criticism appeared in 1903 with a study of the poet Émile Nelligan by Louis Dantin. Camille Roy, however, made one of the first systematic studies; most important among the many works he published was *Manuel d'histoire de la littkrature canadienne-française* (1918). Previous work was rendered obsolete by Gérard Tougas' *Histoire de la littérature canadienne-française* (1960), which inspired other good studies, such as Bessette-Geslin-Parent's *Histoire de la littérature canadienne-française* (1968), Pierre de Grandpré's *Histoire de la littkrature française du Québec* (4 vol., 1967–69), and Guy Sylvestre's *Literatute in French Canada* (1967). Jean Béraud wrote a good account of the theatre in *350 Ans de théâtre an Canada français* (1958), and Séraphin Marion's 10-volume chronicle of the 19th century is entitled *Les Lettres canadiennes d'autrefois* (1939–59).

The *University of Toronto Quarterly* published an excellent annual review of French-Canadian letters, from 1936 onward. Adrien Thério provided a complementary source of information with an annual *Livres et auteurs canadiens* from 1961, and Paul Wyczynski published his scholarly *Archives des lettres canadiennes* from 1961. A number of other serial publications remain active.

English. *From federation to the 1920s.* A growing sense of national identity had preceded the union of the colonies of British North America in the Confederation of Canada in 1867 and was strengthened by its accomplishment. The most national-minded poet in the decade before federation was Charles Sangster, once unofficial Canadian laureate but scarcely remembered since. Charles Mair found themes for poetry on the new frontier and in Canadian history, as used in a verse drama *Tecumseh* (published 1886). Historical romance, always a popular form in Canada, was represented in this period by a costume romance *The Golden Dog* (1877), by William Kirby, about honesty and corruption in New France.

Meanwhile, semiliterary periodicals such as *The Literary Garland*, *Canadian Monthly, Nation* (organ of the Canada First group), and *The Week,* the best of them, appeared.

The Confederation group

The Week's first editor was Sir Charles G.D. Roberts, who, with Bliss Carman, Archibald Lampman, and Duncan Campbell Scott, wrote the best poetry composed in Canada before the 1920s. Of these, Roberts and Carman had much in common; they were first cousins, were educated at the same school and university, and were steeped in English Romanticism and in the Transcendentalism of the 19th-century American writer Ralph Waldo Emerson. Roberts was a memorable regional poet but also expressed the nationalistic optimism of the end of the century. His religious and philosophic poetry seemed less authentic than his rural *Songs of the Common Day* (1893), a pale reflection of Romantic pantheism and Transcendentalism. Carman was best in the romantic or transcendentalist manner in only a few stanzas and brief poems; more often his pseudo-Keatsian beauties were over-sentimental and cloying. The other two poets of the Confederation group, as these nationalistic writers were called, Lampman and Scott, were friends living in Ottawa as members of the civil service. Lampman's poetry of natural description of Ottawa and of social comment were expressions of a lonely and sensitive mind; Scott was also an isolated and slightly austere figure, but more vigorous. Like the others he was a regionalist, but of the remote and vast northern Indian wilderness; Scott's narrative and descriptive poetry depicted man against man or the charged energies of nature against man, but the potentially realistic violence of the themes was subdued and often suggested strangeness and fantasy.

This was the first age of the best seller—romances of Ralph Connor and Yukon ballads of Robert W. Service. Tom MacInnes for 40 years chanted the joys of bohemian vagabondage more plausibly than did Carman; Marjorie Pickthall, a finer artist, combined religious feeling with end-of-the-century aestheticism in beautiful, but often cloying, historical idylls; comic-sentimental dialect poems by William H. Drummond and feminine romances by Lucy M. Montgomery, author of *Anne of Green Gables* (1908), were popular. The best book of the period, *Sunshine Sketches of a Little Town* (1912), by Stephen Leacock, was a regional idyll, though sentiment was laced with ironic satire on the pettiness of life in an Ontario town.

Literature after 1920. The founding of *Canadian Forum* (1920), which championed painters, the Little Theatre movement, and new poets, was a sign of a growing nationalism and interest in cultural arts after World War I. The first of the new poets was E.J. Pratt, whose impulse toward sheer fantasy was evident in *The Witches' Brew* (1925) and *The Titans* (1926); later, most of his longer poems, *The Titanic* (1935), *Brébeuf and His Brethren* (1940), and *Behind the Log* (1947), were narratives of heroism displaying exact historical and technical knowledge and an Elizabethan delight in the vigorous splendour of words. Pratt was an isolated figure, but his break from the traditions of the Confederation poets encouraged younger men, usually influenced by T.S. Eliot. Five of these, with Pratt, in 1936 produced a volume of verse, *New Provinces:* A.J.M. Smith, critic of the group, and Leo Kennedy showed most clearly the influence of Eliot; Robert Finch owed more to Eliot's French masters; F.R. Scott was most effective in brief satirical poems; Abraham Moses Klein was a learned, sensitive Jew, steeped in the history of his people. Later, Raymond Souster, an Imagist poet, and Leonard Cohen, writing morbidly introspective verse, were influential. During the depression of the 1930s and World War II, Earle Birney composed social comment and satire, war poetry (and a comic novel of the war, *Turvey)*, and narrative and regional description. The work of P.K. Page was concerned with economics and politics. *East of the City* (1946), by a critic, Louis Dudek, was occasionally Marxist in social comment but also sometimes lyrically joyful at the industrial landscape of his native Montreal. Irving Layton, also of Montreal, was more aggressively proletarian, Naturalistic, and egocentric, the castigator of Canadian puritanism and intellectualism. James Reaney was also a satirist of small-town life beneath his poetry's complex allusiveness. A preoccupation with symbol and myth, with the English 19th-century poet William Blake, and with the Swiss psychologist Carl Jung informed the poetry of Jay Macpherson.

Rise of the novel

The early novels of Frederick Philip Grove from *Settlers of the Marsh* (1925) to *Fruits of the Earth* (1933) were powerful and sombre studies of pioneering on the prairies, though stiff in form. Two previous volumes of western sketches, *Over Prairie Trails* (1922) and *The Turn of the Year* (1923), were the best of his writings. From *Jalna* (1927) to *Morning at Jalna* (1960), Mazo de la Roche published 16 novels about the turbulent Whiteoak family—the most massive achievement in Canadian fiction. Ethel Wilson wrote with sensitivity and skill about private emotion and conflict; Philip Child, who began with historical romance, turned later to contemporary Realism as in *Ray of Wrath* (1945); Morley Callaghan was a humanitarian Realist. describing with compassion the lives of misfits and the morally bewildered in urban society. Hugh MacLennan was a Realist of Nova Scotia life in *Barometer Rising* (1941) and *Each Man's Son* (1951) and interpreted Canadian national character in *Two Solitudes* (1945). Canadian mores in small cities of southern Ontario provided comic themes for Robertson Davies. In Mordecai Richler's best novel, *The Apprenticeship of Duddy Kravitz* (1959), a juvenile antihero and Jewish Montreal were compounded of the farcical, the realistic, and the pathetic.

Two transient novelists were an Englishman, Malcolm Lowry, who left Mexico to spend 15 years in Canada, chiefly in British Columbia, where he completed his novel *Under the Volcano* (1947); and Brian Moore, who lived briefly in Montreal, the setting of *The Luck of Ginger Coffey* (1960).

AUSTRALIAN AND NEW ZEALAND LITERATURE IN THE 20TH CENTURY

Australian. *Nationalism and expansion.* The last 20 years of the 19th century saw a growth of nationalism and moves toward federation of the separate states. National pride, the values of country life, and sympathy for the struggles of small landholders were common literary themes, particularly in stories by Henry Lawson and Steele Rudd (Arthur Hoey Davis). Another theme was that of early convict days, when prisoners were transported from England to penal servitude in Australia: four volumes of tales by Price Warung (William Astley) exemplify this. Joseph Furphy, in a large novel, *Such Is Life* (1903), described the rural world of the 1880s. It was full of detail of station life, conversations of bullock drivers, nationalistic sentiments, and philosophizings about chance and determinism.

The flourishing tradition of ballad verse was kept alive in the 1890s by Lawson and "Banjo" Paterson.

Federation (1901) was greeted by patriotic fervour in prose and in verse, but generally, in the first 30 years of the 20th century, nationalistic themes diminished in the face of a growing awareness of the complexity of Australian life, and attention was drawn to the plight of ordinary people and need for social reform. Prominent among writers about the life of the city and its workers were Katherine Susannah Prichard, whose novels spanned over 40 years; Louis Stone; Edward Dyson; and Kylie Tennant (Mrs. L.C. Rodd). The life of the Aboriginal and his relationship with the white man were explored by Mrs. Aeneas Gunn, Xavier Herbert, and Katherine Susannah Prichard. Two writers more interested in history and earlier colonial days were Miles Franklin and Eleanor Dark. Notable novels, short stories, and criticism were produced by Vance Palmer. Henry Handel Richardson (Ethel Florence Lindesay Robertson) wrote about her schooldays in Melbourne (*The Getting of Wisdom*, 1910), student days in Leipzig (*Maurice Guest*, 1908), and her father's life (in the trilogy *The Fortunes of Richard Mahony*, 1917–29), in the last of which she combined description of an immigrant's life and conditions on the gold fields with a powerful character study.

The success of poets was slighter but showed diversity of aims and interests. John Shaw Neilson wrote fine, delicate lyrics; Victor Daley created romantic poems and some sharp satires; C.J. Dennis continued the ballad tradition with popular verse; Christopher Brennan, in his Symbolist poems, owed much to European tradition; and Kenneth Slessor wrote powerful dramatic lyrics.

The modern period. During and after World War II, literary magazines proliferated and the reading public grew. Though factual and descriptive writing remained prominent, from the 1950s onward Australian writers became increasingly speculative and searching. The most influential of mid-20th-century novelists was Patrick White, whose major novels were distinctively Australian but whose treatment had a largeness of vision surpassing nationalistic limitations; the Australia reflected in *The Tree of Man* (1955), *Voss* (1957), *Riders in the Chariot* (1961), *The Solid Mandala* (1966), and *The Vivisector* (1970) and in short stories was the product of an individual, critical, poetic imagination. Other novelists included Martin Boyd, who offered a penetrating analysis of English and Australian life; Randolph Stow, who set his symbolic novels in a finely evoked Australian landscape; and Thomas Keneally. The short story continued but retained a strong rural bias—a strange disparity between life as most Australians knew it and as literature represented it. Of particular interest was the autobiographical writing of Hal Porter, which showed the changes in Australian society from the 1930s to the 1960s.

The novels of Patrick White

A tradition of descriptive poetry remained, but mid-20th-century poets ranged more widely than their predecessors. Robert D. FitzGerald was noted for strenuously argued lyrics; A.D. Hope for searching, allusive, witty poetry; Douglas Stewart for forthright, vigorous verse; Judith Wright for sensitive lyrics; and James McAuley for meditative lyrical poetry.

Australian literature had always been deficient in drama: only after the late 1950s and 1960s did any plays achieve real success. Best known were Douglas Stewart's *Ned Kelly* (published 1943) and *Fire on the Snow* (performed 1941); Ray Lawler's *Summer of the Seventeenth Doll* (1955); Alan Seymour's *One Day of the Year* (1961); and Patrick White's *Four Plays* (published 1965). Recently several young playwrights have come into prominence.

The rapid expansion of Australian society in the 20th century made it more difficult for writers to keep pace with their world. A spirit of self-criticism that became pronounced in the 1950s and '60s was exemplified by searching studies of Australian history and society and by reinterpretations of Australian literature.

New Zealand. In organized settlement, Australia had a 50-year start over New Zealand, and for most of the century from 1840 to 1940 a similar gap in national consciousness seemed to prevail. Yet from about 1930 there was a flowering of literature that made it possible for the Anglo-Australian novelist Colin MacInnes to argue that New Zealand writers had made a "sharper, deeper revelation" of their country than had the Australians.

Nineteenth-century writers left future generations little real New Zealand literature: even up to the 1920s most writers aimed at English readers and often exploited and distorted exotic and indigenous elements of their surroundings. There were some worthwhile chronicles and autobiographical narratives, such as F.E. Maning's *Old New Zealand* (1863), Samuel Butler's *First Year in Canterbury Settlement* (1863), and Lady Mary Anne Barker's *Station Life in New Zealand* (1870). Among colonist versifiers, the most notable were Alfred Domett and a Scots dialect poet, John Barr.

Katherine Mansfield in New Zealand literature

Katherine Mansfield was one of the great names of this period, though she did not expect to find a readership in her own country. In her most admired stories, "At the Bay" and "The Garden Party" (1922), her debt to New Zealand pioneer society was, however, clear. Two other writers achieved works of distinction in this same period: Jane Mander, in *The Story of a New Zealand River* (1920), and William Satchell, in *The Land of the Lost* (1902).

A few poets and two or three prose writers first weathered the uncongenial literary climate of the 1920s and 1930s. Outstanding were a short-story writer, Frank Sargeson, and the poets A.R.D. Fairburn, R.A.K. Mason, Allen Curnow, and Denis Glover, and what distinguished their work was that they assumed a readership of fellow countrymen. The yarn spinner had always been characteristic of New World writing, and Sargeson used this to great effect in tales of the Depression years. Two novelists to emerge during this time were John Mulgan, whose *Man Alone* (1939) remained a significant New Zealand novel in title and theme, and John A. Lee, with his harrowing *Children of the Poor* (1934).

The years from 1940 saw a flowering in most literary genres and especially in the short story, whose best writers included Roderick Finlayson, Maurice Duggan, O.E. Middleton, and Phillip Wilson. A poet and playwright, James K. Baxter, and a historian-poet, Keith Sinclair, were perhaps the most considerable figures of the postwar period. Among best known novelists were Dan Davin (*No Remittance* [1959]); David Ballantyne (*The Cunninghams* [1949]). Sylvia Ashton-Warner (*Spinster* [1958], *Greenstone* [1966]); Ian Cross (*The God Boy* [1957], *After Anzac Day* [1961]); Janet Frame (*Faces in the Water* [1961], *A State of Siege* [1966], *The Rainbirds* [1968]); Maurice Shadbolt (*Among the Cinders* [1965], *Strangers and Journeys* [1972]); and Maurice Gee (*A Special Flower* [1965], *In My Father's Den* [1972]). The growth in quantity and quality of New Zealand writing and a rapid expansion of theme reflected accelerating national consciousness.

AFRICAN LITERATURE

South African. *English.* Olive Schreiner, a liberal and a powerful writer on local and international affairs, wrote the first great South African novel, *The Story of an African Farm* (1883). William Plomer pioneered race relations as material for fiction in a novel, *Turbott Wolfe* (1925). Pauline Smith's stories, *The Little Karoo* (1925), dealt sympathetically with rural Afrikaners. Laurens van der Post, in a novel, *In a Province* (1934), dealt with the African-coming-to-town theme.

Cry, the Beloved Country (1948) established Alan Paton's international reputation, which later writing, such as the stories in *Debbie, Go Home* (1961), enhanced. Nadine Gordimer made an impact abroad with her first novel, *The Lying Days* (1953); her talent was at its best perhaps in the short story, such as those in *Not for Publication* (1965). Her novel *A Guest of Honour* (1971) is set outside South Africa. Dan Jacobson wrote with restraint and wry humour, as in *The Beginners* (1966) and in his collected short stories, *Beggar My Neighbour* (1964); *The Rape of Tamar* (1970) is a fictional treatment of a biblical theme. Peter Abrahams (*Tell Freedom: Memories of Africa* [1954]) was a prolific expatriate Coloured novelist. In his short stories and novellas (*e.g.*, *A Walk in the Night* [1962]), Alex La Guma uses the racy Coloured English vernacular that owes much to Afrikaans. His other writings include *And a Threefold Cord* (1964), *The Stone-Country* (1965), and *In the Fog of a Season's End* (1972). Other writers of protest include Lewis Nkosi, author of a collection of essays, *Home and Exile* (1965), and Ezekiel Mphahlele, whose autobiographical *Down Second Avenue* (1959) has become a South African classic. Nkosi and Mphahlele, both Africans, worked for *Drum* magazine, which provided an important forum for their ideas; they are now in exile. Among younger novelists, Jillian Becker has written two powerful works, *The Keep* (1967) and *The Union* (1971).

Warm humour, rare in South African writing, is abundant in the work of H.C. Bosman, who deals affectionately with Afrikaners of the old school in *Mafeking Road* (1947) and *A Cask of Jerepigo* (1957). The only drama of note in English is that of Athol Fugard, particularly *The Blood-Knot* (1965) and *People Are Living There* (1970).

Many South African English poets expressed their consciousness of the difficulty of finding language and rhythm appropriate to uniquely South African material; some, too, expressed a sense of being aliens as people as well as poets. Guy Butler expressed much of this in his fine poem "Home Thoughts" (1956). F.C. Slater often evoked uniquely South African experience, as in "Lament for a Dead Cow" (*Collected Poems* [1957]). In the 1920s came the work of major poets Roy Campbell and William Plomer. Campbell, whose adventurous life was reflected in his powerful poetry, often sounded a note of menace. His most famous work was his long symbolic poem *The Flaming Terrapin* (1924). The acutely observant William Plomer wrote with restraint and experimented with rhythm and diction (*Notes for Poems* [1927], *Visiting the Caves* [1936]). Other poets of merit include Mazisi Kunene, Sidney Clouts, Dennis Brutus, F.T. Prince, Roy Fuller, and the young O.M. Mtshali, whose *Sounds of a Cowhide Drum* (1971) is especially noteworthy.

Afrikaans. By 1750 Afrikaans was sufficiently distinct from its parent Dutch to be called a separate language. The first Afrikaans texts were, however, not published until more than a century later. The Association of True Afrikaners, which published the first newspaper, magazine, and books in Afrikaans, was founded in 1875. From defeat in the South African War came a new upsurge to establish Afrikaans as a national language. The second Afrikaans language movement spread north from Cape Province, and Afrikaans gradually won ascendancy over Dutch, replacing Dutch as the official language of the (then) Union of South Africa in 1925.

Outstanding writers of the "second movement" were primarily poets: Eugène Marais; Jan François Elias Celliers, a pastoral poet; Jakob D. Du Toit (Totius), who wrote some of the best elegiacs in Afrikaans; and C. Louis Leipoldt, foremost war poet. In Realism two novelists achieved success—Jochem van Bruggen and Jan Van Melle—while the two foremost romantic novelists were D.F. Malherbe and C.M. van den Heever. In the 1920s, problems of religion and personal relationships came to the fore. F.P. van den Heever was the outstanding poet at this time. His anti-conformism foreshadowed the great upsurge of "new" poetry in the 1930s.

The Dertigers

The supreme event in Afrikaans literature was the appearance of a group of talented poets, the Dertigers (Men of the '30s), begun by W.E.G. Louw with *Die ryke dwaas* (1934; "The Rich Fool"). This sensitive poet, with his searing conflicts between God and Eros, exemplified qualities soon to become the new generation's hallmark. He was followed by his elder brother, N.P. van Wyk Louw, principal creative artist and theoretician of the new movement. Each of Van Wyk Louw's books was a milestone: in every form he achieved mastery; he wrote the best odes, sonnets, modern ballads, and love lyrics in Afrikaans; his dramatic monologue "Die Hond van God" (1942; "The Hound of God") was unsurpassed in the Dutch literatures, and his epic poem *Raka* (1941) became a timeless and matchless symbol.

The poet D.J. Opperman emerged in 1945. His technique superimposes different historical levels, intermingled with a fascinating mosaic of themes, images, and allusions from a common Western heritage. Of the poets writing in the vernacular of the Cape Coloureds, Adam Small is the most talented.

Afrikaans moved from being an essentially pastoral language to being an urban one, expressing the frustrations and stresses of the city dweller. From the 1950s onward young writers came to the fore in the novel, short story, prose poem, experimental drama, and poetry. Generally, the avant-garde novels of "men of the '60s" caused indignation among the more conservative. Young writers such as Breyton Breytonbach and Ingrid Jonker, both poets, and Etienne Leroux and André Brink, both novelists, revolted against more than mere externals of religious or sexual orthodoxy; their revolt was part of a new way of life. Leroux explored the dilemma of man in search of a myth, the inexhaustible fantasy and satire of his work making it unique in Afrikaans.

Afrikaans drama lagged behind. The two most important dramas—N.P. van Wyk Louw's *Germanicus* (1956) and Opperman's *Periandros van Korinthe* (1954)—were both traditional in form and theme. Of the younger school, P.G. du Plessis is the most promising.

Ethiopian. In the Muslim incursion of 1527–43 Ethiopian literary activity ceased and many manuscripts perished; Islāmicization was widespread, and, even after the repulsion of the invaders, the country never fully recovered. A Muslim merchant who had been converted to Christianity and, as Enbaqom (Habakkuk), became prior of a monastery, Dabra Libanos wrote

Anqas'a amin ("Gate of Faith") to justify his conversion and to persuade apostates to recant. Other similar works were produced, and several were written to defend the Monophysite branch of the Christian faith; meanwhile the arrival of Roman Catholic missionaries constituted a further danger to the Ethiopian Church.

The ancient language of Ge'ez had by now lost its vigour and became a liturgical language with which few people were thoroughly conversant. During the 16th century, Amharic, the principal spoken language, was beginning to be used for literary purposes, and Amharic expressions even appeared in royal chronicles. About 1600, nevertheless, a few substantial works in Ge'ez appeared, including ***Hawi,*** an enormous theological encyclopaedia translated by Salik of Dabra Libanos, a ***History*** of Johannes Madabbar, bishop of Nikiu, containing an account of the Arab conquest of Egypt, valuable since the Arab original has been lost, and ***Fetha Negast*** ("Justice of the Kings"), a compilation of canon and civil law. Ge'ez poetry (*qene*) flourished, at Gonder particularly, in the 18th century and has since continued to be practiced at many monasteries. Some poems of Alaqa Tayya were printed in Asmera in 1921, and an important anthology compiled by Heruy Walda Sellāsē was published at Addis Ababa in 1926.

A Jewish population, known as Falashas, who live mostly in regions north of Lake Tana, still use Ge'ez as their sacred language. Besides the Old Testament (including the ***Book of Jubilees),*** they have a few books peculiar to themselves, notably ***Te'ezaza Sanbat*** ("Ordinance of the Sabbath"), of uncertain date and perhaps mostly translation from Arabic of the 14th century. A ***Falasha Anthology*** was published by Wolf Leslau in 1951.

Works in Amharic

The earliest known Amharic compositions are songs celebrating the victory of Amda Tseyon I (1314–44). From the 16th century onward theological works were produced. A translation of the Bible was made in Cairo early in the 19th century (though probably not by a true Ethiopian, to judge by the quality of the Amharic) and from this version missionary societies composed their editions and revisions by foreigners with an inadequate knowledge of Amharic. A more scholarly version of the New Testament was printed in Addis Ababa in 1955, followed by the Old Testament in 1961. The first official chronicles wholly in Amharic were those of Tewodros (Theodore) II (1855–68). A translation of John Bunyan's ***Pilgrim's Progress*** made in 1892 pointed the way to a new popular form—the allegorical novel, often partly in verse, with a religious bias, of which the first was ***Libb wallad tarik*** (1908; "Imaginative Story") by Afawarq Gabra-Syasus. During the regency of Ras Tafari (1916–20; afterward Emperor Haile Selassie), Blatengeta Heruy Walda-Selase (died 1938) became the leading Amharic writer, especially notable for allegorical compositions such as ***Wadaje lebbe*** ("My Heart as My Friend").

With the restoration of Ethiopian independence after the Italian occupation of 1936–41, a great impetus was given to Amharic literature, the Emperor encouraging authors to produce many types of books, especially on moral and patriotic themes. Writers of merit were Bitwadded Makonnen Endalkachaw (allegorical novels and plays), Kabbada Mika'el (verse dramas, some history and biography), and Takla-S'adeq Makurya (histories).

CARIBBEAN LITERATURE IN ENGLISH, FRENCH. AND SPANISH

Unlike the literatures of Africa, that of the Caribbean has no indigenous tradition. The pre-Columbian American Indians left few rock carvings or inscriptions (petroglyphs), and their oral traditions did not survive 16th-century Spanish colonization. The West Africans who replaced them were also without a written tradition, so that for 400 years Caribbean literature was an offshoot and imitation of the models of the colonial powers—Spain, France, Britain, and The Netherlands. Caribbean writers, however, were not unaware of their environment. The letters and speeches of Toussaint-Louverture, Haitian general and liberator, indicated that from at least the end of the 18th century the Caribbean was conscious of its cultural identity. But it was not until the 1920s that the challenge of a distinctive literary form was accepted. Then, as part of a Latin-American *modernismo* movement, Spanish and French Caribbean writers began to break away from European ideals and to identify themselves with their mainly Negro masses.

The *modernismo* movement

The leaders of this movement, mainly poets, were Luis Palés Matos (Puerto Rico), Jacques Roumain (Haiti), Nicolás Guillén (Cuba), Léon Damas (French Guinea), and Aimé Césaire (Martinique). Jean Price-Mars, a Haitian ethnologist, in ***Ainsi parla l'oncle*** (1928), declared that his purpose was to "restore to the Haitian people the dignity of their folklore." The achievement of this Négritude, finely expressed in Césaire's poem ***Cahier d'un retour au pays natal*** (1939; ***Return to My Native Land,*** 1968), was the construction into poetic forms of the rhythmic and tonal elements of the islands' rituals and speech patterns, using Symbolist and Surrealist techniques.

The British Caribbean, developing its national literature after 1945, made its own contribution in the folk dialect novel: Vic Reid's ***New Day*** (1949), Samuel Selvon's ***A Brighter Sun*** (1952) and ***The Lonely Londoners*** (1956), George Lamming's ***In the Castle of My Skin*** (1953), V.S. Naipaul's ***Mystic Masseur*** (1957) and ***House for Mr. Biswas*** (1961), among others; and the poetry of Louise Bennett ***(Jamaica Labrish,*** 1966). But paradoxically, anglophone Caribbean development was formally conservative, working toward an "open" rather than an autochthonous, or indigenous, expression in the work of C.L.R. James (Trinidad) and the poetry of Derek Walcott (St. Lucia). In the novels of Wilson Harris (Guyana), the Symbolist and Surrealist techniques of the *modernismo* movement reappear; and the poetry of Edward Brathwaite ***(Rights*** of ***Passage*** [1967], ***Masks*** [1968], ***Islands*** [1969]) attempt to reassert the place of Africa in the Caribbean.

FKENCH LITERATURE IN THE 20TH CENTURY

War and political and social change broadened the gap between the literature of the 19th and 20th centuries in France, but more particular influences included the reform of the university system, with its emphasis on modern studies (1502); the arrival of the cinema; the emergence of strong publishing houses and the invention of the ***livre de poche*** ("pocket," or paperback, edition); the creation of such important reviews as ***La Nouvelle Revue Française***, ***Les Temps Moderrzes, Le Mercure de France,*** the ***Cahiers du Sud, Esprit, La Table Ronde, Les Lettres Modernes, Tel Quel***; and the new conceptions of man suggested by Freud and his successors.

Although change in French literature in the 20th century proceeded by phases of revolt, destruction, innovation, and revaluation, the French genius for continuity remained unimpaired, with the same respect for the French language, moral preoccupation, and care for form that has always characterized French literature. Though Paris remained the crucible and experimental centre, literature preserved its personal, often pastoral qualities; the provinces were still a main setting for comedy, tragedy, fiction, poetry. Literature continued to be impeccably French in essence, but because of a concentration on universal problems it was also so successfully outward looking as to integrate itself into the world picture and to share, with English and American literature, leadership of Europe.

Fiction. Of writers whose reputation was made before 1900, Pierre Loti, Henry Bordeaux, Paul Bourget, Léon Bloy, Pierre Louys, Marcel Prévost, Henri de Régnier, and Anatole France suffered some eclipse. A few significant novels appeared before 1914. Romain Rolland, in a great novel cycle ***Jean-Ckristophe*** (1904–12), synthesized the outlook of his time through a central character. Valery Larbaud, after ***Fermina Marquez*** (1911) and ***A.O.*** *Barnabooth* (1913), was a pioneer and theorist of the internal monologue, which, based on Édouard Dujardin's ***Lauriers sont coupks*** (1887), influenced his friend James Joyce. Jules Romains invented what he called the ***cink-roman*** ("cinematic novel") with ***Mort de quelqu'un*** (1911; ***The Death of a Nobody,*** 1914), and the idea (not the book) later influenced the *nouveau roman* ("new novel"; see below). He developed this technique on a vast scale in *Lcs Hommes de bonne volonté* (27 vol., 1932–47; "Men of Goodwill" series, 1933–46), which influenced Jean-Paul Sartre's manner.

Romains's generation was obsessed by the Balzac–Zola–Rolland example of "monumental" structures, variously adapted, even in works so diverse as Sidonie-Gabrielle Colette's "Claudine" saga (1900–03) and Georges Duhamel's domestic *Chronique des Pasquicr* (10 vol., 1933–44; *The Pasquicr Chronicles*, 1937). A more personal adaptation of the monumental structure was Marcel Jouhandeau's caricatural "Monsieur Godeau" series (1926--50). Other domestic cycles were Roger Martin du Gard's work *Les Thibault* (8 vol., 1922–40) and Maurice Druon's *Grandes Familles* (1948; translated as *The Rise of Simon Lachaume*, 1952, and, as part of the trilogy *La Fin des hommes*, 1951, as *The Curtain Falls*, 1952–59). François Mauriac explored problems of conscience in his "Thérèse Desqueyroux" series (1927–35), and the application became political and polemical with Sartre's *Chemins de la liberté* (1945; "Roads to Freedom").

Proust and Gide

None of these writers, however, created as monumental a work of beauty and complexity as Marcel Proust's *A la recherche du temps perdu* (1913–27; *Remembrance of Things Past*, 1922–31j, the basis of which, *Jean Santeuil*, remained unpublished until 1952. The seven novels making up the work represent an allegorical search for truth, in that the narrator recalls the experiences of his past in an attempt to recapture time lost. Proust and André Gide were leading figures of the period up to 1940. Gide distrusted himself as a "novelist," but already, in *Lcs Caves du Vatican* (1914; *The Vatican Swindle*, 1925), he had advanced the concept of the "gratuitous act," which contributed something to Surrealism, to the "philosophy of the Absurd," and (negatively) to Existentialist literature. At the same time the book focussed attention on the amoral adolescent, more carefully portrayed in *Les Faux-Monnayeurs* (1926; *The Counterfeiters*, 1927), which, in its composition, together with the *Journal des Faux-Monnayeurs* (1926; *The Logbook of the Coiners*, 19-52), prepared the way for the "antinovel," although Gide himself thought in terms of a "pure" novel (see below).

The disillusioned adolescent was over-exploited by modern French writers, but Alain-Fournier's *Grand Meaulnes* (1913; *The Wanderer*, 1928), Raymond Radiguet's *Diable au corps* (1923; *Devil in the Flesh*, 1932), Jacques de Lacretelle's *Silbermann* (1922), Jean Giraudoux's *Simon le pathétique* (1918), and Hervé Bazin's *Vipère au poing* (1948; *Viper in the Fist*, 1951) and Françoise Sagan's blasé *Bonjour tristesse* (1954) remained interesting variations on the theme.

Among the few novels about World War J to have survived were Henri Barbussé's *Feu* (1916; *Under Fire*, 1917); Duhamel's *Vie des martyrs* (1917; *The New Book of Martyrs*, 1918) and *Civilisation* (1918); André Maurois's *Silences du Colonel Bramble* (1918; *The Silence of Colonel Bramble*, 1919); and Maurice Genevoix's documentary *Sous Verdun* (1916; *'Neath Verdun*, 1916). A realist–revolutionary crisis in the novel was seen in the work of Louis Aragon, from *Les Cloches de Bâle* (1934; *The Bells of Basel*, 1936) to *Les Communistes* (1949), and in the novels of Louis-Ferdinand Céline (who was later to turn violently anti-Semitic)—notably in *Voyuge au bout de la nuit* (1932; *Journey to the End of the Night*, 1934), *Mort à credit* (1936; *Death on the Installment Plan*, 1938), *Guignol's Band* (1943), and *D'un château à l'autre* (1957; "Castle to Castle"). Observation and fantasy combined in the work of regionalist novelists of this period such as Marcel Aymé; but the best of them, Jean Giono, adopted a dryer, Stendhalian style in *Le Hussard sur le toit* (1952; *The Hussar on the Roof*, 1953) and subsequent works.

Novels exploring the psyche

Meanwhile, other important novelists preferred to explore the psyche—the most important being Mauriac, whose stress on ineffectual guilt and spiritual gloom was also found in Julien Green and Georges Bernanos and in Henri Bosco's *Mas Théotime* (1946; *The Farm Théotime*, 1946). Henry de Montherlant analyzed humanity with contempt in *Les Célibataires* (1934; *Perish in Their Pride*, 1936). Pierre-Jean Jouve, strongly influenced by psychiatry, wrote "poetic" novels that could stand beside Raymond Queneau's *Chiendent* (1933; "The Snag"), *Pierrot, mon ami* (1942), and *Zazie duns le métro* (1959); or those of Cocteau, from *Le Potomak* (1919) to *Les Enfants terribles* (1929); or Andre Breton's *Nadja* (1928) and Aragon's *Le Paysart de Paris* (1926; "The Peasant of Paris"); or the slightly faded preciosity of Giraudoux.

Of many novelists of adventure, Andre Malraux, especially in *La Condition humaine* (1933; *Mart's Fate*, 1934), was the only one to give this form philosophical significance.

The pre-1940 novel seemed less original by the 1970s than it did in its day, despite the formal innovations of Rolland, Proust, and Gide; the linguistic experiments of Céline, Giraudoux, and Queneau; and the monologue techniques of Valery Larbaud, Pierre-Jean Jouve, Jean Schlumberger, and Léon-Paul Fargue. 'The real "explosion" came only after 1940.

Sartre and Existentialism

On the one hand, the traditional novel had been threatened by Gide's concept of the "pure" novel, in which motivation and probability (*vraisemblance*) were questioned: on the other, with Malraux and Aragon it had come so close to documentary as to resemble propagandist journalism, so that Sartre could even argue that literature must become journalism. Fiction practically disappeared in the heroic reportage of Antoine de Saint-Exupéry's *Vol de nuit* (1931; *Night Flight*, 1932). During and after World War II, fiction was at first a *témoignage*, or "inspired newsreel," as in the work of Vercors (pseudonym of Jean Bruller; *Le Silence dc la nzcr* [1942]; *Put Out the Light* [1944]), Jean-Louis Curtis, Pierre Gascar, Roger Vailland (*Drôle de jeu* [1945], *Playing with Fire* [1948]), Romain Gary, André Schwartz-Bart, and Jules Roy. The margin between fiction and documentary also narrowed in Sartre and Simone de Beauvoir. Albert Camus, in his enigmatic book *L'Étranger* (1942; *The Stranger*, 1946), in an allegorical novel, *La Peste* (1947; *The Plague*, 1948), and in *La Chute* (1956; *The Fall*, 1957), was, however, primarily a moralist in the great French tradition. Sartre, a novelist despite himself, used every trick of the trade, but sometimes clumsily. His brilliant *Nausée* (1938; *Nausea*, 1949) was a philosophical tract sublimated by insight and style. *Les Chemins de la liberté* might be laborious propaganda but showed a flair for situation, arrangement, and significant detail: its weakness was that the reader was expected also to know Sartre's philosophical work *L'Être et le néant* (1943; *Being and Nothingness*, 1956), his contribution to Existentialist thought. (Sartre's Existentialism, with its emphasis on man "condemned to be free" and on man's life as "an ineffective passion," was based on the distinction between unconscious beings who exist only in themselves and conscious beings who exist for themselves and are therefore necessarily free. Sartrian Existentialism had a profound influence on many French writers.) Sartre, like Proust, showed, however, that fiction could be philosophy in action; and this was true, too, of Maurice Blanchot, Jean Cayrol, and other of his contemporaries. Simone de Beauvoir, obsessed with ambiguity, mortality, and feminism, began well with *L'Invitée* (1943; *She Came to Stay*, 1954) but barely maintained her reputation with the didactic *Mandarins* (1954; *The Mandarins*, 1956) and the somewhat unscientific *Le Deuxième Sexe* (1949; *The Second Sex*, 1953) before finding her proper medium in autobiography. As a novelist she achieved a lighter touch in *Les Belles Images* (1966). These writers were not concerned to experiment with the novel as a form; for political reasons they regarded it primarily as a means of communication. Alongside their earnestness went the more experimental work of Queneau and of Julien Gracq (*Le Rivage des Syrtes* [1951; "The Banks of the Syrtes"]), André Pieyre de Mandiargues (*Le Lis de mer* [1956; *The Girl Beneath the Lion*, 1959]), Boris Vian (*L'Automne à Pekin* [1947; "Autumn in Peking"]), Serge Groussard, and Maurice Blanchot *(Thonzas l'obscur* [1941, new version 1950; "Thomas the Obscure"]).

The beginning of the "new novel," or "anti-novel"

The postwar novel was strongly influenced by Franz Kafka, Virginia Woolf, William Faulkner, Ernest Hemingway, Erskine Caldwell, and John Dos Passos. Despite this influence there emerged the French blow to the established—or traditional—novel, variously called the "anti-novel" and the *nouveau roman* ("new novel"). It might be said to have begun in 1951 with the Irish-born Samuel Beckett's *Molloy* and *Malone meurt* (*Malone Dies)*, written in French and translated by Beckett in 1955 and 1956, though Nathalie Sarraute's *Tropismes (Tropisms,* 1963) had appeared in 1938 and her *Martereau* followed in 1953. The "new novelists" attacked story, plot, character, form, style, meaning, chronology, analysis of thought and feeling—decomposing reality, flouting time. What was left was a "do-it-yourself" novel, in which elite readers worked from data: direct communication between novelist and reader was taboo. Thus the "hero" was opaque in Michel Butor's *Modification* (1957: *Second Thoughts*, 1958); chronology was fuddled in Claude Simon's *Routes des Flandres* (1960; *The Flanders Road*, 1961); necessary facts and interpretations were withheld in Alain Robbe-Grillet's novels *Le Voyeur* (1955) and *La Jalousie* (1957); time was transposed in Claude Mauriac's *Dîner en ville* (1959; *The Dinner Party*, 1960); psychological "authenticity" was disregarded in Nathalie Sarraute's *Portrait d'un inconnu* (1947; *Portrait of a Man Unknown*, 1958). The term *école du regard* ("school of the glance"), applied to Robbe-Grillet, Jean Ricardou, and Claude Ollier, meant that they showed only the flat surfaces of things; but if the "new novel" had been merely a film script and its masterpiece the film called *L'Année dernière à Marienbad (Last Year at Marienbad),* there would have been no future in it. Jean Cayrol, Robert Pinget, and Marguerite Duras were sometimes associated—the latter wrongly—with the movement.

The general public, meanwhile, continued to encourage writers who were in many respects not so experimental. Among these were Henri Thomas; Jean Cau; a Belgian, Françoise Mallet-Joris; Marguerite Yourcenar; José Cabanis; Bernard Pingaud; Edmonde Charles-Roux; and Roger Peyrefitte.

The theatre. At the turn of the century the French theatre was torn between neo-Symbolism and Social Realism and broke free only 30 years later. Of the lighter comic writers, Georges Feydeau revived (and translated) well. The drawing-room (parlour) "problem" playwrights had also dated, but the form persisted-even in Sartre's *Huis-clos* (performed 1944; *No Exit*, 1946)—and seemed likely to persist so long as people had problems and parlours. The sentimental Symbolism of Maurice Maeterlinck (*Monna Vanna* [1902], *L'Oiseau bleu* [1908; *The Blue Bird,* 1909]) retained a period charm, but the only lastingly great dramatist in the Symbolist tradition was Paul Claudel. The historical theatre of Romain Rolland and the verse drama of Edmond Rostand had no successors.

Alfred Jarry and modern French theatre

The 20th-century theatre began with a time bomb under it, one that exploded only 50 years later: Alfred Jarry's *Ubu Roi* (1896), which could later be seen as a forerunner of Surrealism and the Theatre of the Absurd. Otherwise the period was notable mainly for reforms by great manager-producers: Jacques Copeau, Louis Jouvet, Charles Dullin, Gaston Baty, and Georges Pitoëff. After them came Jean-Louis Barrault, André Barsacq, Jean Vilar, Jean Dasté, Georges Vitaly, and Roger Blin. After 1944 the state theatres—the Comédie-Française, the Théâtre de l'Odéon (also known as the Salle Luxembourg), the Théâtre Nationale Populaire, and the Théâtre des Nations (Sarah Bernhardt)—were reorganized; the state grants were made to provincial troupes.

Claudel's religious plays, deeply rooted in Shakespeare and in classical Greek and later Spanish theatre, tended to be chronicles or panoramic works overflowing with theatrical conventions, so that they long seemed unactable. *L'Annonce faite à Marie (Tidings Brought to Mary)* and *L'Otage (The Hostage)* were staged in Paris in 1912 and 1914, respectively, and soon after in the United States and London, but others were not staged until the 1940s. Claudel, with his grotesque sense of pageantry and hierarchy, buffoonery, brutal psychology, creative imagination, cruel fantasy, and gusts of rhetoric spiced with gross peasant humour, approached the Elizabethen synthesis. He had no apparent heirs, but writers such as Michel de Ghelderode and Gide *(Oedipe* [published 1931: *Oedipus and Theseus*, 19501) could not escape his influence.

Generally speaking, the stage play was less experimental from 1920 to 1940 than sircumstances permitted. Guillaume Apollinaire's *Mamelles de Tirésias* (1917; "The Breasts of Tiresias") was neo-Jarry, but Dada and Surrealism failed to mark the theatre. Giraudoux's witty plays lacked depth and density: his comedies—*Amphitryon 38* (1929), *Intermezzo* (1933; *The Enchanted*, 1950), *La Folle de Chaillot* (1945; *The Madwoman of Chaillot*, 1947)—like his more serious *Guerre de Troie n'aura pas lieu* (1935; *Tiger at the Gates*, 1955), soon lost their topical savour, though they remained actable. This also applied to Cocteau, whose poetic plays, remarkably exploiting every scenic device, came to seem too clever and abstract. Armand Salacrou's nihilistic, cinematographic plays, ranging from almost neo-Surrealistic works to commercial theatre, were ignored. He reached his best with *Les Nuits de la colère* (1946; adapted as *Men of Darkness*, 1948) and *L'Archipel Lenoir* (1947). Jean Anouilh began as an extremely serious dramatist who entertained and shocked the public just enough to remain popular. Though his genius for traditional theatre sometimes made him appear facile, his reputation rested solidly on *Antigone* (1942), *L'Alouette* (1953; *The Lark*, 1955), and *Becket ou l'honneur de Dieu* (1959; *Becket; or, The Honour of God*, 1962).

The "theatre of cruelty"

Meanwhile, Antonin Artaud and Roger Vitrac's Théâtre Alfred Jarry (1927) passed almost unnoticed, though it produced the *théâtre de cruauté* ("theatre of cruelty"), in which Artaud theorized on "events" and "happenings." This was responsible for the production of plays by Artaud, Vitrac, and Raymond Aron, all neo-Surrealist works.

The commercial theatres at this period were dominated by Marcel Achard, Stève Passeur, Henri Lenormand, Édouard Rourdet (*Les Temps difficiles*, 1934), Sacha Guitry (author and often director of and star actor in innumerable minor comedies), and the delightful Marcel Pagnol (*Topaze*, 1928; *Marius*, 1929). Jules Romains was a brilliant humorist whose later devotion to the novel was a loss to the theatre. Other novelist-playwrights, such as Mauriac, Giono, and Gide, lacked this stage sense. The greatest dramatist of that generation, and in some ways Claudel's successor, was Henry de Monthcrlant, last survivor of the classical Symbolist tradition. His postwar plays—La *Reine morte* (1942; *Queen After Death*, 1951), *Le Maître de Santiago* (1947; *The Master of Santiago*, 1951), *Port-Royal* (1954), *Le Cardinal d'Espagne* (1960)—gained a worldwide reputation.

The post-1940 theatre presented a confused picture. At first the Existentialist left wing produced ardent political and philosophical allegories: Sartre with *Les Mouches* (1943; *The Flies*, 1946), *Les Mains sales* (1948; *Dirty Hands*, 1949), *Les Séquestrés d'Altona* (1959; *Loser Wins*, 1960), and *Nekrassov* (1955); Simone de Beauvoir with *Les Bouches inutiles* (1945; "Useless Mouths"); and Camus with *Caligula* (1945), *Le Malentendu* (1944; *Cross Purpose*, 1948), and *Les Justes* (1949). But for all these, the stage was a propagandist platform, so that, with all their talent, they looked back to the problem plays of the turn of the century. More original was Jean Genet, who, in *Haute Surveillance* (1949; *Deathwatch*, 1954) and *Les Bonnes* (1947; *The Maids*, 1954) to *Les Paravents* (1961; *The Screens*, 1962), created a complex anti-Naturalist and antisocial theatre, aggressively poetic but less tied to theatrical convention. More traditional was the poetic theatre of (among others) Jacques Audiberti (*Quoat-quoat* [1946]); Michel de Ghelderode, whose 30 plays (written 1918–37) only began to reach the stage in 1947; Henri Pichette (*Les Epiphanies*, 1947); Jean Vauthier (*Le Capitaine Bada*, 1952); Georges Chéhad6 (*Le Voyage,* 1961); and François Billetdoux (*Va donc chez Torpe,* 1961), though many of them had experimental qualities.

The year 1950 was to prove a turning point, with the production of Eugène Ionesco's ***Cantatrice chauve (The Bald Soprano,*** 1958) and Arthur Adamov's ***Grande et la petite manoeuvre*** and ***Invasion.*** The "antitheatre" thus initiated was really a "pure" theatre, stripped of convention, naked and abstract, crudely poetic, arbitrary, and imaginative. Ionesco, influenced by Kafka and Jarry, caricatured the real and the "absurd" together, exteriorized secret fantasies, often achieved profound social relevance, and was a chief creator of the Theatre of the Absurd. Adamov, an Expressionist, did likewise in ***Le Ping-pong*** (1955). Beckett, in ***En attendant Godot*** (1953; ***Waiting for Godot,*** 1954), ***Happy Days*** (1961; performed in French as *Oh! les beaux jours,* 1963), and ***Play*** (1963; performed in French as ***Come'die,*** 1964), perhaps achieved a greater universality, although he refused tragic catharsis. Other successful dramatists in this new stream were Marguerite Duras (***Le Square,*** 1955), Robert Pinget (***Lettre morte,*** 1960; ***No Answer,*** 1961), and Roland Dubillard, with his baroque tragedy La *Maison d'os* (1962; "The House of Bone"). French experimentalism, in the theatrical vanguard since perhaps 1950, seemed by the early 1970s to be losing touch with the "common" audience.

Poetry. Late-19th-century poetry was already split between a Neoclassical "pure" poetry, written in traditional forms, and a poetry of common speech, usually in *vers libre* ("free verse"). These tendencies persisted in the 20th century, with an occasional call to order. The best poets distrusted rhetoric or eloquence, although the leading poets of the early part of the century—Gmile Verhaeren, Claudel, Francis Jammes, Paul Fort, and Charles Péguy—not only failed to give a lead but themselves suffered from a Hugoesque verbosity. Influenced initially by Arthur Rimbaud, Claudel used a form that became known as the *verset claudélien*—a long, unscanned and usually unrhymed line—which he perfected but which few later poets adopted. The more restrained José Maria de Heredia, Henri de Régnier, Jean Moréas, Albert Samain, and Fernand Gregh all wrote into the 20th century, but classical Symbolism (as opposed to the Romantic Symbolism of Jules Laforgue, Rimbaud, etc.) found its last great representatives in Paul Valéry (***La Jeune Parque,*** 1917; *Charmes,* 1922), who purged poetry of rhetoric and social contingency while using it as a field of research into the creative act; and in Saint-John Perse. Influenced by Homer and Lucretius, Saint-John Perse tightened Claudel's versicle and used it to produce profound, metaphysically inclined poetry.

Reconciliation of conflicting styles in the poetry of Apollinaire

Guillaume Apollinaire synthesized contemporary tendencies while pointing in new directions with *Alcools* (1913) and ***Calligrammes*** (1918), and his posthumous collections added to his reputation. His formula reconciled classical intellectuality and form with Romantic enthusiasm, intuition, Expressionism, and experimentalism of every kind. Alongside him, Max Jacob, André Salmon, Pierre Reverdy, and others, under the influence of Cubism, gave new forms to personal utterance. Apollinaire also helped to unleash the forces of Dadaism and Surrealism by exalting the Marquis de Sade, Jarry, and Rimbaud.

Although the Surrealists numbered among them some readable poets—such as André Breton, Louis Aragon, Philippe Soupault, and Benjamin Péret—Paul Éluard became the movement's leader when, in 1931, he established the difference between the "Surrealist text," or document, and the "Surrealist poem," the text elaborated and formalized. His own work was intensely personal, interior, restrained, yet universal in its deceptive simplicity. Whereas Breton and Robert Desnos never overcame rhetorical extravagance, Aragon unexpectedly did so, with *Le Crève-coeur* (1941; "Heartbreak"). So also did Pierre Reverdy and Pierre-Jean Jouve, in his collections ***Sueur de sang*** (1933; "Bloody Sweat") and ***Gloire.*** Nonetheless, the liberating movement of Surrealism proved disappointing, showing that manifestos, a philosophy of the irrational, even a new imagery could not produce poems if form, style, intention, communication were despised. Thus Surrealist prose poetry is often more satisfying. Here, the marginal Surrealist Henry Michaux made an interesting Expressionist contribution with La *Nuit remue* (1934; "The Night Stirs"), while René Char developed a more classical style in his verse (***Les Matinaux,*** 1950; "The Early Risers"), prose poems, and aphorisms. Of others standing somewhat apart from the current, Jean Cocteau, despite his great gifts, was too much of an opportunist to mature properly and did not surpass his ***Plain-Chant*** (1923); while Jules Supervielle cultivated a certain malaise with wit and feeling but failed to produce really satisfactory verse.

The period after 1940 produced no French poet of international stature, although Jacques Prévert, Francis Ponge, and Raymond Queneau were all interesting. If the French poetic muse has seemed to be in the doldrums, more traditionalist work of high quality continued to be written by Maurice Fombeure, Yves Bonnefoy, Alexandre Toursky, Jean Tardieu, André Frénaud, Jean Follain, Henri Thomas, Georges Clancier, Alain Bosquet, and Marc Alyn. Contemporary French poets show little of the adventurousness that is to be seen in the novel and the theatre, and in the early 1970s there was little sign of their becoming more European or world minded. The unfortunate truth was that ever since Apollinaire, anti-intellectualism had undermined French poetry, and what was lacking was "the shaping spirit of imagination."

Literary criticism. French criticism had a strong historical bias. The "scientific" history of literature formulated by Gustave Lanson and backed by his useful bibliographies was continued by numerous scholars. A major achievement in this regard was the Pléiade *Histoire des littératures* (1955–58; "History of Literatures"), a world survey edited by Raymond Queneau. Against the historical approach stood the impressionistic, beginning with Rémy de Gourmont and Jules Lemaître, and related to it was the amateur criticism that had been an important part of the work of creative writers such as Proust, Gide, Valéry Larbaud, Paul Valéry, Paul Léautaud, Maurice Blanchot, Alain Robbe-Grillet, and Sartre. In the 1960s the opposition between experimental and academic criticism came to a head in the dispute between the "new" critics—Roland Barthes, Georges Poulet, Jean-Pierre Richard, Jean-Paul Weber, and others who had been influenced by the psychological stylistics and structuralist approaches of Leo Spitzer and Gaston Bachelard—and Raymond Picard, a leading representative of traditional academic criticism. Although the dispute remained unresolved, it was clear that some of the younger critics, such as those associated with the review ***Tel Quel,*** were already making important contributions to literary discussion in France. Unfortunately, no modern critical school in France was able to establish, as was done in Britain and the United States, a proper critical terminology. The most accessible new technique was that of the critical "Panorama," which combined criticism with texts; it was created by Gaetan Picon and followed by Georges Clancier, J. Rousselot, and Jeanine Moulin. See above ***Canadian literature in the 20th century: French*** and, below, ***Belgian literature in the 20th century: French.***

Provencal. Among the Provençal writers who became famous during the first half of the 20th century, although many of them had published a good portion of their work before 1900, should be mentioned Valère Bernard of Marseilles (*La Legenda d'Esclarmonda,* 1936), Marius André, Prosper Estieu, Antonin Perbosc, the Languedocian poet of ***Lo Got Occitan*** (1903), and Philadelphe de Gerde, a Pyrenean, and Miqeu Camelat, also a Pyrenean, whose pastoral poem ***Beline*** (1899) was a Gascon counterpart to Mistral's ***Mirkio.*** Contemporary with them were the Provencal writers of the Camargue region of the Rhône Delta: Joseph d'Arbaud, notable author of such stories as ***La Bestio dóu Vaccarès (The Beast of the Vaccarks),*** written in an epic vein, and such verse as ***Lou Lausié d'Arles*** (1913); Folco de Baroncelli-Jaron; and Marius Jouveau. A curious case was that of Sully-André Peyre, writing in Provencal, in French, and in English. Had Paul Froment (1875–98) not died young, he might well have become the great poet of Quercy; his place was taken by a bucolic and religious poet Jules Cubaynes, author of *L'Ome **de Dieu,*** a masterpiece of sacerdotal poetry. P.-L. Grenier was a notable artist who revived the classical language of Limousin. J.B. Chèze, in the same region, wrote the delicate *Una princessa dins **la Tor*** (1932). Albert Pestour of Limousin was the inspired poet of *Rebats sur l'autura* and *L'Autura **enviblada.*** Among other noteworthy Provençal writers born toward the end of the 19th century are P. Eyssavel, R. Rouquette, and their feminine compatriot Marcelle Drutel, as well as J.S. Pons of Rousillon. Contemporary writers are divided between those who remained faithful to the Mistral tradition—the "Provençals"—and those who have been won over to more modern ideas—the "Occitans." Yet such poets as J. Mouzat, C. Campreux, C. Galtier, L. Cordes, M.-P. Delavouet, R. Méjean, R. Nelli, R. Lafont, B. Manciet, and many others less well known, bear witness to the permanent vitality, throughout every langue d'oc-speaking province, of Occitanian poetry.

ITALIAN LITERATURE IN THE 20TH CENTURY

Gabriele D'Annunzio's nationalism. In Italy at the beginning of the 20th century, aggressive nationalism superseded peaceful internationalism; the attempt at the unification of Italy, the Risorgimento, was replaced by the selfishness of the middle classes; and primacy of society was replaced by glorification of the individual. The chief representative of this reaction to the 19th-century thinking was Gabriele D'Annunzio, whose life and art were a blend of Jacob Burckhardt's idea of the Renaissance "complete man" and the superman created by Friedrich Nietzsche, but his literary work was characterized by an inability to distinguish between life and art; he was at his best when writing about himself, as in an autobiographical novel, ***Il piacere*** (1889; ***The Child of Pleasure,*** 1898), in his mature poetry, and his late memoirs and confessions; it was worst when projecting his superhuman ideals as characters. D'Annunzio's "erotic-heroic" ideology (as he defined it) fired younger contemporaries, who identified it with the Fascist mystique.

Benedetto Croce's 'criticism. Although D'Annunzio's fame was worldwide, the function of organizing Italian culture and modernizing intellectual life fell mainly to Benedetto Croce in almost 70 books and in a bimonthly review ***La Critica*** (1903–44). Perhaps his most influential work was literary criticism, which he expounded and continually revised in articles and books spanning nearly half a century. He created a method of aesthetic appreciation independent of social and economic background and practical implications, which became so self-contained as to be insensitive and lend itself to misinterpretation by believers in "pure art" and critics who indulged in unsubstantiated accounts of their aesthetic impressions. Crocean criticism succeeded perhaps beyond Croce's intentions, stifled an earlier school of "historical criticism," and suppressed attempts to evolve stylistic or sociological theories.

Croce's beliefs implied condemnation of Fascist ideology, and he parted with a lifelong friend, Giovanni Gentile, whose theory of "absolute idealism" in his political development justified the "absolute state" and, by extension, the Fascist regime. Nevertheless, Croce was never seriously molested by Fascism, perhaps because the dictator Benito Mussolini wanted a showpiece of toleration of opponents and because Croce's enlightened conservatism did not seriously antagonize the ambitions of the Italian upper middle class, which Fascism manipulated. After the fall of Fascism, Croce's ideology proved unsuited to changed conditions, and he lost much of his authority. But his monumental corpus of philosophical, critical, and historical works, of great scholarship, humour, and common sense, remains the greatest intellectual feat in the history of modern Italian culture.

Literary trends before World War I. While Croce was starting his arduous task, literary life revolved mainly around reviews such as ***Leonardo*** (1903), *Hermes* (1904), ***La Voce*** (1908), and ***Lacerba*** (1913), founded and edited by relatively small groups. The two main literary trends were: *crepuscolarismo*, which favoured a colloquial style to express memories of sweet things past, as in the work of Guido Gozzano and Sergio Corazzini; and *futurismo*, loathing of traditional art and demanding complete freedom of expression, whose leader was Filippo Tommaso Marinetti, editor of ***Poesia,*** a fashionable cosmopolitan review. Both *crepuscolari* and *futuristi* were part of a complex European tradition of disillusionment and revolt, the former inheriting the sophisticated pessimism of French and Flemish "decadents," the latter taking part in an episode in the history of western European avant-garde developed from the French poets Stéphane Mallarmé and Guillaume Apollinaire to the Cubist, Surrealist, and Dada movements. In spite of being apparently at opposite extremes, both shared a feeling of revulsion against D'Annunzian flamboyance and rhetoric, from which they attempted to free themselves and, paradoxically, from which they derived elements of their style (the "crepuscular" mood of D'Annunzio's *Poema paradisiaco* [1893] and most Futuristic "new theories"—identification of art with action, heroism, and speed, free use of words—being implied in his *Laus **Vitae).***

Crepuscolarismo
and
futurismo

The "return to order." The wish for a "return to order" after World War I made it easier for the Fascist regime to impose its own form of "order." The review ***La Ronda***, founded in 1919 by Vincenzo Cardarelli and others, advocated a return from chaotic eclecticism to the saner aspects of Italian tradition; its Neoclassicism started a new literary genre, variously labelled "artistic prose," *elzeviro*, and ***capitolo*** and consisting in stylistic exercises to which considerations of content were almost irrelevant. As Fascist dictatorship thwarted the free expression of ideas, how to write became more important than what to write. Apart from the works of a few good storytellers, such as Riccardo Bacchelli, and the award of the 1926 Nobel Prize to the Sardinian novelist Grazia Deledda, the most notable event of the late 1920s was the "discovery" of Italo Svevo, whose talents would have won earlier recognition had he not been operating outside the literary establishment. His best known novel, a masterpiece of psychological insight, Jewish humour, and pessimism, was ***La coscienza di Zeno*** (1923; ***The Confessions of Zeno,*** 1930). By the time new promising novelists, such as Alberto Moravia, Corrado Alvaro, and Carlo Bernari, appeared in the 1930s, Italy was intellectually isolated, and interest in modern Italian literature was kept alive only by an anti-Fascist expatriate Ignazio Silone. But when his novels ***Fontamara*** (1930; Eng. trans., 1934) and ***Pane e vino*** (1937; ***Bread and Wine,*** 1937) appeared in Italy after the war, the public was unreceptive to their style and mood.

Luigi Pirandello. Drama, which a few playwrights and producers were trying to extricate from old-fashioned realistic formulas and more recent superhuman theories, was increasingly dominated by Luigi Pirandello. His own experience of the "unreal," through his calamitous family life and his wife's insanity, enabled him to see the limitations of realism. From initial short-story writing, in which he explored the incoherence of personality, the lack of communication between individuals, the uncertain boundaries between sanity and insanity or reality and appearance, and the relativity of truth, he turned to drama when—as he said—his characters urged him to give them life on the stage. He realized that life's absurdity could be better expressed by moving from the immediacy of narrative prose to drama, a literary genre already removed from reality, in which the ambiguous relationship between actors and the characters they represent, characters and real people, fact and fiction provided a metaphor of the contradictions of human existence.

To multiply the fragmentation of the levels of reality, Pirandello tried to destroy conventional dramatic structures and to adopt new ones: a play within a play in ***Sei personaggi in cerca d'autore*** (1921; ***Six Characters in Search of an Author,*** 1952) and a scripted improvisation in ***Questa sera si recita a soggetto*** (1930; ***Tonight We***

Pirandello's
play
within
a play

Improvise, 1932). This was a way of transferring the dissociation of reality from the plane of content to that of form, thereby achieving an almost perfect unity between ideas and dramatic structure. Pirandello's plays, including perhaps his best, *Enrico IV* (1922; *Henry IV*, 1960), often contain logical arguments: several, including Croce, were misled into thinking that he intended to express in this way a coherent philosophy, whereas he used logic as a dramatic symbol, a fixed landmark in the flux of human existence.

The Hermetic movement. Poetry in the Fascist period underwent a process of involution, partly influenced by French Symbolism, with its faith in the mystical power of words, and partly under the stress of changed political conditions after World War I, during which literature had declined. Many poets of the wartime generation, weary of tradition and rhetoric, had been seeking new expression: some, like the *futuristi*, had tried to work rhetoric out of their system by letting it run amok; others, such as Camillo Sbarbaro (*Pianissimo* [1914], *Trucioli* [1920; "Shavings"]), cultivated a style purified of unessential elements. Out of those efforts grew a poetry combining the acoustic potentialities of words with emotional restraint and consisting mainly of fragmentary utterances in which words were enhanced by contextual isolation and disruption of syntactic and semantic links. The resultant obscurity compensated poets for loss of influence in a society subservient to dictatorship by turning them into an elite and allowed some, notably Eugenio Montale, to express their pessimism covertly. The name of this movement, *ermetismo* ("Hermeticism"), hinted at both its aristocratic ambitions and its esoteric theory and practice. Its leader, Giuseppe Ungaretti, tried to charge each word of his early poems with such intensity of meaning that concern with technical problems often overshadowed emotion, thus producing supremely stylized forms. Thus, what in the 1920s had appeared revolutionary proved later to be only another facet of the formalistic tradition. Against this background of refinement, obscurity, and unreality, only the simple and moving poems of Umberto Saba preserved an immediate appeal.

The Neorealist movement and after. During World War II the walls of the hermetic ivory tower began to crumble. Ungaretti's style became so articulate as to be almost unrecognizable. Salvatore Quasimodo adopted a new *engagé*, or committed, style, which won critical admiration, and others followed suit in a drift toward social realism.

Social realism

This development had been foreshadowed by some writers under Fascism. In 1929 Alberto Moravia had written a scathing indictment of middle class moral indifference, *Gli indifferenti* (*The Indifferent Ones,* 1932). Carlo Bernari wrote a novel about the working classes, *Tre operai* (1934; "Three Workmen"); Cesare Pavese produced *Paesi tuoi* (1941; *The Harvesters,* 1961); and Elio Vittorini wrote *Conversazione in Sicilia* (1941; *Conversation in Sicily,* 1948); all definitely promised a new literary development. From these and from certain aspscts of U.S. literature (William Faulkner, Erskine Caldwell, John Steinbeck, John Dos Passos, Ernest Hemingway, translated mainly by Vittorini and Pavese), postwar Neorealism took its cue. Writers were confronted with a wealth of unused material. There were the social and economic problems of the south, described in all their magnitude by Carlo Levi in his social, moral, and poetic portrait of the people of the primitive southern Italian province of Lucania, *Cristo si è fermato a Eboli* (1945; *Christ Stopped at Eboli,* 1947) and by Rocco Scotellaro (*Contadini del sud,* 1954) and Francesco Jovine (*Le terre del Sacramento,* 1950). The problems of the urban proletariat were touched by Vasco Pratolini in his portraits of the Florentine working classes (*Cronache di poveri amanti* [1947; *A Tale of Poor Lovers,* 19491, *Metello* [1955]) and exposed by P.P. Pasolini in his tales of the Roman underworld (*Una vita violenta* [1959]). There were memories of the industrial north's struggle against Fascist and Nazi domination, which excited Vittorini, Pavese, and Beppe Fenoglio; and there were sad tales of a lost war recounted by Giuseppe Berto and Mario Rigoni Stern. No longer needing to invent a plot, many became preoccupied with technical details.

Carlo Cassola, in many ways a forerunner of the French *nouveau roman,* or "new novel," emptied his stories of action in a quest for narrative objectivity exemplified in *Il taglio del bosco* (1959; "The Felling of the Forest"); and he was often unimaginative in recording every minute event. A similar tendency was apparent in Moravia's later work, although curiosity about human psychology prevented him from drawing uninteresting characters. Pavese began to use recurrent images (misty hills, winding country roads, empty town streets, sterile seashore) as symbols of a writer's struggle to come to terms with his world; they were used later by film director Michelangelo Antonioni as symbols of man's alienation. Italo Calvino turned to writing fantastic tales such as *I nostri antenati* (1960; three tales translated as *The Baron in the Trees,* 1959, and *The Nonexistent Knight* and *The Cloven Viscount,* 1962) and, later, moralizing science fiction such as *Le Cosmicomiche* (1965; *Cosmicomics,* 1968) and *Ti con zero* (1967; *T Zero,* 1969). Vittorini, paralyzed by stylistic anxiety, turned to criticism and the encouragement of younger writers. Thus "engaged" literature became inward looking and lost its grip on reality. This tendency also affected drama; from practical issues of moral justice and responsibility with which he had dealt earlier, Ugo Betti was drawn to representation of his characters' hidden metaphysical problems, which gave his later plays their ambiguous fascination.

The influx of foreign ideas

As a result of the return to normality after intellectual poverty, Italy was flooded with foreign ideas; intellectual debates gained so much in breadth and subtlety that artists found it more rewarding to feed on them rather than on real-life problems, and writers increasingly interpreted in the light of new critical ideas. Dino Buzzati's charming tales *Il deserto dei tartari* (1940; *The Tartar Steppe,* 1952) and *Sessanta racconti* (1958; "Sixty Stories") were seen as reminiscent of Kafka. Carlo Emilio Gadda's language, full of subtle and unsubtle puns, and rare, vulgar. and invented words, was seen as the influence of the Irish author James Joyce. The need for a return to native sources was shown by the relish with which readers turned in the late 1950s to old-fashioned historical novels, among which perhaps only *Il gattopardo* (1958; *The Leopard,* 1960), written earlier by a Sicilian nobleman, Giuseppe Tomasi di Lampedusa, and *Cinque storie ferraresi* (1956; *A Prospect of Ferrara,* 1962), edited by Giorgio Bassani, were considered of lasting value. Trends in the 1960s confirmed literature's preoccupation with itself, symbolized by Moravia's *Attenzione* (1966; "The Lie"), about a writer writing about the writing of his novel.

Such literary narcissism does not seem to have produced many new ideas. In the same way as Neorealism failed to overcome the ideological and aesthetic limitations of 19th-century realism, the Neo-avant-garde and Neo-experimentalist movements of the 1960s and early 1970s appeared unable to say much that had not already been said by Surrealism and Dada over half a century before. Such poetry as that of Edoardo Sanguineti, Elio Pagliarani, Alfredo Giuliani, and Georgio Manganelli proved to be just as inaccessible to the public of the early 1970s as was the Hermetic poetry of the 1930s.

These movements are caught in a painful contradiction. On the one hand, their members reject class-bound "establishment" literature in the name of democratic and revolutionary ideals. On the other, their attempts to reject the standard language, seen as the incarnation and the instrument of bourgeois cultural and political hegemony, cut them off from the vast majority of the reading public and turn them into a literary caste even more exclusive than the establishment they wish to destroy.

Some novelists, however, less given to introspection and to extreme ideological stances, have been on the whole more successful than the avant-gardes in expressing the crisis of industrial capitalism; among them such novelists are Ottiero Ottieri (*Donnarumma all'assalto,* 1959), Paolo Volponi (*Il memoriale,* 1962), Giovanni Testori (*Il fabbricone,* 1963), and Lucio Mastronardi (*A casa tua ridono,* 1971).

SPANISH LITERATURE IN THE 20TH CENTURY

The generation of 1898. For over two decades before 1900, a mood of seething political and social analysis developed in Spain that gave in Ángel Ganivet's *Idearium español* (1897; *Spain, an Interpretation*, 1946) one of the most searching analyses of the Spanish character ever written. The imperial cycle begun in 1492 ended in ignominy with the Spanish–American War of 1898, and thinking Spaniards embarked on a diagnosis of their country's ills and an attempt to shock the national mentality out of its *aboulia,* or "lack of will." The novel was injected with a new seriousness of purpose, and the critical, psychological, and philosophical essay rose to new importance. Novelists and essayists constituted what Azorin (pen name of José Martinez Ruiz) called the "generation of 1898," a group that regained respect for Spanish letters abroad.

Unamuno's analyses of the national problem

Miguel de Unamuno, who dominated the literary scene for a generation, studied the national problem acutely in the five essays in *En torno al casticismo* (1895; "On Spanish Purism") and in the *Vida de Don Quijote y Sancho* (1905; *The Life of Don Quixote and Sancho,* 1927). He examined the problem of immortality in his most important work, *Del sentimiento trágico de la vida* (1913; *The Tragic Sense of Life in Men and Peoples,* 1921). A provocative rather than a systematic thinker, he aimed at sowing spiritual disquiet. The novel was to him a medium for discussion of the fundamentals of personality; his own include *Niebla* (1914; *Mist,* 1929), *Abel Sanchez* (1917), and *Tres novelas ejemplares y un prólogo* (1920; "Three Cautionary Tales and a Prologue").

Azorin concerned himself with the reinterpretation of earlier literary values and of the Spanish countryside in, for example, *El alma castellana* (1900; "The Castilian Soul"), *La ruta de Don Quijote* (1905; "Don Quixote's Route"), and *Clásicos y modernos* (1913). An artist in criticism and a finely sensitive miniaturist, he contributed powerfully to the deflation of the rhetoric that had vitiated much 19th-century writing. A philosopher, José Ortega y Gasset, developed from criticism and psychology (*Meditaciones del Quijote* [1914; "Meditations on Quixote"]) to national problems (*La España invertebrada* [1921; *Invertebrate Spain,* 1937]), then to international (*El tema de nuestro tiempo* [1923; *The Modern Theme,* 1931] and *La rebelión de las masas* [1929; *The Revolt of the Musses,* 19321).

Pio Baroja repudiated tradition, religion, and the cult of the individual and advocated social action. *La raza* ("The Race"), *La lucha por la vida* (1904; "The Struggle for Life"), and *Agonías de nuestro tiempo* (1926; "Agonies of Our Time") were fiercely vigorous attempts to arouse discontent with material conditions. As vigorous, but possessing greater narrative skill, was Vicente Blasco Ibáñez, who wrote on contemporary social problems from the standpoint of an anarchist, as in *La bodega* (1905; *The Wine Vault,* 1923) and *La horda* (1905; *The Mob,* 1927). He won international renown with *Los cuatro jinetes del apocalipsis* (1916; *The Four Horsemen of the Apocalypse,* 1918), a novel of World War I. Ricardo León, a conscious stylist, sought to recapture the traditional Christian values of a society threatened with materialism.

*No*ecentistas*

'The term *novecentistas* is applied to writers of the early 20th century who sought to renew intellectual and aesthetic standards after the passionate involvement of their immediate predecessors.

The novel. In Ramón Pérez de Ayala, the novel was at once a polished art form and a forum for philosophical discussion. *Belarmino y Apolonio,* a projection of the old debate between faith and reason, made its characters almost symbolic, as did *Tigre Juan,* on the traditional theme of honour. Gabriel Miró's perfect descriptive prose retarded the action of his novels, but he remained a supreme artist in words. The novel as a literary form fell under the influence of Ortega y Gasset, who in *La deshumanización del arte* (1925; *The Dehumanization of Art,* 1948) propounded principles of a pure, depersonalized art. Analyzing the novel as an art form, he predicted its decline. In the following decade Benjamin Jarnés and others attempted, without complete success, to apply a technique of pure art to the novel; Jarnés' works were outstanding examples of the Surrealist novel in Spain. The publication in 1929 of Ramón José Sender's *Imán* (*Earmarked for Hell,* 1934), based on military experiences in Morocco, indicated a return to realism and social criticism. It was attentive to the human situation in all its horror and to a search for a satisfying philosophical and ethical system.

Novels after the Spanish Civil War of 1936–39 often showed emptiness and disillusionment. Camilo José Cela created a realistic style depicting rural violence in horrific detail in *La familia de Pascual Duarte* (1942; *Pascual Duarte's Family,* 1946): this early style was named *tremendismo. La colmena* (1951; *The Hive,* 1953), much more mundane, described the post-Civil War poverty and misery of Madrid.

Drama. With the new century, drama achieved new vigour under the stimulus of Jacinto Benavente y Martinez, winner of the Nobel Prize for Literature in 1922, whose plays were skillful and witty. After a period of effervescence and irony, Benavente showed a preoccupation with aesthetics and later with ethics. By contrast, Gregorio Martinez Sierra remained true to a more traditional Spanish atmosphere, in which idyllic sentiment, though always deep-rooted in Spanish realities, breathed a sensuous joy in nature and life (*Tearro de ensueño,* 1905; "Dream Theatre"; and *El reino de Dios,* 1916; *God's Kingdom*). If the brothers Serafín and Joaquín Alvarez Quintero contributed nothing new to dramatic art, the vivacity of their many Andalusian comedies never failed to entertain. Federico Garcia Lorca stood far above his contemporaries. His drama was poetic in more than the usual sense, presenting elemental passions with an intensity that made the characters poetic symbols of man's tragic impotence to arrest his fate. His dramatic poetry was modern yet traditional, personal yet universal, surrealist yet childlike. His plays *Bodas de sangre* (1933; *Blood Wedding,* 1939), *Yerma* (1934; translated 1941), and *La casa de Bernarda Alba* (1936; *The House of Bernarda Alba,* 1941) depicted extremes of passion involving the traditional Spanish theme of honour.

Garcia Lorca's drama

An impulse of renovation visible in the novel after the Civil War failed to reach the theatre, although Antonio Buero Vallejo attempted to revitalize the drama by using social criticism and Miguel Mihura by introducing the Theatre of the Absurd. The *género chico,* or "little genre" (the short sketch, lively and often with music), continued to charm ordinary people, whose way of life supplied it with setting and theme. Ricardo de la Vega and Carlos Arniches were outstanding in this form.

Poetry. Rubén Dario, Latin America's greatest poet, took *modernismo* to Spain in 1892. In general, modernism was a reaction against 19th-century bourgeois materialism and a search for other and more specifically aesthetic values of life. Dario's Modernism greatly enriched the musical resources of Spanish verse, notably by a daring use of new rhythms and metres; and under his influence poetry in Spain underwent a renaissance.

Modernism had scarcely any contact with the generation of 1898. The Modernists, apart from the novelist Ramón Maria del Valle-Inclán, whose four *Sonatas* (1902–05) typified the four seasons of his hero's life, were almost all Andalusian poets in pursuit of self-realization, esoteric beauty, and refinement of style; the generation of 1898 were mostly northerners and chiefly concerned with content.

Juan Ramón Jiménez, another Nobel Prize winner (1956), soon discovered poetry to be for him not a withdrawal but reality itself; and, as he strove to capture the poetic essence of each passing moment, his style, initially ornate, became more and more bare. Using and gradually winnowing all the resources of Modernism, he became a powerful influence on the younger generation. The brothers Machado well exemplified the Andalusian–Castilian dichotomy: Manuel, always loyal to his native Seville, was the Modernist; Antonio lost his heart to Soria, in Old Castile, and wrote of the recesses of the poet's soul or of the unchanging soul of Castile.

The poetic reputation of Garcia Lorca grew steadily after his death. He was a poet, as he was a dramatist, of fundamentals and showed of what starkness of feeling and effect the traditional song and ballad forms were still capable. His mature poems revealed the potency of the inspiration that he drew from Andalusian ways and folklore.

The Civil War drove many notable poets into exile, Luis Cernuda one of the greatest. In later verse the tone of most major poetry was political or tending toward "social realism," as in the works of Angel González, José Angel Valente, and Blas de Otero.

PORTUGUESE LITERATURE IN THE 20TH CENTURY

The "Integralist" school

The passage from monarchy to republic in Portugal in 1910 saw a revisionary urge in literature associated chiefly with Oporto and the *Renascença Portuguesn* ("Portuguese Renaissance"). Leonardo Coimbra was its philosopher, and Antbnio Sérgio its critic and historian. Its poets—Mário Beirão, Augusto Casimiro, and João de Barros—adopted the *saudosismo* of Teixeira de Pascoais as the key to the nation's recovery of greatness, though the inadequacy of this nostalgia was soon realized. The "Integralist" school reacted from 1913 onward in favour of Catholic monarchist tradition, led by António Sardinha, a historian and poet. Fernando Pessoa, author of *Mensagem* (1934), was posthumously regarded as the most inspired poet of his generation, and his influence enriched poetry without imposing a movement. Another prominent writer was José Régio, an outstanding dramatist and a powerful religious poet. Miguel Torga's verse, like his stories and volumes of autobiography, revealed a faith in the primitive virtues of mankind.

Among novelists, Aquilino Ribeiro was a prolific writer whose themes often centred on his native Beira; his delight in life stood close to awareness of decay and death. J.M. Ferreira de Castro, author of *A Selva* (1930; Jungle, 1934), a brilliant evocation of his time as a worker in an Amazonian rubber plantation, was deeply concerned with social problems. J. Paço de Arcos, a widely travelled and penetrating observer, wrote *Ana Paula* (1938), *Ansiedade* (1940), and *Tons Verdes em Fundo Escuro* (1946). Social and psychological trends were united in J. Rodrigues Miguéis' *Pdscoa Feliz* (1932) and *Onde a Noite se Acaba* (1946). Novelists who treated African colonial questions included Castro Soromenho, and an important woman novelist of the postwar period was Agustina Bessa Luis, whose *Sibila* (1955) was a complex psychological study.

Notable work came out in history and literary criticism, and through this work Portugal's history and literature at last became more adequately known.

LATIN-AMERICAN LITERATURE IN THE 20TH CENTURY

Rubén Dario and the Modernists (1888–1910). A measure of political and economic security vouchsafed the emergence of a cosmopolitan awareness of life and letters that resulted in the revolt against the sentimental romantic versifiers filling the bulging pages of national anthologies. The revolt began when young writers, independently and across the Americas, withdrew temporarily from the local whirlpool to revitalize themselves in the mainstream of world thought and writing. The young despondents and pessimists attempting to shut themselves out from their immediate environment in artificial worlds of their own making were the transition group disparagingly labelled *modernistas* (Modernists) by the older generation. They wrote on exotic themes—the ancient past, the distant Orient, lands of childhood fancy and opiate creation. Beauty was their goddess, "art for art's sake" their creed.

Foremost among the early Modernists were a Mexican, Manuel Gutiérrez Nájera, whose elegiac verse and restrained rhythmical prose sketches and tales best represented the trend from Romanticism to Modernism; a Colombian, José Asuncibn Silva, who wrote a small but influential body of poems; the morbid Julián del Casal and his compatriot José Martí, martyr and symbol of Cuba's struggle for freedom from Spain, whose inspired prose style and deceptively simple, sincere verse set his work above and apart from all schools and movements.

But the full flowering of Modernism came under the leadership of Rubén Dario (died 1916) of Nicaragua, whose collection of verse and prose, *Azul*, published in Chile in 1888, was the cornerstone of Latin-American Modernism. *Prosas profanas* (1896; "Profane Prose") represented the high point of the escapist, cosmopolitan phase of the movement. When Spain's empire crumbled in 1898 and mutual sympathy allayed the centuries-old distrust between Spain and its former colonies, Dario created new songs of life and hope for Spanish-speaking peoples. His *Cantos de vida* y *esperanza* (1905; "Cantos of Life and Hope") were a blending of the best of Modernist form with the more enduring themes of Hispanic lands. Dario's imitators, particularly of his early experimental, escapist phase, were responsible for the emergence of a cult that bore his name, *rubendarismo.* But many of Dario's contemporaries were individuals of exceptional merit: Mexico's Amado Nervo, whose mysticism was reflected in *Serenidad* (1914) and *Elevación* (1917); Peru's José Santos Chocano (1875–1934), whose exalted Americanism gave birth to *Alma América* (1906); Bolivia's Ricardo Jaimes Freyre (1872–1933), who drew upon Scandinavian mythology for *Castalia bárbara* (1899); Colombia's Guillermo Valencia, whose classic bent was manifest in *Ritos* (1898); and Uruguay's philosopher and essayist José Enrique Rodb, whose *Ariel* (1900; Eng. trans., 1921) distinguished him as the leading critic and exponent of Modernist ideals.

After Dario's death, Leopoldo Lugones of Argentina was a leader from early Modernist days to the later phases of an arbitrary nationalism; Enrique Gonzalez Martinez was a sensitive pacesetter of later poetic trends in his native Mexico; and Julio Herrera y Reissig of Uruguay was the outstanding Symbolist of Modernism. But the vast majority of Spanish-American Modernists continued to be spellbound by the verbal magic and brilliance of Dario's *Prosas profanas.*

Brazilian works

In Portuguese-speaking Brazil, reaction against Romantic verse never produced the rich mosaic of Spanish-American Modernism. The Brazilian Parnassians with their formalistic, detached poetry were challenged by poets attracted to French Symbolism, but a Symbolist movement as such never materialized in Brazil. Parnassianism, as epitomized in the poetry of Raimundo Corréia, Alberto de Oliveira, and Olavo Bilac, was unchanged until the "stirring 1920s" of Brazilian Modernism.

Vanguard literature (1918–30). The horror and bloodshed of the Mexican Revolution (1911–17) shocked a complacent intellectual minority into a realization of the plight of their country's submerged masses: the Indian and mestizo (or person of European and Indian parentage) found able champions of their cause. Youthful liberals rejected Modernist tenets and wordy descriptions of local scenes and types, largely in imitation of European Realist and Naturalist techniques, to create the themes, form, spirit, and language that were unequivocal testimony that Latin-American literature had at last come of age. The Spanish Civil War and World Wars I and II gave continental and international perspective to America's attempt to define its place in the world. Worldwide depression and dollar diplomacy fostered radical economic philosophies offering deliverance from imperialism. Latin-American literature was a dramatic tribute to this ferment of self-discovery and self-assertion.

Women lyricists

Shortly after 1900 the Modernist leaders Dario, Nervo, and Lugones initiated a return from their more formal and exotic innovations to traditional forms and sincerity and simplicity. Their followers, however, never wholly rejected the enriched versification and language and countless new thematic sources of Modernism. Of these, an extraordinary band of women stood out whose lyrics were more widely enjoyed than most of the work of male poets in the first half of the 20th century. Love in its impassioned and transcendental manifestations, maternal longing, and social protest were the themes of Delmira Agustini, Alfonsina Storni, Juana de Ibarbourou, and

Gabriela Mistral, winner of the 1945 Nobel Prize for Literature. Stemming directly from the Modernist tradition were the unpretentious lyrics of Enrique Banchs; the experimental verse of a subjective Mexican, Ramón López Velarde; the mystical expression of a Guatemalan, Rafael Arévalo Martinez: the brilliant sonnet sequence of the tropical scene in *Tierra de promisión* (1921; "Land of Promise"), by a Colombian, José Eustasio Rivera; and the intensely racial rhythms of Cuba's nativist mulatto poet Nicolás Guillén. With the post-Modernist return to more traditional forms and themes, other groups, following the nihilistic waves of post-World War I "isms," experimented with free verse, often daring to use obscure imagery that gave a mistaken impression of a coldly intellectual mood. Many *ultraistas,* as these vanguard experimenters were labelled, wrote with socio-political overtones.

Borges and the *ultraísta* movement

Vicente Huidobro of Chile launched one of these ephemeral movements, *creacionismo,* in Spain in 1918. His denial of the poetic past was offset somewhat by the classical leanings and love of the national past that rendered the novel imagery of Argentina's Jorge Luis Borges less vulnerable to the charges of obscurity leveled at vanguard groups, including those of the *ultraísta* movement launched by Borges in Buenos Aires in 1921. César Vallejo of Peru represented a shift from Modernism to the Marxist emphasis of the *ultraístas* of the 1930s. An intensely subjective Chilean, Pablo Neruda, winner of the 1971 Nobel Prize for Literature, also adopted Marxism. His influence bade fair to equal that of Dario and of the 17th-century poet Luis de Góngora in their time. At a Modern Art Week program in São Paulo in 1922, the vanguard revolt in Brazil, usually referred to as Modernism, broke away noisily from academicism and colonial cultural bondage, until the world depression of 1929 and a new political order (1930) under Getúlio Vargas sobered its aggressive unconventionality. Although incited by European iconoclasts, the Brazilian Modernists' primary aim was to modernize national thought and life. This high-pitched, often theatrical, self-searching period of aesthetic re-evaluation and of analysis of the immediate Brazilian present served as a sorely needed purge and produced men of importance in Brazilian literature, such as the movement's high priest, a gifted poet and musicologist, Mario de Andrade; his lieutenant, Oswald de Andrade; a critic and bard, Ronald de Carvalho; and Manuel Bandeira, who has been acclaimed the country's greatest modern lyric poet. Preoccupation with social and metaphysical problems and an imperative urge toward untrammelled self-expression characterized the poetry of Modernist contemporaries or followers—Jorge de Lima, Cecilia Meireles, and Augusto Frederico Schmidt.

Literature of social protest. Contrary to its exceptional development in Brazil, the Spanish-American novel had been left largely to inspired followers of the 19th-century French Realist Émile Zola. Federico Gamboa of Mexico, who wrote *Santa* (1903), about a provincial prostitute reminiscent of Zola's *Nana,* and Baldomero Lillo of Chile gave artistic balance to their social-protest fiction. Later novelists, essayists, and short-story writers developed new and more effective techniques from such stylistic innovators as the Irish novelist James Joyce and the Austrian Franz Kafka.

Martí and González Prada were hailed as the intellectual progenitors of men aware of their responsibilities in guiding the Americas in a rapidly changing world. Rufino Blanco Fombona of Venezuela assailed his country's tyrants with more passion than poise as in *El hombre de hierro* (1907; "The Man of Iron") and *El hombre de oro* (1916; *The Man of Gold,* 1920). He was also among the first to attack "Yankee imperialism," abetted by a militant Argentine, Manuel Ugarte, whose *Porvenir de América latina* (1911; "The Future of Latin America") and *Destino de un continente* (1923; *The Destiny of a Continent,* 1925), together with the writings of a Peruvian, Francisco Garcia Calderón, envisioned Latin America as the future guardian of the Latin tradition. Ricardo Rojas, Argentinian literary historian and critic, and José Vasconcelos, controversial Mexican philosopher and educator, were more concerned with racial and cultural aspects within the American family of nations. Alfonso Reyes, Mexican poet, scholar, and literary critic, made the essay a vital, intimate force.

It was in the short story and the novel that the panoramic grandeur of the American scene in all its drama was best described. The disheartening years following the abolition of slavery in 1888 and the establishment of a republic in 1889 made serious-minded Brazilians analyze their troubled homeland as an extraordinary amalgam of man, land, and climate. An army engineer and journalist, Euclides da Cunha, revealed the bedrock of Brazilian life in an unclassifiable epic *Os Sertões* (1902; *Rebellion in the Backlands,* 1944), which was the first written protest on behalf of Brazil's forgotten man, the emerging Brazilian of the backlands. In *Canaã* (1902; *The Canaan,* 1920), a novel of ideas, José Pereira da Graça Aranha focussed on the effects of recent European immigration upon this evolving Brazilian type. A remarkable collection of short stories, *Urupês* (1918), by José Bento Morteiro Lobato, showed that intellectuals were still probing for native traits in a search that gave direction to the Modernist outburst of the 1920s.

Regionalist Brazilian literature

A new cultural regionalism of Brazil's "northeastern school" that flowered after 1930 produced gifted and dedicated prose writers, including a sociologist, Gilberto Freyre, whose *Casa Grande e Senzala* (1933; *The Masters and the Slaves,* 1946) was fundamental to an understanding of the region. José Lins do Rêgo, in a highly personal, evocative style, depicted the clash of the old and new way of life in his classic "sugarcane" cycle and in *Pedra Bonita* (1938); Jorge Amado gave Brazil some of America's best "proletarian" literature in *Cacau* (1933), *Jubiabá* (1935) and *Terras do Sem-Fim* (1942; *The Violent Land,* 1945). *Angústia* (1936), by Graciliano Ramos, attested to the fact that the individual inner struggle had also been adroitly plumbed by these new "regionalists." Érico Veríssimo was one of Latin America's few cosmopolitan writers of distinction, as demonstrated in *Olhai os Lirios do Carnpo* (1938; *Corzsider the Lilies of the Field,* 1947) and *O Tempo e o Vento* (1950; *Tinze and the Wind,* 1951). The most widely acclaimed work resulting from the Mexican Revolution was *Los de abajo* (1915; *The Under Dogs,* 1929), a novel conceived over the campfire by Mariano Azuela, an army doctor of one of the bands of the revolutionary Pancho Villa. Azuela alone chronicled the revolution from its preliminary rumblings to its later social implications. From the late 1920s on, Martin Luis Guzmán, Gregorio López y Fuentes, and José Rubén Romero gave breadth and scope to Azuela's treatment.

The 20th-century Indian was not the uncompromising hero of the epic *La araucana,* nor the symbol of colonial revolt against tyrannical Spain, and much less the "noble savage" of America's untamed romantic wilderness: he was the victim of centuries-old political and economic forces that kept the masses in abject bondage to colonial institutions. This new *Indianista* literature had already found favour in fiction in the novel *Aves sin nido* (1889; *Birds Without a Nest,* 1904) by a Peruvian, Clorinda Matto de Turner; and the Indian's cause had been advanced by Gonzalez Prada, a precursor of the militant pro-Indian, social reform party APRA (Alianza Popular Revolucionaria Americana) in 1923. A spokesman of the brutally realistic school of Ecuadorean writers, Jorge Icaza, skirted dangerously close to aesthetic suicide in mass-minded, vernacular-ridden novels, *Huasipungo* (1934; *The Villagers,* 1964) and *En las calles* (1935).

The effects of the tensions between the pampa and the city

The clash between the forces of nature and powerful economic forces, the pampa and the city, had by no means died with the passing of 19th-century local chiefs. Land and nature's gifts were more zealously sought than ever before: cattle on the pampas, tin and copper and nitrates in the mountains and deserts, rubber and bananas in the tropics. And the struggle between the two became more violent as man fought against man, selfishly, to possess nature alone. In the growing urban centres, modern industrial economy exerted an even more insidious control over the destiny of the masses. The clash between the old and the new on the pampas of Argentina and Uruguay created the sombre descriptive pages of Uruguay's short-story writer Javier de Viana, the psychological portrayal of rural types in *El terrufio (1916;* "The Native Country"), by Carlos Reyles, also of Uruguay; a simple, humorous narrative of gaucho life in *El inglés de los giiesos* (1924), by Argentina's Benito Lynch, and an image-studded prose epic of the gaucho, *Don Segundo Sombra (1926;* Eng. trans., 1935), also by an Argentine, Ricardo Giiiraldes. In *Doña Barbara (1929;* Eng. trans., *1931*) Rómulo Gallegos gave a dramatic poetic depiction of similar forces at work on the Venezuelan *llanos,* or "plains."

From the tropics appeared two artists who spoke for a growing number of younger writers who had discovered the *selva,* or "jungle": Horacio Quiroga of Uruguay, consummate short-story artist, who excelled in dramatic descriptions of the Chaco and the jungle of Misiones, Argentina; and a Colombian poet, José Eustasio Rivera, whose sole prose work, *La vorágine (1924; The Vortex,* 1935), was a powerful denunciation of exploitation in the upper Amazon during the rubber boom of the early *1900s.*

In *La maestra normal (1914)* an Argentinian novelist, Manuel Galvez, captured forever the universal pettiness and monotony of smaller provincial centres before modern mechanized manners shattered old colonial ways. Chile's underprivileged had two champions in a novelist, Joaquín Edwards Bello, and a short-story writer, Manuel Rojas. Man's struggle with the deeper forces within and beyond himself was re-created in the psychological novels of a Chilean, Eduardo Barrios, and in the "clinical" tales of the Cuban Alfonso Hernandez Catá.

In their analysis of individual and collective existence in the Americas, writers often turned to the past: *La gloria de Don Ramiro (1908; The Glory of Don Ramiro,* 1924), by Enrique Rodriguez Larreta of Argentina, a stylist, was generally considered the best historical novel in Spanish.

Recent trends. The literature of the third quarter of the 20th century was characterized by an increased preoccupation with the problem of man as the victim of alienating forces, of solitude, identity, anguish, and evil and by a marked determination to create new forms and new techniques. Above all, it displayed a new language more responsive to the demands imposed by increasingly complex spiritual, social, and ideological concerns. These concerns found a more direct expression in the essay, a form that has been cultivated with signal distinction by writers equally competent in other literary genres. Individual and collective preoccupation with analysis ran the gamut from the ruthless dissection of his country's ills by the Argentine Ezequiel Martinez Estrada (*Radiografia de la pampa, 1933; X-Ray of the Pampa, 1971)* to the more varied, deeper soundings of the Mexican Octavio Paz *(El laberinto de la soledad, 1950; The Labyrinth of Solitude, 1961*). Poetry veered from experimental and hermetic emphasis to seek answers to the same concerns in a more limpid and more personalized style. But it was the revitalized manner of the older poets that still held captive a younger generation not yet prepared to share fully in the honours bestowed on masters such as Neruda *(Selected Poems, 1970)* and Paz *(Selected Poems, 1963).*

The drama—hitherto the least significant genre in spite of its success in calling international attenion to the Uruguayan Florencio Sánchez *(Representative Plays, 1961)*—emerged from the traditional nativist theatre of local customs and types to embrace new approaches and techniques. These resulted in a mad ensemble of experimental thrusts reflecting and appropriating innovations as diverse as those of the American playwrights Edward Albee and Tennessee Williams, a theatre indulging in Surrealist vagaries, social and political satire, "angry" protest, and "dead pan" portrayal of the "Absurd." The first of the moderns were the Brazilian Joracy Camargo *(Deus lhe Pague, 1932)* and the Mexican Rodolfo Usigli (El *gesticulador, 1937,* and *Corona de sombra, 1943 [Crown of Shadows, 19461).* Of the younger generation, Sebastián Salazar Bondy of Peru, Emilio Carballido of Mexico, Osvaldo Dragún of Argentina, Jorge Díaz of Chile, and José Triana of Cuba were rapidly winning international recognition. Ariano Suassuna of Brazil stood out above all with his *Auto da compadecida (1955; The Rogue's Trial, 1963*), a highly sophisticated adaptation of Brazil's northeast folk theatre.

Prose fiction occupied the centre of the Latin-American literary scene. This "new" fiction gave full voice to the concerns of the generation through a disconcerting maze of techniques that launched an unremitting attack on both the form of the novel and the structure of the language. This "new" narrative prose was already in evidence in the work of a small band of earlier writers who had distinguished themselves in other genres, exemplified especially in the intellectual and "extraordinary" tales of the Argentine Jorge Luis Borges *(Labyrinths, 1962; The Aleph and Other Stories, 1933–1969, 1970);* in the Surrealist, terrifying novel *El señor presidente (1948; The President,* 1963), of the Guatemalan Miguel Angel Asturias, winner of the *1967* Nobel Prize for Literature; in the Joycean fiction of the Mexican Agustin Yáñez (*Al filo del agua, 1947; The Edge of the Storm, 1963);* in the ideological and philosophical novels of the Argentine Eduardo Mallea *(La bahía de silencio, 1940* [*The Bay of Silence,* 1944], and *Todo verdor perecerá, 1941 [All Green Shall Perish, 19661);* in the "magic realism" of the Cuban Alejo Carpentier *(Los pasos perdidos, 1953; The Lost Steps, 1956);* in the earthy, enigmatic prose of the Brazilian João Guimarães Rosa *(Grande Sertão: Veredas, 1956; The Devil to Pay in the Backlands, 1963);* and in the Rabelaisian romances of Brazil's Jorge Amado *(Gubriela, Cravo e Canela, 1958; Gabriela, Clove and Cinnamon, 1962).* It was on the bedrock of these pioneers of the new fiction that the younger generation planted the invigorating and challenging novel of the *1960s.* They include Argentina's Ernesto Sábato (*Sobre héroes y tumbas, 1961)* and Julio Cortázar *(Rayuela, 1963; Hopscotch, 1966);* the Mexican Carlos Fuentes *(La muerte de Artemio Cruz, 1962; The Death of Artemio Cruz, 1964);* the Colombian Gabriel Garcia Márquez *(Cien años de soledad, 1967; One Hundred Years of Solitude, 1970);* and the Peruvian Mario Vargas Llosa *(La ciudad y los perros, 1963 [The Time of the Hero, 19661,* and *La casa verde, 1965 [The Green House, 19681).*

GERMAN LITERATURE IN THE 20TH CENTURY

Through the 20th century, German literature has reflected the social, political, and spiritual uncertainty of its surroundings. Early dissatisfaction with conventional literary forms led to experiments with new forms in an attempt to avoid sterility and revitalize the language aims that have emerged as dominant forces in modern literature. Much early 20th-century writing developed out of the 19th-century style of Naturalism, which made the smallest details of life a subject of literature.

Impressionism. Impressionism had beginnings closely allied to Naturalism, and writers at the turn of the century reflected the influence of Naturalism on their development. Impressionism evokes a mood or state of mind by emphasizing the impression made by an object on its observer. The poet Detlev von Liliencron provided an early example of this, as did Richard Dehmel, who, despite a Naturalistic subject matter, evoked mood through sound patterns. Writers influenced by Symbolism also had elements of Impressionism in their work. A successor of French Symbolism who had considerable effect on other writers was Stefan George, whose solemn, carefully composed verse aimed at asserting the lofty stature of poetry, which, for him, had a religious character. Because of their obscurity, George's poems can only be understood in terms of the framework to which they belong. He founded the journal *Blatter fur die Kunst (1892;* "Journal for Art") to publish his followers' poetry. Among those attracted by his work were the critics Friedrich Gundolf and Ernst Bertram and a poet, Karl Wolfskehl.

Hugo von Hofmannsthal, an Austrian whose Impressionistic elements had their roots in Romanticism, declined to join George's circle. In his melodious poetry he delicately analyzed his sensibilities and was haunted by his obsession with the inadequacy of language completely to convey feeling. An essay, the "Chandos-Brief" (1902; "Letter by Lord Chandos," 1952), records this sense of the inadequacy of words. The plays of the 1890s concerned the aesthete faced with the reality of this inadequacy. Later dramas, such as *Jedermann* (published 1911), an adaptation of ***Everyman,*** and ***Das Saliburger grosse Welttheater*** (1922; "The Great Salzburg Theatre of the World"), were religious in tone and borrowed from Baroque and medieval drama; a comedy, ***Der Schwierige*** (1921; "The Awkward Man"), analyzed a sophisticated mind inhibited by the weight of social tradition. His greatest public successes were his librettos, such as ***Der Rosenkavalier*** (1911), which Richard Strauss set to music. Conscious of the heritage of European culture and of ethical responsibility, Hofmannsthal conveyed a strong awareness of moral issues in his work.

The end-of-the-century analyses of decadence

Arthur Schnitzler, another Austrian writer of note, depicted the decadence of pre-1914 Vienna while giving a penetrating analysis of man's inner life. A searching analyst of Austrian culture was Robert Musil, whose monumental novel *Der Mann ohne Eigenschaften* (1930–43; ***The Man Without Qualities,*** 1953) examined modern incertitude and exposed sham values in society; stories such as ***Drei Frauen*** (1924; "Three Women") and ***Die Verwirrungen des Zöglings Törless*** (1906; ***Young Törless,*** 1955) dealt with the inner decay of man in face of cultural collapse. Musil's rise to fame was slow. More than a decade after his death elapsed before he was recognized as a major figure in German literature. Hermann Broch's novels, too, were appreciated only after his death (1951); his trilogy ***Die Schlafwandler*** (1931–32; ***The Sleepwalkers,*** 1932) is an account of the decline of values in modern society. Elias Canetti's *Die Blendung* (1935; ***Auto da Fé,*** 1946) continued the Symbolist tradition and analyzed the impact of mass society on the individual.

Prague was another centre of writing: Franz Kafka, Rainer Maria Rilke, Franz Werfel, Kafka's friend Max Brod, and Gustav Meyrink wrote and lived in Prague. Their work inclined toward the esoteric and was strongly influenced by Symbolism.

Rilke's view of God

Symbolism. Like most great lyric poets of the time, Rilke was indebted to Symbolism. His melodious verse, which made him one of Germany's great lyric poets, gave his poetic vision of reality compelling power. This was clearly evident in ***Das Stunden-Buch*** (1905; "The Book of Hours"), which described a search for spiritual health in a hostile urban civilization. Rilke's conception of the universe, God, and death was determined by a quest for artistic fulfillment. In ***Neue Gedichte*** (1907–08; "New Poems") he no longer sought to depict the poet's subjective response to spiritual isolation but wanted to convey an objective appraisal of reality by recreating the "things" that he encountered in art and nature. In ***Die Aufzeicknungen des Malte Laurids Brigge*** (1910; ***The Notebooks of Malte Laurids Brigge,*** 1930) he studied artistic sensitivity by analyzing a neurotic personality. The ten poems contained in the volume ***Duineser Elegien*** (1923; ***Duino Elegies,*** 1939) summed up his spiritual struggles in complicated severe verse. In ***Die Sonette an Orpheus*** (1923), Orpheus was symbolic of poetry's power to transmute problems of existence, and of its attempt to justify reality.

Mann's analyses of his age

Symbolism also influenced prose writers, not least Thomas Mann, who attempted to use symbol and myth in narratives that started from a clinical analysis of modern man's mental and physical state. Mann's characterization was Impressionist, but impressions became leitmotivs conveying the power of the subconscious. His work was influenced by the philosophers Schopenhauer and Nietzsche; yet his portrayal of social change, of the impact of ideology, was an organic part of the story and he was occupied, as always, with the status of the artist in society. In his early works art was symbolic of decadence, an over-refinement no longer acceptable; but with ***Der Zauberberg*** (1924; ***The Magic Mountain,*** 1927), Mann, like Rilke, emphasized the constructive qualities of art. His great novels examined different facets of his age: *Buddenbrooks* (1900), bourgeois society; ***Der Zauberberg,*** intellectual corruption; ***Doktor Faustus*** (1947), the German mind and character during the Third Reich. His novellas—*e.g.*, ***Tonio Kroger*** and ***Tristan*** (both 1903) and ***Der Tod in Venedig*** (1912; ***Death in Venice,*** 1925)—depicted the same themes.

Irony and resultant ambiguity were characteristic of Mann. His increasingly complex style reflected the complexity of his mind and his study of the history of ideas. Mann's last work, ***The Confessions of Felix Krull*** (1954), humorously stated the doubt that ran through all his work—whether the pursuit of beauty and, hence, pursuit of culture and art were not in the end a great deception.

Hermann Hesae was influenced by Neoromanticism and concentrated upon man's spiritual conflict. His novels, from his first success, ***Peter Camenzind*** (1904), portrayed the struggle of individuals in a world hostile to sensitivity or explored the subconscious and the balance between sensuality and the spirit. In ***Der Steppenwolf*** (1927), he examined the conflict between the bourgeois world and the sensitive outsider, here resolved by self-abandonment to fantasy; in ***Das Glasperlenspiel*** (1943; ***Magister Ludi,*** 1949) he questions the whole purpose of civilization. Hesse wrote poetry and also short stories, of which the best known was "Knulp" (1915).

Ricarda Huch's novels, like those of Thomas Mann and his brother Heinrich Mann, emphasized the individual's independence and dignity. Heinrich Mann savagely attacked social and political abuses in novels such as ***Professor Unrat*** (1905) and ***Der Untertan*** (1918; ***The Patrioteer,*** 1921).

Expressionism. The often ecstatic style of Heinrich Mann anticipated Expressionism, the key movement in German literature during and immediately after World War I. Expressionism emphasized the inner significance of things and not their external forms and therefore did not depict life in action but evoked states of mind.

Frank Wedekind's dramas, in which bourgeois morality was pilloried, were forerunners of this style, but the first fully Expressionist drama was Johannes Reinhard Sorge's ***Bettler*** (1912; "The Beggar"), in which characters appeared as abstract functions in each other's lives. This play, like those of Walter Hasenclever, Paul Kornfeld, Fritz von Unruh, Ernst Barlach the sculptor, and Oskar Kokoschka the painter, was characterized by a quest for the essence of things, for the ideas behind personality and spiritual meaning in life. Ernst Toller wrote political plays employing an Expressionist technique in *Die Maschinenstürmer* (published 1922; ***The Machine-Wreckers,*** 1923). Georg Kaiser, the leading Expressionist playwright, moved from Naturalism through Expressionism to a mature traditional style, while Carl Sternheim unmasked bourgeois pretensions by means of a shrill satire of contemporary language.

Expressionist poetry was equally non-referential, attaining coherence through its associative power. The chief poets were Ernst Stadler, Georg Heym, Georg Trakl, Franz Werfel, Johannes Becher, August Stramm, Theodor Daubler, Gottfried Benn, and Else Lasker-Schiiler. The leaders, Heym, Trakl, and Stadler, died young. Trakl's poetry expressed disgust with modern civilization; Heym, Stadler, and Stramm were dominated by fears erupting from man's urban existence and often expressed their vision of doom in grotesque, concentrated verse. Benn's macabre early verse clinically described human suffering; the nihilism of his later poetry took the form of obscurity in language. Like Jean (Hans) Arp and Yvan Goll, Benn later abandoned Expressionism, as did many major figures within its orbit.

The nightmare vision of Franz Kafka

Kafka, famous only after his death, depicted the individual in a world of anxiety, crushed by unfathomable forces. His work, such as the posthumously published novels *Der Prozess* (1925; ***The Trial,*** 1937) and ***Das*** *Schloss* (1926; ***The Castle,*** 1930), exposed the incongruities of aspiration and the weakness of reason with grotesque humour. He was deliberately ambiguous about the characters' sense of guilt and their protests against the established order.

Post-Expressionism and Social Realism. After 1918 Expressionism gave way to Social Realism through which writers hoped to gain objectivity. The first subjects chosen were World War I and its aftermath: Arnold Zweig's ***Streit um den Sergeanten Grischa*** (1927; ***The Case of Sergeant Grischa,* 1927**), Erich Maria Remarque's ***Im Westen Nichts Neues*** (1929; ***All Quiet on the Western Front,* 1929**), and Hans Fallada's (Rudolf Ditzen's) and Erich Kästner's works documented the war and postwar society.

This new objectivity continued in the work of Anna Seghers and in Carl Zuckmayer's ***Hauptmann von Köpenick*** (1931; ***The Captain of Kopenick,*** 1932) and ***Des Teufels General*** (produced 1946; "The Devil's General"). Alfred Döblin combined documentation with psychological analysis, whereas Ernst Jünger explored the philosophical implications of technology and modem civilization. Junger's ***Auf den Marmorklippen*** (1939; ***On the Marble Cliffs,* 1947**), a criticism of the Third Reich, revealed his inability imaginatively to conceive individual character; the poetry and prose of his brother Friedrich exhibited greater lyrical and narrative skill.

The Third Reich. The Third Reich disrupted the continuity of literary life in Germany. Under the Nazis, good writers either left Germany, were driven out, were forced into silence, or were exterminated, and good writing did not come to the surface until after World War II. Of the vaunted school of blood and soil, the talent of E.G. Kolbenheyer, Friedrich Blunck, and Hans Grimm was limited. Hans Carossa's autobiographical novels were significant for their emphasis on the individual's needs to develop inner resources by following tradition.

Literature after World War II. After World War II, established writers continued to publish, reputations such as those of Kafka and Musil were definitely established, and the poetry of Gertrud Kolmar, Jesse Thoor, and Lasker-Schüler came to light. Attempts to free lyric language from customary forms encouraged unconventionality of expression, and Ingeborg Bachmann, Giinter Eich, Walter Hollerer, Karl Krolow, and Heinz Piontek expressed the chaos of the age through experimental forms. Paul Celan was an important poet who emerged after 1945; his poetry owed much to Trakl and Expressionism and delineated reality in a deliberately complex and obscure manner.

The plays of Bertolt Brecht

The giant of European theatre was the poet, playwright, theatrical producer, and theoretician Bertolt Brecht, much of whose work surprisingly had been completed almost 30 years before he received international acclaim. His first successful play, ***Trommeln in der Nacht*** (performed 1922; "Drums in the Night"), the poems and songs in ***Hauspostille*** (1927; ***A Manual of Piety*),** and the satirical opera ***Die Dreigroschenoper*** (1928, music by Kurt Weill; ***The Three-penny Opera,*** 1949) all dated from the 1920s, when he was developing his theory of "epic theatre" and his social criticism which led him to Marxism. During the Third Reich he lived first in Scandinavia and then in the United States, and in those countries he wrote such famous works as ***Mutter Courage und ihre Kinder*** (1941; ***Mother Courage and Her Children,*** performed 1949), ***Leben des Galilei*** (1943; ***The Life of Galileo,* 1947**), and ***Der kaukasische Kreidekreis*** (first performed in English as ***The Caucasian Chalk Circle;*** performed in Berlin, 1954). As a poet he was characterized by simplicity of diction and suggestive power, as a dramatist by acute visual perception and sardonic wit, while as a theoretician he sought to blend didacticism with entertainment.

The German political dramatists who followed Brecht received international attention: Rolf Hochhuth ***(Der Stellvertreter* [1963; *The Deputy,* 1964]);** Heinar Kipphardt ***(Joel Brand* [1965]),** and the two best known, Peter Weiss and Gunter Grass. ***Die Ermittlung*** (1965; ***The Investigation,* 1966),** by Weiss, dealt with the Auschwitz trial of 1964–65. In ***Die Verfolgung und Ermordung des Jean-Paul Marats . . .*** (1964; often known as ***Marat/Sade***) Weiss portrayed the conflict between individualism and socialism simultaneously on several levels of reality. A left-wing intellectual, increasingly extreme, Weiss embraced political beliefs that several (though by no means all) members of the so-called Group 47 shared. This group included many well-known writers, such as Grass, Hans Magnus Enzenberger, Ingeborg Bachmann, Heinrich Böll, and Alfred Andersch. Boll analyzed postwar psychological problems in ***Haus ohne Hiiter*** (1954; ***The Unguarded House,*** 1957) and ***Und sagte kein einziges Wort*** (1953; ***Acquainted with the Night,*** 1954) and satirized bourgeois life in ***Doktor Murkes gesammeltes Schweigen*** (1958; "The Collected Silences of Dr. Murke") and ***Ansichten eines Clowns*** (1963; ***The Clown,*** 1965). In 1972 he was awarded the Nobel Prize for Literature. Grass's plays, poems, and novels, in particular ***Die Blechtrommel*** (1959; ***The Tin Drum,*** 1962) and ***Hundejahre*** (1963; ***Dog Years,* 1965),** focussed attention on the grotesque and absurd elements in life. Other talented writers were Hans Nossak, Rudolf Hagelstange, Siegfried Lenz, Peter Handke, and Johannes Bobrowski.

Group 47

SWISS LITERATURE IN THE 20TH CENTURY

German Swiss. Many 20th-century authors in German, guided by the Swiss distrust of one-sided attitudes, believed that the values of tradition and emancipation were not irreconcilable; and novelists such as H.A. Moser, Robert Faesi, and Heinrich Federer, among others, sought to have their native country reflect the world at large.

A similar universal humanism was at work in history and letters—in Carl J. Burckhardt, Fritz Ernst, and others. Although Carl Spitteler, with an epic, ***Der Olympische Friihling*** (1900–06; "Olympian Spring"), gained recognition early in the century and won the Nobel Prize in 1919, Robert Walser (who in his stories and novels explored the feelings of the underdog and of minds out of tune with reality) and Albin Zollinger were appreciated at large only after World War II. Curt Gotz's comedies were keenly witty, but the outstanding experiments in dramatic form were made by Max Frisch and Friedrich Dürrenmatt. Their work had a profound influence on the German theatre and German writing.

Frisch's ***Nun singen sie wieder*** (1946; ***Now Sing Again)*** and ***Andorra*** (1961) and Dürrenmatt's ***Besuch der alten Dame*** (1956; ***The Visit,*** 1960) and ***Physiker*** (1962; ***The Physicists,*** 1963) were modern morality plays. Like Eich and Bachmann, Diirrenmatt experimented with radio plays, criticizing, in the manner of Frisch's novels ***Stiller*** (1954) and ***Homo Faber* (1957),** the sterility of modem life. His inventive power, genuine sense of comedy, and grotesque vision of life's absurdities reflected awareness of a world out of joint.

French Swiss. In the 20th century the leading Swiss author in French was C.-F. Ramuz, who wrote strongly imaginative novels about life among vinegrowers and Alpine herdsmen. Gonzague de Reynold demonstrated the inward connection between German- and French-speaking Switzerland in the 18th century and depicted the whole of his country in his ***Cités et pays suisses*** (1914–20).

As much 20th-century literary activity was carried on in the canton of Geneva, contemporary French Swiss literature had a cosmopolitan air, to distinguish what was peculiarly Swiss—as, for example, the prose of Henry Vallotion and the poetry of Gilbert Trolliet. A typically European outlook was to be found in the essays of Denis de Rougemont. The prose of Ramuz and Léon Savary made use of more particularly Swiss themes.

DUTCH LITERATURE IN THE 20TH CENTURY

The writers of the Dutch revival of the 1880s were essentially individualistic, but in the next generation a new concern for philosophical and social problems became apparent.

The poetry of a foremost socialist writer named Henriëtte Roland Holst-van der Schalk was characterized by a desire for justice and charity. The socialist dramas of Herman Heijermans were internationally successful; he and Israel Querido described life in Amsterdam in novels concise in narrative style. A group of Naturalist-Realist novelists, including **Marcellus Emants,** P.A. Daum, and Augusta de Wit, was also writing at this time. Two great novelists were the Neoromantic Arthur van Schendel and Louise Marie Anne Couperus.

Significant early poets were Adriaan Holst, J.C. Bloem, and P.N. van Eyck, a philosophical poet and essayist. Immediately after World War I two poets emerged: Hendrik Marsman, an advocate of free verse and representative of the Vitalist movement; and the pessimistic J.J. Slauerhoff, who roamed the world in search of the unattainable.

The literary periodical ***Forum*** was founded in 1932 by Menno ter Braak and Edgar du Perron, leaders of a movement that aimed to replace superficial elegance with greater sincerity and warned against the German threat before the war.

The most important mid-20th-century Dutch writer, Simon Vestdijk, was originally associated with the ***Forum*** group, and other noteworthy authors were Ferdinand Bordewijk and Albert Helman. The most original poet was Gerrit Achterberg, whose poems explored the boundary between life and death.

During the Nazi occupation, free literature either stopped or was published secretly. The new generation that arose after 1945 adopted an experimental and irrational style verging on the absurd: typical of these were the poets Leo Vroman, Hans Lodeizen, Remeo Campert, and Ellen Warmond—whose experimentation was in turn rejected by the poets of the 1960s.

Postwar novelists

Postwar novelists showed the influence of the Nazi occupation in various ways. Anna Blaman treated existential solitude, while Adriaan van der Veen showed concern for human relationships. The novels and short stories of Maria Dermout and H.J. Friedericy were based on memories of life in the former Dutch East Indies, and the work of Simon Carmiggelt and Anton Koolhaas revealed bitter humour. The influence of Existentialism was seen in the prose of the "angry young men" Willem Frederik Hermans and Gerard Kornelis van het Reve, whose works of the 1960s expressed despairing self-revelation. (For Dutch literature in Flemish, see below.)

BELGIAN LITERATURE IN THE 20TH CENTURY

Flemish. At the turn of the century a group of writers strove through Neoromanticism for an art in which individual feelings would be given universal significance. In masterly essays and a symbolical novel, ***De wandelende Jood*** (1906; "The Wandering Jew"), their leader, August Vermeylen, advocated rationalism infused with idealism. Emmanuel de Bom published the ***fin de siècle*** novel ***Wrakken*** (1898; "Wrecks"), and Alfred Hegenscheidt produced the Wagnerian drama ***Starkadd*** (1898).

The poetry and prose of Karel van de Woestijne formed a symbolical autobiography of a sophisticated, world-weary sensualist striving after spiritual detachment. His work, a passionate confession of human frailty, represents one of the achievements of European Symbolism. Stijn Streuvels made the West Flemish landscape his microcosm, while the polished work of Herman Teirlinck, novelist, dramatist, and essayist, was characterized by imagination, sensuality, and a sonorous vocabulary.

Naturalism reached its height in the robust tales and pithy plays of Cyriel Buysse. The regional novel attained its summit in the charming Bruges stories by Maurits Sabbe, and in the vivid treatment of Antwerp life by Lode Baekelmans. But the writer who, apart from Vermeylen, represented the strongest intellectual force of his time was the priest Cyriel Verschaeve: his poems, plays, and essays were passionate expressions of his intensely Augustinian outlook.

The Boomgaard ("Orchard") group, which included André de Ridder and Paul Gustave van Hecke, introduced a new, individual note, strove to be cosmopolitan, and defended a more dilettante attitude to culture. To this generation belongs the amiable stoic Jan van Nijlen.

After World War I. During World War I there was a new flowering of the picturesque regional tale: ***Pallieter*** (1916), by Felix Timmermans, and the roguish ***De witte*** (1920; "Whitey"), by Ernest Claes, became known outside Flanders. From the poetry of August van Cauwelaert and the prose of Franz de Backer it could be seen that the generation that fought in the war stressed life rather than literature. But a trend first revealed during the German occupation found its most direct outlet in revolutionary Expressionism, and a review, ***Ruimte*** (1920–21; "Space"), published its manifesto: ethics must take priority over aesthetics, and the art of the community over that of the individual. Expressionism was most apparent in lyrical poetry and drama. Wies Moens' poetry is representative of this trend. The outstanding lyricist of the movement was Paul van Ostaijen, who expressed faith in humanity in ***Het sienjaal*** (1918; "The Signal") but soon went through a crisis of Dadaism, adopted rhythmical typography (***Bezette stad***, 1921), wrote pure poetry concentrated on word and sound, and grotesque verse and prose. His principal opponent, Urbain van de Voorde, wrote and advocated traditional "eternal poetry." Poets of a review ***'t Fonteintje*** (1921–24; "The Little Fountain"), including Richard Minne and Maurice Roelants, reacted against Expressionism, proclaiming love of life in traditional forms. Expressionism gave new life to drama. In the 1920s the Flemish Popular Theatre became one of the foremost avant-garde theatres in Europe. The standard of drama was raised by Herman Teirlinck with ***De man zonder lijf*** (1925; "The Bodiless Man") and Anton van de Velde with ***Tijl*** (1925).

Flemish Expressionism

By 1930 the tide of Expressionism had run out and the novel came into its own. The regional novel was replaced by the psychological novel, introduced by Roelants with ***Komen en gaan*** (1927; "To Come and Go"). Gerard Walschap wrote dynamically about man's social, religious, and moral problems, and Lode Zielens about the lives of the poor.

The New Realism movement included Marcel Matthijs, Norbert Edgard Fonteyne, and André Demedts. Since 1930 the focal point of the literature has been man. Refined stylists such as Filip de Pillecijn and Maurice Gilliams (the latter also a poet of value) are literary psychologists as well. The caustic Willem Elsschot, the skeptical Raymond Brulez, and the disenchanted humanist Marnix Gijsen are all more or less detached observers of human weaknesses.

Poetry also changed in character around 1930. The new tone was set by the "personalistic" poets of the Vormen ("Forms") group, of whom Pieter Geert Buckinx and René Verbeeck were representative: they had read Rimbaud and expressed their vitalism in a new imagery. There were also such independent writers as the religious poet Albe (Renaat A. Joostens), the dynamic Bert Decorte, and the "noble pagans" Karel Jonckheere and Herwig Hensen. All of them restored syntax and even the regular verse.

After World War II. The outstanding writers of World War II and the postwar period were novelists. The range of subjects and styles in the novel was remarkable. The Catholic novels by Paul Lebeau and Gaston Duribreux, the "magic-realist" work of Johan Daisne and Hubert Lampo, the Social Realism of Louis-Paul Boon (***De kapellekensbaan,*** 1953; "Chapel Road"), the Existentialism of Jan Walravens, and Hugo Claus's experimental novels are but a sample. The last three writers belonged to a review group called ***Tijd en mens*** (1949–55; "Time and Man"), which was marked by postwar chaos, rebellion, and Experimentalism. From the 1950s, Experimentalism has had a great influence on literary production. Poetry developed, on the one hand, first into post-Surrealism and linguistic experiment (Hugo Claus and Albert Bontridder), then into "concrete" poetry (Paul de Vree and his avant-garde review ***De tafelronde*** ["The Round Table"]); on the other hand, the development was toward engaged poetry and toward New Realism. The experimental trend in the novel, already anticipated by Daisne and Boon, led to new prose either based on splitting and stream-of-consciousness association (Hugo Raes and Ivo Michiels) or consisting of introverted "texts" dwelling largely on the act of writing itself (Willy Roggeman and Daniel Robberechts). Nevertheless, the tradition—invigorated by new techniques—proved to be fertile; ***e.g.***, in the

The avant-garde reviews

poetry of Jos de Haes and Paul Snoek, in the humanistic, satiric novels by Ward Ruyslinck, and in Jef Geeraerts' violent colonial novels inspired by his experiences in what was then the Belgian Congo.

In drama three developments were notable: the staging of the literary plays of Johan Daisne and Herwig Hensen; the popularity of such massive productions as ***Het heiligbloedspel*** (1938; "The Play of the Holy Blood"), by Jozef Boon; and the success of small, avant-garde theatres and of new plays, some of them experimental, by Claus, Pie Sterckx, and Tone Brulin.

French. ***Between World Wars I and II.*** A new generation of Belgians who wrote in French arose between World Wars I and II. The novels of André Baillon showed keen yet compassionate observation of life. Jean Tousseul was concerned with human suffering, whereas Charles Plisnier wrote powerfully analytical novels (***Mariages***, 1936; ***Nothing to Chance***, 1938). Other talented novelists were Marie Gevers and the visionary Franz Hellens. Somewhat later, Georges Simenon imbued the detective story with exceptional psychological penetration.

The poetry of this period was characterized by increased stylistic experiment and the development of a neoclassical poetry of great fluency. Henri Michaux, who, influenced by Surrealism, revolutionized the poetic language; Marcel Thiry, whose lyrical style harmonized traditional and modern; and Géo Norge were chief exponents of an experimental use of words. To another group, headed by Odilon-Jean Périer, an original poet of unusual clarity who aimed to convey his awareness of the interpenetration of daily life by the divine message, a poet must free himself from personal involvements and achieve a transparent medium through which what lies behind life's apparent confusions may be clearly seen.

In drama Fernand Crommelynck wrote savage farces laced with poetry. Michel de Ghelderode astonished audiences by a love of anachronistic character and situation and puppetlike characters, and Herman Closson reinterpreted historical events and characters.

Developments after World War II. After World War II the novel became less regional than formerly, often set in foreign surroundings; generally it subordinated action to detailed psychological analysis of characters' reactions in somewhat contrived situations. Albert Ayguesparse, Pierre Demeuse, Hubert Juin, Maud Frere, and Marie Nicolaï were all ultimately concerned with the theme of man. Françoise Mallet-Joris won instant success with ***Lc Rempart des beguines*** (1951; ***Into the Labyrinth***, 1953); ***Les Personnages*** (1961; ***The Favourite***, 1964), about France in the reign of Louis XIII, showed her bare economical style and widening of range.

Postwar poetry reverted to regular metrical forms in poems on the eternal themes of life, love, birth, and death and topical subjects, including that of contemporary Europe. Outstanding dramatists were Georges Sion, who wrote comedy, religious and historical plays, and translations from Shakespeare; Charles Bertin, whose ***Prétendants*** (1947), a modern version of the Ulysses story, was played abroad as ***Love in a Labyrinth;*** Jean Mogin, whose plays, successful in both Paris and Brussels, explored "being" and "nonbeing"; José-André Lacour; and Romain Sanvic (1883–1957), who adapted English and American plays for the Brussels theatre.

Literary periodicals, which were usually organs for political or literary groups of writers, were influential, especially ***Le Journal des Poètes***, which in 1952 founded a biennial international conference of poets and critics at Knocke-le-Zoute.

Walloon. Walloon literature explored new paths in the 20th century. Dialect studies were undertaken by a whole army of scholars, and the literary possibilities of the dialect were extended as a result of a standardizing of rules of spelling and grammar, as well as of attempts by Émile Lempereur and some other writers to renew the sources of inspiration. Alongside several veteran authors, such as a talented prose writer Joseph Calozet of Namur, the younger generations sought to achieve a strict unity of thought and technique. Among poets the following were to be noted: Frans Dewandelaer, Robert Grafe, Willy Bal, Henri Collette, Émile Gilliard, Jean Guillaume, Marcel Hicter, Albert Maquet, Georges Smal, and Jenny d'Invérno. Widely praised storytellers and novelists were Léon Mahy, Dieudonné Boverie, and Léon Maret, among many others. The dramatists included François Roland, Jules Evrard, Georges Charles, Charles-Henri Derache, François Masset, and J. Rathmès.

The work of dialect writers continued to be assisted by the Société de Littérature Wallonne, with its associations and publishing centres at Liège, Namur, Mons, La Louvière, Nivelles, and Brussels.

SCANDINAVIAN LITERATURE IN THE 20TH CENTURY

Norwegian. In the 1890s established Norwegian writers came under fire from the new generation. The manifesto of new ideas was an essay published in 1890 in the periodical *Samtiden* ("The Present Age") by Knut Hamsun, "Fra det ubevidste Sjaeleliv" ("From the Unconscious Life of the Mind"), which demanded attention to what was individual and idiosyncratic rather than typical. Hamsun was impatient with contemporary emphasis on social problems, and his early novels exemplified these ideas; *Sult* (1890; ***Hunger***, 1899), ***Mysterier*** (1892; ***Mysteries***, 1927), and ***Pan*** (1894); his later novels, such as *Markens grøde* (1917; ***Growth of the Soil***, 1920), were less extreme but still showed a strong, sometimes savage irony. Hamsun won the Nobel Prize for Literature in 1920.

Knut Hamsun's emphasis on the individual

Lyric poetry at this time flourished with Sigbjørn Obstfelder, who had a close affinity with the Symbolist movement, and Nils Collett Vogt, who produced some of the best lyrics of the 1890s. In drama Gunnar Heiberg, who combined a sharply satirical wit with a lyric deftness, expressed the new spirit in ***Kong Midas*** (1890), ***Gerts have*** (1894; "Gert's Garden"), *Balkonen* (1894; "The Balcony"), and ***Kjaerlighetens tragedie*** (1904; "The Tragedy of Love"). Sharing Hamsun's preoccupation with the irrational side of human conduct was Hans Kinck, a writer of power and penetration. His verse drama ***Driftekaren*** (1908; "The Drover") and long novel *Sneskavlen brast* (1918–19; "The Avalanche Broke") showed him to be a more reflective and analytical writer than Hamsun.

The real achievements of Norwegian literature in the first half of the 20th century were in the novel and lyric poetry. Drama was not conspicuous, except for the plays of Gunnar Heiberg and Nordahl Grieg. In the early decades of the century, regionalism was a strong element, particularly in the novel; and authors adopted language coloured by dialect, thus becoming identified with their region. Kristofer Uppdal, of the midnorth region of Trøndelag, wrote a remarkable work—a 10-volume novel cycle, ***Dansen gjenom skuggeheimen*** ("The Dance Through the Shadow World"). The novel also treated of conflicts arising from the spread of industrialism, which Norway underwent later than did other European countries. The most proletarian writer was Oskar Braaten, but superior as an artist was Johan Falkberget, who wrote with understanding and historical insight about the miners in Røros in *Christianus Sextus* (1927–35) and in *Nattens brød* (1940; "Bread of Night"). Sigrid Undset, who won the Nobel Prize for Literature in 1928, set her novels in many different ages, and their concern was to examine women's loyalties within the framework of their role in society. A long historical novel, ***Kristin Lavransdatter*** (1920–22), was a masterpiece of Norwegian literature. Her later novels, ***Gymnadenia*** (1929; ***The Wild Orchid***, 1929) and ***Detz braendende busk*** (1930; ***The Burning Bush***, 1932), were greatly influenced by her conversion to Roman Catholicism. Olav Duun, again of the midnorth region, revealed his insight into life as endless conflict in a six-volume novel cycle about the development of a peasant family through four generations—***Juvikfolke*** (1918–23; ***The People of Juvik***, 1930–35).

Shortly before World War I, there were several good lyric poets: Herman Wildenvey, Olaf Bull, Tore Ørjasaeter, and Olav Aukrust. Between World Wars I and II, there were many socially committed writers: the poet

Arnulf Øverland; a novelist and critic, Sigurd Hoel; a dramatist and critic, Helge Krog: and Nordahl Grieg. After World War II, Tarjei Vesaas wrote a remarkable series of novels, including a symbolic *Huset i mørkret* (1945; "The House in the Darkness") and *Bruene* (1966; "The Bridges"). Cora Sandel, who had made a major contribution to literature with her "Alberte" trilogy (1926–39), continued to write, as did Aksel Sandemose, an experimental writer, and Johan Borgen, who was noted for his early short stories, his *Lillelord* trilogy (1955–57), and the autobiographical *Barndommens rike* (1965; "Childhood's Realm").

Swedish. The early years of the 20th century were a period of decadence and pessimism in Swedish literature. Representative of this mood were Hjalmar Soderberg and Bo Bergman. Soderberg's forte was the short story (*Historietter* [1898]), in which psychological subtlety and irony were happily combined and in which, as in his novels *Martin Bircks ungdom* (1901; "Martin Birck's Youth") and *Doktor Glas* (1905), he appeared as a master of Swedish prose. Bergman also produced memorable short stories, but his real medium was the lyric; he developed his talent in a series of collections from *Marionetterna* (1903; "The Marionettes") to *Riket* (1944; "The Kingdom").

The development of the novel was associated with Gustaf Hellstrom, Ludvig Nordstrom, Elin Wagner, and Sigfrid Siwertz. Hellstrom's work as journalist in Europe, the United States, and England greatly influenced him. Irony and careful detail emerged in his best known novel, *Snormakare Lekholm får en idé* (1927; *Lacemaker Lekholm Has an Idea,* 1930). Siwertz was a more elegant stylist—he began as a lyricist—and a decisive influence upon him was the philosophy of Henri Bergson, reflected in *En flanör* (1914; "An Idler"); but his weightiest work was a family saga, *Selambs* (1920; *Downstream,* 1922), a novel of Stockholm during World War I. Nordstrom, overflowing with vitality and keen but grotesque humour, accomplished some of his best work in *Landsorts-bohème* (1911; "Small-Town Bohemia") and in his short stories, as in *Fiskare* (1907; "Fishermen") and *Öbacka-bor* (1921). Elin Wagner was an ardent pacifist and feminist; her most powerful work was a peasant novel, *Åsa-Hanna* (1918). The outstanding novelist of the 1920s was Hjalmar Bergman: with vivid imagination and restless energy, Bergman wrote a long series of stories, many set in "Wadkoping" (his native Örebro), others in Italy. In *Loewenhistorier* (1913) he depicted an irrational, impulsive, unsuccessful hero; in *Farmor och vår Herre* (1921; *Thy Rod and Thy Staff,* 1937) he portrayed one of the dominating female personalities that fascinated him: The satire *Markurells i Wadkoping* (1919; *God's Orchid,* 1924) and *Swedenhielms* (performed 1925), one of the few Swedish comedies, were his most widely known works.

Meanwhile, the "proletarian" novel had been developed by writers concerned with the miseries of the working class, particularly Martin Koch and Ivar Lo-Johansson. There was particularly harsh criticism of working class conditions in stories by Jan Fridegård. Vilhelm Moberg wrote novels of peasant life but achieved his greatest success with the four-part prose epic about a group of Swedish emigrants to North America, *Utvandrarna* (1949–59; *The Emigrants,* 1951–61). The development of the Swedish autobiographical novel was helped by Eyvind Johnson, with the series "Romanen om Olof" (1934–37); Harry Martinson, with *Nässlorna blomma* (1935; *Flowering Nettle,* 1936) and *Vägen ut* (1936; "The Way Out"); and Agnes von Krusenstjerna. In her novel cycles, the "Tony" trilogy (1922–26) and the "Froknarna von Pahlen" series (1930–35), she described her own aristocratic environment and analyzed a degenerate psychology. Harry Martinson was one of a group, formed *c.* 1930, of five primitivist writers. He later developed into one of the finest lyricists of the century. Sensuous imagery and a feeling for nature characterized his work. He attempted to revive the verse epic in his *Aniara* (1956), a symbolical story of a voyage of a spaceship.

Par Lagerkvist's work

The internationally best known Swedish writer of the 20th century was Par Lagerkvist, who won the Nobel Prize for Literature in 1951. In his youth a bold innovator, he later developed an admirably pure prose style, as in the allegorical novel *Dvärgen* (1944; *The Dwarf,* 1953); Expressionistic in style were collections of poems, *Ångest* (1916; "Anguish") and early plays—for instance, *Himlens hemlighet* (1919; "The Secret of Heaven")—reflecting the years of World War I. Lagerkvist wrote an autobiographical novel and love poetry, but the dominant theme throughout his work was a search for vital, often outspokenly religious values.

Several of the best Swedish writers were connected with the development of lyric poetry, notably Vilhelm Ekelund, in his youth the chief exponent of Symbolism in Sweden, and later, as an author of aphorisms, a weighty influence on the development of literary modernism. Among the most popular poets were Dan Andersson, Birger Sjoberg (who set many of his songs to his own music), and Hjalmar Gullberg. In Gullberg's poetry religious commitment and classical learning are balanced by irony and wit. A more esoteric style in modernism was introduced by Bertil Malmberg and developed by the group of poets called the generation of the 1940s, which included Erik Lindegren and Karl Vennberg. Stylistically influenced by T.S. Eliot, they often expressed an anguish and disbelief close to French Existentialism. Lindegren's *Mannen utan väg* (1942; *The Man Without a Wuy,* 1969) was typical of this generation's search for meaning in life. The most distinguished novelist of the 1940s was Lars Ahlin; who was concerned with man's search for grace through love and humiliation in works such as *Min dod ar min* (1945; "My Death Is Mine").

The greatest lyric poet of the century was Gunnar Ekelof. His first collection of poems, *Sent på jorden* (1932; "Late on Earth"), was heralded as the first specimen of Surrealism in Swedish literature. Ekelof's later development passed through successive phases of romanticism and anti-poetic skepticism resolved in a trilogy of books blending autobiography and Eastern mysticism. The first part was called *Diwan over Fursten av Emgion* (1965; "Diwan of the Prince of Emgion").

In reaction to the literature of the 1940s and 1950s, which was much concerned with artistic form and the individual approach to life, the 1960s was a period of political and social commitment in poetry and fiction alike. Recurrent topics were the war in Vietnam and bitter onslaughts on the Swedish welfare state. Independent lyric poetry, however, continued to be written by Östen Sjostrand and Thomas Transtromer, and a tortured experience of life, coloured by Roman Catholicism, was forcefully expressed in the novels of Birgitta Trotzig.

Finnish literature in Swedish. The second flowering of Swedo-Finnish literature was in the 1920s, with the development of modernism in poetry, initiated by Edith Sodergran, who wrote visionary, dreamlike poems. After her came Elmer Diktonius, whose deliberately rough-edged lyrics had a revolutionary political accent; Gunnar Bjorling, with a technique of broken-off sentences; Rabbe Enckell, who wrote impressionistic nature poetry; and Hagar Olsson, writer of poetical novels and essays supporting modernism. Poetry became characterized by free rhythms, unrhymed lines, and powerful images. A similar style was adopted by a second generation of modernists: Solveig von Schoultz, Ralf Parland, Eva Wichman, and Bo Carpelan. During the 1960s there was a reaction against modernism, notably by the FBT review and the group that formed around it, who disliked the modernists' aestheticism and individualism and introduced more critical and social discussion. The most prominent FBT poets were Lars Huldén and Claes Andersson.

Among 20th-century prose writers, Tito Colliander and Göran Stenius wrote from a religious standpoint. Drama was presented by Walentin Chorell, and the psychological novel by Oscar Parland, while Tove Jansson was internationally known as a children's writer. Among younger authors Christer Kihlman and Anders Cleve stood out. Essayists included the critic and literary historian Yrjo Hirn and Göran Schildt.

Georg Brandes' radical ideas

Danish. About 1870 in Denmark, a new movement began, led by Georg Brandes, from which a modern (that is, a naturalistic or realistic) literature emerged. His *Hovedstrømninger **i** det 19de aarhundredes litteratur* (1872–90; ***Main Currents in 19th Century Literature,*** 1901–05), describing the growth and defeat of reaction, caused a great sensation. He influenced Henrik Ibsen and August Strindberg and wrote many scholarly and critical works illustrating radical ideas, while later biographies of Shakespeare, Goethe, Voltaire, Julius Caesar, and Michelangelo revealed how he was influenced by Nietzsche into developing a philosophy of aristocratic radicalism. Among his followers were Jens Peter Jacobsen, whose short story "Mogens" and novel ***Fru Marie** Grubbe* are the supreme examples of Danish Naturalism, while his other novel ***Niels** Lyhne* (1880) and some of his short stories dealt with dream as against reality; and Holger Drachmann, greatest lyric poet of the period, who later reacted strongly against Brandes and whose poetry and prose was often about the sea.

Henrik Pontoppidan, one of Denmark's greatest novelists, dealt at first with social injustices and contemporary political, moral and religious problems in his short stories. The Denmark of his day was also the subject of his greatest work, three long novel cycles, ***Det** forjaettede **land*** (1891–95; ***The Promised** Land*, 1896), ***Lykke-Per*** (1898–1904; "Lucky Peter") and ***De** døcles **rige*** (1912–16; "The Realm of the Dead"); and in these he makes penetrating, if unflattering, analyses of Danish national character. Herman Bang was another novelist interested in the outsiders of life and in insignificant people. His skillful. mainly impressionistic technique was displayed in his best novels, *Ved vejen* (1886; "By the Way-Side"), ***Tine*** (1889), and ***Det** hvide **hus*** (1898; "The White House"). Karl Gjellerup was another of Brandes' followers who later opposed him and one of whose best novels, *Møllen* (1896; "The Mill"), had a poetic beauty; but a metaphysical idealism made his later stories unreadable.

Other notable writers at the end of the century were Gustav Wied, whose "satyr plays" and whose novels *Livsens Ondskab* (1899; "Life's Malice") and ***Knagsted*** (1902) were full of malicious humour; Vilhelm Topsøe, a conservative realist; Peter Nansen. who wrote stories reminiscent of those of Guy de Maupassant; Carl Ewald, whose nature stories were based on Darwinian philosophy; Karl Larsen, who caught the atmosphere of Copenhagen and its inhabitants with fine precision; and several playwrights, including Edvard Brandes, Otto Benzon, Gustav Esmann, Sven Lange, Einar Christiansen, and Henri Nathansen.

In the 1890s a Neo-Romantic poetic revival occurred, reinstating the value of emotion and fantasy. The leader of these Symbolist poets was Johannes Jørgensen, since 1896 a convert to Roman Catholicism, whose prose works were also widely read. They included *Lignelser* (1898; "Parables"), books of travels, saints' lives, and an autobiography. Other poets at this time were Viggo Stuckenberg, who expressed sad resignation; Sophus Claussen, whose poems, often obscure, show sensuality, pantheistic love of nature, and sophisticated aestheticism; Helge Rode, a mystic who also wrote plays and criticism attacking intellectualism; and Ludvig Holstein, who was more akin to the classical tradition of Romantic poetry.

Several women contributed to literature at the turn of the century: Gyrithe Lemche, who wrote a novel cycle. ***Edwardsgave*** (1900–12); Agnes Henningsen, a brilliant writer who was often concerned with experiences of the emancipated woman; and Karin Michaelis, a fine psychologist, best known for her novel ***Den** farlige **alder*** (1910; ***The Dangerous Age,*** 1911).

Nexø and Jensen

The two greatest early 20th-century novelists were Martin Andersen Nexø and Johannes Vilhelm Jensen. Nexø's works described the lives of poor people; ***Pelle Erobreren*** (4 vol., 1906–10; ***Pelle the Conqueror,*** 1913–16) and ***Ditte Menneskebarn*** (3 vol., 1917–21; ***Ditte: Daughter of Man,*** 1920–23) were great epics of proletarian life, and his reminiscences were among the finest in the language. Jensen, who was also a great and original lyric poet and prolific essayist, wrote ***Den** lange rejse* (6 vol., 1908–22; ***The Long Journey,*** 1922–24), an ambitious epic of man from the baboon stage to the discovery of America; he was also noted for ***Himmerlandshistorier*** ("Tales from Himmerland"), based on his childhood memories of North Jutland; ***Kongens fald*** (1900–01; ***The Fall of the King,*** 1933); and nine volumes of ***Myter*** (1907–44; "Myths"). Other novelists of this period included Jakob Knudsen, whose interest, taking account of the inequality of man and the need for authority, was with Christian and moral problems; Harald Kidde, an introspective and melancholy writer; and Knud Hjortø, a keen and intelligent writer of psychological novels.

Three important poets, all with a Jutland background, were Jeppe Aakjaer, whose popularity and appeal recall that of Robert Burns in Scotland; Johan Skjoldborg; and Thøger Larsen. Also the novels of Marie Bregendahl and Harry Søiberg draw upon Jutland settings. Other prominent novelists were J. Anker Larsen, Otto Rung, Johannes Buchholtz, Thit Jensen, and Astrid Ehrencron-Kidde.

Significant poets were Tom Kristensen, whose novels *Livets* ***arabesk*** (1921; "Life's Arabesque") and ***Haervaerk*** (1930; ***Havoc,*** 1968) were also important; Ottc Gelsted; Emil Bønnelycke; Hans Hartvig Seedorff; Harald Bergstedt; Kai Friis Møller; and Per Lange. Among interesting novelists after World War I, Jacob Paludan wrote some very good novels—Fugle ***omkring fyret*** (1925; ***Birds Around** the **Light,*** 1928) and *Jørgen Stein* (1932–33; Eng. trans., 1966)—as also did Hans Kirk, whose ***Fiskerne*** (1928; "The Fishermen") was social realism at its best. Harald Herdal, a disciple of Nexø, exposed society's hypocrisy in his proletarian novels. Jørgen Nielsen's themes were suppressed hatred, sin, and fear among Jutland peasants. H.C. Branner, an important writer of novels, plays, and short stories, spoke of the loneliness of men and the danger of power. Another writer of these three genres was Knud Sønderby, who had a brilliant style and deep understanding. Nis Petersen, poet and novelist, was famous for ***Sandalmagernes gade*** (1931; ***The Street of the** Sandalmakers,* 1933) and ***Spildt** maelk* (1934; ***Spilt Milk,*** 1935). Karen Blixen (Isak Dinesen), an aristocratic writer with subtle irony and unusual sensitivity, wrote ***Seven Gothic Tales*** (1934) in English, ***Out of Africa*** (1937), ***Winter's Tales*** (1942), and ***Last Tales*** (1957). Two other notable novels by Faeroese writers were ***Barbara*** (1939; Eng. trans., 1948), by Jørgen-Frantz Jacobsen, and ***De** fortabte spillemaend* (1950; ***The Lost Musicians,*** 1972), William Heinesen's masterpiece. Among the distinguished novelists of the 1930s were also Mogens Klitgaard, whose first novel was about the "white-collar proletariat"; Hans Scherfig, a great humorist and social satirist; Martin Alfred Hansen, a psychological novelist, whose best known novel is *Løgneren* (1950; ***The Liar,*** 1954); Knuth Becker; Aage Dons; Leck Fischer; and Peter Freuchen, most of whose novels are about his Arctic experiences.

Danish playwrights of the post-World War I generation such as Sven Clausen and Svend Borberg were influenced by German Expressionism, Symbolism, Luigi Pirandello, and Sigmund Freud. Kaj Munk, who revived the heroic drama of Shakespeare and Schiller, showed unusual qualities in his best plays—En ***idealist*** (1928; ***Herod the King,*** 1953), and ***Ordet*** (1932; ***The Word,*** 1953)—which were concerned with problems of God and man. The work of Kjeld Abell marked a severance from Naturalist drama, and a radical perspective underlay his witty dialogue. His most important plays were ***Melodien, der blev vaek*** (1935; ***The Melody that Got Lost,*** 1939), ***Anna Sophie Hedvig*** (1939; Eng. trans., 1944), ***Dage på en sky*** (1947; ***Days on a Cloud,*** 1964), and ***Skriget*** (1961; "The Scream"). C.E. Soya is the third important playwright of the period, and also a novelist and fine short-story writer; some of his daring experiments with the theatre have been very successful.

Many postwar poets found an aesthetic manifesto in ***Fragmenter af en dagbog*** (1948; "Fragments of a Diary"), by Paul la Cour, who was influenced by contemporary French poetry. Jens August Schade, a sophisticated naivist, also had an important influence, as did the prematurely deceased poets Gustaf Munch-Petersen and Morten Nielsen. A revival of poetry followed the liberation (1945), and an Existential periodical *Heretica* (1948–53) became the voice of a group of young writers who regarded a Christian philosopher, Vilhelm Grønbech, as their spiritual progenitor. Two outstanding poets apart from the *Heretica* group were Halfdan Rasmussen, who also wrote excellent nonsense verse, and Erik Knudsen, also a brilliant satirical playwright. Both studied contemporary problems and reacted against the anti-rationalism and anti-intellectualism of the *Heretica* movement. Tove Ditlevsen was another important poet, as well as novelist and short-story writer, unattached to any group. Though belonging to a previous generation. Albert Dam published his best prose work in the 1950s and 1960s. Among the young modernist writers of those two decades, the best known was Klaus Rifbjerg, both as a poet and a novelist. Other important prose writers of the same period were Hans Lyngby Jepsen, Leif Panduro, Poul Ørum, Villy Sørensen, Cecil Bødker, Peter Seeberg, Erik Aalbaek Jensen, Thorkild Hansen, Ulla Ryum, Sven Holm, Tage Skou-Hansen, Peter Ronild, Svend Age Madsen, Per Højholt, Christian Kampmann, Anders Bodelsen, Hans Jørgen Nielsen, Vagn Lundbye, and Henrik Stangerup.

Poetry also flourished in those decades, and the moat noteworthy poets were Frank Jaeger, Thorkild Bjørnvig, Ole Sarvig, Ole Wivel, Jørgen Gustava Brandt, Ove Abildgaard, Robert Corydon, Ivan Malinovski, Lise Sørensen, Jørgen Sonne, Benny Andersen, Poul Borum, and Inger Christensen.

The best playwrights of the 1950s and 1960s were Ernst Bruun Olsen, Finn Methling, Erik Knudsen, Leif Panduro, and Klaus Rifbjerg.

Icelandic. Modern Icelandic prose writing did not really develop until the late 1870s, when a group of young men, influenced by the theories of the Danish critic Georg Brandes, began their literary careers. Unfortunately, they had absorbed Brandes' ideas uncritically, which resulted in introspective, self-pitying works believed by their authors to be realistically written. The early works of Einar Kvaran were like this, but he later proved to be a novelist of skill and power.

Several writers of this time showed a keen eye for character and an understanding of human feelings and of the stark life of rural Iceland: Jón Trausti (Gudmundur Magnússon), who wrote the cycle *Heidarbýlid* (4 vol., 1908–11; "The Mountain Cot"); Gunnar Gunnarsson, whose *Kirken på bjerget* (1923–28; "The Church on the Mountain") was written in Danish; and Gudmundur Hagalín. The outstanding modem prose writer was Halldór Laxness, who was awarded the Nobel Prize for Literature in 1955. His mature works, written after several years' travelling abroad, were influenced by his conversion to Roman Catholicism and his identification with the basic ideas of Communism. His major works were *Salka Valka* (1936), *Sjálfstaet fdlk* (1935; *Independent People,* 1945), *Íslandsklukkan* (1943; "The Bell of Iceland"), and *Gerpla* (1952). He helped restore his native language as a sensitive medium for storytelling.

The works of Halldór Laxness

A writer of independence—egocentric, individualistic, acidly satirical—was Thórbergur Thórdarson, who wrote essays, *Bréf til Láru* (1924; "Letters to Laura"), and an autobiographical fragment, *Ofvitinn* (1940–41; "The All-Too-Wise"). Other novelists included Ólafur Jóhann Sigurdsson, whose first hook, *Skuggarnir áf baenum* (1936; "The Shadows of the Farm"), was immediately well received, and Agnar Thórdarson, who also made an impressive debut with *Haninn galar tvisvar* (1949; "The Cock Crows Twice"). During the 1960s a new writer of force emerged in Svava Jakobsddltir, whose *Leigjandinn* (1969; "The Lodger") revealed a stylist of vigour and a writer of great insight into the murkier side of the human psyche.

At the beginning of the 20th century, poetry had lyricists in Thorsteinn Erlingsson, whose early delicacy later developed into a more powerful note in "Aldaslagur" (1911; "Sound of the Ages") and in an incomplete epic, "Eidurinn" (1913; "The Oath"); in Einar Benediktsson, who wrote in an ornate style sometimes capable of greatness, as "I dísarhöll" ("In the Hall of the Muses") shows; and in Stephan G. Stephansson, an expatriate farmer in Canada, who was a more bitter poet, influenced by the "realism" that passed for Georg Brandes' ideas in Icelandic literature—but *Andvokur* (1909–38; "Sleepless Nights") revealed a sensitive spirit and a consummate control of words.

David Stefánsson was a major Icelandic poet, a traditionalist who wrote with feeling and simplicity in *Kirkja fyrirfinnst engin* ("No Church To Be Found") and *Svarti daudi* ("Black Death"). His approach was shared by Tómas Gudmundsson and Jón Helgason, whose book *Úr landsudri* (1939, revised 1948; "From the South") was outstanding. Steinn Steinarr (Adalsteinn Kristmundsson) made innovations: he introduced the techniques of free verse in *Raudur loginn brann* (1934; "Red Burned the Flame") and was deeply influenced by Surrealism, and his nihilistic outlook was reflected in *Ferd án fyrirheits* (1942; "Journey Without Promise"). Om Arnarson (Magnús Stefánsson) possessed a rare humour, as *Rimur af Oddi sterka* (1938; "The Rhyme of Oddi the Strong") shows. Sigurdur Ivarsson also had a similar quality, and *Vjer brosum* (1929; "We Smile") and *Verkin tala* (1930; "Works Speak") became classics. Two young poets, Hannes Pétursson and Einar Bragi (Fridriksson), published work in the 1960s that showed great sensitivity and skill in adapting Icelandic to new, European metres.

Icelandic drama really started to develop through Jóhann Sigurjónsson, whose first success was *Fjalla-Eyvindur* (1911; *Eyvind of the Hills,* 1916), followed by *Galdra-Loftur* (1915; "Loftur the Sorcerer"); both plays were based on powerful folktales. Gudmundur Kamban's *Hadda-Padda* (1914) was highly praised by Georg Brandes, and he remained important in Scandinavian drama for the next quarter of a century. After Kamban, there were few plays of tasting value, though David Stefánsson's *Gullna hlidid* (1941; "The Golden Gate"), Jakob Jónsson's *Tyrkja-Gudda* (published 1948), and Agnar Thórdarson's satirical comedy of modern Reykjavik life, *Kjarnorka og kvenhylli* (1957; "Nuclear Force and Female Popularity"), had considerable merit. In *Ganksklukkan* (1962; *The Cuckoo Clock,* 1972) the latter produced a powerful play on the dehumanizing effect of modem life.

RUSSIAN AND SOVIET LITERATURE IN THE 20TH CENTURY

Soviet literature in the first years of the 20th century was dominated by the attempts of writers and politicians to come to terms with a new ideology. Early writers, reared in the literary traditions of the 19th century, were opposed, before and after the 1917 Revolution, by writers concerned to support the new social and political ideas. With the installation of the Communist Party, a left-wing, party-backed movement in literature was established whose aim was to produce literature in support of Communist theories. At the same time, anxiety about freedom of expression increased, and these two seemingly opposed interests remained central issues, earning any given writer either official approval or disapproval of his work in the Soviet Union.

The early 20th century in Russia. *Symbolism.* The revival of poetry in Russia stemming from the 19th-century Symbolist movement had as its leader Vladimir Solovyov. His poetry expressed a belief that the world was a system of symbols expressing metaphysical realities. Konstantin Balmont and Valery Bryusov were more talented poets, and the greatest poet of the movement was Aleksandr Blok, who in *Dvenadtsat* (1918; *The Twelve,* 1920) united the Revolution and God in an apocalyptic vision in which 12 Red Army men became apostles of the New World, headed by Christ. Other important Symbolist poets were Vyacheslav Ivanovich Ivanov, Fyodor Sologub, and Andrey Bely. Each brought his particular field of knowledge to his conception of reality: Ivanov, a Greek scholar, proclaimed the identity of Christ and Dionysus, and Bely, outstanding theoretician of the movement, worked out his philosophy in novels strongly reminiscent of Gogol and Dostoyevsky.

Futurism. Viktor Khlebnikov founded a most revolutionary poetic movement in 1910. A manifesto announcing the establishment of Futurism, also signed by Vladimir Mayakovsky, the future "poet laureate," appeared under the title of *Poshchochina obshchestvennomu vkusu* (1912; "A Slap in the Face of Public Taste"). Futurists wanted to wake everyone up by shock tactics. Mayakovsky rallied at first to the Soviet regime. His later satires, *Klop* (1929; *The Bedbug,* 1960), and *Banya* (1930; "The Bathhouse"), showed disillusion with Soviet bureaucracy. His suicide in 1930 was ascribed to this political disillusionment. Boris Pasternak also began as a Futurist poet and by 1917 had published most of the poems included in *Sestra moya zhizn* (1922; *Sister: My Life,* 1967).

Post-Revolutionary literature. The impetus of the Revolution led to brilliant experimentation, and contending literary groups sprang up, in particular, over rejection of the 19th-century literary heritage in favour of a proletarian culture. Styles and themes became increasingly realistic violent stories of the Civil War; Isaac Babel's Cossack tales *Konarmiya* (1926; *Red Cavalry,* 1929) and Aleksandr Fadeyev's *Razgrom* (1927; translated as *The Nineteen,* 1929) shared an emphasis on realism.

Among authors concerned with the struggle between old and new—whom Trotsky called mere "fellow travellers of the Revolution"—were Konstantin Fedin, Leonid Leonov, Yury Olesha, and the satirist Valentin Katayev.

The abolishment of RAPP

Five-Year-Plan literature and Socialist Realism. The struggle between "fellow travellers," who demanded creative self-determination, and the left-wing writers, arguing for proletarian literature subservient to the Communist Party, became so bitter that the Central Committee refused to endorse any one group. When the First Five-Year Plan was launched, however, the party backed the Russian Association of Proletarian Writers (RAPP) and other left-wing groups, thereby supporting themes of construction and collectivization. RAPP was abolished by the Central Committee in 1932, and it was suggested that all literary groups form a single union. A national Union of Soviet Writers was thereupon organized in 1934, and a new doctrine of "Socialist Realism" propounded to guide creative efforts. Literature was expected to show reality in its "revolutionary development," which in practice meant it was supposed to propagate the party's ideological objectives. Some tolerance was permitted during this period of economic success and Popular Front agitation. The ablest writers, such as Mikhail Sholokhov and Leonov, took advantage of this. Melekhov, for example, hero of Sholokhov's *Tikky Don* (1928–40; vol. I and II translated as *And Quiet Flows the Don,* 1934, and continued in *The Don, Flows Home to the Sea,* 1940), remains a tragic "White" Cossack throughout. It remained for Nikolay Ostrovsky to create the "new Soviet man" in *Kak zakalyalas stal* (1932–34; *How the Steel Was Tempered,* 1952), and most writers took up this model. The party required "positive heroes" and optimistic themes that romanticized heroic efforts to achieve Communism. Literary critics imitated one another in insistence on the party's version of Socialist Realism. Some authors were "liquidated" in Stalinist purges, and others, among them Anna Akhmatova, a major poet of exquisite skill, and Pasternak, perhaps the most celebrated Soviet poet, fell virtually silent.

War literature, 1943–45. During World War II many writers shared fervent national patriotism and were assigned to the front as correspondents. Party control relaxed and authors felt free to concentrate on emotions evoked by the tragedy of war. The most popular novels were *Nepokorcnnye (Semya Tarasa)* (1943; "The Taras Family"), by Boris Gorbatov; *Dni i nochi* (1943–44; *Days and Nights,* 1945), by Konstantin Simonov; and Fadeyev's *Molodaya gvardiya* (1946; *The Young Guard,* 1959), about the underground activities of Young Communist League members under German occupation. The best plays were Leonov's *Nashestviye* (1942; "Invasion") and *Lyonushka* (1943), and the most popular were sensational exposures of Red Army weaknesses in *Front* (1942), by Aleksandr Korneychuk, and Simonov's *Russkiye lyudi* (1942; *The Russians,* 1944). War was most sincerely expressed in poetry. The finest lyrics were by Pasternak, Olga Bergolts, Margarita Aliger, and Simonov, and the best long poems by Aleksandr Tvardovsky, Nikolay Tikhonov, and Vera Inber.

Zhdanovism, 1945–53. Any hope that the party would extend the creative freedom of the war was squelched by the Central Committee's 1946 decree that all art must be politically inspired and understood. The edict was the work of Andrey Zhdanov, heir apparent to Stalin and unofficial commissar of culture. The works of Akhmatova and Mikhail Zoshchenko were singled out for savage attack. Although Socialist Realism was not to be rejected, the principal tenet of literature became party spirit. The word Zhdanovism (Zhdanovshchina) was given to both this policy and the sterile writing it generated. Some authors took refuge in war themes, on which many novels were based, notably *Povest o nastoyashchem cheloveke* (1946; "A Story About a Real Man"), by Boris Polevoy (Kampov); Ilya Erenburg's *Burya* (1947; *The Storm*); *Za vlast sovetov* (1949; "For the Power of the Soviets"), by Valentin Katayev; and Mikhail Bubennov's *Belaya brryoza* (1947; "The White Birch"). Novels such as Pyotr Pavlenko's *Schastye* (1947; "Happiness") and Semyon Babayevsky's *Kavaler zolotoy zvezdy* (1947; "Cavalier of the Golden Star") complied with the official demand for literature dealing with peaceful reconstruction. Fedin's trilogy *Pervye radosti* (1945–46; *Early Joys,* 1960), *Neobyknov ennoye leto* (1948; *No Ordinary Summer,* 1950), and *Kostyor* (1962; "The Bonfire") was far superior and perhaps the best postwar novel.

Protesters against party control

Literature after Stalin. The number and directness of articles critical of literature's stagnation increased considerably after Stalin's death in 1953, and writers were quick to take up the protest. Leonid Zorin's play *Gosti* (1954; "The Guests") challenged the claim that class structure had vanished. Poets such as Bergolts and Inber attacked the "steam shovel" school of verse, and Pasternak, in 1954, brought out his first original verse since the war, namely, ten poems later included in his novel *Doctor Zhivago* (1957). Erenburg, however, gave name and focus to the dissent in a play, *Ottepel* (1954; *The Thaw,* 1955), which highlighted the tragedy of the artist under Stalin. The same year, at the Second Writer's Congress, offending authors and critics were reprimanded. Premier Nikita Khrusbchev's exposure of Stalin's crimes in 1956 revived the protest against literary control, and some of the best writing of 1956 and 1957 was an indictment of Stalinism and hypocrisy in Soviet life. Soviet reaction to the 1956 Hungarian Revolution made these works seem part of an organized attack on authority, and administrators of the Writers' Union demanded recantations from authors. Publication of Pasternak's *Doctor Zhivago* in western Europe in 1957 provoked another outburst of denunciation, but writers continued to press for greater freedom. These "angry young men," including the poet Yevgeny Yevtushenko, were strongly opposed by the "conservatives," who forced several leaders to recant at a 1963 meeting of the Board of the Writers' Union. When Khrushchev left office in 1964, the party stated it would insist on an identification of literature and party ideology, although freedom would be allowed in matters of form and content. Under these guidelines a new liberal generation strove for an individuality of expression alien to Socialist Realism.

The most talented prose writers were Aleksandr Solzhenitsyn, Viktor Nekrasov, Yury Kazakov, Vasily Aksyonov, Yury Nagibin, and Vladimir Tendryakov. Solzhenitsyn became internationally known for an authentic description of prison camps under Stalin in *One Day in the Life of Ivan Denisovitch* (1962), and later works, such as *The First Circle* (1968) and *Cancer Ward* (1968), established him as a mature and talented author and resulted in his winning the Nobel Prize for Literature in 1970. He was not permitted to receive the Nobel Prize, however, and the work he considered to be his greatest, *August 1914,* was also not permitted to be published. The novel was published in the West in the early 1970s in both Russian and English. In 1971 his *August 1914* was published in Paris (Eng. trans., 1972), the first of a projected series of novels about World War I. Aksyonov published his first story, *Kollegi* ("Colleagues"), in 1960. His work was marked by an ability to write pithy colloquial Russian and unusual powers of description. Among many good poets, Andrey Voznesensky and Yevtushenko were perhaps best known. The brilliant concentration of verbal experiments in Voznesensky's poetry gave it a deceptively witty appearance. He was, in fact, centrally concerned with problems of modern society. Yevtushenko was a more erratic and less talented poet, but his verse, like Voznesensky's, showed involvement with social and personal problems. He travelled widely after 1960 and was perhaps the best known of the modern Russian poets. Among less well-known but talented poets were Yevgeny Vinokurov and Bella Akhmadulina. But the degree of nonconformity allowed writers hung in the balance after two minor writers, Andrey Sinyavsky and Yuly Daniel, were imprisoned (1966) for publishing work abroad that criticized the Soviet Union. The pattern of periods of partial relaxation of control followed by periods of severe ideological disciplining established itself as a regular feature of Soviet literary life and gave every indication of continuing.

Ukrainian. In the first three decades of the 20th century Ukrainian literature experienced a renascence, characterized by a variety of quickly succeeding and often strongly competing literary movements. Realism, with a distinctly decadent strain, was the most notable characteristic of Volodymyr Vynnychenko's prose. Pavlo Tychyna was the leading Symbolist poet; others included D. Zahul, P. Tereshchenko, and O. Slisarenko. Neoclassicism produced outstanding poets in Mykola Zerov, Maksym Rylsky, and Mikhaylo Dray-Khmara. Futurism was initiated by M. Semenko and produced one of the Ukraine's greatest 20th-century poets, Mykola Bazhan.

After the Russian Revolution, during a period of relative freedom between 1917 and 1932, a host of talented writers emerged: Mykola Khvylovy's prose was imbued with revolutionary and national Romanticism, Hryhoriy Kosynka's prose was Impressionistic, while Y. Yanovsky's stories and novels were unabashedly romantic, and V. Pidmohylny adhered to the principles of realism. Other writers of note include the novelist and film maker Aleksandr Dovzhenko as well as the novelists Borys Antonenko-Davydovych, Volodymyr Gzhytsky, Mikhaylo Ivchenko, Oles Dosvitny, the poet Mike Yohansen, and the humorist Ostap Vyshnia. The outstanding dramatist was Mykola Kulish.

In 1932 the Communist Party enforced Socialist Realism as the required literary style. Typical representatives of this official literature were the dramatist Oleksandr Korniychuk and the novelist Mikhaylo Stelmakh.

The post-Stalinist period saw the emergence of a new generation that rejected drab Socialist Realism. Known as the "Writers of the Sixties," they included Lina Kostenko, Vasyl Symonenko, V. Korotych, Ivan Drach, M. Vinhranovsky, V. Holoborodko, and I. Kalynets. Repressive measures taken in the early 1970s silenced many of them or else turned them back to Socialist Realism.

Lithuanian. In 1918 Lithuania regained independence. Writers began to concentrate on developing national culture and a greater degree of sophistication in literature.

V. Kreve-Mickevitius, novelist and dramatist, was regarded by some as the greatest Lithuanian writer, and J. Baltrušaitis achieved distinction as a lyrical poet. Other prominent names were those of V. Mykolaitis-Putinas, who pioneered the modern Lithuanian romance; B. Sruoga and K. Binkis, both poets and dramatists; J. Savickis, novelist and short-story writer; a poet, F. Kirša; P. Vaičiūnas, dramatist; I. Šeinius, novelist and short-story writer; and S. Čiurlionienė, novelist and dramatist. B. Brazdžionis was a leading poet of the younger generation.

When Lithuania was occupied by the Soviet Union, writers were compelled to follow the Communist line. Those working in the West (about 100) tried to further the development of the national literature. New modes of expression were successfully attempted in the philosophical poetry of A. Nyka-Niliiinas, in the idylls of J. Mekas, and in the novels of M. Katiliškis. The genres most favoured have been the short story and the lyric.

Latvian. A great emotional experience was the revolution of 1905 when the Latvians tried to break away from imperialistic Russian and local German tutelage. Lyricism then began to predominate. In the verse and fairy tales of a great poet, Kārlis Skalbe, the ethical world of folk poetry was reborn. A new generation of authors arose when Latvia became independent in 1918. J. Akuraters portrayed himself or romantic heroes with aesthetic ideals in the spirit of Nietzsche, and his lyrics were powerful but improvised. A. Upītis, inspired by French and Russian Naturalism, idealized working class heroes. E. Virza (E. Lieknis) created lyrics in strict classical forms; his prose poem *Straumēni* (1933) praised the patriarchal farmstead. Lyrical emotionalism was disciplined in J. Jaunsudrabiņš, whose best novel was a trilogy, *Aija*, *Atbalss*, and *Ziema*. World War I provided many themes for works such as K. Štrāls' *Kari* (1922–27), Anna Brigadere's *Kvēlošā loka* (1922), and A. Grins's *Dvēseļu putenis* (1932–34); the postwar atmosphere found expression in well-composed short stories by J. Ezeriņš and K. Zarinš. J. Veselis tried to harmonize the spirit of the age with that of Latvian folk poetry; this is successfully realized in the poetry of Zinaida Lazda and Andrejs Eglitis and also in that of Veronika Strēlerte.

Latvians found it difficult to achieve a unified view of the world in the 20th century, however, and so turned to psychological detail. The stories of M. Bendrupe show Freudian influence, and E. Ādamsons depicted the neuroses of modern man. Anšlavs Eglitis delighted in caricaturing and intensifying one particular human quality at a time. M. Ziverts, the best modern Latvian dramatist, evolved a long, one-act play culminating in a great monologue, as in his historical tragedy *Vara* (1944).

Several poets were still influenced or inspired by folk songs, but A. Čaks (A. Čadarainis) created a new tradition, describing in free verse, with exaggerated images, the atmosphere of the suburbs. His outstanding work was a ballad cycle, *Mūžības skārtie* (1938; "Marked by Eternity"), about the Latvian fusiliers of World War I. His influence was felt in a new generation of poets who migrated to the West after World War II.

The poetry of Velta Sniķere contains certain elements of Surrealism in verse reminiscent of ancient Latvian magic formulas. A fusion of A. Čaks's imaginist poetry and the experience of big American cities led to the poetry of L. Tauns and G. Saliņš. Caks's verse may have appeared too avant-garde to find an echo in the work of poets in present-day Latvia; but three gifted poets there, Vizma Belševica, O. Vācietis, and I. Ziedonis, gave individual expression to their inner worlds of experience, constrained by external pressures. In the West, new vistas were opening up in the poetry of Astride Ivaska, Aina Kraujiete, and Raiba Bičole. In the field of prose, A. Bels, a noteworthy writer in Latvia, portrays a many-facetted reality; in the West, Ilze Sķipsna has moved from Existentialism to profound Symbolism working at various levels, as in her novel *Neapsolitas zemes* (1971).

The Siuru and Arbujad groups

Estonian. During the russification of 1890–1906 the style of Realism introduced was best represented in Estonian literature by Juhan Liiv. It was superseded by the Neoromantic Young Estonia group, whose leader, a poet, Gustav Suits, devised the slogan "More European culture! Be Estonians but remain Europeans!" For Suits and his followers this meant greater attention to form. With the 1917 Revolution emerged the Siuru group (named after a bird in Finno-Ugrian mythology). These Neoromantic poets reacted against Suits's emphasis on formalism. Their emotional intensity was well illustrated by Henrik Visnapuu, who, with Marie Under, developed the lyrical potential of Estonian to the full. By the 1930s a renewal of Realism brought poetry closer to life, but the only outstanding poetry of this revival was descriptions of modern urban life in the work of Juhan Siitiste (Schiitz). The Arbujad group (which also took its name from a word with origins in mythology) of the mid-1930s, on the other hand, stressed intellectual and aesthetic aspects of literature. Leading poets were Betti Alver, whose skillful use of symbolic imagery was shown in ***Tolm ja tuli*** (1936; "Dust and Fire"); Heiti Talvik, who in ***Kohtupäev*** (1937; "Doomsday") predicted the coming holocaust; Uku Masing, a religious mystical poet; and Bernard Kangro, later the leading lyrical poet in exile.

After World War II more than half of Estonia's writers went into exile, and their poetry reflected either pessimism, like Kangro, or longing for Estonia, as in Visnapuu's exile poetry. Gradually a new generation of ironic poets emerged, best exemplified by Kalju Lepik, experimental author of ***Kollased nõmmed*** (1965; "Yellow Heaths"); a skeptical poet, Arno Vihalemm, whose work was spiced with self-irony; and the author of the epic ***Peetri kiriku kellad*** ("The Bells of St. Peter's"), Ivar Grunthal. In Estonia little poetry appeared under Stalin's Socialist Realism, but new poets, adopting Western styles, appeared in the 1960s. Among these were Jaan Kross, Ellen Niit, Ain Kaalep, and Mats Traat.

Prose writing was equally influenced by movements current in Europe. The Realism of the beginning of the century was exemplified in the social criticism of Liiv's *Kümme lugu* (1893; "Ten Tales") and in Ernst Peterson's criticism of social injustice, ***Boils*** (1899–1901). An outstanding Realist novelist was Eduard Vilde, who wrote a historical trilogy attacking the Balto-Germanic feudal system and in *Mäeküla piimamees* (1916; "The Dairyman of Maekula") again treated the relationship between landowner and serf. Friedebert Tuglas, who introduced Impressionism and Symbolism, belonged to Young Estonia, while August Gailit was a leading ***Siuru*** prose writer. Among the Neoromantics who became Realists were Anton Tammsaare, who wrote an ethico-psychological chronicle, ***Tdde ja õigus*** (1926–33; "Truth and Right"); Albert Kivikas, whose ***Nimed marmortahvlil*** (1936; "Names on the Marble Tablet") was about the war of liberation; August Malk, who concentrated on nautical subjects; and August Jakobson.

Estonian novelists in exile

Novelists in exile found inspiration in the very fact of their exile. Two principal themes were wartime experiences and the problem of adapting to new environments. Among writers in exile were Gailit, Malk, Kivikas, Ristikivi, Pedro Krusten, Karl Rumor, Juhan Jaik, Evald Mand, and Valev Uibopuu. New writers included a critic, essayist, and dramatist, Arvo Magi, and the novelists Ilmar Talve, Ilmar Jaks, Helga Nõu, and Elin Toona. Of these, the last three showed an increasing internationalism in their work. In Estonia postwar fiction decayed in the way poetry did. The deadening effect of Socialist Realism gradually gave way to greater subtlety, and younger novelists, such as Arvo Valton, Enn Vetemaa, and Mati Unt, were able to examine some of the problems of Communism and begin stylistic experimentation.

Dramatic works were few, but two early playwrights stood out: August Kitzberg, author of both comedies and serious plays, and Hugo Raudsepp, whose realistic and symbolical plays were social satires.

OTHER LITERATURES EASTERN EUROPE AND THE N MEDITERRANEAN IN 20TH CENTURY

Polish. The literary trend of the period following the 1863 uprising was called positivism and reflected a practical mood and reaction against Romanticism, the rise in Russian and Austrian Poland of an urban intelligentsia acting as a leaven in this ferment. Periodicals were of particular importance in disseminating new ideas, especially the ***Tygodnik Ilustrowany*** ("Illustrated Weekly"), founded in 1859. The natural consequence of a positivist outlook was a predominance of prose. With other writers of the Warsaw school, Aleksander Świętochowski voiced anticlerical and anti-aristocratic views in his weekly ***Prawda.*** Bolesław Prus (Aleksander Głowacki), a journalist, ranked high among Polish novelists: ***Lalka*** (1890; "The Doll") gave a complex picture of bourgeois life in Warsaw, and ***Faraon*** (1897; "The Pharaoh and the Priest") was an ambitious evocation of ancient Egypt, masking political problems that could not be published in the form of a modern novel.

Eliza Orzeszkowa, a campaigner for social reform. wrote about women's emancipation, the ignorance of the peasants, and the Jewish problem. Her books showed psychological penetration and a fine sense of style.

In 1905 Henryk Sienkiewicz was a winner of the Nobel Prize for Literature; his early critical works propounded positivist aesthetic theories, and his early short stories showed an acute awareness of the contemporary situation and a stark Realism. His famous ***Quo Vadis*** (1896; Eng. trans., 1898), a historical novel of Rome under Nero, has been widely translated.

A more flexible Naturalism gained ground toward the end of the period, as seen in the stories of Adolf Dygasiński, famous for portrayals of animal life—such as *Zając* (1900; "The Hare")—which could be compared with those of Rudyard Kipling. Gabriela Zapolska, a critic of social hypocrisy in Naturalistic novels and lively comedies, excelled in dialogue and dramatic situations, in such plays as ***Moralność Pani Dulskiej*** (1906; "The Morals of Mrs. Dulska").

The period produced two important poets: Adam Asnyk, a reflective lyricist of formal dexterity; and Maria Konopnicka, writer of sentimental verses on the underprivileged.

The "Young Poland" movement

The "Young Poland" movement describes several different groups and tendencies united by opposition to positivism, and a desire to return to imagination in literature; hence its other name, Neoromanticism. Among its pioneers were Antoni Lange, a poet, and Zenon Przesmycki (Miriam), editor of a Symbolist review, ***Chimera.*** Both made translations from a number,of other languages and expressed aesthetic theories in critical essays. Przesmycki's most influential contribution to a development of modern literature was his discovery of Cyprian Norwid, the chief innovator of Polish poetic diction in the previous century, whose experiments and originality could be compared with those of Ezra Pound and, more particularly, with Gerard Manley Hopkins.

Kazimierz Tetmajer achieved popularity with his nostalgic *Poezje*, but his prose had greater vigour and precision of observation, ***Na skalnym Podhalu*** (1903–10; "Tales of the Tatras") containing some effectively stylized folk material. An interesting poet, Jan Kasprowicz, was of peasant origin. His principal contribution to Polish literature lay in the structure of his longer lyrical poems; those in the volume ***Ginqcemu światu*** ("To the Dying World") employed a technique of association, quotation, musical repetition, and free metre that anticipated the early style of such poets as T.S. Eliot. Tadeusz Micinski, a forerunner of Expressionism and Surrealism, wrote philosophical and mystical poems and plays, notably a collection of poems, ***W mroku gwiazd*** (1902; "In the Twilight of the Stars"), and a play, ***Kniaź Patiomkin*** (1906). The lyrical poet Leopold Staff, whose work shows great variety and technical dexterity, was at this period associated with the "Young Poland" movement, although some of his finest work was written later.

Stanislaw Przybyszewski was a leading exponent of the movement's new aesthetic theories and edited a literary magazine, *Życie* ("Life"). Stephen (Stefan) Żeromski expressed passionate concern for social justice and national freedom in widely read works, but an excess of Realist documentation frequently vitiated the power of his later work. Władysław Stanisław Reymont, of peasant stock, adapted the Naturalistic technique to create a vision of peasant life in an epic novel cycle, *Chłopi* (4 vol., 1902–09; "The Peasants"), for which he received the Nobel Prize in 1924. One of the most effective novels of the period, ***Żywot i myśli Zygmunta Podfilipskiego*** (1898; "The Life and Thoughts of Zygmunt Podfilipski"), by Józef Weyssenhoff, presented an ironical portrait of the egoist in society. Wacław Berent's *Próchno* (1903; "Rotten Wood") portrayed with biting irony the late-19th-century decadence of life and art. His ***Ozimina*** (1911; "Autumn Sowing"), a Symbolist novel foreshadowed the associative structure and narrative technique of James Joyce's *Ulysses* (1922). His ***Zywe kamienie*** (1918; "The Living Stones") stressed the unity of medieval culture and Poland's place within it. A bold experiment antedating by several years the psychoanalytical novel in western Europe was *Pałuba* (1903) by Karol

Irzykowski. In it, motivation and behaviour were presented from different viewpoints; ingeniously cemented by the author's own analyses, as in a scientific study. Irzykowski was also a critic and, in Dziesiqta muza (1924; "The Tenth Muse"), the first to give attention to the cinema as an art form.

Stanislaw Brzozowski, an outstanding critic, insisted that a critic must represent the moral consciousness of his age; in *Legenda Młodej Polski* (1909; "The Legend of Young Poland") he analyzed the weakness of the period's literature and expounded his view of the unity of all work — physical, technical, intellectual, and artistic.

Stanisław Wyspianski was an artist and dramatist of genius. In his plays he reforged elements from classical tragedy and mythology, Polish Romantic drama, and national history in a complex whole. *Wesele* (1901; "The Wedding") was a visionary parable of Poland's past, present, and problematical future, cast in the form of the traditional puppet-theatre play. It was a masterpiece of evocative allusion, tragedy, and humour.

The literature of the period was characterized by close contact with western European literatures, but writers such as Wyspianski turned back to the Polish Romantics in a search for new poetic language.

The Skamander poets

Literature *in* restored Poland. The restoration of independence in 1918 decisively affected Polish literature. The period between 1918 and 1939 was characterized by richness and variety and increasing contact with other European literatures, especially through publication of many translations. For nearly a decade after 1918 lyrical poetry predominated. A periodical, *Zdrój* ("The Fountainhead"), showed affinities with German Expressionism. In Warsaw several poets formed a group called Skamander, from the name of their monthly publication; it was united by a desire to forge a poetic language, attuned to modern life. Julian Tuwim, a poet of emotional power and linguistic sensitivity, headed the group, and during World War II, in Brazil and the United States, wrote a long, discursive, autobiographical poem, *Kwiaty* polskie (1949; "Polish Flowers"). Among Skamander members were Jan Lechoń, Kazimierz Wierzyński (both died abroad after World War 11), Antoni Slonimski, and Jaroslaw Iwaszkiewicz, who was also a prolific prose writer; among sympathizers of the group were two eminent woman poets, Maria Sasnorzewska-Pawlikowska and Kazimiera Iłłakowiczówna, and Władysław Broniewski, a poet with strong left-wing sympathies, who became a master of the revolutionary lyric, expressing involvement in current social and ideological problems.

A writer of importance was Bolesław Leśmian, whose symbolic, Expressionist poetry was remarkable for the inventiveness of its vocabulary, sensuous imagery, and philosophic content, all anticipating Existentialism. He published only three notable collections—*Łąka* (1920; "The Meadow"), Napdj *cienisty* (1936; "The Shadowy Drink"), and *Dziejba leśna* (posthumous, 1938; "Woodland Tale") — but has been recognized by his admirers as the most outstanding 20th-century Polish lyrical poet.

Other experimental movements (*e.g.*, Futurism) followed revolutionary trends in poetry — particularly in Italy and Russia. More original was a group called Awangarda Krakowska (of Cracow). Led by Tadeusz Peiper, it produced few works but had widespread influence in the regeneration of poetic technique. Two of its adherents, Julian Przyboś and Adam Wazyk, who was loosely connected with the movement, ranked among the outstanding poets of the post-World War II period. Also noteworthy was Józef Czechowicz, who assimilated traditional and regional elements to the new style.

Prose writing reached its ascendancy in the second decade of independence. Early stories by Zofia Nalkowska belonged to the "Young Poland" movement and aimed to expose the feminine psyche; later she turned to other themes, striving for narrative objectivity and technical simplicity. Two other women writers of distinction were Zofia Kossak-Szczucka, noted for historical novels, and Maria Kuncewiczowa, who wrote psychological novels. Juliusz Kaden-Bandrowski used experimental Realism in Czarne skrzydta (1928–29; "Black Wings") and Mateusz *Bigda* (1933), which treated social and political themes. Michał Choromański's *Zazdrość i medycyna* (1933; Jealousy and Medicine, 1946) used experimental methods of narrative sequence and was remarkable for its clinical analysis of character. A writer skilled in reflecting subtleties of perception was Bruno Schulz, author of Sklepy *cynamonowe* (1934; Cinnamon Shops, 1963), whose prose was reminiscent of that of Kafka.

Tadeusz Żeleński Boy was a literary critic who showed interest in biographical and psychological rather than literary considerations and provided revaluations of most of the great figures of the past. The essay form was represented by Jan Parandowski, whose main theme was the classical culture of Greece and Rome. He was also known for a novel about Oscar Wilde, Krdl zycia (1930; "King of Life"). A subversive attack on intellectual and social conventions was launched in Ferdydurke (1937; Eng. trans., 1961) by Witold Gombrowicz, who displayed in it a satirical talent reminiscent of Alfred Jarry. The taste for the cyclic novel was satisfied by Maria Dąbrowska with *Noce* i *dnie* (4 vol., 1932–34; "Nights and Days"), an outstanding modern Polish example of a chronicle novel in epic style, about the development of the Polish intelligentsia of upper-middle class origin.

Stanislaw Witkiewicz and "catastrophism"

The drama was the weakest of the literary forms during this period and returned to the Young Poland's especially symbolic form (Karol Hubert Rostworowski, Jerzy Szaniawski). The experimental dramas of Stanisław Ignacy Witkiewicz were of interest chiefly for their expression of anti-Realist aesthetic theories; he developed many ideas of the Awangarda, applied the principles of "pure form" to painting and drama, and was the main exponent of a movement in Polish literature known as "catastrophism." Obsessed with the idea of a disintegration of European culture, endangered by totalitarian ideologies, and an attempt to impose the uniformity of a "mass society," he developed his ideas in plays combining elements of Surrealism, grotesque misrepresentation, and what later became known (in the plays of Eugène Ionesco, for example, whose work Witkiewicz to some extent foreshadowed) as the Theatre of the Absurd. His novels Pozegnanie *jesieni* (1927; "Farewell to Autumn") and Nienasycenie (1930; "Insatiability") expressed the same philosophy. After World War II his work attracted interest abroad and appeared in translation (*e.g.*, The *Madman* and the Nun and *Other* Plays, 1968).

Literature after 1945. The impact of World War II, the experience of occupation, and the establishment in 1945 of the People's Republic decisively affected the character of literature in Poland and also produced a number of émigré writers who had become famous between World Wars I and II: lyrical poets of the Skamander group and former associates of the Awangarda movement, among them Czesław Milosz, who won the Prix Littéraire Européen for Zdobycie wtadzy (1953 in French, 1955 in Polish; "The Usurpers"). Many émigrés wrote of experiences in German or Soviet prisons and forced-labour camps. The most accomplished as literature were *Inny świat* (1953; A World Apart, 1951) by Gustaw Herling-Grudzinski and Sprawa *Józefa* Mosta (1953) by Herminia Naglerowa. Jerzy Pietrkiewicz, whose early work was published in Poland during World War II, later wrote mainly in English (the novels The Knotted Cord, 1953; Isolation, 1959). Witold Gombrowicz also published his postwar work abroad and became famous with the novels Trans-Atlantyk (1953), *Pornografia* (1960; Eng. trans., 1966), and *Kosmos* (1965; Cosmos, 1967), which won him the 1967 International Prize for Literature. He also published abroad a play, *Ślub* (1953; The Marriage, 1970), and three volumes of diaries (*Dziennik*, 1957–66). In all these works, especially the novels, Gombrowicz treated philosophical and psychological themes in a satirical narrative style in which, by emphasizing the grotesque and irrational elements in human nature, he presented an expos6 of the conventions of modern life and culture.

In Poland the years immediately after the war saw publication of works written during the occupation or by writers who had been in concentration camp or prison. A frequent theme was the attempt to come to terms with Fascism and war. This was exemplified in the short stories of Tadeusz Borowski, a prisoner at Auschwitz, who, in collections published in 1948, *Pozegnanie z Mariq,* ("Farewell to Mary") and *Kamienny Swiat* ("The World of Stone," published in *This Way for the Gas, Ladies and Gentlemen, and Other Stories,* 1967), explored human depravity and degradation; and in those of Adolf Rudnicki, who treated moral and philosophical themes in lyrical prose and described, in *Szekspir* (1948; "Shakespeare") and *Ucieczka z Jasnej Polany* (1949; "Flight from Jasna Polana"), the wartime fate of the Jewish community in Poland. Another theme, shown, for example, by a novel *Popidt i diament* (1948; *Ashes and Diamonds)* by Jerzy Andrzejewski, was the examination of the moral controversies that accompanied political and social changes of the postwar period, especially the tragic situation of young conspirators involved in the struggle against the new regime.

The literature of Social Realism and the reaction against it

During 1949–54 the literature of Social Realism gained ground, and writers attempted to emulate the great 19th-century Realists. A new type of hero was created—the ordinary man or woman actively engaged in "productive" work, and those elements in the social scene that served to present an idea of revolutionary progress were accentuated. The main writer was Leon Kruczkowski, a prewar exponent of the Marxist novel whose outstanding plays *Niemcy* (1949; "The Germans") and *Pierwszy dzieli wolnosci* (1960; "The First Day of Freedom") were successfully performed abroad. Kazimierz Brandys, whose development typified postwar tendencies in Polish literature, published an epic novel cycle, *Między wojnami* (1947–51; "Between the Wars"), and a Social-Realist novel, *Obywatele* (1954; "Citizens").

Among writers of the Social-Realist period who stood aside from political involvement were Konstanty Ildefons Gałczyński, who combined lyricism with grotesque fantasy, and the reflective Mieczyslaw Jastrun, who in later work—for example, the essay collection *Mit Śródziemnomorski* (1962; "The Mediterranean Myth")—moved toward Existentialism. A group of Catholic writers opposed to Social Realism included Antoni Gołubiew, author of an epic novel cycle, *Bolesław Chrobry* (1947–54), a prose writer and dramatist Jerzy Zawieyski, and a historical novelist, Hanna Malewska. Teodor Parnicki used a background of conflict between cultures for an analysis of contemporary problems in historical novels set mainly in the early Christian period: *Koniec "Zgody Naroddw"* (1955; "End of the Agreement Between the Nations"), *Nowa baśń* (6 vol., 1962–70; "A New Fairytale"), and *Tylko Beatrycze* (1962; "Only Beatrice").

The weakness of the Social-Realist movement—its attempt to impose a political pattern on creative writing, its avoidance of themes arising from contemporary conflicts—resulted partly from the stranglehold of the Stalinist regime. In the period beginning in 1954–55, writers began to criticize these weaknesses and to oppose them. Andrzejewski, for example, presented contemporary ideas and problems in two novels combining historical and metaphorical treatment, *Ciemności kryją ziemie* (1957; *The Inquisitors,* 1960) and *Bramy raju* (1960; *The Gates of Paradise,* 1962), and Brandys criticized Stalinism in the novel *Matka Królów* (1957; *Sons and Comrades,* 1961).

The political "thaw" after 1956 made it possible for writers to renew contacts with the West. As a result, a period of development and experiment began in which the essay was a vehicle for philosophical and intellectual discussion and comment, and there was an increase of satire. Tadeusz Breza analyzed contemporary life and ideas in a prose of Proustian subtlety in, for example, *Spiżowa brama* (1959; "The Bronze Gate"), a keen description of life in the Vatican. Writers continued to be concerned with World War II, as shown by such novels as *Czarny potok* (1954; *Black Torrent,* 1969) by Leopold Buczkowski, *Kolumbowie-rocznik* 20 (1957; "The Columbuses-Generation of 1920") by Roman Bratny, and *Tren* (1961; "Threnody") by Bohdan Czeszko. Many chose as their theme consequences of wartime experience, as, for example, Tadeusz Konwicki in *Sennik współczesny* (1963; *A Dreambook for Our Time,* 1969). His new novels, *Wniebowstqpienie* (1967; "Ascension") and *Nic albo nic* (1971; "Nothing or Nothing") project those themes on contemporary problems. In *Gtosy W ciemności* (1956; "Voices in the Dark") and *Austeria* (1966; "An Inn"), Julian Stryykowski restated the feeling of the Orthodox Jewish Polish community that the world has already ended and gave it universal application.

Several prewar writers continued to publish: the poets Leopold Staff and Kazimiera Iłłakowiczówna; Maria Dąbrowska, who enhanced her reputation with a collection of short stories, *Gwiazda zaranna* (1955; "Morning Star") and a series of critical essays on Joseph Conrad; the novelist Maria Kuncewiczowa; and the novelist, poet, and dramatist Jarosław Iwaszkiewicz, who published the epic novel *Slawa* i *chwała* (1956–62; "Fame and Praise").

The works of Mrożek

Many writers of the 1950s and 1960s wrote prose works ranging from the political novels of Jerzy Putrament to the psychological novels of manners of Stanislaw Dygat, while the novels of Stanislaw Lem, leading Polish representative of serious science fiction, looked to the future. Young writers such as Marek Nowakowski, in their search for a moral basis for life, often looked into the worlds of those on the fringes of society—outcasts and misfits. An interesting younger writer was Sławomir Mrożek, who went into exile after the political·crisis of 1968; both in his plays—Policja (1958; "The Police"), *Na pełnym morzu* (1961; *Out at Sea,* 1967), *Karol* (1962; *Charlie,* 1967), *Striptease* (1962), and above all in *Tango* (1965), his most widely known work—as well as in stories collected in *Słoń* (1957; *The Elephant,* 1962), he displayed an acute sense of satire and the grotesque, which he used to express a philosophy of life both topical and timeless. His comedy belonged to the Theatre of the Absurd and was distinguished by a subtlety of parody and highly stylized language.

Poetry of the "second" postwar period was notable for expression of philosophical thought. Stanislaw Jerzy Lec was a satirical poet noted for skeptical philosophical aphorisms in *Myśli nieuczesane* (published in series from 1957, collected posthumously 1968; *Unkempt Thoughts,* 1967). Zbigniew Herbert, whose most notable collection was *Studium przedmiotu* (1961; "Study of a Subject"), was a representative of intellectual poetry. Perhaps the most interesting, however, was Tadeusz Różewicz, who had a profound influence on younger lyrical poets; from *Niepokdj* (1947; "Trouble"), his first collection, to *Głos Anonima* (1961; "The Nameless Voice"), his work was occupied with moral themes. He also wrote plays resembling those of Ionesco: *Swiadkowie albo nasza mała stabilizacja* (1962; *Faces of Anxiety,* 1969), and one published with poems in *Kartoteka* (1961; *The Card Index and Other Plays,* 1969). Modem problems also dominated the intellectual, ironic poetry of Wisława Szymborska, collected in *Sól* (1962; "Salt") and *Sto pociech* (1967; "The Hundred Amusements").

The lyrical poetry of the generation of poets born about 1930 was characterized by a variety of aims and styles. On the one hand, the controversial work of such poets as Miron Bialoszewski showed extreme experimentalism; on the other hand, a poet such as Ernest Bryll reasserted traditional poetic forms. Some poets—Tadeusz Nowak and Jerzy Harasymowicz—turned for inspiration to the peasant culture; others, among them Jarosław M. Rymkiewicz, an outstanding translator of English and American poetry, based their poetic program upon the example of T.S. Eliot, in a return to Baroque and Neoclassical forms, and developed an erudite, allusive poetry. Most representative of the poets of this new generation was perhaps Stanislaw Grochowiak, who created an expressive poetic style based on sudden contrasts and deliberate emphasis on grotesque aspects of life.

Critics and essayists included Artur Sandauer; Jan Kott (now in the U.S.), whose *Szkice o Szekspirze* (1961; *Shakespeare, Our Contemporary,* 1964) was widely translated and discussed; Kazimierz Wyka; and Jan Błoński.

Political events in 1968 forced several Polish writers to emigrate and publish abroad, thus joining those who had left Poland previously, like Czesław Milosz (now in the U.S.) and Witold Gombrowicz (who died in Switzerland). Some writers, such as Jerzy Andrzejewski, who published his novel *Apelacja* (1970; *The Appeal,* 1971) in Paris, defied the increasing control; others, like Paweł Jasienica and Stefan Kisielewski, were banned from publishing their works. Although the change in political and cultural life occurring after December 1970 promised more creative freedom, the basic pattern of limited freedom of expression, experimentation, and search for new literary forms established in 1956 continued throughout the period.

The *Lumir* and *Ruch* groups

Czechoslovak. *Czech.* By the 1870s Czech literature had fully established itself in poetry and the novel but not yet in the drama. Literary life crystallized around two groups associated with the periodicals *Lumir* and *Ruch* ("Stir"). The former stressed the necessity to Europeanize Czech literature, the latter looked to the strength of native traditions and themes. The leading representative of the cosmopolitan tendency was Jaroslav Vrchlický (Emil Frida), probably the most prolific of all Czech writers. His lyrics showed an amazing mastery of language, with a vast cycle of historical epics containing his best work. Perhaps his greatest influence was exercised by numerous translations of major European writers, and he was also instrumental in forming a modern Czech poetical language. Cosmopolitanism found expression, too, in the poetry of Vrchlický's Neoromantic lyrics and epics, though his novels and stories, particularly the semi-autobiographical *Jan Maria Plojhar*, were more original.

The principal figure of the "Stir" group was Svatopluk Cech, whose work was inspired by love of his nation and its traditions and by a liberal humanism. One of his most popular works, *Ve stínu lípy* ("In the Shade of the Lime Tree), gave an idyllic picture of Czech country life; in his prose satires he created Mr. Brouček, who satirized the Czech middle classes' philistinism.

The Czech novel of the last quarter of the century exhibited a trend toward realism of description, which was apparent in two major historical novelists, Alois Jirásek and Zikmund Winter (whose novels, such as *Mistr Kampanus*, were by-products of his work as a historian). Both men presented a romanticized version of Czech history, but the details of reconstruction were based, in both instances, on scholarly research. The novels of Jirásek presented an entire history of the Czechs up to his own time, and his mature craftsmanship was seen at its best in a novel of the period of national decline, *Temno* ("Darkness").

In the last decade of the century new trends shaped poetry. Lyrics of more subtle content and more complex form were written by Antonín Sova, and the turn of the century saw the best work of two remarkable Czech poets: Otakar Biezina (Václav Jebavý) and Petr Bezruč (Vladimir Vašek). Biezina expressed a personal religion in a language of subtlety and originality; his metrical structure (including free rhythms) influenced later poets. Bezruč was a regional poet, obsessed by the oppression of the Czechs of Silesia.

In the 1890s, critical reaction to the romantic idealism of the national revival had inspired the historical and sociological writings of T.G. Masaryk. A more sophisticated approach to literary criticism showed itself in the work of František Xaver Šalda, whose sensitive interpretations of his country's greatest writers have not been surpassed.

The satire of Karel Capek

With the establishment of an independent Czechoslovak state in 1918, new possibilities lay before Czech writers. Some of the best achievements of Czech literature sprang from this period, especially in lyrical poetry of the utmost variety. The Czech drama came into its own, notably, in the idealistic and satirical plays of Karel Capek and in the versatile talent of František Langer. The realistic novel of Prague life had been successfully cultivated by Karel Matěj Capek-Chod and in the lighter vein by Ignát Herrmann, but Czech narrative prose reached a peak in the work of three very different writers: Karel Capek, Ivan Olbracht (Kamil Zeman), and Vladislav Vančura. Capek wrote Wellsian fantasies, subtle psychological studies, and stories, such as a trilogy *Hordubal, Povětroň,* and *Obyčejný život* (1933–34), *Krakatit* (1924), and *Válka s mloky* (1936). His best plays exposed the problems of a centralized, machine society of the 20th century: *R.U.R.* (1920), from which the word robot was introduced into English, and *Ze života hmyzu* (1921; *The Insect Play).* In his later plays, pacifism gave way to an acceptance of patriotic duty. Olbracht's most successful works had a Ruthenian background. Vančura cultivated a complex, mannered style with success. The novels of Jan Otčenášek stand out as the most effective from a mass of fiction dealing with World War II and the establishment of a Communist government. The highest achievement of Czech 20th-century literature was the lyrical poetry of the period 1918–45. Outstanding was the work of Josef Hora, František Halas, Vítězslav Nezval, and Jaroslav Seifert.

Slovak. The chief strength of Slovak literature continued to be in the lyric, and the pre-eminent prewar poet was Hviezdoslav (Pavol Országh). He enriched Slovak poetic language by original work and by translations, especially from English, Russian, German, and Hungarian. Other notable poets were the Russophile Svetozár Hurban Vajanský and Ivan Krasko (real name, Jan Botto), whose volumes of verse, *Nox et solitudo* (1909) and *Verše* (1912), were among the finest achievements of Slovak literature.

After World War I leading lyric poets were Martin Rázus, Janko Jesenský, Emil Boleslav Lukáč, Jan Smrek (real name, Ján Cietek), Ján Poničan, and Laco Novomeský. In the novel, country tales by Timrava (real name, Božena Slančíkova), a vast chronicle of 20th-century Slovakia by Milo Urban, and the lyrical prose of Margita Figuli were outstanding. The problems of World War II and its aftermath of Communist government found vivid, personal expression in the work of Ladislav Mňačko and Alfonz Bednár.

Yugoslavian. *Serbian.* Serbian writers between the wars, like Miloš Crnjanski, followed major European literary movements (Expressionism and others); the union of the realistic tradition and modernism was reached in the work of Nobel Prize winner Ivo Andrić. Other writers who gained reputations as exponents of the realistic and psychological orientation included Isak Samokovlija and Veljko Petrović. In the 1930s, when political and social problems intensified, formal experiment and introspective writing met with strong opposition.

Serbian literature during and after World War II dealt predominantly with war themes, particularly in the immediate postwar period. As far as its procedural and ideological orientation went, the literature continued the trends of the 1930s.

In the 1950s and 1960s, however, new writers emerged with a range of themes and styles that reflected the tendencies of western European literature. A poet and novelist, Oskar Davičo, wrote about the individual in Communist society and the problems involved in creating a Socialist state. Branko Ćopić was one of the most popular writers of adults' and children's books to emerge after 1945, and the novels *Prolom* (1952; *The Break,* 1952) and *Gluvi barut* (1957; *The Deaf Gunpowder,* 1957) were important works. Other noteworthy writers were Mihailo Lalić, Dobrica Ćosić, highly original Miodrag Bulatović, together with the poets Vasko Popa, nature poet Stevan Raićković, Miodrag Pavlović, and Dušan Kostić.

Croatian. In the first part of the 20th century, poetry was the chief form of Croatian literature. It dealt with re-creation of natural beauty, spiritual struggle, and a search for deeper meaning, to be seen in the work of Vladimir Vidrić and Vladimir Nazor and of prose writers such as Milutin Cihlar, Dinko Šimunović, and the dramatist Ivo Vojnović. By the middle of the century, the influence of Expressionism was reflected in a frequent adoption by writers of inner monologue, psychoanalytical exploration, and stream-of-consciousness technique. Socialist Realism grew up among the Partisans of the national liberation struggle in World War II; and after 1948 a reaction against uniformity of expression and literature strengthened, and literature became less restricted. Chief representatives of the modern movement in Croatian poetry were Miroslav Krleža, Tin (Augustin) Ujević, and Antun Branko Šimić, a poet of deep contrasts and social awareness. Modern prose was represented by Slavko Kolar, a short-story writer of peasant life and the Partisan struggle, and August Cesarec, who describes working class life. The eminent lyricists are D. Cesarić, D. Tadijanović, and satirist and dramatist Ranko Marinković.

The Slovene modernist movement

Slovene. The emergence of a Slovene modernist movement in the early 20th century resulted in a second peak in Slovene literature (the first being in the time of Prešeren, 1800–49). With rich metaphor and lively rhythm the moderns gave fresh life to poetry, while in prose they replaced Naturalistic description by greater sensitivity and a more subjective attitude toward the world. In addition to the young poets Dragotin Kette and Josip Mum, the two main founders of this movement were Oton Župančič, a vigorous poet of luxuriant diction and native sentiments, and Ivan Cankar, a masterful prose stylist who combined themes of human suffering and longing with severe social criticism. Naturalistic and Realistic elements developed in the prose works and plays of Ksaver Meško, Alojz Kraigher, Zofka Kvedrova, and Fran S. Finžgar. Župančič's verbal enthusiasm was counterbalanced by the less colourful but expressive and reflective lyrical poetry of Alojz Gradnik, especially his love poems.

After World War I social and spiritual tension was evident in the predominantly Expressionistic poetry of Tone Seliškar, Miran Jarc, and Anton Vodnik, as well as in the plays of Slavko Grum. Expressionism was less clearly manifest in fiction: the historical novels of Ivan Pregelj had essentially a Romantic basis in spite of their Baroque stylization, while the prolific story writer France Bevk showed psychological, realistic vision. The most remarkable representative of this period was a young poet, Srečko Kosovel, who in direct, simple language described the horrors of being and of dying and the eternal division of modern man, crucified between two historical eras.

With the development of Social Realism (*c.* 1930), the novel and short story revived, especially with the works of Juš Kozak, Miško Kranjec, Prežihov Voranc, and Ciril Kosmač, who turned to the everyday life of simple country people. Socio-historical conflicts provided plots and themes for the dramatic works of Bratko Kreft and the poetry of Igo Gruden and Mile Klopčič; and if some poetry (*e.g.*, that of Edvard Kocbek) pointed the way to a more inward-looking subjective attitude, other verse (*e.g.*, the caustic sonnets of Božo Vodušek) manifested a sober disillusionment. Between World Wars I and II, moral and psychological criteria were introduced into criticism by Josip Vidmar and socio-ideological criteria by I. Brnčič.

During and after World War II, writers who had been active before it remained rooted in the realistic tradition, although many deepened, or at least broadened, the scope of their themes. This is particularly true of the poet Karl Destovnik-Kajuh, who was killed in the war, and of poet and dramatist Matej Bor, as well as of Ivan Potrč and Mira Mihelič, both prose writers and dramatists. The prose of Danilo Lokar and Boris Pahor does not differ essentially from this movement in spite of its innovations. A more subjective, metaphorical trend has developed in the poetry of Cene Vipotnik, France Balantič, and especially Jože Udovič, to whom the postwar generations look in their new endeavours.

Macedonian. The Macedonian republic established after 1944 in the Yugoslav federation charged such scholars as B. Koneski with standardizing Macedonian as the official literary language, and a national culture began to develop freely. Writers who became prominent in the 1950s and 1960s include the poets G. Todorovski, S. Ivanovski, M. Mateski, and P. Popovski, and the playwrights M. Čašule and T. Arsovski.

Bulgarian. By the 1890s the school of older writers began to be challenged by a younger group, the "Europeans," intent on freeing art from parochialism and socio-political militancy. Leading this was the review *Misŭl* ("Thought," 1892–1908), founded by Krŭst'o Krŭstev, the first Bulgarian critic to stress the importance of the aesthetic conscience. A member of the *Misŭl* group, Pencho Slaveykov, broadened the Romantic tradition of Bulgarian poetry and helped to create a complex poetic language. Influenced by Nietzsche, he glorified the heroism of spiritual achievement and wrote his ***Epicheski pesni*** (1896–98; "Epic Songs") on the giants of the human spirit he revered—Dante, Beethoven, Shelley, and Giacomo Leopardi. His ideas were expressed in his essays and in his autobiographical anthology of "apocryphal" verse by fictitious poets, ***Na ostrova na blazhenite*** (1910; "On the Isle of the Blessed"). His narrative poems ***Boyko*** (1897) and ***Ralitsa*** (1893) interpreted folk themes psychologically, and his greatest, though unfinished, work, *Kŭrvava pesen* (1913; "Song of Blood"), was an epic on Bulgaria's history and destiny. Even more than Slaveykov, Petko Todorov, originator of the Bulgarian Romantic short story, believed that literature was sufficient unto itself; both in his *Idilii* (1908), prose poems inspired by folklore, and in several dramas based on Balkan mythology, notably ***Zidari*** (1906; "Masons"), he displayed a delicate poetic talent.

Symbolist poetry in Bulgaria

With the beginning of the 20th century, avant-garde literary currents encouraged a "modernist" phase related to the Symbolist movement in Western poetry. An anacreontic permissiveness and lyrical power distinguished the poetry of Kiril Khristov, as in ***Himni na zorata*** (1911; "Hymns to the Dawn"). P. Yavorov, a member of the ***Misiil*** group, did most at this time to develop the musical and evocative potentialities of Bulgarian in poetry. His work closely reflected his restless spiritual development, and although his plays showed great promise, lyric poetry was his real achievement. Echoes of Yavorov are found in the melodious, sensuous stanzas of Dimcho Debelyanov, whose death in World War I made him a symbol of tragic frustration for intellectuals. Symbolism inspired the postwar poetry of Nikolay Liliev and Teodor Trayanov.

Meanwhile, the Realist tradition continued in the work of such writers as Anton Strashimirov and G. Stamatov, whose cynical stories denigrated Sofia's society. Strashimirov was an acute observer of the contemporary social scene; one of his best stories of peasant life was "Kochalovskata kramola" (1895; "The Kochalovo Quarrel"), and he also wrote the novels ***Yesenni dni*** (1902; "Autumn Days"), ***Krŭstopŭt*** (1904; "Crossroad"), and ***Sreshta*** (1908; "Meeting") and the dramas ***Vampir*** (1902) and *Svekŭrva* (1906; "Mother-in-Law"). His contemporary Elin Pelin portrayed his native rural province with wit and humanity in ***Razkazi*** (1904 and 1911; "Stories") and in the tragic novellas ***Geratsite*** (1911; "The Gerak Family") and ***Zemya*** (1928; "Land"). Yordan Yovkov, novelist and playwright, excelled at describing the effects of war, the subject of his early masterpiece, ***Zemlyatsi*** (1915); his short stories "Staroplaninski legendi" (1927) and "Vecheri v Antimovskiya khan" (1928; "Evenings in the Antimovo Inn") display deep insight into the Bulgarian mind and a classical mastery of narrative prose.

In the aftermath of World War I the literary left was represented by a number of poets who died tragically young: Geo Milev, a convert to revolutionary Marxism; Khristo Smirnenski; and later, the young, gifted Nikola Vaptsarov, who died a martyr in the anti-Nazi resistance, but not before he had hailed the dawn of Socialism and the machine age in his poems ***Motorni pesni*** (1940; "Motor Songs") and ***Izbrani stihotvoreniya*** (1946; "Selected Verses").

Pre-eminent as the prose of Elin Pelin and Yovkov was between the World Wars, the younger generation brought artistic refinement to realistic portrayal of Bulgarian life, and high literary standards were maintained in such authoritative reviews as the Symbolist

Hypcrion (1920–31) and *Zlatorog* (1920–44). The mystical-fantastic evocations of medieval Bulgaria by the art historian Nikolay Raynov represented Bulgarian Neo-romanticism at its best. A satisfying fusion of traditional and experimental poetry was to be found in the work of Elisaveta Bagryana.

The Communist regime set up in 1944 encouraged only writing of "Socialist Realism" as defined by Soviet "aesthetic" theory. The resulting uniformity of purpose made it difficult to evaluate the work of many writers, though the novels of D. Dimov and D. Talev received universal acclaim, especially Talev's work on 19th-century Macedonia. By the 1960s literary critics were still engaged in a Marxist re-evaluation of the past, but this, with the demise of the *kult* (Stalinist "personality cult"), seemed decreasingly mandatory amid a wealth of young talent. The originality of Blaga Dimitrova, Lyubomir Levchev, Asen Khristoforov, Nikolay Khaytov, and Yordan Radichkov augured well for the future.

Greek. The great social and economic changes that Greece underwent at the end of the 19th century soon found reflection in its literature with publication in 3838 of *To taxidi mou*, by Ioannis Psicharis, and development of the Demotic movement, through which modern Greek prose and poetry changed momentously. The outcome of a long linguistic and literary battle was victory for the vernacular.

Prose. Writers now turned to the life of the Greek village for inspiration; as in the 1880s, there were no large towns apart from Athens, so this new prose became predominantly pastoral—a tendency that continued till the turn of the century, when urban life developed. Although there was an early pastoral novel in *Thanos Vlekas* (1855), by Pavlos Kalligas, it was not until after the 1880s that Alexandros Papadiamadis and Andréas Karkavitsas wrote tales of villages and fisherfolk. Papadiamadis was the greatest modern Greek short-story writer, and his short novel I *Fonissa* ("The Murderess") was outstanding; in a slightly archaic idiom he examined the psychology of island people and moving incidents in their lives. The short stories of Karkavítsas were of an almost equal emotional power, but his greatest achievement was a realistic novel, I *zitiános* ("The Beggar"). Another important work at this time was an autobiographical novel, I *Zoï kai o thanatos tou Karabela*, by Konstantinos Theotokis, notable for its realism and insight.

The urban novel

The development of the urban novel was aided by translations of French novels, particularly those of Émile Zola, and was fostered by Psicharis and Gregorios Xenopoulos. Psicharis' novels (the most successful of which was *To Oneiro tou Yianniri* [1897]) had a great influence in spite of faults in construction and characterization. Xenopoulos serialized many works in Athenian newspapers and often had to compromise his art in order to earn a living. His best sellers were *Foteini Sandri* (*O kokkinos Vrakhos*) and *Stella Violanti*. There followed Konstantinos Christomanos, who wrote *Kerenia koukla*, and Konstantinos Chatzopoulos, who produced the first Greek Symbolist novel, *Fthinoporo* (1917; "Autumn").

After World War I, when vernacular prose was reaching maturity, Stratis Mirivilis took the literary scene by storm with the first of his "war trilogy" novels, I *zoï en tapho* ("Life in the Tomb"). The second was I *daskala me ta khrisa matia* ("The Schoolmistress with the Golden Eyes"), and finally came I *Panagia i Gorgona* (1955). He also wrote shorter stories of power and originality. His style was a lively demotic with dialectal and lyrical elements. Elias Venezis also had a striking first success, with *To Noumero* 31328, about life as a prisoner of war; his major writings include a long novel, *Aeolia*, full of childhood memories.

A most gifted author of the 1930s was Kosmas Politis (P. Taveloudis), who in his *Eroica* (1938), about a gang of children, *Lemonodassos*, and *Gyri*, proved himself a master of long narrative. Georgios Theotokas' writings covered many genres in a flowing and simple demotic, and included a long novel, *Argo* (1936).

After World War II the earlier generation still strongly influenced younger writers, of whom Nikos Kazantzakis was pre-eminent. He turned to the novel after a long career writing poetry, travel books, dramas, and essays and showed full mastery of the medium in *Zorba the Greek*, *Christ Recrucified*, and *Freedom and Death*. His autobiographical novel *Report to Greco* was published posthumously.

Real literary merit was often achieved by the *Chronografema*, a column in certain Athens newspapers treating some daily event in a humorous or lyrical manner. In the hands of such writers as P. Nirvanas (Petros Apostolides) or Spiros Melas, it transcended journalism.

In the 1960s some writers stood out for their use of prose and imaginative powers, among them Vasilis Vassilikos, Rodis Provellengios, Nikos Kazdaglis, and Kostas Takhtsis.

Poetry. Kostis Palamás was a central figure of the Demotic movement; his many important works portrayed modern Greek life, the continuity of Greek history, and the social and spiritual convulsions of the late 19th and early 20th centuries. The long philosophic poem *Dodecalógos tou Giftou* (1907; "Twelve Lays of the Gypsy") was perhaps his greatest achievement; a powerful lyric epic, *Í flógera tou Vasiliá* (1910; "The King's Flute"), and the collection of lyrics *Asalefte Zoé* (1904; *Life Immovable*, 1904) established his reputation throughout the Western world, and his influence on all subsequent Greek literature was profound. The poets of this school explored the expressive and metrical possibilities of the spoken idiom and introduced Symbolism and free verse.

After Palamás, a leading figure in modern Greek poetry was Angelos Sikelianós. His vigorous verse was rooted in the new school of Athens, but his thought followed a more obscure course. Greek nature and history were seen in the light of a Dionysian mysticism.

Modern tendencies reflected in Cavafy's verse

One great Greek poet who remained untouched by the influence of the new school of Athens was Constantine Cavafy (Konstantínos Kaváfis), who dealt with the tragic glory of Hellenistic Greece and its decadence. In his work, historical memories and personal experiences were inextricably blended. His tragic outlook, sensuality, and irony reflected modern tendencies in Western culture, so that his writings have been widely acknowledged.

The many poets who wrote in Greece after the end of World War I include Nikos Kazantzakis, better known as a novelist, whose philosophical epic poem *Odyssey* stretches to 33,333 lines; George Seferis, whose poems were awarded the Nobel Prize for Literature in 1963; and other contemporaries Odysseus Elytis and John Ritsos, who compose in Greek.

Romanian. Romanian literature of the 20th century was rooted in the traditions of the 19th. Important figures spanned both centuries and established genres, literary groups, and methods of criticism that continued into the 20th century. Thus Titu Maiorescu, founder of a literary circle Junimea in 1863, which reacted against interest in form at the expense of content, pointed toward a later reassessment of the uses of literature. Ion Luca Caragiale died in 1912 but was relevant to the 20th century as the creator of Romanian social comedy. Similarly Barbu Ştefănescu established the historical national drama and Moses Gaster was important as a pioneer of research into Romanian folklore.

Literary movements of eastern and western Europe at the beginning of the century were reflected in Romania. The periodical *Viaţa Românească* (1901), based on the Russian model of "populism," had a social and political ideology. A critic Constantin Dobrogeanu Gherea's theories followed Karl Marx, although western modernism also influenced Romanian writers. Ovid Denuşianu clearly followed Symbolism, as did the poets Ion Minulescu and Georghe Bacovia, while Impressionism was taken up by E. Lovinescu and Niculae Davidescu, whose epic *Cântecul omului* (1928–37; "The Song of Man") aimed at re-creating world history.

After World War I. In the period of national unity in Romania following World War I, the novel began to compete with lyric poetry. Writers took inspiration from society or recent events, principally the war. Liviu Rebreanu wrote of the peasants' need for land and independence and in *Răscoala* (1932; ***The Uprising,*** 1934) described the 1784 peasant uprising in Transylvania. His best work, inspired by Romanian participation in World War I, was the *Pădurea spînzuraţilov* (1922; ***The Forest of the Hanged,*** 1967).

Cezar Petrescu and the poet Minulescu also dealt with the war, while other writers examined different areas of society: Ionel Teodoreanu described the disappearance of patriarchal life, Victor Popa wrote about rural subjects, G.M. Zamfirescu depicted the Bucharest suburbs, and D.D. Pătrăşcanu wittily described political life. A leading realist writer early in the century was Mihail Sadoveanu, who together with I.A. Brătescu-Voineşti represented a link with the older generation and was extremely influential for the development of prose. He concentrated on the place of the peasant in society and in 1924 was awarded the Golden Peace Medal for a description of the peasants' part in the war in ***Mitren Cocor.***

Nonfiction writers of the period

Scholars, philosophers, critics, and translators also made contributions to literature of the period. An archaeologist, Vasile Pârvan, commemorated the sacrificed war generation in ***Parentalia;*** a historian, Nicolae Iorga, founded literary periodicals and wrote plays, poetry, and criticism; a geographer, S. Mehedinţi (Soveja), edited a periodical and wrote village stories; Lucian Blaga was a philosophic essayist and poet, while Gala Galaction translated the Bible and wrote mystical poems and novels on biblical subjects. Lyric poetry was the most cultivated genre in modern Romanian literature. The diversity of styles was illustrated by Nichifor Crainic's religious traditionalist tendency, the mathematical form of I. Barbu's poems, and the influence of French and German lyric in Ion Pillat's verse. Tudor Arghezi was noteworthy for the expression he instilled into lyric poetry. Both his prose essays of 1935–36 and his verse, ***Cuvinte potrivite*** (1927; "Suitable Words"), ***Flori de mucegai*** (1931; "Mold Flowers"), ***Versuri de seară*** (1935; "Evening Verses"), were landmarks in Romanian letters.

Developments after World War II. Several famous writers continued to write after World War II. Arghezi reached new lyrical heights in *1907* and in a hymn praising mankind's will to live and struggle for freedom. Geo Bogza joined the Social Realist movement. Mihail Beniuc became (as he said) "the drummer of the new age" and, in stirring lyrics, celebrated achievements of the postwar period. Demostene Botez, whose prewar poetry described the sadness of provincial life, later revealed a vigorous optimism, and a poet Eugen Jebeleanu protested on contemporary events and themes. Among those who came to the fore during and after World War II were the poets Maria Banuş, who expressed the struggle for peace, Miron Paraschivescu, a lyrical poet taking themes from folklore, and Marcel Breslaşu, a complex writer on a wide range of themes.

Dramatists included Aurel Baranga, who dealt with the problems of contemporary life; Horia Lovinescu, whose plays depicted changing intellectual attitudes; and M. Davidoglu, author of plays set in mines and factories.

The postwar novel also dealt with topical themes. Zaharia Stancu wrote novels evoking Romanian village life in a vanished age. Eusebiu Camilar, in his novel ***Mist,*** bitterly indicted Fascism.

Essays and criticism were written by Călinescu, by Mihai Ralea, who also published travel books and philosophical and psychological works, and by Tudor Vianu, who revealed a materialistic and methodological approach after first adhering to the aesthetic school. Among critics becoming known in the 1950s were Paul Georgescu, Ovid Crohmălniceanu, Silvian Iosifescu, author of a study of the Romanian novel, and Dumitru Micu, who wrote a book on the modern novel in Romania (1959). In England, Miron Grindea was editor of the international literary review ***Adam.***

Hungarian. The last third of the 19th century in Hungary was an era of literary decline in which writers based their work on social and political ideals which were becoming sterile. The great majority of Hungarian writers came from the nobility and were part of the middle class: only at the end of the century did lower middle class writers come to the fore. The periodical ***A Hkt*** ("The Week"), founded in 1890 by József Kiss, became the organ of a number of gifted writers, including Zoltán Ambrus, a disciple of Maupassant and a theatrical expert.

The influence of the periodical ***Nyugat***

The year 1906, when Endre Ady burst into Hungarian life "with the new songs of a new epoch," marked a turning point. Ady's revolutionary political ideas made it difficult for sections of society to accept that he was a poet of unusual genius. In form and content Ady was an original lyrical poet who rejuvenated Hungarian language and poetical thought. In 1908 the left-wing periodical ***Nyugat*** ("The West") began to appear, under the editorship of Hugo Ignotus and Ern6 Osvát. The influence of this was not to be overestimated: it brought together vigorous talents and lifted the level of Hungarian letters. Among writers associated with ***Nyugat*** was Mihály Babits, who became editor in 1929; others were Dezső Kosztolányi, foremost impressionist in Hungarian poetry and an accomplished short-story writer, and Árpád Tóth and Gyula Juhász, who voiced the distress of the solitary and oppressed in society. Zsigmond Móricz was a great novelist whose descriptions of provincial life were masterly portrayals of peasants and gentry. His powerful works present a troubled society full of human conflicts and dramatic passions. Margit Kaffka, a first major woman writer in Hungary, was another significant novelist. From 1908 three main streams were discernible in Hungarian literature: the ***Nyugat*** group already mentioned; an official literature written by conservative-nationalist writers; and a superficial "boulevard" literature to entertain. The principal writer of the second group was Ferenc Herczeg, author of novels and plays of theatrical skill. The third type included several playwrights: an impressionist, Dezso Szomory; Menyhért Lengyel, whose *Tájfun* (1907; "Typhoon") had worldwide success; and the exceptionally gifted Ferenc Molnár. The importance of the ***Nyugat*** diminished considerably after World War I, and leading writers and poets of the group gained official recognition, and the conflict that opposed them to the traditionalist men of letters became less violent. A solitary writer, Dezso Szabó, author of powerful and bewildering novels and stories, alienated sympathy with venomous, though often pertinent, attacks on almost everything. Lajos Kassák was the first notable working class writer whose novels and autobiography accurately depicted working class life. He was, at the same time, the father of Hungarian avant-garde poetry.

The period between World Wars I and II saw a flowering in Hungarian letters. In poetry the chief figure was that of Attila József, whose great intellectual poems voice the anguish of modern man in a technological era and his passionate, socialistic ideas. Lőrincz Szabó was a master of fine observation and a gifted translator of foreign poets. The poetry of Miklós Radnóti (died 1944) reached a tragic climax in splendid poems written in a German camp and during a death march in which he himself perished. Gyula Illyés, considered by many as the leading Hungarian poet of the mid-20th century, found inspiration in the life of the peasantry.

In Hungary, as elsewhere, the novel became the principal form of literary expression. Between World Wars I and II Lajos Zilahy, Ferenc Kormendi, and Sándor Marai vividly described the Hungarian bourgeoisie, and Áron Tamási wrote on Transylvanian life. The realistic novels of Pál Szabó, János Kodolányi, and József Darvas and firsthand accounts of peasant life by Péter Veres were probably among the better works laden with a social "message." Problems of Socialism influenced some writers even before World War II; later, social and political considerations became overwhelming, and they constituted the materials of literature for some years. Almost all major novelists and poets tried writing drama, but the most satisfying achievements in this field were the historical plays of Lászlo Németh.

During the 20th century, literary criticism and the history of literature kept pace with growing production and an increased interest in letters. The two most outstanding personalities were János Horváth, a dignified scholar who brought Hungarian literary research to a summit, and

Gyorgy Lukács, a great Marxist literary theorist whose critical and aesthetic works assure him an international reputation.

Finnish. Finnish writers at about the turn of the century included Teuvo Pakkala, whose stories of childhood became classics; and Johannes Linnankoski, who wrote the best seller *Laulu tulipunaisesta kukasta* (1905; *The Song of the Blood-Red Flower,* 1920), describing the amatory adventures of a Finnish Don Juan, and a well-knit novel of peasant life; Aino Kallas's works had an Estonian setting, and her best were prose ballads, written in an archaic style, about illicit love: *Barbara von Tisenhusen* (1923), *Reigin pappi* (1926; *Eros the Slayer,* 1927), and *Sudenmorsian* (1928; *The Wolf's Bride,* 1930).

Many writers continued the tradition of "folk portrayal" but in a more critical spirit; after the civil war of 1918 an attitude of self-criticism became general. A leading Finnish prose writer was Joel Lehtonen, whose *Putkinotko* (1919–20) was a colourful, humorous, bitterly critical study of the lives of rural poor. His last novel, *Henkien taistelu* (1933; "Struggle of Souls"), bitterly satirized contemporary conditions. Volter Kilpi, in an important novel, *Alastalon salissa* (1933; "In the Parlour at Alastalo"), used interior monologue, long "flashback" episodes, and exact detailed description, spreading events of six hours over more than 900 pages. Kilpi was an exponent of the experimental novel; his interest in problem of time and in re-creation of the past linked him with Marcel Proust and James Joyce. The stories of Heikki Toppila set people's lives against a background of superstition, and his writing was grimly effective. F.E. Sillanpaa viewed his characters from a biologist's standpoint, as an integral part of their surroundings; for example, in his most important novels, *Hurskas kurjuus* (1919; *Meek Heritage,* 1938) and *Nuorena nukkunut* (1931; *The Maid Silja,* 1933). He received the Nobel Prize in 1939. Notable poets of the period include Veikko Antero Koskenniemi, contemplative, pessimistic, and academic; and Otto Manninen, a master of laconic compression and a brilliant translator.

Tulenkantajat—the Torchbearers

In the mid-1920s a group of young writers emerged called Tulenkantajat (Torchbearers), who took as their slogan "Open the windows to Europe!" Through them, free verse, exotic themes, and a new urban romanticism became acclimatized. Their original ideals were realized in the early verse of Katri Vala; at first a prophetess of sensual joys, she later turned to social criticism and socialism. One of the Torchbearers' leaders was an essayist, Olavi Paavolainen, a brilliant travel writer and analyst of the times, who in books published in the 1930s criticized Hitler's Germany and whose last published work was a war diary, *Synkkä yksinpuhelu* (1946; "Sad Soliloquy"). Mika Waltari's achievements were short stories and a novel, *Sinuhe, egyptiläinen* (1945; *Sinhue, the Egyptian,* 1949), set in ancient Egypt but reflecting postwar disillusionment.

Among the chief poets of the years between World Wars I and II were Uuno Kailas and Kaarlo Sarkia, both of whom returned to classical ideals of poetry and traditional metres. The former wrote *Sleep and Death* and upheld a rigid moral code; the latter was a fastidious stylist and sensitive seeker after beauty. Aaro Hellaakoski and P. Mustapaa were recognized as major poets only after World War II; both broadened the traditional style, especially in rhythm.

Leading prose writers included Pentti Haanpaa and Toivo Pekkanen. In his short stories Haanpaa observed the life of the pool, revealing himself as a skillful stylist possessing an individual vein of irony. Pekkanen portrayed the industrial worker; many regarded his account of his own childhood in a working class family, *Lapsuuteni* (1953; "My Childhood"), as a masterpiece.

In the years immediately before World War II many literary trends were discernible: colourful romanticism, depth psychology, bitter social criticism. In 1936 a group of left-wing writers known as Kiila (The Wedge) was formed, most of their important work appearing after the war (*e.g.,* Elvi Sinervo's novel *Viljami Vaihdokas* [1946]). Left-wing ideas were also expressed in the work of Haanpaa and in that of the period's most notable dramatist, Hella Wuolijoki, who collaborated in 1940 and 1941 with Bertolt Brecht in writing a play, *Herr Puntila und sein Knecht Matti* (performed 1948).

World War II marked a period of transition. The generation that began writing before or during the war suffered a crisis survived by few, among them the poets Aaro Hellaakoski, P. Mustapaa (pseudonym of Martti Haavio), Aale Tynni, Viljo Kajava, and Arvo Turtiainen. A school of younger poets soon emerged whose work partook of the free rhythms, lack of rhyme, symbolic imagery, and unpoetical themes of modernism. Its leaders (who included Paavo Haavikko *[Selected Poems,* 1968], also a prose writer and author of experimental and eccentric plays) shared an avoidance of political or religious commitment, an often skeptical outlook, and an interest in problems of lyrical expression. In lyrical poetry the 1950s were a rich and creative period, thanks to such poets as Helvi Juvonen, Eeva-Liisa Manner, Tuomas Anhava *(In the Dark Move Slowly,* 1969), Pentti Holappa, and Lassi Nummi.

A similar development took place, more slowly, in prose. In fiction a restrained, objective style became customary; for example. that of Eila Pennanen and Eeva Joenpelto. Occasionally, as in Antti Hyry's *Kevatta ja syksyä* (1958; "Spring and Autumn") and *Alakoulu* (1965; "Primary School"), this technique resulted in effective prose. (Hyry's characters were depicted behavioristically.) Other writers explored new paths, notably Veijo Meri in his grotesque, Chaplinesque war novels, as *Manillakoysi* (1957; *Manila Rope,* 1967); and Marja-Liisa Vartio, who blended realism and fantasy. A more traditional narrative style was retained by Vaino Linna, author of a war novel, *Tuntematon sotilas* (1954; *The Unknown Soldier,* 1957), and of *Täällä Pohjantähden alla* (1959–62; "Here Beneath the Pole-star"), a long novel of social criticism.

Poets of the 1960s mainly adopted a frank, urbane, conversational style and took a firm stand on social and political questions. The most notable were Pentti Saarikoski and Matti Rossi. In the poems of Anselm Hollo and Vaino Kirstina, emphasis was on linguistic experiment. The most outstanding theatrical work of the 1960s was the *Lapualaisoopera* (1966; "Lapua Opera"), a political musical about the Finnish Fascist movement of the 1930s, by Arvo Salo. In prose, intellectual experiment, sexual romanticism, documentary fiction, traditional narrative styles, and naturalistic treatment of everyday life were obvious features. The most important new writers were Hannu Salama and Alpo Ruuth. (For Finnish literature in Swedish, see above *Scandinavian literature in the 20th century: Finnish literature in Swedish.)*

Modern literature in Hebrew. *Formative influences.* The first formative influences on 20th-century Hebrew literature belong to the late 19th century. The middle classes of eastern European Jewry that read Hebrew books turned to Jewish nationalisms and Zionist activity, coupled with the movement for speaking Hebrew, widened the circle of Hebrew readers. Hebrew daily papers began to appear in 1886. Writers borrowed extensively from medieval translators and European languages, and the Hebrew language assumed a new character. A key figure in the transition to modern writing was S.J. Abramovich, who wrote under the pseudonym of Mendele Mokher Sefarim; after his first novel he became convinced that biblical Hebrew was unsuitable for modern subjects and turned to Yiddish. From 1886 onward he turned first to Yiddish, then later returned to Hebrew and by using Hebrew and Aramaic phrases from the Talmud was able to capture the homeliness he prized in Yiddish. His stories depicted life as it really was, and his style and positive attitude to traditional values attracted a wide readership. The popularity he brought to stories of ghetto life ensured that they would remain the most read and written genre of Hebrew literature until the mid-20th century. A group of talented writers adopted "grandfather Mendele" as their model. One of these, Asher Ginsberg (Aḥad Haʿam), wrote, from 1889 onward, a series of articles evolving a philosophy of secular Jewish nationalism and postulating a "cultural centre" in Palestine. His periodical *ha-Shiloaḥ* attained editorial standards previously unknown in Hebrew. In Tel Aviv from 1921, he devoted his last years to the edition of his correspondence, a valuable documentary of a vital period.

The conflict between Judaism and the natural instincts of the Jews

H.N. Bialik, an important poet, essayist, editor, and anthologist of medieval literature, was, for a time, literary editor of *ha-Shiloaḥ* and was much influenced by Aḥad Ha'am. His poetry expressed the inner struggles of a generation concerned about its attitude to Jewish tradition. Saul Tchernichowsky, on the other hand, was untroubled by tradition, and his poetry dealt with love, beauty, and the three places where he had lived, the Crimea, Germany, and Palestine. I.L. Peretz, who wrote both in Hebrew and in Yiddish, introduced the Ḥasidic, or pietistic devotional, element into literature. The emotionalism and simple joy of life of that milieu thereafter strongly influenced writers, and the language absorbed many Ḥasidic terms. A literary historian, R. Brainin, discerned the presence of a "new trend" in literature and foresaw a concentration on human problems. Bialik had already pointed to a conflict between Judaism and the natural instincts of Jews. This psychological interest dominated the work of a group of short-story writers and, in particular, that of the writer and critic D. Frischmann, who, more than anyone else, imposed European standards on Hebrew literature. European literary tendencies thus became absorbed into Hebrew. Uprooted by the pogroms of 1881 and the two Russian revolutions of 1905 and 1917, Jews had emigrated to western Europe and America, and Hebrew literary activity in eastern Europe was disrupted. The Soviet Union eventually banned Hebrew culture, and it also decayed in other eastern European countries and in Germany as the position of Jews deteriorated.

Émigré and Palestinian literature. The writers of this generation were known as the émigré writers. Their work was pessimistic, as the rootlessness without hope of U.N. Gnessin and Y.Ḥ. Brenner exemplified. The majority of writers active in Palestine before 1939 were born in the Diaspora (Jewish communities outside Palestine) and were nostalgically concerned with the past. An exception was Y. Burla, an Oriental Jew who wrote about Oriental Jewry. The transition from ghetto to Palestine was achieved by only a few writers, among them A. Barash, who described the early struggles of Palestinian Jewry. S.J. Agnon, the outstanding prose writer of this generation (and later a joint winner of the 1966 Nobel Prize for Literature), developed an original narrative style that borrowed from the Midrash (homiletical commentaries on the Hebrew Scriptures), popular stories, and ethical writings of earlier centuries. Whereas his earlier stories were set in Galicia, he began in the 1940s to write about Palestine.

In contrast with prose, poetry immediately began to treat of Palestinian life. Among outstanding writers were Rachel (R. Blovstein), who wrote intensely personal poems, many of which have been set to music; U.Z. Greenberg, a political poet and exponent of free verse; and A. Shlonsky, who led the Symbolist school.

Israeli literature. World War II and the Arab-Israeli War of 1948–49 brought to the fore Palestinian-born writers who dealt with the problems of their generation in colloquially flavoured Hebrew. In the State of Israel literature developed on a large scale, mainly along contemporary western European and American lines. The extreme diversity in culture of parts of the population, and the problems of new immigrants, provided the main theme for fiction. Poetry flourished, but original drama was slow to develop. Greenberg's *Rehoboth HaNahar* (1951) traced the process by which the humiliation of the massacred is transmuted by the pride of martyrdom into the historical impulse of messianic redemption. In a long dramatic poem (*Between the Fire and Salvation*, 1957) Aaron Zeitlin envisioned the annihilation of European Jewry in mystical terms, examining the relationship of catastrophe and redemption.

Native Israeli prose writers wrote of their life in the kibbutz, the underground, and the war of 1948–49. S. Yizhar and Moshe Shamir emerged as the outstanding representatives of this generation, probing the sensibility of the individual in a group-oriented society. But the establishment of the State of Israel could not allay the anxieties of the individual. The dominant themes of writers who had no access to collective ideals were personal ones—frustration, confusion, and alienation. The later works of Yizhar and Shamir emphasize the dissolution of social coherence and portray people without a sense of historical mission. Aharon Meged's Surrealist novel *The Case of the Simpleton* (1960) traces a picaresque progress through a dehumanized society. In *Living on the Dead* (1965) a hero of the pioneer generation is cast in an ironic light.

Memories of the holocaust haunt the lyrical work of Aaron Applefeld. Flight and hiding are the characteristic situations of his stories. His novel *The Skin and the Gown* (1971) portrays survivors of European prison camps who are unable to exorcise the past. In the stark stories of Abraham B. Yehoshua the tone of the narrator is remote, the people are isolated and drained of emotion. Only occasionally in a solitary forest, desert, or village does an act of feeling or meaning break the mood of boredom and light up the humanity of a character. Personal frustration and religious vision are the subjects of Pinḥas Sadeh's work. In his novel *Notes on the Conditions of Man* (1967) the misery of social failure and spiritual beauty alternate. *My Michael* (1968) by Amos Oz is a study of a young housewife who narrates her own story of psychological disintegration. Amalia Kahane-Carmon explored the subjective impressions of experience and the complexities of time and memory. She used stream of consciousness techniques to reveal the intricate inner life of quite ordinary people.

ADVISERS, CONTRIBUTORS, AND REVIEWERS: (J.A.A. / G.A. / W.C.A. /Ro.Au. / H.-G.B. / E.Be. / B.S.B. / S.Be. / A.By. / Wa.B. / E.L.B. / L.B. / G. Car. / P.-G.Ca. / C.Ck. / W.W.C./ N.Ds. / F.Dr. / F.Do. / M.S.D. / D.G.D. / J.E.E. / J.Ep. / M.V.F. / A.Gi. / G.P.Gi. / D.Gr. / N.K.G. / A.Gn. / E.v.H. / J.Hy. / G.W.H. / T.Jo. / M.E.J. / P.K.K. / T.Kl. / L.J.K. / S.A.Kr. / J.Kr. / K.L.K.L. / N.J.L. / D.Le. / S.Lr. / R.F.Li. / D.W.L. / J.W.McF. / Da.Ma. / E.G.M. / B.Mo. / W.G.Mo. / R.P.C.M. / Ch.R. / S.Ra. / H.S.R. / C.A.Rn. / A.Rs. / F.Se. / D.Se. / R.Sh. / M.F.R.S. / E.J.Si. / D.Si. / Em.Š. / J.S.T. / D.S.T. / C.A.Tr. / E.O.G.T.-P. / V.U. / K.Vy. / E.Vi. / E.Wk. / J.H.Ws. / D.M.Wh. / F.Wh. / D.W. / J.R.We. / B.W. / J.E.C.W. / Ja.W. / S.G.W. / V.Z. / O.E.Z.)

BIBLIOGRAPHY

General introductory works: FERNAND BALDENSPERGER and WERNER P. FRIEDERICH, *Bibliography of Comparative Literature* (1950, reprinted 1960), comprehensive coverage of literatures, classified by author, country, and according to genre, theme, etc.; ANTONY BRETT-JAMES, *The Triple Stream: Four Centuries of English, French, and German Literature, 1531–1930* (1953), contains useful comparative tables arranged chronologically; JOHN M. COHEN, *A History of Western Literature*, rev. ed. (1963), a useful factual and critical account; HORATIO SMITH (gen. ed.), *Columbia Dictionary of Modern European Literature* (1947), contains over 1,000 articles, including surveys, critical essays, and biographies, with bibliographies; WILLIAM F. THRALL and ADDISON HIBBARD, *A Handbook to Literature*, rev. by C. HUGH HOLMAN (1960), good coverage of literary terms and movements; RENE WELLEK, *A History of Modern Criticism: 1750–1950*, 4 vol. (1955–), surveys theory in the major European nations since the Renaissance.

English and Anglo-Irish literature: F.W. BATESON (ed.), *The Cambridge Bibliography of English Literature*, 4 vol. (1941), *Supplement*, ed. by GEORGE WATSON (1957); *New Cambridge Bibliography of English Literature*, vol. 3 (1969) and vol. 2 (1971) ed. by GEORGE WATSON, vol. 4 (1972) ed. by I.R. WILLISON; ALBERT C. BAUGH (ed.), *A Literary History of England*, 2nd rev. ed. (1967; paperback 4-vol. ed., 1967), a standard work; DAVID DAICHES, *A Critical History of English Literature*, 2 vol. (1960; paperback 4-vol. ed., 1968); BORIS FORD (ed.), *The Pelican Guide to English Literature*, rev. ed., 7 vol. (1961–64), stimulating, with a critical rather than historical emphasis, containing helpful bibliographies; SIR PAUL

HARVEY (ed.), *The Oxford Companion to English Literature,* 4th rev. ed. (1967), a basic reference work; GEORGE SAMPSON (ed.), *The Concise Cambridge History of English Literature,* 3rd rev. ed. (1970), to some extent correcting and supplementing *The Cambridge History of English Literature,* 15 vol. (1907–16, 1932), the most important general history of the literature, largely superseded but still useful for some periods; F.P. WILSON and BONAMY DOBREE (eds.), *The Oxford History of English Literature,* projected 15 vol. (1947–); STEPHEN L. GWYNN, *Irish Literature and Drama in the English Language* (1936).

Celtic literatures: DOUGLAS HYDE, *A Literary History of Ireland from Earliest Times to the Present Day,* new ed., introd. by BRIAN O. CUIV (1967), a pioneer work; KURT WITTIG, *The Scottish Tradition in Literature* (1958), perhaps the best general guide; THOMAS PARRY, *Hanes Llenyddiaeth Gymraeg hyd 1900,* 3rd ed. (1953; Eng. trans., *A History of Welsh Literature,* 1955), the only modern and authoritative history; H.M. and N.K. CHADWICK, *The Ancient Literatures of Europe,* vol. 1 (1932), includes a comprehensive treatment of Celtic literature.

English literature in the United States, Canada, Australia, and New Zealand: ROBERT E. SPILLER *et al.* (eds.), *Literary History of the United States,* 3rd ed., 2 vol. (1963), a standard general work; MARCUS CUNLIFFE, *The Literature of the United States,* 3rd ed. (1967); VERNON L. PARRINGTON, *Main Currents in American Thought,* 3 vol. (1927–30; reprinted in 1 vol., 1958), essential background reading; ARTHUR H. QUINN (ed.), *The Literature of the American People* (1951); WALTER BLAIR, THEODORE HORNBERGER, and RANDALL STEWART (eds.), *The Literature of the United States,* 3rd ed., 2 vol. (1966); CARL F. KLINCK (ed.), *Literary History of Canada: Canadian Literature in English* (1965); HENRY M. GREEN, *A History of Australian Literature, Pure and Applied, 2* vol. (1961); ERIC H. MCCORMICK, *New Zealand Literature* (1959).

French and Provencal literature: GEOFFREY BRERETON, *A Short History of French Literature* (1954), a good brief survey with basic bibliography; LOUIS F. CAZAMIAN, *A History of French Literature* (1955), a well-written history containing independent and discriminating evaluations and short biographies; PATRICK E. CHARVET (gen. ed.), *A Literary History of France, 6* vol. (1967–70); and JOHN CRUIKSHANK (ed.), *French Literature and Its Background,* 6 vol. (1968–70), two thorough historical works; W.G. MOORE, *The French Achievement in Literature* (1969); DANIEL C. HASKELL, *Provencal Literature and Language, Including the Local History of Southern France* (1925), with bibliographies.

Italian literature: JOHN H. WHITFIELD, *A Short Hzstory of Italian Literature* (1960), useful, reliable, and readable; FRANCESCO DE SANCTIS, *Storia della letteratura italiana a cura di Benedetto Croce,* 4th ed., 2 vol. (1959; Eng. trans., *History of Italian Literature,* 2 vol., 1960); ROBERT ANDERSON HALL, *A Short History of Italian Literature* (1951); ERNEST HATCH WILKINS, *A History of Italian Literature* (1954).

Spanish, Portuguese, and Ibero-American literature: RICHARD E. CHANDLER and KESSEL SCHWARTZ, *A New History of Spanish Literature* (1961); R.O. JONES (gen. ed.), *A Literary History of Spain,* 8 vol. (1971–72), rapidly becoming a standard work—very readable; GEORGE T. NORTHUP, *An Introduction to Spanish Literature,* 3rd ed. rev. by NICHOLSON B. ADAMS (1960), a good readable history; AUBREY F. BELL, *Portuguese Literature* (1922, reprinted 1970), a standard work, with extensive bibliography; ARTHUR TERRY, *Catalan Literature* (1972); H.V. LIVERMORE and W.J. ENTWISTLE (eds.), *Portugal and Brazil: An Introduction* (1953), contains a literary bibliography of studies and of translations from Portuguese to English; ALFRED COESTER, *The Literary History of Spanish America,* 2nd ed. (1928, reprinted 1971); JEAN FRANCO, *An Introduction to Spanish-American Literature* (1969), a substantial guide to writings, from colonial times to the present.

German, Swiss, Dutch, and Afrikaans literature: JETHRO BITHELL, *Germany: A Companion to German Studies,* rev. 5th ed. (1955), a collection of articles by various authors—good background reading; JOHN G. ROBERTSON, *A History of German Lzterature,* 6th ed. (1970), an essential reference work providing easy access to dates and facts; GILBERT WATERHOUSE, continued by H.M. WAIDSON, *A Short Hrstory of German Literature,* 3rd ed. rev. (1959), the best brief outline; ALEX NATAN (ed.), *Swiss Men of Letters* (1970), 12 essays covering all facets of the literary scene; THEODOOR WEEVERS, *Poetry of the Netherlands in Its European Context, 1170–1930* (1960); C.M. VAN DEN HEEVER and THEODORE J. HAARHOFF, *The Achievement of Afrikaans* (1934); THEODORE J. HAARHOFF, *Afrikaans: Its Origin and Development* (1936).

Scandinavian literature: ELIAS L. BREDSDORFF, BRITA MORTENSEN, and RONALD POPPERWELL, *An Introduction to Scandinavian Literature, from the Earliest Times to Our Day* (1951), a useful short survey; HARALD BEYER, *Norsk litteratur historie* (1952; Eng. trans., *A History of Norwegian Literature,* 1956), probably the best survey in English; ALRIK GUSTAFSON, *A History of Swedish Literature* (1961), the best critical history in English, with bibliographical appendix; PHILLIP M. MITCHELL, A *History of Danish Literature* (1957); STEFAN EINARSSON, *A History of Islandic Literature* (1957).

Slavic and Eastern European literatures: MAURICE B. LINE, *A Bibliography of Russian Literature in English Translation to 1900 (Excluding Periodicals)* (1963); WILLIAM EDWARD HARKINS, *Dictionary of Russian Literature* (1956, reprinted 1971); HELEN MUCHNIC, *An Irztroduction to Russian Literature,* rev. ed. (1964); D.P. MIRSKY, *A History of Russian Literature, Comprising a History of Russian Literature and Contemporary Russian Literature,* ed. by FRANCIS J. WHITFIELD (1949), annotated and updated—thorough and readable, but the author's critical judgments are often highly personal; MARK L. SLONIM, *An Outline of Russian Literature* (1958), a good introduction, with bibliographical notes; JANIS ANDRUPS and VITAUTS KALVE, *Latvian Literature* (1954); ANTANAS VAICIULAITIS, *Outline History of Lithuanian Literature* (1942); GREGORY LUZNYCKY, *Ukrainian Literature Within the Framework of World Literature: A Short Outline of Ukrainian Literature from Renaissance to Romanticism* (1961); ARVO MAGI, *Estonian Literature* (1968); CZESLAW MILOSZ, *The History of Polish Literature* (1969), an excellent work by an outstanding Polish poet and literary historian; ELEANOR EDWARDS LEDBETTER, *Polish Literature in English Translation* (1932); MANFRED KRIDL, *A Survey of Polish Literature and Culture* (1956); FRANTISEK CHUDOBA, *A Short Survey of Czech Literature* (1924, reprinted 1969); PAUL SELVER (ed.), *An Anthology of Czechoslovak Literature* (1929, reprinted 1969); D. SHISHMANOV, *A Survey of Bulgarian Literature* (1932); ANTUN BARAC, *Jugoslavenska književnost* (1954; Eng. trans., A *History of Yugoslav Literature,* 1955); HORACE G. LUNT, *A Survey of Macedonian Literature* (1953); PAVLE POPOVIC, *The Literature of the Southern Slavs* (1917); JANKO LAVRIN and ANTON SLODNJAK, *The Parnassus of a Small Nation: An Anthology of Slovene Lyrics,* 2nd ed. (1965), English translations, with introduction; JACOB STEINBERG (ed.), *Introduction to Rumanian Literature* (1966), selections from representative prose writings by 24 authors; TIBOR KLANICZAY, JOSEPH SZAUDER, and MIKLOS SZABOLCSI, *History of Hungarian Literature* (1964); ILMARI HAVU (ed.), *An Introduction to Finnish Literature: A Book of Readings* (1952), translations, with introduction and notes.

Armenian, Syriac, and Hebrew literature: E. ARCH. DOURIAN, *The History of Armenian Literature* (1933); WILLIAM WRIGHT, *A Short History of Syriac Literature* (1894); MEYER WAXMAN, *A History of Jewish Literature,* 2nd ed., 5 vol. (1960), excellent coverage of Yiddish writing as well as Hebrew.

Ancient Hebrew, Greek, and Latin literature: JULIUS A. BEWER, *The Literature of the Old Testament in Its Historical Development,* 3rd ed. (1962); SANFORD C. YODER, *Poetry of the Old Testament* (1948), a useful introduction and representative collection of biblical poetry; M.J. ANDERSON (ed.), *Classical Drama and Its Influence* (1965), a collection of essays by leading scholars, covering aspects of classical drama and European theatre; J.W.H. ATKINS, *Literary Criticism* in *Antiquity,* vol. 2, *Graeco-Roman* (1934, reprinted 1961), perhaps somewhat out of date, but a worthwhile comprehensive study; MARGARETE BIEBER, *The History of the Greek and Romatt Theatre,* 2nd rev. ed. (1961); GILBERT HIGHET, *The Classical Tradition* (1949); and R.R. BOLGAR, *The Classical Heritage and Its Beneficiaries* (1954), two standard works that trace the posthumous influence of classical authors; SIR PAUL HARVEY (ed.), *The Oxford Companion to Classical Literature* (1937), a basic reference work; MAURICE PLATNAUER (ed.), *Fifty Years of Classical Scholarship* (1954), contains valuable bibliographies; C.M. BOWRA, *Greek Lyric Poetry from Alcman to Simonides,* 2nd ed. (1961), controversial but perceptive; see also his *Landmarks in Greek Literature* (1966); ROBERT FLACELIERE, *Histoire littéraire de la Grèce* (1962; Eng. trans., *A Literary History of Greece,* 1964), a systematic, readable survey of ancient literature; ROBERT GRAVES, *The Greek Myths,* 2 vol. (1957), a learned work, notable for the author's speculative theories; MOSES HADAS, *A History of Greek Literature* (1950), a useful account; GILBERT MURRAY, A *History of Ancient Greek Literature* (1897, reprinted 1966), a valuable general guide in spite of its date; HERBERT J. ROSE, *A Handbook of Greek Literature from Homer to the Age of Lucian,* 4th ed. (1950); and A *Handbook of Latin Literature from the Earliest Times to the Death of St. Augustine,* 3rd ed. (1954), two standard works; WILLIAM BEARE, *The Roman Stage,* 3rd rev. ed. (1964), an authoritative work on drama during the Republic; MICHAEL GRANT, *Roman Literature*

(1954), a good short account; William A Laidlaw, *Latin Literature* (1951), a useful introduction; John Wight Duff, *A Literary History of Rome,* 2nd–3rd ed., 2 vol., ed. by A.M. Duff (1960–64), standard introduction containing comprehensive and scholarly surveys, with supplementary bibliographies.

The Middle Ages: H.M. Chadwick, *The Heroic Age* (1912, reprinted 1967), a comparative study covering Greek, Teutonic, Slavonic, and Celtic literatures and cultures; E.K. Chambers, *The Mediaeval Stage,* 2 vol. (1903), a standard reference work; W.T.H. Jackson, *The Literature of the Middle Ages* (1960), includes a useful bibliography; Charles W. Jones (ed.), *Medieval Literature in Translation* (1950), a good, readable anthology; C.S. Lewis, *The Allegory of Love* (1936, reprinted 1958), a critical study of courtly love in medieval literature; Roger S. Loomis (ed.), *Arthurian Literature in the Middle Ages* (1959), an excellent survey.

Medieval Latin literature: Ernst R. Curtius, *Europäische Literatur und lateinisches Mittelalter* (1948; Eng. trans., *European Literature and the Latin Middle Ages,* 1953), an important study; Charles H. Haskins, *The Renaissance of the Twelfth Century* (1957), a standard work; Maurice Helin, *Littérature d'Occident: histoire des lettres latines du moyen âge* (1943; Eng. trans., *A History of Medieval Latin Literature,* rev. ed., 1949); M.L.W. Laistner, *Thought and Letters in Western Europe, A.D. 500 to 900,* new ed. (1957); Frederick A. Wright and T.A. Sinclair, *A History of Later Latin Literature from the Middle of the Fourth to the End of the Seventeenth Century* (1931), a very useful survey.

French literature: Jessie Crosland, *Medieval French Literature* (1956); Mary D. Legge, *Anglo-Norman Literature and Its Background* (1963).

English literature: Stanley B. Greenfield, *A Critical History of Old English Literature* (1965); William P. Ker, *Medieval English Literature* (1912, reprinted 1962), an established useful introduction; William L. Renwick and H. Orton, *Introductions to English Literature,* vol. 1, *The Beginnings of English Literature to Skelton, 1509,* 2nd rev. ed. (1952), a standard work; Margaret Schlauch, *English Medieval Literature and Its Social Foundations* (1956, reprinted 1971), an excellent critical introduction; Kenneth Sisam, *Studies in the History of Old English Literature* (1953), a good basic critical work; John E. Wells, *A Manual of the Writings in Middle English, 1050–1400* (1916; nine suppl., 1919–52), a basic work of reference, with comprehensive lists, summaries, and descriptions; Dorothy Whitelock, *The Beginnings of English Society,* ch. 9–10 (1952), very useful background reading; Richard M. Wilson, *Early Middle English Literature,* 3rd ed. (1968), a thorough work, and *Lost Literature of Medieval England,* 2nd ed. (1970), a classic in its field; Charles L. Wrenn, *A Study of Old English Literature* (1967), an excellent survey; David M. Zesmer, *Guide to English Literature from Beowulf Through Chaucer and Medieval Drama* (1961); St. John D. Seymour, *Anglo-Irish Literature, 1200–1582* (1929).

Celtic literature: Eleanor Knott and Gerard Murphy, *Early Irish Literature* (1966).

German literature: John Knight Bostock, *A Handbook on Old High German Literature* (1955), an essential reference work; Maurice O'Connell Walshe, *Medieval German Literature* (1962), a very useful reference work; Paul Salmon, *Literature in Medieval Germany* (1967), the best English introduction to the subject.

Scandinavian literature: Gabriel Turville-Petre, *The Heroic Age of Scandinavia* (1951) and *Origins of Icelandic Literature* (1953).

Italian literature: Adolf R. Gaspary, *Geschichte der italienischen Literatur* (1885; Eng. trans. of part of vol. 1, *The History of Early Italian Literature to the Death of Dante,* 1901).

Spanish literature: Otis H. Green, *The Literary Mind of Medieval and Renaissance Spain* (1970), 12 essays by an eminent scholar; A.D. Deyermond, *A Literary History of Spain,* vol. 1, *The Middle Ages* (1971), a very good introduction, with bibliography.

Greek literature: F.H. Marshall and John Mavrogordatol, "Byzantine Literature," in N.H. Baynes and H. St. L.B. Moss, *Byzantium* (1948).

Hebrew literature: Abraham E. Millgram (ed.), *An Anthology of Medieval Hebrew Literature* (1961).

Russian literature: Dmitri Chizhevsky, *History of Russian Literature, from the Eleventh Century to the End of the Baroque* (1960); N.K. Gudzii, *History of Early Russian Literature* (Eng. trans. of 2nd Russian ed., 1949, reprinted 1970).

The Renaissance: Ernst Cassirer, Paul O. Kristeller, and John H. Randall (eds.), *The Renaissance Philosophy of Man* (1948), a collection of excerpts from the writings of Renaissance philosophers; Joseph A. Mazzeo, *Renaissance and Revolution: The Remaking of European Thought* (1966), useful background reading; Henry O. Taylor, *Thought and Expression in the Sixteenth Century, 2* vol. (1920, reprinted 1959), a good standard work.

Latin literature: Roberto Weiss, *The Dawn of Humanism in Italy* (1947) and *Humanism in England During the Fifteenth Century* (1941), among the best scholarly works on the subject.

Italian literature: Bernard Weinberg, *A History of Literary Criticism in the Italian Renaissance,* 2 vol. (1961), an indispensable work; Jefferson B. Fletcher, *Literature of the Italian Renaissance* (1934); John A. Symonds, *Renaissance in Italy,* 7 vol. (1875–86), later reprints worthwhile and interesting, but should be supplemented by more recent historical works.

French literature: Kathleen Butler, *A History of French Literature,* 2 vol. (1923, reprinted 1966); Arthur A. Tilley, *The Literature of the French Renaissance,* 2 vol. (1904) and *Studies in the French Renaissance* (1922); Bernard Weinberg (ed.), *French Poetry of the Renaissance* (1954), good background reading.

Spanish literature: Otis H. Green, *Spain and the Western Tradition: The Castilian Mind in Literature from El Cid to Calderon,* 4 vol. (1963–66), readable and authoritative; R.O. Jones, *A Literary History of Spain,* vol. 3, *The Golden Age: Prose and Poetry* (1971), an excellent introduction; Edward M. Wilson and Duncan Moir, *The Golden Age: Drama* (1971), essential reading.

English and Anglo-Irish literature: Muriel C. Bradbrook, *The Growth and Structure of Elizabethan Comedy* (1955) and *Themes and Conventions of Elizabethan Tragedy* (1935, reprinted 1957), two of the best critical works on the drama of the period; Douglas Bush, *Mythology and the Renaissance Tradition in English Poetry,* 2nd ed. (1963), an important study; Patrick Cruttwell, *The Shakespearian Moment and Its Place in the Poetry of the 17th Century* (1954), a study of thought and expression at the turn of the 16th century—not limited to Shakespeare; Madeleine Doran, *Endeavours of Art: A Study of Form in Elizabethan Drama* (1954), a major critical work on the drama; Una M. Ellis-Fermor, *The Jacobean Drama,* 4th ed. (1961), an excellent critical work; C.S. Lewis, *English Literature in the Sixteenth Century, Excluding Drama* (1954), probably the best volume in the "Oxford History of English Literature" series and an excellent survey; Frank P. Wilson, *Elizabethan and Jacobean* (1946), treats various literary topics; St. John D. Seymour, *Anglo-Irish Literature, 1200–1582* (1929).

German literature: Archer Taylor, *Problems in German Literary History of the Fifteenth and Sixteenth Centuries* (1939).

The seventeenth century: Mario Praz, *Studi sul concettismo,* 2 vol. (1934–46; Eng. trans., *Studies in Seventeenth Century Imagery,* 2 vol., 1939–48), a standard and readable work; Joel E. Spingarn (ed.), *Critical Essays of the Seventeenth Century,* 3 vol. (1909, reprinted 1957), a valuable collection with a long critical introduction.

English and American literature: Cleanth Brooks, *The Well-Wrought Urn* (1947, reprinted 1963), close studies of works by several writers by a leading critic; Douglas Bush, *English Literature in the Earlier Seventeenth Century, 1600–1660,* 2nd rev. ed. (1962), a standard work with an excellent bibliography; H.V.D. Dyson and J.E. Butt, *Augustans and Romantics, 1689–1830,* 3rd rev. ed. (1961), a standard work with introductions with annotated bibliography; T.S. Eliot, *Selected Essays,* 3rd ed. (1951), essential for Renaissance and 17th-century studies; H.J.C. Grierson, *Cross Currents in English Literature of the XVIIth Century* (1929; reprinted as *Cross-currents in 17th Century English Literature,* 1958), essential reading; Victor I. Harris, *All Coherence Gone* (1949, reprinted 1966), on the philosophical background; Richard P. Jones *et al., The Seventeenth Century: Studies in the History of English Thought and Literature from Bacon to Pope* (1965), a good study of the philosophical background and of attitudes to style; Louis L. Martz, *The Poetry of Meditation* (1954), an interesting critical work on metaphysical and Baroque poetry; Marjorie H. Nicolson, *Science and Imagination* (1956); James R. Sutherland, *English Literature of the Late Seventeenth Century* (1969), contains a good bibliography; E.M.W. Tillyard, *The Elizabethan World Picture* (1943, reprinted 1966), somewhat old-fashioned in approach, but good background reading; Rosemond Tuve, *Elizabethan and Metaphysi-*

cal Imagery (1947), the best work on the subject; CICELY V. WEDGWOOD, *Seventeenth-Century English Literature,* rev. ed. (1970), a short survey; BASIL WILLEY, *The Seventeenth Century Background* (1934), an excellent study of the philosophical background; PERRY G. MILLER, *The New England Mind: From Colony to Province* (1953) and *The New England Mind: The Seventeenth Century* (1939, reprinted 1961), authoritative.

French literature: ANTOINE ADAM, *Grandeur and Illusion: French Literature and Society, 1600–1715* (Eng. trans. 1972), a comprehensive and penetrating history of the literature of the period; NATHAN EDELMAN (ed.), *A Critical Bibliography of French Literature,* vol. 3, *The Seventeenth Century* (1961); JEAN JACQUES DEMOREST (ed.), *Studies in Seventeenth-Century French Literature* (1962); WILLIAM D. HOWARTH, *Life and Letters in France,* vol. 1, *The Seventeenth Century* (1965); HENRY C. LANCASTER, *A History of French Dramatic Literature in the Seventeenth Century,* 9 vol. (1929–42, reprinted 1966); PHILIP J. YARROW, *A Literary History of France,* vol. 2, *The Seventeenth Century, 1600–1715* (1967), very good background reading.

German literature: ROY PASCAL, *German Literature in the Sixteenth and Seventeenth Centuries* (1968), the best English introduction to this period.

The eighteenth century: LILIAN R. FURST, *Romanticism in Perspective* (1969), on the Romantic movement in France, England, and Germany; JOHN B. HALSTED (ed.), *Romanticism* (1969), a collection of extracts and key essays by the leading figures of Western European Romanticism, with a long introduction and chronological table; PAUL HAZARD, *La Pensée européenne au XVIIIème siècle* (1963; Eng. trans., *European Thought in the Eighteenth Century,* 1963); DANIEL MORNET, *The Development of Literature and Culture in the XVIIIth Century* (1954), an extensive survey; MARIO PRAZ, *La carne, la morte e il diavolo nella letteratura romantica* (1930; Eng. trans., *The Romantic Agony,* 2nd ed., 1970), a remarkable work, now a classic in this field; JOHN G. ROBERTSON,, *Studies in the Genesis of Romantic Theory in the Eighteenth Century* (1923, reprinted 1962).

English, Scottish, and American literature: JAMES L. CLIFFORD (ed.), *Eighteenth-Century English Literature: Modern Essays in Criticism* (1959), 21 recent articles; H.V. DYSON and J.E. BUTT, *Augustans and Romantics, 1689–1830,* 3rd rev. ed. (1961), introductions with annotated bibliographies; ARTHUR R. HUMPHREYS, *The Augustan World: Life and Letters in Eighteenth-Century England* (1954); SAMUEL JOHNSON, *Lives of the English Poets,* 4 vol. (1781; standard ed. by GEORGE BIRKBECK HILL, 3 vol., 1905, reprinted 1968), still one of the best critical works on the period; MARJORIE H. NICOLSON, *Newton Demands the Muse* (1946), on the effects of science on literature; BONAMY DOBREE, *English Literature in the Early Eighteenth Century, 1700–1740* (1959), contains full bibliography; GEORGE SHERBURN and D.F. BOND, *The Restoration and Eighteenth Century, 1660–1789,* rev. ed. (1967), a very good handbook; IAN P. WATT, *The Rise of the Novel* (1957), a good introduction, mainly on Defoe, Richardson, and Fielding; BASIL WILLEY, *The Eighteenth-Century Background* (1941), an essential background work; DAVID CRAIG, *Scottish Literature and the Scottish People, 1680–1830* (1961), a good, useful survey; MOSES C. TYLER, *A History of American Literature, 1607–1765,* 2 vol. (1879), and *The Literary History of the American Revolution, 1763–1783,* 2 vol. (1897, both reprinted 1966), useful histories of the Colonial Period and early Republic.

French literature: ERNST CASSIRER, *Die Philosophie der Aufklärung* (1932; Eng. trans., *The Philosophy of the Enlightenment,* 1951); LESTER G. CROCKER, *An Age of Crisis* (1959); GEORGE R. HAVENS, *The Age of Ideas* (1955), an interesting account of the period, studied through biography; ROBERT NIKLAUS, *A Literary History of France,* vol. 3, *The Eighteenth Century, 1715–1789* (1970), a good general work.

German literature: WALTER H. BRUFORD, *Germany in the Eighteenth Century: The Social Background of the Literary Revival* (1935, reprinted 1959), a thorough, commendable work; HENRY B. GARLAND, *Storm and Stress* (1952), an adequate introduction; WILLIAM D. ROBSON-SCOTT, *The Literary Background of the Gothic Revival in Germany: A Chapter in the History of Taste* (1965), on the history of art as well as of literature; ERNST L. STAHL and W.E. YUILL, *German Literature of the Eighteenth and Nineteenth Centuries* (1970), the best introduction in English.

Italian literature: VIOLET PAGET, *Studies of the Eighteenth Century in Italy,* 2nd ed. (1907).

Spanish Literature: ROBERT E. PELLISSIER, *The Neo-Classic Movement in Spain During the XVIIIth Century* (1918); NIGEL GLENDINNING, *A Literary History of Spain,* vol. 4, *The Eighteenth Century* (1972), a very good survey.

The nineteenth century: GEORG BRANDES, *Creative Spirits of the Nineteenth Century* (1923, reprinted 1967); BENEDETTO CROCE, *Poesia e non poesia,* 7th ed. (1964; Eng. trans., *European Literature in the Nineteenth Century,* 1924, reprinted 1967); JANKO LAVRIN, *Aspects of Modernism: From Wilde to Pirandello* (1935, reprinted 1968); JOHN LUCAS (ed.), *Literature and Politics in the Nineteenth Century* (1971).

English and American literature: MEYER H. ABRAMS, *The Mirror and the Lamp: Romantic Theory and the Critical Tradition* (1953), a study of English Romantic critical theory with attention to its Germanic roots; RICHARD D. ALTICK, *Victorian People and Ideas* (1972); O. ELTON, *A Survey of English Literature 1780–1830,* 2 vol. (1912), and *. . . 1830–1880,* 2 vol. (1920), deals with writers separately, rather than in a general survey; WALTER E. HOUGHTON, *The Victorian Frame of Mind, 1830–1870* (1957), a useful survey; FRANK R LEAVIS, *Revaluation: Tradition and Development in English Poetry* (1936, reprinted 1963), a major work of criticism; *The Oxford History of English Literature* devotes three volumes to the period: vol. 9, *1789–1815* (1963); vol. 10, *1815–1832* (1964); vol. 11, on the Victorians (in press), helpful as surveys but of little critical interest; G. WILSON KNIGHT, *The Starlit Dome* (1941), an excellent critical work on the Romantics; GEORGE M. YOUNG, *Victorian England: Portrait of an Age,* 2nd ed. (1960), a standard work; FRANCIS O. MATTHIESSEN, *American Renaissance* (1941, reprinted 1968), an essential work on the major writers of the period; JAY MARTIN, *Harvests of Change: American Literature, 1865–1914* (1967), a useful study; ARTHUR H. QUINN, *A History of the American Drama from the Beginning to the Civil War* (1923) and *. . . From the Civil War to the Present Day,* rev. ed. (1936, reprinted 1964), the most thorough treatment; MARIO PRAZ, *La crisi dell'eroe nel romanzo vittoriano* (1952; Eng. trans., *The Hero in Eclipse in Victorian Fiction,* 1956).

French literature: PATRICK E. CHARVET, *A Literary History of France,* vol. 4, *The Nineteenth Century, 1789–1870* (1967), a good and readable introduction; MARCEL RAYMOND, *De Baudelaire au Surréalisme,* rev. ed. (1947; Eng. trans., *From Baudelaire to Surrealism,* rev. ed., 1957), the best critical treatment of this period available in English; ALBERT THIBAUDET, *Histoire de la littérature française de 1789 à nos jours* (1938), the best critical history.

German literature: GEORG BRANDES, *Hovedstrømninger i det 19de aarhundredes Litteratur,* vol. 6, *Det unge Tyskland* (1872–90; Eng. trans., *Main Currents in 19th-Century Literature,* vol. 6, *The Young Germany,* 1901–05), a thorough and basic work on European comparative literature; RONALD D. GRAY, *The German Tradition in Literature, 1871–1945* (1965), an interesting study with particular reference to the more recent authors, Thomas Mann and Rainer Maria Rilke.

Italian literature: ANTONIO CIPPICO, *The Romantic Age in Italian Literature* (1918).

Spanish literature: DONALD L. SHAW, *A Literary History of Spain,* vol. 5, *The Nineteenth Century* (1972), a very good survey; EDGAR A. PEERS, *A History of the Romantic Movement in Spain,* 2 vol. (1940; abridged ed., 1950), a comprehensive account.

Scandinavian literature: BRIAN W. DOWNS, *Modern Norwegian Literature, 1860–1918* (1966), the first book in English about the "classic" period—an excellent survey, with bibliography; RICHARD BECK. *History of Icelandic Poets, 1800–1940* (1950); STEFAN EINARSSON, *History of Icelandic Prose Writers, 1800–1940* (1948).

Russian and Polish literature: M.I. OLGIN, *A Guide to Russian Literature, 1820–1917* (1920, reprinted 1971); JULIAN KRZYZANOWSKI, *Polish Romantic Literature* (1930).

Hebrew literature: JOSEPH KLAUSNER, *A History of Modern Hebrew Literature, 1788–1930* (1932; orig. pub. in Hebrew, 1920); JACOB S. RAISIN, *The Haskalah Movement in Russia* (1913).

The twentieth century: ERICH AUERBACH, *Mimesis* (1946; Eng. trans., 1953), traces the development of Realism from ancient to modern times; CLEANTH BROOKS and ROBERT PENN WARREN (eds.), *Understanding Poetry.* 3rd ed. (1960), and *Understanding Fiction,* 2nd ed. (1959); and CLEANTH BROOKS and ROBERT B. HEILMAN, *Understanding Drama,* rev. ed. (1948), detailed anthologies, with explications of texts—invaluable introductions, especially to the study of modern writing, though not limited to 20th-century literature; GYORGY LUKACS, *Studies in European Realism* (1950), an important work by a major Marxist critic; RENE WELLEK and AUSTIN WARREN, *Theory of Literature,* 3rd ed. (1963), a general and theoretical work, especially useful as a guide to modern literature; CLAUDE MAURIAC, *L'Alittérature contemporaine* (1958; Eng. trans., *The New Literature,* 1959); STEPHEN SPENDER, *The Struggle of the Modern* (1963); EDMUND WILSON, *Axel's Castle* (1931), and *The Triple Thinkers,* rev. ed. (1948).

English, Australian, South African, and American literature: WALTER E. ALLEN, *Tradition and Dream* (1964), covers post-1920 fiction; JEROME H. BUCKLEY, *The Victorian Temper* (1951), a thorough critical survey; see also its sequel, *The Triumph of Time* (1966); DAVID DAICHES, *The Present Age After 1920,* rev. ed. (1958), a useful handbook; RICHARD ELLMANN and CHARLES FEIDELSON (eds.), *The Modern Tradition: Backgrounds of Modern Literature* (1965), discussions of the principal literary figures; GEORGE S. FRASER, *The Modern Writer and His World,* rev. ed. (1964), an extremely good critical guide; FRANK R. LEAVIS, *New Bearings in English Poetry,* rev. ed. (1950), a stimulating introduction; JOHN R. TAYLOR, *Anger and After,* rev. ed. (1963), deals with contemporary British drama; ALEC D. HOPE, *Australian Literature, 1950–1962* (1963), a good survey by a leading Australian poet; NADINE GORDIMER and LIONEL ABRAHAMS (eds.), *South African Writing Today* (1967); ALFRED KAZIN, *On Native Grounds: An Interpretation of Modern American Prose Literature,* rev. ed. (1956), a readable, useful critical history; HEINRICH STRAUMANN, *American Literature in the Twentieth Century,* 3rd ed. (1965); WILLARD THORP, *American Writing in the Twentieth Century* (1960); PAUL WEST, *The Modern Novel,* 2nd ed., 2 vol. (1965), a survey of both European and American writing since 1900; MORTON D. ZABEL (ed.), *Literary Opinion in America,* rev. ed. (1951).

French literature: JOHN FLETCHER (ed.), *Forces in Modern French Drama* (1972); MAURICE NADEAU, *Histoire du Surréalisme* (1945; Eng. trans., *The History of Surrealism,* 1965), with bibliography and extracts from Surrealist authors, reviews, and manifestos; JUSTIN O'BRIEN, *Contemporary French Literature* (1971), essays on various topics; WALLACE FOWLIE, *A Guide to Contemporary French Literature from Valéry to Sartre* (1965); JOHN SIMON, *Modern French Criticism* (1972).

Italian literature: LANDER MacCLINTOCK, *The Age of Pirandello* (1951); SERGIO PACIFICI, *A Guide to Contemporary Italian Literature: From Futurism to Neorealism* (1962).

Spanish and Ibero-American literature: PAUL ILIE, *The Surrealist Mode in Spanish Literature* (1968), a useful account, and the first examination of Spanish Surrealism; G.G. BROWN, *A Literary History of Spain,* vol. 6, *The Twentieth Century* (1972), one of the most thorough surveys; JOHN M. COHEN (ed.), *Latin American Writing Today* (1967), translations of prose and poetry in Spanish and Portuguese, from Argentina, Brazil, Chile, Colombia, Cuba, Mexico, Peru, and Uruguay; JOHN A. NIST, *The Modernist Movement in Brazil: A Literature Study* (1967), a very good survey.

German literature: JETHRO BITHELL, *Modern German Literature, 1880–1950,* 3rd ed. (1959), a useful reference work, from Naturalism to Surrealism—includes postwar Austrian writers; AUGUST CLOSS (ed.), *Introductions to German Literature,* vol. 4, *Twentieth Century German Literature* (1969), the best English introduction; WALTER H. SOKEL, *The Writer in Extremis: Expressionism in Twentieth-Century German Literature* (1959).

Belgian literature: VERNON MALLINSON, *Modern Belgian Literature, 1830–1960* (1966).

Scandinavian literature: HELGE G. TOPSOE-JENSEN, *Den skandinaviske litteratur fra 1870 til vore dage* (1928; Eng. trans., *Scandinavian Literature from Brandes to Our Day,* 1929).

Slavic and eastern European literature: MAX HAYWARD and LEOPOLD LABEDZ (eds.), *Literature and Revolution in Soviet Russia, 1917–1962* (1963), based on contributions of a symposium held in Oxford, 1962—includes an important introduction that is a general survey of the period; VLADIMIR MARKOV, *Russian Futurism* (1969), the first study of its kind and well worth reading; MARK L. SLONIM, *Soviet Russian Literature: Writers and Problems, 1917–1967,* 2nd rev. ed. (1977), a comprehensive work with critical evaluations and bibliographical notes; GLEB STRUVE, *Russian Literature Under Lenin and Stalin, 1917–1953* (1971), the definitive analysis of revolutionary and post-revolutionary Russian literature; *Soviet Russian Literature, 1917–50* (1951), excellent bibliography of works in English translation to 1950—a detailed and comprehensive survey—see also his many studies on individual writers and works; ADAM GILLON and LUDWIK KRZYZANOWSKI (eds.), *Introduction to Modern Polish Literature: An Anthology of Fiction and Poetry* (1964), includes a short historical introduction on Polish literature and brief biographical notes; EWALD OSERS and J.K. MONTGOMERY (eds. and trans.), *Modern Czech Poetry* (1945); CLARENCE A. MANNING and ROMAN SMAL-STOCKI, *The History of Modern Bulgarian Literature* (1960); ANTE KADIC, *Contemporary Serbian Literature* (1964) and *Contemporary Croatian Literature* (1960); C.A. TRYPANIS (comp.), *Medieval and Modern Greek Poetry* (1951), includes extensive bibliography on many modern Greek poets; BASIL MUNTEANO, *Panorama de la littérature roumaine contemporaine* (1938; Eng. trans., *Modern Rumanian Literature,* 1939); RIMVYDAS SILBAJORIS, *Perfection of Exile: Fourteen Contemporary Lithuanian Writers* (1970), includes a brief survey of Lithuanian literature.

Hebrew literature: SIMON HALKIN, *Modern Hebrew Literature: Trends and Values* (1950).

Lithuanian Soviet Socialist Republic

One of the 15 union republics of the Soviet Union, the Lithuanian Soviet Socialist Republic is more popularly known as Lithuania (Lithuanian Lietuva, Russian Litva). It is one of the three Baltic republics, lying on the shores of the Baltic Sea, and is one of the westernmost portions of Soviet territory. Lithuania is relatively small in size, with an area of 25,200 square miles (65,200 square kilometres), and in population, with about 3,398,000 people according to the 1979 census, 80 percent of them Lithuanian. In spite of its small size, an ancient culture and an often wooded landscape largely molded by Ice Age glaciers lend character and variety to the republic.

Since World War II there has been considerable diversification and industrialization of the Lithuanian economy; the national income has risen dramatically, and about two-thirds of the population has become urban, with 14 percent concentrated in the capital, Vilnius. As a result, less than one-fourth of the economically active population of Lithuania still works in the traditionally dominant sphere of agriculture.

Lithuania is bounded on the north by the Latvian S.S.R., on the east and south by the Belorussian S.S.R., and on the southwest by the Kaliningrad oblast (region) of the Russian S.F.S.R. and by Poland. The waters of the Baltic Sea to the west add an important maritime element to the natural environment and economy of the republic. (For related information, see BALTIC STATES, HISTORY OF THE; RUSSIA AND THE SOVIET UNION, HISTORY OF; and BALTIC LANGUAGES.)

THE LAND

Topography and soil types. Underlying rock structures are of little significance for the contemporary terrain of Lithuania, which, basically, is a low-lying plain that is scraped by Ice Age glaciers, which left behind thick, ridgelike terminal deposits known as moraines. The Baltic coast area is fringed by a region that is characterized by geographers as the maritime depression, which rises gradually eastward. Sand dunes line the attractive coast, and the Kuršių Marios (Kurisches Haff), a lagoon almost cut off from the sea by a thin, 60-mile (100-kilometre) sandspit, forms a distinctive feature. The lagoon is bounded on the east by the Žemaičių (Samogitian) Hills, which give way to the flat expanses of the central Lithuanian lowland.

The central lowland

The lowland area of Lithuania consists of glacial lake clays and boulder-studded loams. It stretches in a wide band across the republic from north to south; some portions of it are heavily waterlogged. The lowland is bordered in the southeast by the narrow Žiežmariai Plain, which was formed from an ancient glacial valley filled with sandy, gravelly, and pebbly deposits. The sandy portions have often been whipped up into dunes, which have become overgrown with pine trees. The elevated Baltic Ridge thrusts between these two lowland areas into the eastern and southeastern portions of the republic; its rumpled glacial relief includes a host of small hills and numerous small lakes. The Švenčionių-Naročiaus and the Ašmenos hills—the latter containing Juozapinė, at 958 feet (292 metres) above sea level the highest point in Lithuania—are located in the extreme east and southeast of the republic.

Lithuanian soils range from sands to heavy clays. In the northwest the soil is either loamy or sandy (and sometimes marshy) and is quite heavily podzolized, or leached. In the central region weakly podzolized, loamy peats predominate, and it is there that the most fertile, and hence most cultivated, soils are found. In the southeast there are sandy soils, somewhat loamy and moderately podzolized. Sandy soils, in fact, cover a quarter of Lithuania, and most of these regions are blanketed by woodlands.

Climate and drainage. The climate of the republic is transitional between the maritime type of western Europe and the continental type found farther east. As a result, damp air masses of Atlantic origin predominate, alternating with continental Eurasian and, more rarely, colder Arctic air or air of a southern, tropical origin, Baltic Sea influences dominate only a comparatively narrow coastal zone. The mean temperature for January, the coldest month, is 23.4° F (−4.8° C), while July, the warmest month, has an average temperature of 63° F (17.2° C). The average annual rainfall is 25 inches (630 millimetres), with the heaviest amounts occurring along the coast and the total diminishing inland. Rainfall reaches a peak in August, except in the maritime strip, where the maximum is reached two to three months later.

Lithuanian rivers drain to the Baltic and generally have the slow, meandering characteristics of lowland rivers. The Neman (Nemunas), cutting north and then west through the heart of the country, is the largest. Its main tributaries are the Merkys, Neris, Nevėžís, Dubysa, Jūra, Minija, and Šešupė. The rivers have a total length of some 1,700 miles, nearly a fifth of which is navigable. Most of them can be used for floating timber rafts and for electric-power generation. A distinctive feature of the Lithuanian landscape is the presence of about 3,000 lakes, mostly in the east and southeast. The boggy regions produce large quantities of peat that, dried by air, is used in both industry and agriculture.

Plant and animal life. The natural vegetation cover of Lithuania falls into three separate regions. In the maritime regions, pine forests predominate, and wild rye and various bushy plants grow on the sand dunes. Spruce trees add their colour to the hilly eastern portion. The central region is characterized by large tracts of oak trees, with elegant birch forests in the northern portions, as well as distinctive black alder and aspen groves. Pine forests again prevail in the south; indeed, about a quarter of the whole republic is forested, with a further quarter taken up by meadowlands. Swamps and marshlands account for about 7 percent of the total area.

Three distinctive vegetation zones

Wildlife is very diverse, with about 60 mammalian species. There are wolves, foxes, otters, badgers, ermines, wild boars, and many rodents. The deep forests harbour elk, stag, deer, beaver, mink, and water rats. Common birds include delicate white storks, a variety of ducks, geese, and swans, cormorants, herons, hawks, and even an occasional bald eagle. There are many types of grouse and partridge, and the total number of bird species recorded approaches 300. The more than 50 species of fish found in Lithuanian rivers include include salmon, eel, bream, carp, and trout, with cod, plaice, and herring common off the coast.

THE PEOPLE

Composition and distribution. The 1979 census reported some 3,398,000 people in the republic. Ethnically, about 80 percent of the people are Lithuanians, but there are also Russians (almost 9 percent), Poles (about 7 percent), and lesser numbers of Belorussians, Ukrainians, Jews, Latvians, Tatars, Gypsies, and others. The Lithuanians belong to the Baltic group and have very ancient roots. Linguistic and archaeological evidence indicates that their ancestral populations moved northward from steppe regions centred in the Upper Dnepr Basin and by about 2,500 years ago had differentiated into the modern language groups.

The average density of population in the republic is more than 135 persons per square mile (52 persons per square kilometre), with a concentration in the southeastern region, notably around Vilnius, the capital. Urban dwellers make up almost two-thirds of the total population. Lithuania has more than 90 settlements classified as cities and a further two dozen towns. The largest centres are Vilnius (503,000 population, 1981), a modern city that has spread from a historic old quarter with narrow streets onto the wooded hills around the Vilija and Vilnia River junction; Kaunas (377,000); Klaipeda (178,000); Šiauliai (121,000); and Panevėžys (104,000).

Demographic trends. Natural increase, rather than any inward migration, has accounted for almost 90 percent of recent population growth. Internally, however, there has been rapid movement of people to the cities, accentuated in the case of the new, planned regional centres of Alytus, Kapsukas, Utena, Plunge, and Mažeikiai.

THE ECONOMY

Resources and their development. Lithuania possesses a good range of useful mineral resources, including sulfates, notably gypsum; chalk and chalky marl; limestones; dolomites; various clays, sands, and gravels; peat; some iron ore and phosphorites; and mineral waters. Oil deposits have been detected in the offshore regions. The power potential of the many rivers and the traditional resources of the great forests and the rich agricultural areas have added to the basic wealth of the republic, which is well placed geographically for trade.

During the Soviet period, economic policy has emphasized industrialization, and, since the end of World War II, the machinery, shipbuilding, electronic, electrical- and radio-engineering, chemical, cement, and fish-processing industries have expanded rapidly, partly to meet Soviet goals and partly to satisfy local needs. Traditional industries such as food processing, especially of meat, milk, and fish products, and various branches of light industry such as the manufacture of clothing, footwear, and durable consumer goods have also expanded. The building materials and construction industries have expanded to meet the demands of the growing and increasingly urban population. Gross industrial output in the republic increased more than 3 times during the 1960s and nearly doubled during the 1970s. Worsening labour shortages, however, have made it increasingly difficult to sustain such high growth rates.

Planning goals

Expansion in energy production in Lithuania has been particularly notable. Electrical output expanded from 81,000,000 kilowatt-hours in 1940 to 11,664,000,000 kilowatt-hours in 1980. A 1,800,000-kilowatt thermal station is located at the town of Elektrėnai. Other major plants include a hydroelectric station at Kaunas and a thermal station at Vilnius. A 1,600,000-kilowatt hydroelectric plant near Kaišiadorys was begun in 1979, and construction began near Ignalina on the republic's first nuclear-power station, with a 6,000,000-kilowatt capacity, in 1975. Since 1961 the Lithuanian power system has been part of the network covering the northwestern Soviet Union.

To facilitate the expansion of the basic Lithuanian industries, there has been a marked emphasis on the training of engineers and technicians. Industrial expansion has added such towns as Siauliai and Panevėžys to the major existing industrial centres of Vilnius, Kaunas, and Klaipėda. A number of small towns, notably Kapsukas, Alytus, and Utena, are becoming the focus of the metalworking, mechanical-engineering, and textile industries, while the modest centres of Kedainiai and Jonava now specialize in the chemical industry. The rapidly growing town of Mažeikiai has a major oil refinery, with accompanying fuel and petrochemical production. Lithuanian industrial products are exchanged with other union republics and exported to more than 85 countries, while imports, mostly from the other union republics, include machinery and other equipment, some raw materials, fuels, and the products of light industry.

Agriculture. Lithuanian agriculture still maintains something of its traditional importance, with almost one-fourth of the economically active population engaged in farming. There are some 775 collective and 325 state farms, which occupy 11,508,000 acres (4,656,000 hectares); two-thirds of this land is used for agriculture, about 70 percent being arable, and most of the rest is given over to meadows and pastures, with a small amount to gardens. The development of agriculture has been closely linked to land reclamation and swamp-drainage schemes. The chief agricultural trend is toward the production of meat and milk, together with the cultivation of such crops as flax, sugar beets, potatoes, and vegetables. Half the total production is made up of fodder crops; about 40

Emphasis on livestock breeding

percent is grain and leguminous crops; and most of the rest is potatoes and vegetables. Livestock breeding is still the single most important branch of agriculture, however, with emphasis on dairy cattle and pigs. There are smaller numbers of sheep, goats, horses, and poultry on Lithuanian farms. One feature of the Soviet period has been the introduction of technology in agriculture; most crop cultivation is mechanized, although at fall harvest time large amounts of manual labour are still required.

Economic regions. Lithuanian planners, using as criteria environmental, economic, and transportational indexes, have divided the republic into four economic regions.

Eastern Lithuania. Occupying almost 30 percent of the total area, Eastern Lithuania, including the capital, Vilnius, is characterized by diversified and rapidly growing industry (primarily metalworking and mechanical engineering, woodworking, branches of light industry, and, more recently, energy production from nuclear sources); industry is centred on the main cities, but there is also a substantial rural economy. The region also has a number of health resorts, including Trakai and Ignalina, and is crisscrossed by highways linking it with major Soviet cities.

Southern Lithuania. Occupying slightly more than one-fourth of the republic's territory, the southern portion of central Lithuania contains more than half of the republic's developed waterpower resources. Metalworking, mechanical-engineering, food-processing, and chemical industries predominate; farming is intensive, with a concentration on stock raising and the growing of sugar beets. Kaunas, Alytus, and Kapsukas are the main centres.

Northern Lithuania. Sprawling over the northern portion of the middle Lithuanian lowland and the eastern slopes of the Žemaičių Hills, Northern Lithuania occupies almost 30 percent of the republic. The region is noted for its fertile soils and for its dolomite, gypsum, and limestone reserves. Farming is intensive, with almost half the winter-wheat sowing and most of the sugar-beet and flax crops located there. Industry is not well developed. The main centres are Šiauliai, Panevėžys, and Rokiškis.

Western Lithuania. Occupying the remaining 15 percent of the republic, Western Lithuania lies along the Baltic Sea, which lends a distinctive maritime quality to the region. Shipbuilding, ship repairing, fish processing, and oil refining are the main industries, with Klaipeda, Telšiai, Plunge, and Mažeikiai the main centres. There are many pastures, and horse breeding and the raising of dairy cattle and pigs, together with poultry farming and fishing, are well developed.

Transportation. Railways continue to be the main means of transport in Lithuania. About 1,300 miles of track are in use, and freight and passenger haulage is almost three times greater than it was in 1960. Motor transportation has nevertheless increased sharply, and cars and buses account for about 95 percent of the total passenger traffic. The hard-surfaced road network covers some 13,400 miles, about two-thirds of the total. Sea transport is important, with freight transportation having shown a rapid increase since World War II. River transport is also significant.

Vilnius is the main air-transportation centre, with links to other important Soviet cities and to resort areas. The transportation network also includes a natural-gas pipeline that carries gas from the Ukraine and an oil pipeline that carries crude from western Siberian oil fields to the refinery at Mažeikiai.

ADMINISTRATION AND SOCIAL CONDITIONS

The constitutional and political framework. The constitution of the Lithuanian Soviet Socialist Republic, originally adopted on August 25, 1949, has been modified several times since, always in close accordance with executive, legislative, and judicial patterns prevailing throughout the Soviet Union. The constitution states that political power resides in the workers and peasants of the republic, acting through the local soviets, or councils of workers' deputies. This political foundation rests on a socialist economic system, with communal ownership of the means of production.

The Supreme Soviet, elected for four years and acting through a Council of Ministers, is the highest legislative body of the republic. Its Presidium exercises executive power between sessions. Locally, power is vested in various levels of soviets, down to the smallest village unit; all are elected for two-year periods.

Justice is administered through the Supreme Court and a local network of people's courts, with members elected for five and two years, respectively. Law enforcement is in the hands of the Lithuanian S.S.R. procurator, who is appointed for five years.

The executive, legislative, and judicial bodies established within the Lithuanian Soviet Socialist Republic are substantially administered by, and are accountable to, the authorities of the Soviet Union. In accordance with long-standing practice, control is exercised through central directives and agencies established for that purpose.

The Lithuanian Communist Party

The Lithuanian Communist Party is a constituent of the Communist Party of the Soviet Union and is closely subordinated to its control; it is the sole political organization of the republic. Its 170,000 members and candidates for membership, of whom about 70 percent are of Lithuanian nationality, are supported by the activities of the Komsomol youth movement, which has some 450,000 young men and women in its ranks. As elsewhere in the Soviet Union, trade unions, which have about 1,750,000 members, play roles in formulating and implementing social and economic policy in accordance with party directives.

Social services. Lithuanian society has benefitted from emphasis on the production of consumer goods and the improvement of social services, as well as on the expansion of heavy industry. Both average monthly income and the proportion of the national income set aside for social purposes have shown a steady expansion. The standard of living is augmented by free education and medical services, as well as a range of ancillary services. The latter include the upkeep of kindergartens and day nurseries, sick leave, pension payments, and help in providing vacation and sanatorium accommodations. Pensions and related allowances, for example, covered more than 670,000 people, one-third of them farmers, by the early 1980s. Urbanization has meant a constant, and still unresolved, struggle to keep up with housing needs, and new techniques, including prefabrication, have been introduced.

A typical state budget for the 1970s showed 92.5 percent of the state income coming from the socialist economy. Some two-thirds of budget expenditures went for reinvestment in the economy, about 10 percent went to finance education and culture, and the remainder was spent on various social services. There has been a considerable expansion of health facilities in Lithuania, including increases in the numbers of doctors and hospital beds, since World War II.

The schools of the republic serve more than 1,300,000 students. More than half of these are enrolled in general education schools of all types, and a transition to a compulsory 10-year education was begun in the early 1970s. Some 70 specialized secondary schools enroll about 70,000 students, and universities and other post-secondary institutions in Lithuania have about the same total enrollment. The Lithuanian Academy of Sciences, which was created in 1941, directs the work of 12 research centres and is the main agency of research in the republic.

Cultural life. As elsewhere in the Soviet Union, there is a high level of public interest in various forms of cultural life. The republic has well over 2,000 public libraries and a large number of museums, amateur cultural centres, theatres, and cinemas. A year's output of books amounts to some 16,400,000 copies, all but 3,000,000 or so in Lithuanian, and readership is at a high level. The works of contemporary Lithuanian writers, poets, and playwrights are evolving in a milieu that blends a very old cultural tradition with the relatively new social factors introduced during the Soviet period. Many Lithuanian critics have praised the poetry of Eduardas Mieielaitis, especially his collection *Žmogus* (1962; "Man"), for its deep and optimistic human, as well as social, vision. Related concerns animate much modern historical writing, as in ***Parduotos vasaros*** (1957; "Bartered Summers"), a novel by Juozas

Baltušis, and *Sodybų tuštėjimo metas* (1970; "The Time of Emptying Homesteads"), a novel by Jonas Avyžius. Writings about the bitter conflicts of the World War II invasion include the heroic poem Kraujas ir pelenai (1960; "Blood and Ashes"), by Justinas Marcinkevičius, who also treats national and historical themes in the drama trilogy Mindaugas, *Mažvydas* (both biographical), and Katedra (1968, "Mindaugas"; 1976, "Mažvydas"; 1970, "The Cathedral").

The great majority of Lithuania's newspapers and magazines, which enjoy high circulations, are published in Lithuanian, and there are a number of broadcasting services, including a television service that is part of the East European Intervision network. In spite of such modern incursions, however, Lithuanian folklore continues to hand down a rich and original heritage from generation to generation. Lithuanian songs and a remarkable collection of fairy tales, legends, proverbs, and aphorisms have their roots deep in a language and culture that are among the oldest in Europe. The folk songs (*dainos*) are melodious and lyrical, while the communal folk dances, often related to everyday activities, are characterized by an elegant symmetry of design and motion. Lithuanian folk art is mainly embodied in ceramics, leatherwork, wood carving, and textiles; its colouring, which tends to avoid gaudiness, and its original geometric or floral patterns are characteristic features.

The country's folk heritage

The Vilnius drawing school, founded in 1866, has had a strong influence on the republic's fine-arts traditions, while the composer and painter M.K. Čiurlionis, who died in 1911, also had a considerable influence on contemporary forms. During the Soviet period, a realistic note has been introduced into sculpture and painting, and Lithuanian contemporary drawing, noted for the use of natural colour and a highly refined technique, has won international acclaim. The republic's architecture has been affected by a rich heritage in monuments and old buildings and by styles, ranging from the Gothic to the Neo-classical, that have acquired a distinctive local character. While much attention is given to preservation of historical monuments, much of the energy of modern Lithuanian architects has gone into the designing of new buildings, both industrial and domestic, and especially of cultural and educational centres.

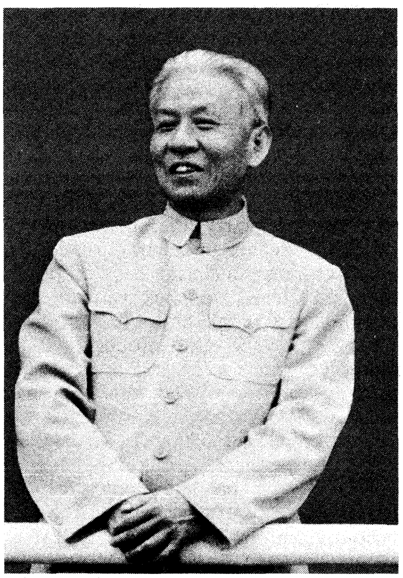
Music, too, has a strong tradition in the republic, a special feature being the dances and singing festivals held in the towns and villages every summer. These build to a climax every five years in national singing festivals during which as many as 40,000 persons may compete.

The outlook. Published data indicate that the rapid economic growth that characterized the previous two decades of development in Lithuania had decelerated by the early 1980s. The decline in demographic growth, exhaustion of labour surpluses, diminishing returns on capital investments in industry and agriculture, rigidities in managerial controls, and decline in rates of labour productivity—all are indicators that point not only to more severe difficulties in increasing production but also to problems in maintaining the living standards that have been achieved.

(K.A.M./B.V.M.)

Liu Shao-ch'i

Liu Shao-ch'i (Pinyin romanization, Liu Shao-qi) was head of state of the People's Republic of China, chief theoretician for the Chinese Communist Party (CCP), and considered the heir apparent of Mao Tse-tung (Mao Zedong) before being purged in the late 1960s. One of the early Chinese Communists, Liu was active in the Chinese labour movement and influential in formulating party and, later, governmental strategy. He played an important role in Chinese foreign affairs after the Communists had gained control of the country.


Liu Shao-ch'i.
Eastfoto

Liu was born in 1898 in Ning-hsiang (Ning-xiang) district, Hunan Province, the youngest son of a rich peasant landowner. He attended middle school, went to normal school in Ch'ang-sha (Chang-sha), and by 1918 was studying French in North China. Two years later, Liu joined the Socialist Youth League and subsequently journeyed to Moscow, where he enrolled in the University for Toilers of the East and became a member of the newly formed CCP. Liu's education and his extensive experience abroad later made him one of the most cosmopolitan of the Chinese leaders.

Investment in China's labour movement

In 1922 Liu returned to China, where he helped organize the First National Labour Congress in May. Soon after this he was assigned to the Hunan Secretariat as an aid to Mao Tse-tung. Late that year Liu went to the An-yiian (An-yuan) colliery to organize miners for what was to become a successful strike. From this time onward he became increasingly more involved in the labour movement—as leader of a sympathy strike in February 1923, as vice chairman of the All-China Federation of Labour in May 1925, and as secretary general of the Third National Labour Congress in 1926.

The year 1926 marked the high point of the Chinese labour movement. When the Nationalist Party (Kuomintang) and the CCP split in April 1927, Liu and his comrades found the urban component of the Chinese Communist movement decimated; they fled underground. In late April, Liu was elected to the Fifth Central Committee of the CCP.

After the turbulence of 1927, Liu's ascendancy in the party was rapid; he was named director of the workers' department in mid-1928, and in the following year he assumed the post of secretary of the Manchurian Provincial Party Committee, a post he later relinquished (early 1931) in order to go to Shanghai. By 1931 the centre of gravity of the Chinese Communist movement had shifted away from the cities to rural areas, but Liu remained in Shanghai, underground, taking part in movements in opposition to the Japanese invasion of Manchuria and encroachment in North China.

In late 1932 Liu joined Mao Tse-tung's forces in Kiangsi (Jiang-xi) Province and was named head of the All-China Federation of Labour; by 1934 he had gained a seat on the Politburo (Political Bureau) of the Sixth Central Committee. Shortly after his elevation to the Politburo, he left Mao's forces, who were in the midst of their Long March, and went to Peking (Bei-jing) to agitate further against the Japanese. In early 1936 Liu was designated secretary of the party's North China Bureau. When the war with Japan had spread to most of China, he was put in charge of the Central Plain Bureau of the CCP, and in this position he worked closely with the Communist New Fourth Army operating in Central China.

Liu's lectures of 1939

In mid-1939 in Yen-an (Yan-an; the Communist headquarters), Liu delivered a famous series of lectures called "How To Be a Good Communist." In these talks he drew upon all his organizational experience as a labour leader and underground figure to define the demands to be made upon all party members; at this point Liu began to assume the role of chief theoretician for the party. He delivered another important talk in July 1941, entitled "On Intraparty Struggle," which further enhanced his position as the party's theoretical spokesman.

In 1943 Liu was named secretary of the Central Secretariat and vice chairman of the People's Revolutionary Military Council, making him one of the most powerful men in the Chinese Communist hierarchy. By June 1945 he was clearly spokesman for party affairs, and at the Seventh Party Congress he delivered the major address, entitled "On the Party." During the closing days of World War II, Mao Tse-tung and Chiang Kai-shek were engaged in talks at Chungking (Chong-qing), and Liu served as acting chairman in Mao's absence. By October 1, 1949, the Communists controlled much of China, the People's Republic of China had been established, and Liu had been named vice chairman of the new government.

Experience with Soviet leadership

Once the Chinese Communists had achieved power, Liu's cosmopolitan outlook and experience with the Soviet leadership enabled him to act as initial spokesman for Peking's new hard-line foreign policy, and it also gave him a central role in the formation of China's industrialization plans; both endeavours were dependent on Moscow in 1949. For these reasons Liu headed the Sino-Soviet Friendship Association in 1949–54 and led the Chinese delegation to the Soviet 19th Party Congress in October 1952. In August 1954, at the First National People's Congress, Liu was elected chairman of the Standing Committee. His influence was not confined to state organs, however, but extended also into the party, where late in 1953 and early 1954 he led a purge of regional power holders. Subsequently his position in the party grew, and by 1956 he was clearly Mao Tse-tung's heir apparent.

The Great Leap Forward

During the second session of the Eighth Party Congress in May 1958, Liu outlined the strategy for the second five-year economic plan (called the Great Leap Forward), which was to lay the foundation for the rapid industrialization of China. Shortly after the initiation of the Great Leap Forward, however, it became apparent that industrialization could not be achieved as rapidly as hoped, and a policy of retrenchment was called for. Partly as a result of the failures of the Great Leap, Mao relinquished his position as chairman of the People's Republic of China, though he remained party chairman, and Liu assumed the chairmanship in April 1959. During this period, Liu tried to revitalize agriculture by initiating policies that permitted peasants to cultivate private plots and spurred them on with monetary incentives; both were policies to which Mao later strongly objected.

Liu's fall from power

In his new post as head of state, Liu began to play a more prominent role in foreign affairs, receiving state visitors from Indonesia, the Soviet Union, Pakistan, Ghana, Cuba, North Vietnam, Cambodia, and North Korea. In addition, he travelled abroad rather extensively during 1959–66. Upon reaching this pinnacle, however, Liu became one of the most important figures to be purged in the Great Proletarian Cultural Revolution (1966–69). Many persons associated with him, such as P'eng Chen (Peng Zhen), mayor of Peking, and Teng Hsiao-p'ing (Deng Xiao-ping), a member of the Politburo, were also purged, decimating what had been viewed as a highly cohesive Chinese leadership. In October 1968 Liu was stripped of party positions and labelled China's Khrushchev, and, by April 1969, a new constitutionally designated successor to Mao had been chosen—Lin Piao (Lin Biao), head of the armed forces. In the autumn of 1971, however, Lin Piao disappeared, and it was announced that he had died in an airplane crash while fleeing from an attempt to assassinate Mao.

During 1974 rumours of Liu's death gained wide circulation, and on October 31, a Communist newspaper in Hong Kong confirmed the fact. No details of date or place of death were revealed, however, until May 1980, when the *Beijing Review* reported that Liu had died on November 12, 1969, in K'ai-feng (Kai-feng), northern Honan Province.

The causes of Liu's fall (and events leading to Lin's death) are not clear. For several years the names of Liu, Teng, and Lin were linked, and the three were condemned in the party press as "capitalist roaders" intent on defeating the revolution. After Mao's death on September 9, 1976, however, his widow, Chiang Ch'ing (Jiang Qing), and the other members of her so-called "Gang of Four" undertook a coup that was quickly aborted. Hua Kuo-feng (Hua Guo-feng), a relatively junior member of the hierarchy, achieved party leadership, and Teng Hsiao-p'ing was rehabilitated. Then, in February 1980, the 11th Central Committee of the CCP decided "to completely rehabilitate" Liu, calling him a "great Marxist and proletarian revolutionary," and to remove the labels of "renegade, traitor, and scab" formerly attached to him. Lin was then identified with the Gang of Four and charged with "concocting false evidence" and subjecting Liu to "political frame-up and physical persecution" while overthrowing other leaders as being Liu's agents.

While little is known of Liu's first four spouses, his fifth wife, Wang Kuang-mei (Wang Guang-mei), achieved great notoriety during the Cultural Revolution for her "bourgeois" life-style. Liu had at least eight children, none of whom achieved political prominence.

BIBLIOGRAPHY. DONALD W. KLEIN and ANNE B. CLARK, ***Biographic Dictionary of Chinese Communism, 1921–1965*** (1971), a monumental, two-volume set that is well written and gives an excellent discussion of Liu's life; LIU SHAO-CH'I, ***The Collected Works of Liu Shao-ch'i***, 3 vol. (1968–69), an anthology of Liu's major statements; "Fifth Plenary Session of 11th C.C.P. Central Committee," ***Beijing Review***, No. 10 (March 10, 1980), pp. 3–10, which documents official rehabilitation measures.

(R.C.N.)

Liver, Human

A pyramid-shaped structure located in the upper right region of the abdomen, with its upper border at about the level of the nipples, the liver is the largest organ in the human body, accounting for about one-eighteenth of the body weight in an infant and one-fiftieth in an adult. It is divided into two parts, or lobes, of which the right is about six times the size of the left.

Functions

The liver serves many functions, including the regulation of blood volume and manufacture of certain blood clotting factors; the storage of several substances, including glycogen, copper, iron, and vitamin B_{12}; the metabolism of proteins, carbohydrates, and fats; the production of heat; the detoxication (removing the poisonous effect) of certain foreign substances in the blood; the destruction of old red blood cells; and the formation of bile.

In the undersurface of the right lobe of the liver is a fissure, known as the porta hepatis—"the door of the liver"—so called because it is the point of entrance for arterial and venous blood and of exit for bile. Venous blood is brought by the portal vein from the intestines and spleen; arterial blood, by the hepatic (liver) artery from the aorta by way of the celiac artery. The hepatic artery and the portal vein both divide into branches to the right and left lobes of the liver within the porta hepatis. There, also, bile ducts from the two lobes unite in the common hepatic duct. The hepatic duct in turn joins with the duct from the gallbladder (the cystic duct) to form the common bile duct, which leads toward the small intestine.

Usually, within the wall of the duodenum, the first portion of the small intestine, the common bile duct and the main channel for pancreatic juice come together to form a channel called the ampulla of Vater, which then empties into the duodenum. The portion of the common duct within the duodenal wall is encircled by muscle fibres called the sphincter of Oddi.

Bile is stored in the gallbladder, a pear-shaped structure resting in a shallow furrow in the right lobe of the liver and extending from the lower edge of the liver to the porta hepatis. The narrow lower end of the gallbladder, called its neck, joins with the cystic duct.

Hepatic network of nerves

The cystic artery, a large, twisting branch of the hepatic artery, brings blood to the gallbladder. The ganglia beside the spine at the level of the chest, the right and left vagus nerves, and the right phrenic nerve contribute fibres to the hepatic network of nerves. (The vagus nerves are the pair of tenth cranial nerves; the phrenic nerves serve chiefly the diaphragm—the muscular partition between the chest and the abdomen—and the pericardium, the sac that encloses the heart.) Branches of the hepatic network of nerves accompany all the branchings of the hepatic artery and of the bile ducts, even into the hepatic parenchyma, or functioning liver tissue.

The hepatic veins carry venous blood from the liver into the inferior vena cava, the large vessel that brings venous blood from the lower half of the body to the right atrium of the heart. There is a deep groove in the back of the liver where the inferior vena cava presses against it.

Most lymph channels from the liver lead to a small group of lymph nodes in the vicinity of the porta hepatis; some superficial lymphatic vessels penetrate the diaphragm and terminate in nodes in the chest. Lymph from the gallbladder flows by way of the node at the gallbladder neck to nodes near the common bile duct; there is a connection at this point with lymphatic vessels near the right end of the pancreas.

Ligaments and pressures exerted by the muscles of the abdominal wall hold the liver in position.

***Origin** and development.* The liver originates as a ventral, or forward, outgrowth from the embryo's innermost germ layer, called its endoderm. The bud divides into two portions, one forming the main body of functioning liver tissue and the hepatic duct and its branches; the other portion forming both the gallbladder and the bile ducts.

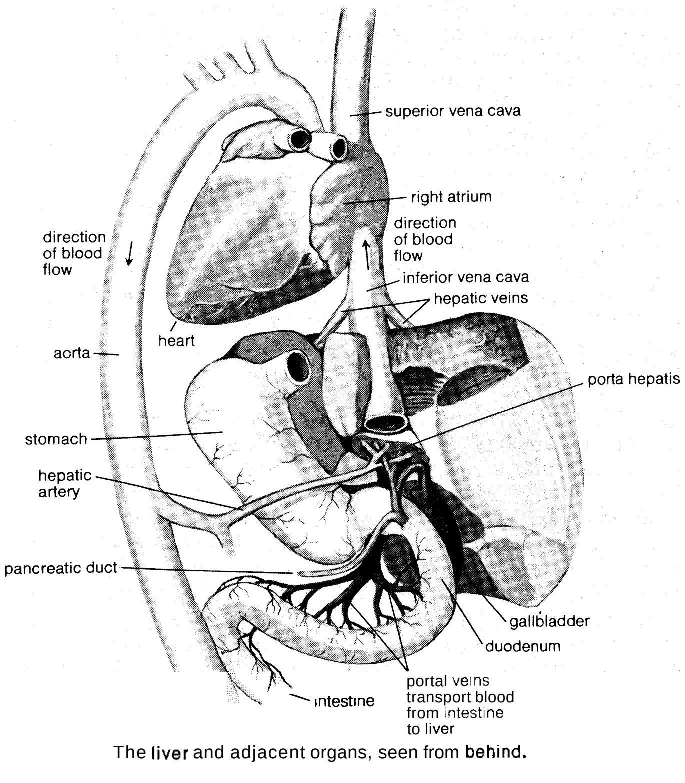
Normal structure and functions

STRUCTURE

The lobules

Grossstructure. The liver has been described as a collection of units called lobules, each of which contains, in its centre, a branch of the hepatic vein and, in its outer areas, a complex known as a portal tract. The portal tract includes a bile duct, a small branch of the portal vein, and a branch of the hepatic artery. Between the tributary of the hepatic vein and the portal tract are columns of hepatic cells and blood channels called sinusoids.

It is more precise to think of the liver as sheets of hepatic cells, the sheets pierced by two types of tunnels, one made up of the portal tracts and the other of the central tributaries of the hepatic vein. The two types of tunnels tend to run in planes perpendicular to each other; they never touch, although their smallest branches may be only about 0.5 millimetre (0.02 inch) apart. The sinusoids are not in any regular order, but tend to run perpendicular to the central veins.



The liver and adjacent organs, seen from behind.

Blood flows from the terminal branches of the portal vein into the sinusoids and thence, by way of hepatic vein branches, into the hepatic vein. This direction of flow, from portal vein to hepatic vein rather than the reverse, is brought about by the greater pressure in the portal vein.

The hepatic cells, the functioning cells that make up about 60 percent of the liver, are polyhedral and about 25 to 35 microns (one micron = 0.001 millimetre or 0.00004 inch) in diameter. Each cell usually has one nucleus, and the cells multiply by mitosis, the cell division that is the usual form of reproduction of body cells.

The walls of the sinusoids are made up of two types of cells, endothelial, or lining, cells and Kupffer cells. The Kupffer cells are capable of phagocytosis—the ingestion of other cells and of foreign particles. They also have important functions in the production of antibodies and in blood formation.

Between the hepatic cells and the walls of the sinusoids there is tissue fluid, which flows outward into the lymphatic vessels. The lymph channels are located in connective tissue around the portal vein.

Arterial blood is supplied to the portal tracts by a network of arterioles (minute arteries) around the bile ducts; the blood then empties into the sinusoids and hence into the central hepatic veins.

In grooves on the surfaces of the hepatic cells are the bile canaliculi, or capillaries, the smallest of the series of bile channels that terminate in the right and left main hepatic ducts. These emerge from the liver at the porta hepatis and carry the bile into the duodenum.

Intracellular structure. The hepatic cell, also called the hepatocyte or liver cell, has more metabolic functions than any other cell of the body and is the producer of bile. In its cytoplasm (substance outside the nucleus) the netlike structure known as the smooth endoplasmic reticulum is the site at which the bile pigment bilirubin is metabolized and many enzymes, such as those necessary to make glucose available to the blood, are synthesized. The bile acids, needed in the digestion of fats, are formed at this point, and drugs are detoxicated.

The rough endoplasmic reticulum produces certain proteins, such as the albumin and clotting factors of the blood. The mitochondria (membranous structures in the cytoplasm) have many functions, including the production of enzymes that play a role in the synthesis of glycogen and enzymes involved in the metabolism of fats.

Other structures of the hepatic cell cytoplasm—the network of fine tubes known as the Golgi apparatus and the minute bodies and channels known, respectively, as the lysosomes and canaliculi—act as the cell's excretory apparatus. The canaliculi and the nearby tissues are especially involved in bile excretion.

FUNCTIONS

Contents of bile

Production and storage of bile. The liver secretes about 800 to 1,000 millilitres (about one quart) of bile each day at a pressure of 15 to 25 centimetres (six to ten inches) of water. In the bile are bile salts and bile products of metabolism, such as cholesterol and phospholipid. Cholesterol, ordinarily insoluble, is in large-molecule compounds that permit it to be carried in solution. Bile salts are needed for the digestion of fats in the diet (see DIGESTION, HUMAN), and bile is the chief vehicle for the excretion of cholesterol and the bile pigment bilirubin. Some drugs and poisons, such as the salts of heavy metals, atropine, strychnine, and salicylates, are also excreted principally in the bile.

The sphincter of Oddi, the ring of muscle fibres that encircles the common bile duct within the wall of the duodenum, cuts off the flow of bile into the small intestine and diverts it to the gallbladder, where it is concentrated and stored. Through the mucous membrane lining the gallbladder, water and salts (a solution of sodium chloride and bicarbonate) are absorbed in such quantities that the bile is reduced to one-tenth of its former volume. At the same time the membrane secretes a thick mucus into the bile.

The principal function of bile is in the digestion of fats. Fat in the duodenum causes the sphincter of Oddi to relax and the gallbladder to contract, so that bile is discharged into the duodenum.

Functional disorders of the biliary tract, particularly spasm of the sphincter of Oddi, sometimes occur after surgical removal of gallbladder (cholecystectomy). These disorders are called biliary dyskinesia.

Bilirubin metabolism. Bilirubin, or bile pigment, which gives bile its characteristic green-orange colour, is formed chiefly in the spleen, the bone marrow, and the phagocytic—cell ingesting—cells of the liver called Kupffer cells, during the destruction of worn out red blood cells (erythrocytes) and the attendant breakdown of the pigment heme from the hemoglobin of these cells. A little bilirubin results from the breakdown of heme-containing enzymes and of young red cells in the bone marrow.

After the bilirubin has been carried in the bloodstream to the hepatic cells, the endoplasmic reticulum of these cells forms the compound bilirubin glucuronide. This compound is excreted into the bile channels and eventually reaches the small intestine. From the duodenum, the compound passes into the colon, where bacteria break some of it down into fat-soluble bilirubin. Limited amounts of the bilirubin are reabsorbed from the colon into the portal-vein system and carried to the liver, where they are re-excreted by the hepatic cells. The intestinal bacteria convert some bilirubin glucuronide into stercobilinogen, and about 100 to 200 milligrams (0.004 to 0.008 ounce) of this is eliminated each day in the feces, giving them their characteristic colour.

Formation of blood clotting factors. Most of the proteins that make up the clotting factors in the blood are produced by the liver. So-called clotting factor VIII (the antihemophilic factor) is made elsewhere, and for the production of certain other factors—factors II (prothrombin), VII, and X—sufficient quantities of fat soluble vitamin K must be absorbed from the intestine. Vitamin K absorption is inadequate if insufficient fat is absorbed—*e.g.*, when obstruction of the bile duct causes the supply of bile in the intestine to be inadequate.

Bleeding and liver disease

A number of elements may be involved, therefore, in the tendency of persons with liver disease to bleed. The liver may fail to produce clotting factors; insufficient fat-soluble K may be absorbed; or excessive destruction of blood platelets in a large spleen may lead to their being in short supply (platelets both produce a clotting factor—platelet factor 3—and themselves form the first plug to fill a gap in a vessel wall).

Metabolism of nutritive substances. Cholesterol, the precursor of the primary bile acids cholic and chenodeoxycholic acid, is synthesized by the liver. The liver also processes cholesterol that is absorbed into the portal-vein system from the intestine and not re-excreted into the bile. (Cholesterol is also the precursor of adrenal cortical hormones and of estrogens.) The liver combines the primary bile acids with glycine and taurine in the production of bile salts. Most of the bile salts are absorbed through the terminal part of the small intestine, called the ileum, and returned to the liver. The bile salts that are not absorbed pass through into the colon—the section of large intestine that is next to the rectum—and there are broken down by bacteria and converted into secondary bile acids—deoxycholic and lithocholic acid. About 0.7 gram (0.025 ounce) of bile acid is eliminated each day in the feces.

Fat that reaches the liver may be converted by it into fatty acids and converted in turn into carbohydrates or ketone bodies. The ketone bodies—acetoacetic acid, acetone, and beta-hydroxybutyric acid—are transported in the bloodstream to the tissues, where they are metabolized, with carbon dioxide and water as the end products.

Much of the fat that is transported is in the form of lipoproteins (fat–protein compounds). In combination with bile salts and cholesterol, lipoproteins form large protein compounds that are excreted in the bile.

Monosaccharide sugars (*e.g.*, glucose, fructose, and galactose), absorbed through the intestine, are stored in the liver. Fructose is converted directly to glucose and galactose, either directly or indirectly through glycogen. Glucose that is not converted to glycogen and stored or used in formation of amino acids is changed into fatty acids, carbon dioxide, and water. When the levels of glucose in the blood are low, glucose is supplied by the liver by conversion of stored glycogen, amino acids, or fatty acids. Lactic, ketoglutaric, and pyruvic acids are removed by the liver and metabolized (see also METABOLISM).

The synthesis by the hepatic cells of blood-serum proteins, including albumin and several clotting factors, has already been touched upon. Disease of these cells results in a lowering of the blood levels of the proteins, especially of albumin. At the same time amino acids, which the liver normally uses in the production of proteins, are excreted in the urine. Disease of the hepatic cells also lowers the blood levels of urea, normally produced by the cells from amino acids.

Ammonia, which is produced (1) by bacteria in the colon, (2) by the action of the enzyme urease in the stomach and small intestine, and (3) in the kidneys, is converted by the liver to urea. A high level of ammonia in the blood from abnormal liver function is believed to be partially responsible for the coma and the aftereffects on the brain that may be complications of hepatic-cell disease.

Drug metabolism and detoxication. In the liver, enzymes oxidize, reduce, hydrolyze, or combine many substances, metabolizing them and making them more soluble in water for elimination in the feces or the urine. The enzymes, which are usually bound to microsomes (structures that are part of the smooth endoplasmic reticulum of the hepatic cell), develop gradually in the fetus, attaining adult levels after birth. These enzymes detoxicate many compounds from outside the body and also work on various substances that are produced in the body—*e.g.*, bilirubin and thyroxin; they have a limited effect on the steroid hormones. Some drugs are detoxicated in the mitochondria of the hepatic cell. Detoxication, which is usually thought of as removal of the poisonous properties of a substance, also includes the modification of a drug to make it less active and more easily eliminated.

Drug tolerance

The effect of certain drugs, particularly barbiturates, is decreased and of shorter duration if the drugs are used repeatedly. This is due in part to a decrease in the sensitivity of the brain but also to metabolic processes in the hepatic cells. The barbiturates, and many other drugs soluble in fats, alter the microsomal enzymes and make them active; the effect is not confined to the particular enzymes that are involved in the detoxication of the drugs. This is called enzyme induction.

Among the substances that activate enzymes are alcohol, coffee, food additives—such as butylated hydroxytoluene (BHT), an antioxidant used to preserve fats—and DDT (dichlorodiphenyltrichloroethane), which may contaminate food that is eaten.

Because diseased livers metabolize drugs more slowly than normal, caution is exercised in the administration of sedatives to persons with liver disease, especially if they have shown any tendency to develop hepatic coma.

Malfunction and diseases

JAUNDICE

Jaundice, the yellowness of the eyes and skin that comes from excess amounts of bile pigment—bilirubin—in the blood, may arise in any of a number of ways. (1) The hepatic cells may form more bilirubin than usual because of an abnormally high level of red blood cell destruction. This type is called hemolytic jaundice. (2) Impaired uptake or transport of bilirubin by the hepatic cells may occur (this congenital disorder is known as Gilbert's syndrome). (3) A defect within the hepatic cell itself or an obstruction in the bile duct system may prevent the excretion of bilirubin glucuronide into the bile. (4) A number of these defects may occur at one time. In hepatocellular jaundice, for example, there may be abnormalities in the transport of bilirubin, in its combining with glucuronic acid to form bilirubin glucuronide, and in its excretion in the bile.

In hemolytic jaundice the skin has a characteristic lemon-yellow colouring, and there is usually anemia. In the urine there are excess quantities of urobilinogen, a product of the reduction (removal of oxygen) of bilirubin. The urine contains little or no bilirubin. The feces are dark brown because of the presence of excess quantities of urobilinogen, also called stercobilinogen. There are signs in the peripheral blood—*e.g.,* blood from the fingertips or ear lobes—that red blood cells are being destroyed and regenerated in abnormal numbers. Hemolytic jaundice and its varieties are covered in detail in BLOOD DISEASES.

In hepatocellular, or liver-cell, jaundice the hepatic cells are unable to function as a result of hepatitis (inflammation of the liver) or cirrhosis (see below) or because drugs that damage the liver have been taken. The feces are ordinarily pale, and the urine contains both urobilinogen and bilirubin.

Obstructive jaundice

Obstructive jaundice, also called cholestatic (bile-stopping) jaundice, results from interference with the flow of bile from the liver into the duodenum. The causes include gallstones, cancer of the head of the pancreas, and cancer of the main bile ducts. The jaundice may come from the reaction of the liver to a drug such as chlorpromazine or to the hepatitis virus.

Obstructive jaundice causes the skin to itch. If the jaundice continues for some time, cholesterol may be deposited in the skin, particularly the skin of the eyelids. If the obstruction is a gallstone, the affected person may be feverish or may feel pain in the abdomen. If cancer is the cause there may be loss of weight. The feces are pale because they lack normal amounts of bile pigment and contain abnormal quantities of fat (steatorrhea). The steatorrhea occurs because inadequate amounts of bile salts have reached the intestines, so that fats in the diet are not digested.

TESTS OF LIVER FUNCTION

Tests of liver function help in identifying liver disease, in estimating the damage to the liver and the chances of recovery, and in evaluating the treatment.

The levels of bilirubin in the blood are determined to confirm the presence of jaundice and to follow progress. Increased quantities of the enzyme alkaline phosphatase are found in the blood serum when the bile ducts are obstructed or when the liver is cirrhotic or contains tumours or abscesses. Because high serum levels of the enzyme may be caused by bone disease, further tests may be necessary; levels of the serum 5-nucleotidase enzyme are normal when there is disease of bone and elevated in the presence of liver or bile duct disease.

Obstruction of bile ducts may elevate blood cholesterol levels. If the obstruction continues for some time, cholesterol may be deposited in the skin. These deposits, called xanthomas, are found especially in the upper eyelid.

When the hepatic cells are damaged by disease, various enzymes, including glutamic oxalacetic transaminase (GOT, also called aspartate aminotransferase) and glutamic pyruvic transaminase (GPT, also called alanine aminotransferase), leak into the blood. Determination of transaminase levels in the blood serum helps in early identification of acute virus hepatitis, especially during epidemics, and in uncovering cases not made evident by the presence of jaundice.

Levels of glutamic pyruvic transaminase are especially high in active chronic hepatitis and are elevated in liver disease due to alcoholism.

Low serum levels of albumin suggest the presence of chronic liver disease such as cirrhosis, and abnormally high serum levels of the protein globulin are suggestive of chronic liver disease such as active chronic hepatitis and cirrhosis of the alcoholic when in relapse. The type of globulin that is present in abnormally large amounts also is significant as a clue to the type of abnormality: increase of gamma globulin, for example, suggests hepatic-cell disease, while an increase of alpha and beta globulin is indicative of bile-duct obstruction.

Bromsulfalein test

A bromsulfalein (BSP) test is of value in evaluating liver function in the absence of jaundice. The dye bromsulfalein is injected into a vein, and blood samples taken from veins 30 to 45 minutes later. After 45 minutes persons whose livers function normally retain in their blood 4 percent or less of the dye administered. Retention of greater amounts of the dye is observed in a person whose liver is fatty but not cirrhotic. The test may suggest that a drug that is being administered is injuring the liver.

Needle biopsy, a method of obtaining a sliver of liver tissue for examination, is performed in a hospital and is carried out only on persons whose blood clots normally. A small hollow needle is inserted into the liver after administration of a local anesthetic, and a core of tissue is obtained. The fragment of tissue is examined under a microscope, and further investigations may be made into whether disease organisms are present. Diagnosis of cirrhosis, tumours, infections, and jaundice may be reached by this method. A series of such biopsies is sometimes carried out to follow the results of treatment.

In scintiscanning, a small amount of radioactive substance is introduced into the body and the scintillations, or flashes, on a screen sensitive to the radiation are recorded by an automated process. The apparatus required for scintiscanning requires a heavy investment, but the method does not tax the patient. The positions of tumours, cysts, and abscesses can be determined in this way, and liver deformities can be indicated. If the liver as a whole shows a patchy uptake, cirrhosis may be suggested.

LIVER DISEASE

Symptoms and signs of liver cell failure. The hepatic, or liver, cell may fail to function because of hepatitis or cirrhosis, less commonly, because tumour cells have spread through the liver or because blood flow through liver arteries or veins is obstructed. The hepatic cells may be poisoned by such substances as carbon tetrachloride, phosphorus, or certain drugs.

The most common symptoms are weakness and a tendency to become fatigued. Jaundice is another indication. The skin is usually warm, and the blood pressure low. There is often a slight fever. The breath has a characteristic sweetish, slightly fecal odour, known as fetor hepaticus. Inability of the hepatic cell to synthesize blood-clotting factors may lead to easy bruising and nose bleeds. In the skin of the face, shoulders, arms, and hands, lesions can be seen that are called vascular spiders because of their resemblance in shape to spiders. Each lesion is made up of a central arteriole (minute artery), which forms the spider body, and numerous small vessels radiating from the arteriole like spider's legs. The larger vascular spiders may pulsate. The palms of the hands may be bright red, an appearance that may also be present in rheumatoid arthritis, in pregnancy, and when the thyroid gland is overactive.

Mental and neurologic changes

Liver disease of almost every type may be complicated by mental changes. The affected person may be apathetic, slow to respond, and then sleepy and finally may go into a coma. Mental capacity may decrease in the chronic case. Speech may be slow and slurred. The so-called flapping tremor is the most characteristic neurologic symptom. This is seen when the person affected stretches his arms out in front of him. A person with this tremor may tend to drop things and may experience difficulty in handwriting.

The brain waves, as recorded in the electroencephalogram, are symmetrical, large, and slow—changes similar to those found in other metabolic disturbances of the brain.

The mental disturbances in liver disease have not been completely explained. A meal heavy in proteins may precipitate the symptoms, and they may be decreased by oral antibiotics. These characteristics suggest that some toxic substance formed in the intestines by the action of bacteria on proteins may be responsible, at least in part, for the mental disturbances. The toxic substance, which may be ammonia, may reach the brain because the liver has failed to remove the substance from the bloodstream. Other possible causes include low levels of electrolytes (*e.g.,* sodium, chloride, bicarbonate, in solution) in the blood as a result of efforts to remove excess fluids from the tissues or of bleeding into the gastrointestinal tract.

Treatment directed toward relief of the mental disturbances includes restriction of dietary protein, maintenance of calorie intake, oral administration of antibiotics, and measures to maintain free bowel action.

Ascites, the accumulation of fluid in the peritoneal cavity (between the membrane lining the abdominal cavity and that covering the organs and structures in the abdomen), is usually preceded, in persons with liver disease, by accumulation of fluid in the ankles. Ascites causes the abdomen to feel unpleasantly distended and may interfere with breathing. With the increased body weight there may be abdominal hernias. The muscles become extremely wasted. The fluid is examined by taking a sample through a puncture in the abdominal wall (paracentesis), and the affected person is placed on a low-sodium diet and given diuretics—drugs that increase urine production.

Acute hepatitis. Acute virus hepatitis (inflammation of the liver), which includes the varieties known as infectious hepatitis (IH) and serum hepatitis (SH), is believed to be caused by more than one virus, although the viruses have not been conclusively identified. In the 1960s an antigen (a substance, most often a protein or a polysaccharide, that causes the formation of antibodies) was found in 20 to 40 percent of persons who were in the acute stage of virus hepatitis. The antigen, known as Australia antigen, or hepatitis-associated antigen, is identified in persons with serum hepatitis and may actually be part of a virus that causes hepatitis, for persons given transfusions of blood containing the antigen often develop recognizable virus hepatitis.

The only proven sources of infection with virus hepatitis are human blood, urine, and feces. That the virus is present in high concentration is suggested by the fact that as little as 0.0005 millilitre (0.00002 ounce) of whole blood is infectious. Immunity probably develops with age, since the disease is less common in older persons than in children and young adults, but susceptibility to serum hepatitis does not decline in the same manner.

Spread of infectious hepatitis

Infectious hepatitis is usually spread by fecal contamination of food or drink, although it can be transmitted by infection of blood. The disease may occur as sporadic cases after person-to-person contact or in large epidemics after widespread contamination of water or food. At least three epidemics have been caused by the consumption of raw clams or oysters from polluted waters. The disease becomes manifest about 15 to 40 days after exposure.

Serum hepatitis, which causes symptoms about 50 to 160 days after exposure, is usually transmitted by injection. The virus may be in whole blood, plasma, fibrinogen, or the globulin used to treat hemophiliacs, or infections may come from the use of instruments such as syringes, needles, or lancets that have been contaminated with human blood containing the virus.

A person is usually immune for the rest of his life after an attack of infectious hepatitis and probably always gains immunity from an attack of serum hepatitis. Neither form of virus hepatitis brings immunity to the other form, however.

Serum and infectious hepatitis cause the same abnormalities in the liver tissue. These are essentially inflammation and destruction of hepatic cells. The framework of the liver is not affected, and, with recovery from attack, the liver tissues usually return to normal.

Attacks of virus hepatitis vary from slight distress to a serious illness that culminates in coma and death. Usually there is jaundice after three days to two weeks of ill health marked by failure of appetite, nausea, aversion to smoking, and a dull pain in the right upper part of the abdomen. There is moderate fever. Just before the appearance of jaundice, the liver is tender and perceptibly larger. The urine takes on an orange-brown colour, the feces lighten, and jaundice shows in the whites of the eyes. When jaundice spreads to the skin some of the physical distress disappears and the appetite improves. For a few days the skin may itch. During the illness there may be a loss in weight of about ten pounds (five kilograms).

About three to ten days after the onset of jaundice the convalescence begins. Feces regain their usual colour, and the jaundice disappears.

An adult may be ill for two to six weeks and may experience lassitude and a tendency toward fatigue for as long as six months.

From 2 to 15 percent of persons who recover from virus hepatitis have relapses. Some of the relapses are like mild recurrences of the original attack, but most are manifested merely by chemical abnormalities in the blood. Usually when a relapse occurs it is because of a too early return to activity or to drinking alcohol. After relapses, as after the original attack, there is usually complete recovery.

Children sometimes have acute attacks of hepatitis without showing jaundice. The symptoms are like those of gastroenteritis; that is, like the effects of inflammation of the stomach and intestinal lining. Tenderness and enlargement of the liver may be noted. Identification of the disease may be based upon a number of factors, including a history of association with a person who had the disease, the finding of bilirubin in the urine, evidence that the liver is not functioning normally, or, occasionally, the results of a needle biopsy of the liver. Cases of hepatitis not accompanied by jaundice may account for the way in which immunity to the disease spreads in a community.

Fulminant hepatitis

Rarely, a hepatitis attack is of a severe type, sometimes called acute fulminant hepatitis, that makes the patient gravely ill usually within ten days of the onset of the disease.

In the cholestatic form of jaundice the decrease in the flow of bile is similar to that which occurs in obstructive jaundice after surgery. The skin itches, and jaundice is persistent. Recovery is usually complete after an illness of eight to 20 weeks.

In major outbreaks of hepatitis, the death rate is about one to two persons per 1,000. Infectious hepatitis causes fewer deaths than serum hepatitis. An average young adult usually is ill for about six weeks and rarely is ill longer than three months. In children the attacks tend to be milder. Cirrhosis as a result of hepatitis is extremely rare, and, to judge from follow-up investigations of epidemics in the Middle East in World War II, in the Korean War, and in Delhi, an attack of hepatitis does not increase the likelihood of chronic liver disease.

To prevent the spread of infectious hepatitis, great care must be exercised in the practice of sanitation, especially with respect to the disposal of feces. The feces of an infected person contain virus for as much as two weeks before jaundice appears, so that the disease may be widespread before a diagnosis is made. Consequently, isolation of ill persons and of contacts is unlikely to have much effect on the spread of the hepatitis.

Persons (and particularly pregnant women) who are in close association with one who is ill with hepatitis are sometimes given an injection of gamma globulin to prevent spread of the disease.

The spread of serum hepatitis is prevented by avoiding unnecessary use of blood products and by taking care that any injections that are made are sterile. Blood to be transfused must be screened for Australia antigen. Whole blood that is contaminated with hepatitis virus cannot be sterilized. Blood donors who have previously given blood a number of times without infecting recipients with hepatitis are the safest donors. It has not been definitely proved that gamma globulin is effective in preventing serum hepatitis.

Current forms of treatment do not hasten recovery from virus hepatitis. A person with the disease should stay in bed as long as he has jaundice; his convalescence lasts about twice as long as the period of bed rest. Early in the disease, when there is a tendency toward nausea, a diet low in fat is recommended. A patient is advised to increase his exercise gradually after recovery and to abstain from alcohol for at least six months and preferably for a year.

In newborn infants a number of organisms can cause hepatitis, including German measles (rubella), Coxsackie and ECHO (enteric cytopathogenic human orphan—

which means that the virus attacks the cells of the human intestine but was once thought not to cause disease) viruses, and the organisms that cause toxoplasmosis and syphilis.

Drug-associated liver injury. Jaundice may result from injury to the liver by such substances as carbon tetrachloride, cytotoxic (cell-destroying) drugs used in treatment of psoriasis or leukemia, phosphorus, or the tetracycline antibiotics, if extremely large doses are given by injection into a vein.

Drugs such as the newer drugs used in treatment of tuberculosis or the anesthetic halothane may, in very rare instances, have effects almost indistinguishable from those of virus hepatitis. The illness is not affected by the size of the dose and seems to be caused by individual sensitivity to the drug. Treatment is similar to that for virus hepatitis.

An obstructive type of jaundice can result from taking certain steroid compounds, including some that are contained in some of the oral contraceptives. The jaundice is more severe if the affected person already has some form of liver disease, such as cirrhosis or hepatitis. The jaundice tends to disappear as soon as taking of the drug is stopped.

Causes of cirrhosis

Cirrhosis of the liver. Cirrhosis, a chronic liver disease that is irreversible, is characterized by widespread fibrosis and nodule formation (replacement of functioning liver tissue with bands and small lumps of scar tissue). It has a number of possible causes, including virus hepatitis, alcoholism, obstruction of bile channels, heart failure, obstruction of the hepatic vein, deposition of iron (hemochromatosis) or copper (Wilson's disease), and some forms of metabolic disease. In many instances the cause is not known.

Cirrhosis causes hepatic-cell failure and abnormally high pressure in the portal vein—portal-venous hypertension. The effects of hepatic-cell failure have already been touched upon. They include fatigue, jaundice, easy bruising and bleeding, mental changes, ascites, and the appearance of vascular spiders.

The portal-venous hypertension is the result of fibrosis and formation of nodules; these interfere with portal blood flow through the liver. The portal vein brings venous blood to the liver from the intestines and the spleen. Obstruction to the portal-venous system and resultant hypertension in the portal vein cause back-pressure enlargement of the spleen, and this, in turn, causes the spleen to be overactive in destruction of red and white blood cells and platelets. Blood in the portal vein, being prevented from flowing through the liver, is diverted to important veins in the area of the stomach and the esophagus (gullet). Rupture of these swollen veins leads to hematemesis and melena—vomiting of blood and passing blood in the stools. This loss of blood is a common cause of death in persons with cirrhosis. The swollen veins in the esophagus and the state of the portal system can be detected by X-rays. The blood pressure in the portal system can also be measured directly.

Enlargement of the spleen or liver, jaundice, gastrointestinal bleeding, and ascites are among the early effects of cirrhosis. In treatment, the cause of the cirrhosis is eliminated if possible. For example, heart failure is controlled and obstruction of bile ducts is removed. Avoidance of alcohol, restriction of salt (sodium) in the diet, and administration of diuretics are other measures taken. If coma threatens, the amount of protein eaten is reduced and oral antibiotics (*e.g.*, neomycin) may be given. If there is bleeding, blood volume may be restored by transfusion and measures may be taken to reduce the hypertension and stop the bleeding.

Alcoholism is clearly associated with cirrhosis of the liver. There is a direct relationship, in Western countries, between the amount of alcohol consumed and the prevalence of cirrhosis. Daily ingestion of at least a pint of spirits (hard liquor) for more than ten years seems necessary for severe liver damage to occur. It is the quantity of alcohol that is important, whether it is in beer, spirits, or wine. The incidence of cirrhosis is, however, only about one in 12, even among severe alcoholics.

The injurious effect of alcoholism on the liver is principally from its interference with the metabolic processes of the liver. The tendency of the alcoholic to eat inadequate amounts of protein probably contributes to the damage. In the liver are gross fatty changes, and inflammation surrounds areas in which hepatic cells are destroyed (alcoholic hepatitis). Later, genuine cirrhosis develops.

Effects of alcoholism

In early stages of the disease, the alcoholic shows no loss of weight and may even be obese. Abnormalities present may include anemia, vitamin deficiencies, inflammation of the pancreas, and inflammation of the peripheral nerves. Chronic gastritis (inflammation of the stomach lining) results from loss of appetite, early morning nausea, and retching. The affected person has difficulty in concentrating and exhibits a tremor. The ankles are swollen and there may be nosebleeds. Enlargement of the liver and vascular spiders on the skin are other features.

The cirrhosis resembles that resulting from other causes but the prospects for survival in alcoholic cirrhosis are better than in other types. If the alcoholic abstains, he has two chances in three of surviving five years. He has two chances in five of surviving this long even if he continues to drink.

An alcoholic with cirrhosis shows striking improvement in condition when he has rested in bed and stopped consumption of alcohol. If coma does not threaten, a diet high in protein hastens recovery. Such people must avoid all forms of alcohol for the rest of their lives.

Active chronic hepatitis. Active chronic hepatitis, sometimes called lupoid hepatitis, is a form of cirrhosis that attacks primarily young persons, usually girls and young women between the ages of ten and 30. Most cases have no obvious cause, although some seem to progress from an acute virus hepatitis. There is persistent jaundice and usually there is failure to menstruate. About 50 percent of persons with the disease also have other abnormal conditions that seem to have no relation to the liver; these may include pain in the joints, ulcerative colitis, diabetes, and thyroiditis and nephritis (inflammation of the thyroid and the kidney).

In the liver there is inflammation and formation of scar tissue, resulting finally in true cirrhosis. There is substantial evidence that there is an immunological disturbance involved (a reaction against a foreign protein or against a protein in the body's own tissue). For this reason corticosteroid drugs are given. They do not prevent the progression to cirrhosis but do prolong life, bringing relief from fatigue and decrease in jaundice.

The mean survival from this type of cirrhosis is about five years, but some affected persons survive as long as 16 years.

Miscellaneous forms of cirrhosis. Hemochromatosis, a type of cirrhosis in which there are abnormal deposits of iron in the liver and other organs, causes diabetes because of the iron deposits in the pancreas, and sexual incapacity because of deposits in the endocrine glands. In treating persons afflicted with hemochromatosis, iron is removed from the body by multiple bloodlettings (venesections). This is very successful.

Wilson's disease

Copper deposits in the liver and in other tissues are characteristic of a rare form of cirrhosis known as Wilson's disease, or hereditary hepatolenticular degeneration. Deposits of copper in the brain may cause tremor and other abnormalities related to the nervous system. A brownish-green coloration of the cornea of the eye, called Kayser–Fleischer rings, is also caused by the copper deposits. Wilson's disease is treated by administration of penicillamine. This drug removes copper from the tissues and causes its excretion in the urine.

Obstruction of the bile ducts, with pigmentation and itching of the skin, may lead to biliary cirrhosis. There are two types, called respectively primary and secondary biliary cirrhosis. The primary type, a rare disease that occurs chiefly in middle-aged women, begins with inflammation of the small bile ducts in the liver, which are finally destroyed. Secondary biliary cirrhosis may occur after a chronic obstruction of the bile ducts, as from gallstones or the constriction of bile ducts after surgery.

Tumours of the liver. Both benign and malignant growths affect the liver, but the benign growths are of little significance. Cancers of the liver are common. Of these, secondary cancers (growths that originate outside the liver) are 25 times more common than primary cancers.

Primary carcinoma of the liver cells (malignant hepatoma). Malignant hepatoma usually follows upon cirrhosis, although about one case in three occurs spontaneously. This type of cancer occurs most frequently in tropical areas. The prevalent theory of its causation is that it is due to ingestion of food contaminated with carcinogens from fungi. It may also be related to Australia antigen-positive virus hepatitis.

Among the early manifestations of the cancer are pain in the area of the liver and enlargement of the organ. There may be ascites and some of the features of cirrhosis.

No cure is currently available. On the average, the terminal illness lasts about seven months.

Secondary tumours of the liver. Metastases—secondary tumours spreading through the bloodstream—occur more frequently in the liver than in any other organ. One-half of the cancers of the stomach, breast, lung, pancreas, and colon—and about one-third of all metastasizing cancers—metastasize to the liver. The liver may contain only one or two microscopic growths, or the secondary tumour may involve the whole liver, which becomes enormously enlarged. Treatment is only palliative, there being no cure available.

Infections of the liver. *Pyogenic liver abscess.* Pus-forming abscesses of the liver are less common than formerly, as a result of the development of antibiotic treatment of infection and the earlier diagnosis of acute appendicitis, which is the most frequent cause of liver abscess.

Effects of the presence of the abscess include shivering, sweating, a high temperature, and weakness. Tenderness and enlargement of the liver are noted. The location of the abscess is determined by X-rays and scintiscanning, antibiotics are administered, and the abscess may need to be drained by surgery.

Abscesses, of a type called cholangitic, develop when pus forms in the bile ducts inside the liver above a mechanical obstruction in the main bile ducts. Gallstones and narrowing of the duct after surgery are the principal causes. Symptoms include pain in the liver, jaundice, and fever. The person affected is given antibiotics and the common bile duct is drained by surgery.

Amebic liver abscesses

Amebic liver abscesses are caused, as the name indicates, by an amoeba *(Entamoebn histolytica)*, a parasite which lives in the wall of the large intestine and is brought to the liver in the portal vein. An enzyme produced by the amoeba destroys liver tissue and causes formation of an abscess. Amebic liver abscess is ordinarily a disease of the tropics and subtropics.

About one-fourth of the persons with amebic liver abscess have had amebic dysentery previously. Sufferers are ill, but their fever is moderate (their temperature rarely is higher than 103" F, or 39° C). There is pain over the liver, and the liver itself is tender and enlarged. If the abscess ruptures and contaminates the lungs or the abdominal cavity, there may be a secondary pyogenic (pus-forming) infection that causes the patient to be gravely ill, with a high fever.

The abscess is located by use of X-rays and scintiscanning, and the pus may need to be aspirated (drawn out with a syringe and needle). Amebic dysentery can be diagnosed by the appearance of the pus, which is characteristically thick and brown. The response to drainage of the pus and treatment with anti-amebic drugs is often striking.

Tuberculosis of the liver. Generalized tuberculosis usually affects the liver, but ordinarily this aspect of the disease is not important and is not even discovered during the patient's lifetime. The diagnosis of tuberculosis of the liver may be made by needle biopsy of the liver. Jaundice may be present, not as the result of tuberculosis but because of the effect of antituberculosis drugs or as a result of a simultaneous attack of virus hepatitis.

Syphilis of the liver. Syphilis of the liver is now rare. When it occurs it may be as a result of heavy involvement in congenital syphilis, with the ultimate result of scars around the hepatic cells. Jaundice in newborn infants may be one result of this involvement. The liver may also be infected in the secondary stage of syphilis, and in tertiary syphilis there may he liver abscesses or gummas (rubbery, tumourlike formations). Treatment is by antibiotics.

Weil's disease. Weil's disease is a form of hepatitis caused by infection with *Leptospira icterohaemorrhagiae* from the urine of infected rats. It occurs most often in men engaged in such pursuits as farm work or coal mining. The disease results in deep jaundice; but death, when it occurs, is usually from heart and kidney involvement.

Schistosorniasis (bilharziasis). Schistosomiasis, infestation with the parasite *Schistosoma mansoni* or, more rarely, with *Schistosonza japonicum,* is found in Africa, the Far East, and areas of South America and the Caribbean. The organisms penetrate the liver by way of the portal vein and cause inflammation within the liver in the terminal branches of the portal vein. The inflammation results in obstruction of these small veins and a rise in blood pressure within the portal vein. The portal hypertension in turn causes enlargement of the spleen and development of varicose veins in the esophagus (gullet). If the veins rupture, there is vomiting of blood. Treatment is with preparations of antimony (tartar emetic).

Hydatid (echinococcal) disease. Hydatid disease, also called echinococcal disease and echinococcosis, is infestation with a small tapeworm, *Echinococcus granulosus,* the larvae of which form cysts (hydatids) in the liver and lungs. The disease is important in sheep-raising areas, especially Australia, New Zealand, Africa, South America, and southern Europe, where the infestation spreads to man from the excreta of dogs that have eaten the entrails of infected sheep. The fluid of the hydatids contains a foreign protein that sensitizes the host. The sensitization may develop into severe anaphylactic shock but more often causes recurrent itching wheals on the skin. The cysts are dealt with surgically.

Effects of hydatid fluid

Liver flukes. Flukes that invade the liver include the Chinese liver fluke *(Clonorchis sinensis)* and the common sheep fluke *(Fasciola hepatica). Clonorchis sinensis* is found in the inhabitants of East Asia and in immigrant Chinese elsewhere. It enters the body in raw or partially cooked fish and causes inflammation of the bile ducts (cholangitis). Of the instances of primary cancer of the liver observed in Hong Kong about one case in six occurs after infestation with liver flukes.

Fasciola hepatica is most common in middle and western Europe and in the Caribbean. Infestation in man comes principally from sheep by way of an intermediate host, a snail that contaminates watercress and other greens, particularly in wet summers.

Glycogen-storage disease. The glycogen-storage diseases, resulting from the excessive retention of glycogen in the tissues, are inherited and of several types, which have differing enzymatic and structural defects in the glycogen and vary in the severity and course of the disease. In all types there is failure to maintain glucose at normal levels in the blood by the liver's production of glucose and the influx of glucose from the intestines.

The first sign of glycogen-storage disease in the infant is usually an enlarged liver. The spleen is not enlarged as it is in the lipid-storage diseases.

The diagnosis is confirmed by noting the failure of glucose levels in the blood to rise sufficiently in response to stimulation of liver breakdown of glycogen.

BIBLIOGRAPHY, C. ROUILLER (ed.), *The Liver: Morphology, Biochetnistry, Physiology, 2* vol. (1963–64), a modern review of liver function and structure in health and disease; L. SCHIFF (ed.), *Diseases of the Liver,* 3rd ed. (1969), a standard text covering liver diseases in man, with extensive bibliography; H. POPPER and F. SCHAFFNER (eds), *Progress in Liver Diseases,* vol. 3 (1970), one of a series dealing with present-day research topics; S. SHERLOCK, *Diseases of the Liver and Biliary System,* 4th ed. (1968, rev. 1971), an up-to-date text, with comprehensive bibliography.

(S.S.)

Liverpool

A city of barely 520,000 inhabitants and the core of Merseyside, a metropolitan area containing some 1,500,000 people, the Liverpool of the early 1980s remains an enigma in modern Britain, a unique blend of cosmopolitanism and parochialism. Three centuries of far-flung maritime trade may have dominated the life of Liverpool and broadened the horizons of its citizenry, but a fierce loyalty and pride, far in excess of the normal "hometown" feeling, remain interwoven with these outgoing attitudes. A visitor may find this local affection for a grimy commercial centre, from which some of the worst British examples of overcrowding and poverty are only gradually being eradicated, difficult to understand. The feelings in question may stem, in part, from the city's long-standing social isolation from adjoining Lancashire, but they are primarily derived from the warmth and vibrance of a folk culture second to none in Britain. There is little modern significance in the distinction between the Liverpool man of commerce and the Manchester industrialist expressed in the 18th-century aphorism "Liverpool gentleman, Manchester man." The separate character of Merseyside stems from a 19th-century immigration whereby impoverished Celtic (and particularly Irish) migrants contributed a distinctive accent and cultural style to the area. A paradox was thus created: the confirmation of the social isolation of Merseyside at a time when the economic integration of the port with its industrial hinterland was proceeding apace.

The special character of the city

Although the flat catarrhal accents, blunt aggressive style, and deadpan expression of the typical "scouse" or "wacka" (informal variants of the more official "Liverpudlian") are restricted to a few miles around the docks, the continued popularity of the seemingly endless stream of local comedians and other entertainers who have won national acclaim has made these attributes familiar to the rest of Britain. Such national fame may also be attributed to the success of two of the premier soccer teams in the country, Liverpool ("the Reds") and Everton ("the Blues"), who are urged on each week during the football season by regular local crowds of more than 45,000 enthusiastic—and often choral—supporters. It was during the 1960s, however, that local characteristics received their widest exposure: under the leadership of the Beatles, the vibrant sounds of the "Mersey Beat," a rhythm-and-blues derivative with marked local features, made Liverpool, for a few hectic years, the Mecca of the international world of popular music. By the early 1980s, the peak of such adulation had passed, but folk culture, in music and in verse, remains very much alive in Liverpool public houses and clubs.

In more formal terms, Liverpool is an urban centre of northwest England with most of the economic and cultural attributes of a large city: two daily newspapers, the larger of which, the ***Liverpool Daily Post,*** has an extensive circulation outside the city in North and Mid Wales; one of the largest provincial universities and medical schools in the United Kingdom, with an enrollment exceeding 7,000 students; a professional symphony orchestra (the Royal Liverpool Philharmonic), two large choirs (Philharmonic and Welsh Choral), and many other smaller orchestral and choral groups. Liverpool entered the 1980s with generally well-supported theatres, and the city is one of the few in Britain that does not depend on massive infusions of government aid for the arts. Strong circulation for a local monthly magazine in this field reflects this interest.

Architecture

Such commitment to the arts is not, however, reflected in the physical fabric of the city. Liverpool does possess small features of architectural work, notably the restored Georgian elegance of Abercromby Square and other precincts around the university, together with the solid porticoed magnificence of the Albert Dock and other early 19th-century granite quays and wharves. Apart from these individual buildings and the extensive parks that occupy almost one-twelfth of the city, Liverpool lacks any architectural or environmental areas of distinction. Even the Neoclassical buildings comprising the central cultural precinct of St. George's Hall, the Merseyside County Museum and Library, and the Walker Art Gallery present a confused, rather than a dramatic, perspective due to an orientation along rather than across their sloping site. The visual effect of this area is further lessened by the litter of traffic signs and parking lots lying close to the Mersey Tunnel, which is the major routeway under the Mersey to the Wirral Peninsula, nearby Cheshire, and beyond. Only in the two cathedrals is any sense of civic drama restored, and even these are of recent origin. The red sandstone Anglican Cathedral, the largest church of its denomination in England, was begun in 1904 by King Edward VII. Built in a massive neo-Gothic style, the cathedral was completed in 1978. Its nearby Roman Catholic counterpart, the Metropolitan Cathedral of Christ the King (1962–67), is constructed in contrasting materials of concrete, glass, and aluminum. It has an ultramodern design: the central altar is surrounded by a cone-shaped building topped with a pillared cylindrical core representing a crown of thorns.

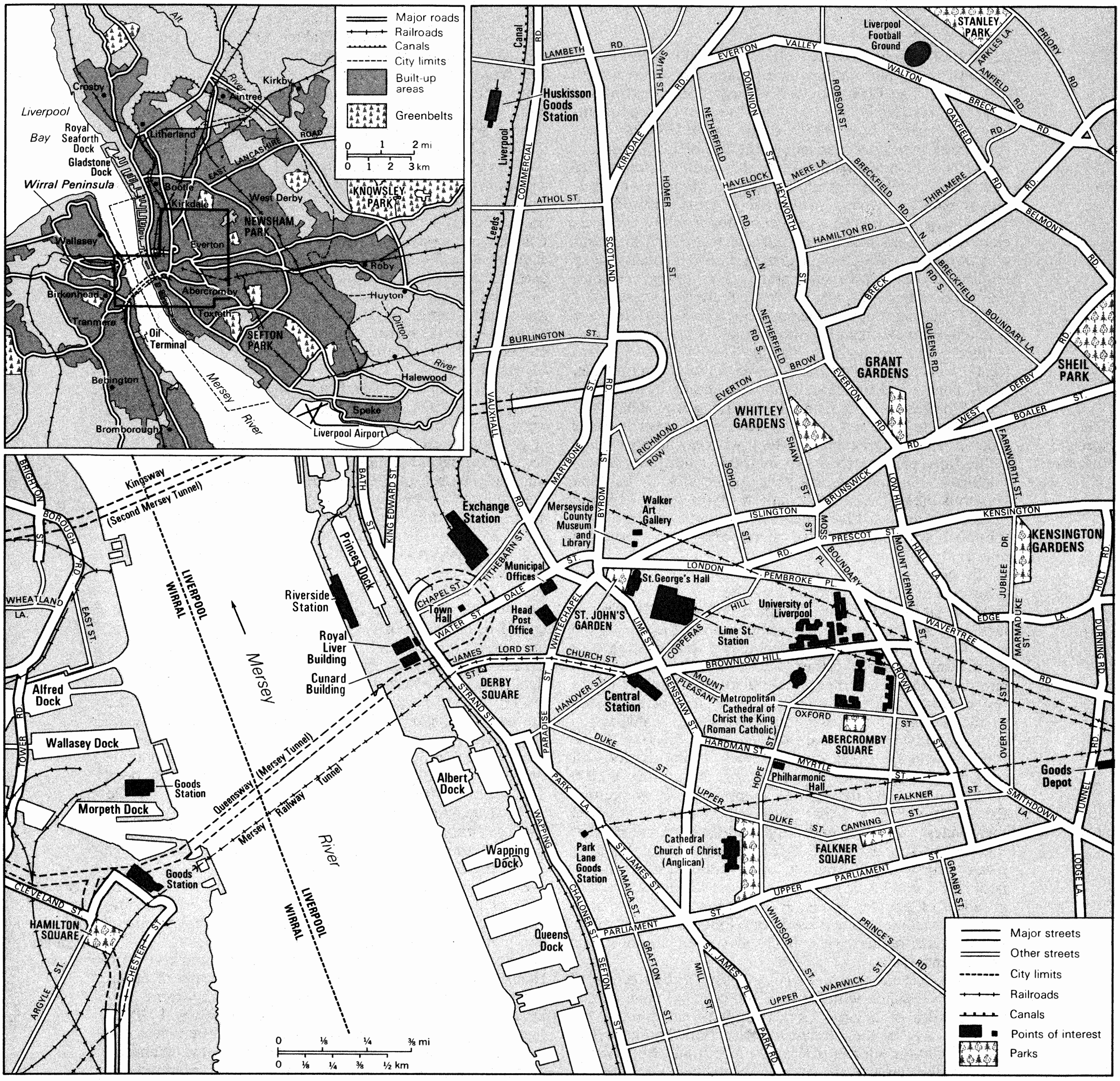
History. The first significant date in the history of Liverpool is 1207, when King John of England granted a charter for a planned new town. Although it was located at the narrowest part of the funnel-shaped Mersey Estuary, the initial advantage the town derived from its site was the presence of a short but widemouthed tidal creek, draining one of the sandstone ridges that rise 200 feet (60 metres) on both sides of the river above a gently undulating plain. Only in the last two centuries, with the technological development of ships with a deeper draft, has the tidal scour created by the shape of the estuary and the depth of water (28 feet at low spring tide) become an asset to development. In all probability, the town acquired its name from a hamlet lying on the other side of the "pool" created in the creek at high tide, but this partial continuity of settlement was of little economic significance. The stimulus to the growth of a small coastal trading centre came from the physical protection of the site and the legal privileges provided by royal charters.

Medieval growth was slow and halting. Not only was the town located in what was a remote and sparsely populated corner of England but also the area had long been dominated by the fortress and port of Chester on the River Dee. As a result, the young settlement did not reach a population of 2,000 until 1660. By then, the silting of the Dee, allied with the development of larger ships, tilted the balance of the Irish and coastal trade from Chester to Liverpool. It was the profitable return in 1667 of the ship "Antelope" from a trading journey to the West Indies that may be taken to signify the first major boom in the importance of Liverpool.

Development of the city as a port

During the next 70 years, the growth of commerce with North America, combined with the illegal spoils of privateering, made Liverpool the second most important port in Britain, with a population reaching 15,000 by 1730. A significant element in the general trading pattern was the famous, or infamous, Liverpool Triangle—the exchange of manufactured goods from the Mersey hinterland for slaves in West Africa who were in turn traded for sugar, molasses, spices, and other plantation crops in the West Indies. Although modern research has called into question the continuity implied by the threefold links, these activities were still of considerable economic importance. Not only were three sets of profits possible but the trade provided a stimulus to the manufacturers of an economically poor region, and it is the gradual industrialization of its hinterland that provided the key to Liverpool's progress.

Such developments placed additional demands upon the initial site of the town, which proved inadequate. The pool was filled in by 1715, creating the first dock, and by the end of the century four other docks were established along the Mersey, enclosing 30 acres (12 hectares) of water and forming two miles (three kilometres) of quays, so that the port outranked even London in dock space. Yet capitalization upon the advantages of the estuary rather than the pool was only one of the preconditions for the city's growth. A more important one was the construction of an economic transport network to the hinter-



Central Liverpool and (inset) its metropolitan area.

land, especially to the Lancashire coalfield and the Cheshire saltfields. Local improvements in river navigation and canals were at the forefront of world development and culminated, in 1830, in the Liverpool and Manchester Railway, the first in England to link two major cities. By the mid-19th century, a rail network providing easy and cheap access to all major British industrial centres had been created. The establishment of steam ferry links between Liverpool and the expanding centres of the Wirral was a parallel local development.

Nineteenth-century expansion

These changes were reflected in the growth of population. In 1801 Liverpool had 78,000 inhabitants and was the only town of any size in the area. During the next 50 years the city's population increased fourfold, but the extension of the docks system along the estuary, and later on the Wirral shore, enabled the Merseyside region as a whole to attain a population of 500,000. This growth was characterized by persistently high levels of immigration from surrounding areas and from Ireland, especially during and after the Irish famine (1845–48). Indeed, a measure of the effect of this disaster is the fact that no fewer than 300,000 Irish emigrants disembarked at Liverpool in 1847, many to settle permanently in the area.

Accommodation of such a rapid population increase proved difficult and led to overcrowding and squalor. An 1846 survey, for example, revealed that 40,000 people were living in 8,000 cellars in Liverpool. Expansion of the built-up area beyond the city—at Everton, Kirkdale, and West Derby, for example—did occur, with monotonous rows of terraced houses springing up behind the docks. In the second half of the 19th century many of these small townships were annexed by the city, but subsequent extension of the city limits took place to the south and east, leaving the contiguous built-up areas of Bootle, Litherland, and Crosby to the north as independent administrative entities.

Although immigration continued at high levels, the growth of population until the beginning of the 20th century was dominated by the natural increase of a fertile and youthful people. By 1901 Liverpool had reached a population of 147,000, with more than 1,000,000 in Merseyside as a whole. Clearly dominating its surroundings—and doing so in a much more extensive way than is the case in the six other major British metropolitan regions—Liverpool had become the centre of seven miles of docks extending along the Mersey from Hornby (1884) in the north to Herculaneum (1866) in the south, with additional developments at Birkenhead on the Wirral shore. It was also the principal exporting port in the country, handling no less than one-quarter of Britain's foreign trade, as well as a major part of the extensive overseas emigrant movement.

The 20th century has seen the continuation of many of these trends, particularly after the addition of the Gladstone Dock (1927) at the seaward edge of Merseyside. A more decisive theme, however, has been the internal readjustment of the city, with an intensive effort being applied to the eradication of the problems caused by the rapidity of earlier development. All this has been set against a stagnating local economy, which since 1918 has been reflected in a new demographic trend, namely net emigration. Although 81,000 people left the metropolitan area between 1921 and 1939, 55,000 between 1951 and 1961, and 170,000 during 1966–76, levels of natural increase sufficient to offset these losses continued until the early 1970s. Thus, population gains, combined with decreasing household sizes through the early 1970s, served only to aggravate urban sprawl along both sides of the Mersey. During the decade of the 1970s, however, accelerating emigration and a declining rate of natural increase were estimated to have caused in Merseyside a loss of more than 150,000 persons.

Internal population movements. Within Merseyside, considerable population relocation has occurred. Some has been the product of spontaneous growth, as in the case of the development of largely middle-class suburbs outside Liverpool, on the west of Wirral Peninsula, and to the north of the city, in Crosby and Southport. Other trends—notably the decline of population in the central city—have resulted from the deliberate policy of the city's government. Although most of the inner wards of Liverpool were experiencing population loss by 1871, it was not until after World War II that major changes took place. In 1966 the city council drastically upgraded the scale of its previous attempts at housing improvements by the adoption of a seven-year program for the demolition of 33,000 slum dwellings in the inner city, a policy involving the relocation of 100,000 people. Although, by 1970, about 2,000 houses per year were being modernized with the help of government improvement grants, this process, together with the redevelopment of cleared sites at only half their previous residential densities, did not immediately offset the effects of wholesale demolition. By 1980, however, the city and Merseyside as a whole had a housing surplus. By the early 1980s, the city's population had been declining at the rate of almost 10,000 per year for nearly a decade. The most significant decline was in the four inner wards of Kirkdale, Everton, Toxteth, and Abercromby, which lost 25 percent of their population during the 1960s; it was estimated that they would decline by another 30 percent by 1981.

Slum clearance

Clearance of substandard housing has ensured that Liverpool Corporation plays a major role in the city's housing program. Although this trend is not new—Liverpool constructed the first municipally owned housing block in Britain in 1869—the scale of building by the public sector is important: between 1945 and 1970, 85 percent of new housing was publicly owned, and by the early 1980s, the corporation owned about a third of the total existing housing stock. This proportion is even higher if all those areas of council housing on Merseyside that are the products of Liverpool's earlier housing needs are considered. The housing surplus of the late 1970s, however, reduced the public sector to less than 50 percent of new housing constructed during 1976–78. In the late 19th and early 20th centuries, shortage of land for housing drove the council to seek new territory in districts beyond its boundaries, some of which were subsequently incorporated into the city. Most of the more recent large council housing developments remained outside the political control of Liverpool, for example the former Huyton-with-Roby and Kirkby. The latter was created an urban district in 1958 when its population was already about 30,000. Even these peripheral expansions proved inadequate, and, in the 1960s, more than 7,000 families were moved even farther afield, to such "new towns" as Skelmersdale (in Lancashire) and Runcorn (in Cheshire), specifically designated to receive Liverpool "overspill," and to town-development schemes in Cheshire at Winsford, Ellesmere Port, and Widnes and in Lancashire at Burnley.

George's Landing Stage and the Pierhead, Liverpool. The large office blocks are the Royal Liver Building, the Cunard Building, and the Mersev Docks and Harbour Board.
By courtesy of the City of Liverpool Corporation

The economy. These decentralization trends have not been reflected in any comparable major shift in the location of commercial activity. Liverpool, and particularly its central business district, exercises a greater degree of dominance over the surrounding metropolitan county than is found in other British cities.

The traditional employment categories of transport, communication, distribution, and shipping are still important in the economic life of the city, representing in the early 1970s about 30 percent of the labour force. This, however, was considerably below the 50 percent level of 1929. The change was mainly because of the decreasing local influence of the port, a feature paralleled by national trends. The trend continued during the 1970s, and by 1978 Liverpool was only fourth among the principal exporting ports in the country: by the early 1980s exports were less than a third of London's, and imports were running at four times the level of exports. Because of this national trend, Liverpool had become the sixth largest port in the country by the late 1970s, with a correspondingly smaller percent of the national dock labour force. It controls only about 5 percent of Britain's foreign trade, and only 1.5 percent of the coastwise trade. In addition, the halcyon days of its role as a major passenger port have passed: by the 1970s it was a poor third to Southampton and London, handling only one-seventh of the passengers travelling to centres outside Europe. Such changes could be attributed mainly to the decreasing significance, in the economic life of Britain, of Liverpool's industrial hinterland and its traditional trading partners, the United States and West Africa. Low capital investment and labour strife in the docks intensified the situation.

Economic significance within the nation

Within the port, the long-established dependence on dry cargo has been maintained, with exports dominated by chemicals, iron, and steel; and imports by cereals, ores, scrap, sugar, wood, and pulp. With the completion in

1969 of a new oil terminal at Tranmere, on the Wirral shore south of Birkenhead, capable by 1980 of handling tankers of 275,000 dead weight tons, crude oil became by far the most important commodity imported. About 13,000,000 tons were imported during the terminal's first year of operations, making the Mersey at the time the third largest oil port in Britain; by the late 1970s imports had declined to only about 3,500,000 tons a year, however. Further improvements included the Royal Seaforth Docks, opened in 1971, and extension of the dock system northward beyond the Gladstone Docks into the town of Crosby.

Manufacturing changes

The economic difficulties of the port have been compounded by major structural changes in the manufacturing base of the area. There was a brief spurt of post-World War II prosperity, followed, toward the end of the 1940s, by a drastic decline in the indigenous manufacturing industries, particularly those associated with the processing of imported goods. A total of 18,000 jobs were lost in the process. Merseyside was designated a development area in 1949, and the government assistance that followed resulted in the provision of about 29,000 new jobs, making a net overall gain of about 11,000 jobs. A persistently high rate of population increase combined with these circumstances to give the region an unemployment rate that approached two and a half times the national average during the course of the 1950s. In the 1960s increased powers given to the government in the matter of industrial location stimulated the establishment of new motor-vehicle plants on Merseyside. These were situated at Ellesmere Port (Vauxhall), Halewood (Ford), and Kirkby (British Leyland Motor Corporation) and involved an investment of £65,000,000 generating 33,000 new jobs. Unfortunately, this did not result in the eradication of local unemployment, which remained above the national average, but it did act as a catalyst in the restructuring of the local economy. No less than 17 percent of the insured workers in the area changed their employment categories in the 1960s, compared with 12 percent in the previous decades. Manufacturing industry, which formerly provided over 40 percent of the employment in Merseyside County as a whole, is no longer dependent on the port but rather upon more modern growth industries.

Locational changes followed in the wake of such structural changes, and a general decentralization of manufacturing industry took place. As in the case of population movement, it was the inner areas of the city that experienced the greatest loss; the dock areas saw both a change in industrial type and an absolute loss. So extensive was the inner city's landscape of empty factories, abandoned buildings, and derelict land that, by the early 1980s, planners could contemplate the use of miles of land adjoining the Mersey for development projects. The moribund national economy could not provide investment capital, however, and local economic managers undertook a campaign to attract foreign, particularly U.S., capital.

Most of the new industry attracted during the 1960s was located on the periphery of the Liverpool district, in the new industrial estates at Kirkby, Aintree, Huyton-with-Roby, Speke, and Halewood, as well as over the Mersey at Bromborough and Ellesmere Port, in Cheshire. It was also possible to recognize the creation of a major industrial axis running from Kirkby along the East Lancashire Road to the Mersey Tunnel and down the east Wirral to Ellesmere Port.

Transportation and administration. The creation of this industrial axis, together with the overall decentralization trends in industry and housing, has placed additional strains upon the communication systems of the area. Despite the series of transportation improvements designed to ease movement across the Mersey—ranging from the early ferries to rail tunnels and culminating in the opening of the Queensway Road Tunnel in 1934—the decreasing significance of the port has meant that the estuary is regarded increasingly as the major barrier to movement in the city, rather than its sole artery. A survey made in the 1960s revealed that 130,000 trips were made across the river on an average weekday, 34 percent by road, 40 percent by rail, and 26 percent by ferry. Even with the relief provided by low car ownership levels in Liverpool (72 percent of households are without cars, compared with 40 percent for the rest of England and Wales), congestion necessitated the opening of a new road tunnel, the Kingsway, in 1971.

Transport in the central city

The historical dominance of the central city within the economy of the region as a whole represented an allied problem. In the early 1970s an inner motorway was being constructed to encircle the commercial heart of the city, but public transportation continued to account for more than 70 percent of trips made to the central area. The Merseyside Passenger Transport Executive, set up in 1969, has as its function overall integration of the existing transportation services, but vital physical links are still missing in the central area: the suburban rail services that enter it terminate at separate stations, Liverpool Central and Exchange, neither of which connects with the major national rail terminal at Liverpool Lime Street. In regional terms, the electrification of the main line to London has eased one problem; but the standard major road links to the rest of the country are still considered inappropriate for a city of Liverpool's size, whereas the airport at Speke remains a subsidiary to the regional airport at Manchester Ringway.

The attention paid by the transportation authorities to Merseyside as a whole, rather than to the city of Liverpool alone, represents the continuation of a trend that began with the creation, in 1857, of an overall ports authority, the Mersey Docks and Harbour Board. A more formal expression of these tendencies may be seen in the work of the Maud Commission on Local Government in England (1969), which recommended the creation of a Greater Merseyside administrative entity. Successive governments modified the original proposals, but, effective April 1974, the metropolitan county of Merseyside came into existence, with a series of distinct authorities exercising wide powers within the overall unit.

Problems and prospects. It remains a matter of conjecture as to whether the land-use changes associated with the creation of a major metropolitan county around Liverpool would alter the distinctive character of the city. The economic problems of the post-World War II era have, however, ensured that Liverpool has not experienced the major influx of Commonwealth citizens, a trend that has altered, in some cases drastically, the ethnic balance in more prosperous British cities. On the contrary, in social terms, the heritage of past immigration streams can still be seen in one of the highest concentrations of Roman Catholics in Britain, whereas the influence of the port may be responsible for the high proportion of unskilled and semiskilled workers (particularly among the unemployed) in the makeup of the city. In spite of these changes, the continuity with its past is still expressed in the attachment to the mythical stone liver birds at Pierhead and the folk song exploits of Maggie May, symbolic features which still characterize the city in the minds of residents and visitors alike.

BIBLIOGRAPHY. Although dated, W. SMITH (ed.), *A Scientific Survey of Merseyside* (1953), remains a classic work on all aspects of the city, containing a valuable historical summary of the town, which has been complemented in recent years by a portrayal of the development of the area through the medium of maps by J.A. PATMORE and A.G. HODGKISS (eds.), *Merseyside in Maps* (1970). More specific historical studies have been usefully collected together by J.R. HARRIS, *Liverpool and Merseyside: Essays in the Economic and Social History of the Ports and its Hinterland* (1969). R. LAWTON and C.M. CUNNINGHAM (eds.), *Merseyside: Social and Economic Studies* (1970); and the DEPARTMENT OF ECONOMIC AFFAIRS, *Problems of Merseyside* (1965), are useful accounts of the socioeconomic conditions of Liverpool and the Merseyside conurbation; D.M. SMITH, *Industrial Britain: The Great North West* (1969), puts these issues in a wider regional context. The port of Liverpool is carefully documented in J. BIRD, *The Major Seaports of the United Kingdom* (1963); the architectural character of the city is vividly portrayed by J.Q. HUGHES in *Seaport: Architecture and Townscape of Liverpool* (1964); whereas D.M. MUCHNICK provides an intriguing account of the process of urban renewal in *Urban Renewal in Liverpool* (1970). More recent works, concentrating on the analysis and solution of

the social and economic problems of the city, include the DEPT. OF THE ENVIRONMENT, ***Change or Decay, Final Report of the Liverpool Inner Area Study*** (1977); a report of Oxford University's DEPT. OF SOCIAL AND ADMINISTRATIVE STUDIES, ***Government Against Poverty***, a report on the 1970–75 Liverpool Community Development Project (1977); and PLANNING RESEARCH APPLICATIONS, ***Liverpool Social Area Study***, 2 vol. (1975).

(W.K.D.D.)

Livestock and Poultry Farming

Livestock is a general term that refers to all animals kept on a farm for use or for pleasure. Poultry includes birds raised on a farm for meat or eggs, In this article, the discussion of livestock includes both beef and dairy cattle, pigs, sheep, goats, horses, mules, asses, buffalo, and camels, while the discussion of poultry comprises chickens, turkeys, ducks, geese, guinea fowl, and squabs. Fur farming is treated in the article FURS. For further information on dairy-cattle breeds, feeding and management, as well as on milk and other dairy foods, see DAIRYING AND DAIRY PRODUCTS. For a further discussion of thoroughbred horses, see HORSE RACING and HORSES. Draft-horse breeds are discussed in DRAFT ANIMALS. Poultry eggs are covered in EGG PRODUCTION, COMMERCIAL. For other general information, see also ANIMAL FEED and DISEASES OF ANIMALS.

An efficient and prosperous animal agriculture historically has been the mark of a strong, well-developed nation. Such an agriculture permits a nation to store large quantities of grains and other foodstuffs in concentrated form to be utilized to raise animals for human consumption during such emergencies as war or natural calamity. Furthermore, meat has long been known for its high nutritive value, producing stronger, healthier people.

Ruminant (cud-chewing) animals such as cattle, sheep, and goats convert large quantities of pasture forage, harvested roughage, or by-product feeds, as well as nonprotein nitrogen such as urea, into meat, milk, and wool. Ruminants are therefore extremely important; more than 68 percent of the world's farmland is in meadows and pasture. Poultry also converts feed efficiently into protein; chickens, especially, are unexcelled in meat and egg production. Milk is one of the most complete and oldest known animal foods. Cows were milked as early as 9000 BC. Hippocrates, the Greek physician, recommended milk as a medicine in the 5th century BC. Sanskrit writings from ancient India refer to milk as one of the most essential human foods.

Significantly, peoples of the developed countries today consume six times as much animal protein per person per day as do those in underdeveloped countries. The United States leads the world in total per capita meat consumption (including poultry and fish) at 304 grams (10.7 ounces) per day (1979), and Lebanon is lowest at 15 grams (0.5 ounce) per day (1979). U.S. per capita daily milk consumption amounts to 405 grams (14.3 ounces; 1979), compared with 772 grams (27.2 ounces) for Finland (1979) and 16 grams (0.6 ounce) for Peru (1979). Israel and Hungary lead the world in per capita egg consumption with 1.1 eggs per day (1979). Of the world's red-meat supply (*i.e.*, excluding poultry and fish), 51 percent is beef and veal, 42 percent is pork, 6 percent is mutton, goat, and lamb, and less than 1 percent is horsemeat. Argentina leads the world in per capita red-meat consumption with 227 pounds (103 kilograms) consumed in 1979.

Farm technology, especially in the U.S. and other developed countries, has steadily improved the rate and efficiency of food production, permitting the average consumer to spend less of his disposable income for foods than he once did. In 1870 the average American spent 58 percent of his disposable income for food. Today he spends 21.7 percent, including 4.4 percent for red meat. In the Soviet Union the average citizen spends 32 to 36 percent of his disposable income for food. Italians spend 20 percent, Frenchmen 26 percent, and Englishmen 30 percent. Less than 4 percent of the U.S. population produces food for the other 96 percent, whereas in colonial times practically all able-bodied people spent the major part of their time in the production of food and fibre.

World livestock numbers. The Food and Agriculture Organization of the United Nations in 1979–80 reported a world population of 61,800,000 horses, 42,800,000 asses or donkeys, 11,600,000 mules, 1,212,000,000 cattle, 763,500,000 pigs, 1,084,000,000 sheep, 445,900,000 goats, 130,500,000 buffalo, and 16,800,000 camels. The United States leads the world in horse population. Brazil has the largest mule population, and China owns more than one-fourth of the world's asses. China also leads the world in hog and goat numbers. India has the most cattle and buffalo and Somalia the most camels. Australia is the leading sheep producer of the world. During the period 1947 to 1979 world horse numbers declined 19 percent, while world mule numbers increased 21 percent, asses increased 17 percent, cattle 59 percent, sheep 40 percent, goats 54 percent, buffalo 62 percent, camels 77 percent, and hogs 160 percent. Mere numbers, of course, do not indicate animal quality. Thus, while India possesses more cattle than any other nation, many of its animals are poorly nourished, sick, or of substandard breeding stock.

World poultry numbers. China was the leading country in number of chickens in 1979 with 1,410,000,000. Vietnam led the world in number of ducks with over 30,300,000, while Poland had the most geese at 8,450,000 and Mexico the most turkeys with 13,000,000. Worldwide egg production that year totalled 335,126,000,000 with North America accounting for one-fourth of this number and Europe slightly more than a fifth.

Fears that increased population pressures will cause human beings to compete with livestock for the world's grains and other edible plant production appear unfounded because modern technology has proved that it is possible to increase plant production manyfold while simultaneously developing types and strains of animals that will produce food for people with ever-greater efficiency. Problems of world-food supply, however, are political, economic, and social, as well as simply technological. Because almost 70 percent of the world's farmland and ranchland is in pasture and forage crops that are valuable only as animal feeds, and because ruminant animals can utilize easily synthesized nonprotein nitrogen for the production of high-quality animal proteins, it appears that, as a country develops economically, animal numbers and quality increase (see also ANIMAL BREEDING).

LIVESTOCK

Cattle. Cattle were probably first domesticated during the New Stone Age (*c.* 10,000 BC) from the aurochs (now extinct), which stood six to seven feet (1.8–2.1 metres) high at the withers (the ridge between the shoulder bones). Modern cattle are of the family Bovidae and the genus ***Bos***. The European breeds are of the species B. ***taurus***, whereas the humped cattle of the tropics, including the Brahman, or zebu, of India, and the Africander of Africa, belong to the species ***B. indicus***. Identifiable breeds throughout the world number 277, with 33 generally classified as beef breeds, 18 as draft breeds, 39 as meat-draft, 54 as meat-dairy, 21 as dairy-draft, 61 as meat-dairy-draft, and 51 as dairy breeds. Most of these are quite limited in distribution and importance. No cattle are native to the North American continent, only bison, or buffalo. Cattle used for draft purposes are usually oxen—that is, castrated males at least two or three years old. Though long supplanted by the horse and the tractor in the developed countries, oxen are still used in Africa and tropical Asia.

The cattle that Christopher Columbus and others brought to the New World were valued mainly for their milk, butter, hides, and work. Because of the plentiful supply of wild game, it was then thought wasteful to slaughter cattle at less than seven to nine years of age. This view was in accord with the practice in many Continental European countries, where few cattle have ever been kept solely for beef production.

Early cattle breeding in America

Cattle raising expanded rapidly in the New World. The need for fresh and larger grazing areas drew cattle farming west into Ohio and Kentucky, where corn for fattening purposes could be raised at low cost. Cattle were driven overland to seaboard markets by 1805. The first shipment of cattle marketed by rail originated in Lexington, Kentucky, in 1852. These cattle were driven to Cincinnati, Ohio, freighted to Cleveland on Lake Erie, boated to Buffalo, N.Y., freighted to Albany, and then boated down the Hudson River to New York City. By 1860 railroads had extended beyond the Mississippi River, opening the Southwest to range production and the Midwest to cattle feeding.

In the U.S. today there are four times as many beef cattle as dairy cattle. Production of milk per dairy cow in the U.S. more than doubled after 1930 and reached 11,194 pounds (5,078 kilograms) in 1977. Elsewhere in the New World, the vast pampas of Argentina—an area as large as France—provide excellent conditions for grazing of cattle throughout the year without need for shelter. In Australia, cattle have always ranked second in importance among livestock after sheep. Today's Australian beef cattle are raised mostly in the east and northeast portions of the continent, where feed and climate are unsuitable for sheep.

British breeds

Beef cattle breeds. The British Isles led the world in the development of the principal beef breeds; Herefords, Angus, beef Shorthorns, and Galloways all originated in either England or Scotland. Other breeds of greatest prominence today originated in India (Brahman), France (Charolais; Limousin; Normandy), Switzerland (Simmental), and Africa (Africander). The Hereford breed, considered to be the first to be developed in England, probably descended from white-faced, red-bodied cattle of Holland crossed with the smaller black Celtics that were native to England and especially to Herefordshire. By the middle of the 18th century the slow process of selective breeding that resulted in the smooth, meaty, and prolific Herefords had begun. The United States statesman Henry Clay of Kentucky imported the first purebred Herefords to America in 1817.

The Hereford, which became the most popular beef breed of the United States, is distinguished by its white face, white flanks and underline, white stockings and tail, and white crest on the neck. Its body colour ranges from cherry to mahogany red. It is of medium size, with present-day breeders making successful efforts to increase both its rate of weight gain and mature size, in keeping with the demand for cheaper, leaner beef.

The Polled Hereford is a separate breed of cattle originating from hornless mutations in 1901. It has the same general characteristics as the horned Hereford and has gained substantial favour because of its hornlessness and often faster rate of weight gain.

The Aberdeen Angus breed originated in Scotland from naturally hornless aboriginal cattle native to the counties of Aberdeen and Angus. Solid black, occasionally with a spot of white underneath the rear flanks, the breed is noted for its smoothness, freedom from waste, and high quality of meat.

Although the native home of the Galloway breed is the ancient region of Galloway in southwestern Scotland, it probably had a common origin with the Angus. The Galloway is distinguished by its coat of curly black hair. Though the breed has never attained the prominence of other beef breeds, it has been used extensively in producing blue-gray crossbred cattle, obtained by breeding white Shorthorn bulls to Galloway cows.

The beef, or Scotch, Shorthorn breed developed from early cattle of England and northern Europe, selected for heavy milk production and generally known as Durham cattle. These were later selected for the compact, beefy type by the Scottish breeders. Emphasis on leaner, high-quality carcasses in the second half of the 20th century has diminished the popularity of this breed. The Polled Shorthorn originated in 1888 from purebred, hornless mutations of the Shorthorn breed. The milking, or dual-purpose, Shorthorn, representing another segment of the parent Shorthorn breed, was also developed in England to produce an excellent flow of milk as well as an acceptable carcass, therefore resembling the original English type of Shorthorn. Shorthorns range in colour from red through roan, to white- or red-and-white-spotted.

The Brahman breed originated in India, where 30 or more separate varieties exist. Preference is given to the Guzerat, Nellore, Gir, and Krishna Valley strains, which are characterized by a pronounced hump over the shoulders and neck; excessive skin on the dewlap and underline; large, droopy ears; and horns that tend to curve upward and rearward. Their colour ranges from near white through brown and brownish red to near black. Their popularity in other areas such as South America and Europe, into which they have been imported, is attributable mainly to their heat tolerance, drought resistance, and resistance to fever ticks and other insects. The Santa Gertrudis was developed by the King Ranch of Texas by crossing Brahman and Shorthorn cattle to obtain large, hearty, tick-resistant, red cattle that have proved to be popular not only in Texas but in many regions along the semitropical Gulf Coast. Until the tick was eradicated in the southern and southwestern U.S., Brahman crosses were raised almost exclusively there.

Cross-breeds

Several lesser breeds have been developed from crosses of the Brahman on other beef breeds such as: the Charbray (Charolais), Braford (Hereford), Brangus (Angus), Brahom (Shorthorn), and Beefmaster (Brahman-Shorthorn-Hereford).

The Charolais breed, which originated in the Charolais region of France, has become quite popular in the United States for crossing on the British breeds for production of market cattle. The superior size, rate of gain, and heavy muscling of the pure French Charolais and the hybrid vigour accruing from the crossing of nonrelated breeds promise an increased popularity of this breed. Many American Charolais, however, carry significant amounts of Brahman blood, with a corresponding reduction in size, rate of gain, and muscling. Important in France, the Charolais is the foremost meat-cattle breed in Europe.

The Limousin breed, which originated in west central France, is second in importance to the Charolais as a European meat breed. Limousin cattle, often longer, finer boned, and slightly smaller than the Charolais, are also heavily muscled and relatively free from excessive deposits of fat.

The most prevalent breed of France, the Normandy, is smaller than the Charolais or Limousin and has been developed as a dual-purpose breed useful for both milk and meat production. A fourth important breed is the Maine-Anjou, which is the largest of the French breeds.

The Simmental accounts for nearly half of the present-day cattle of Switzerland, Austria, and West Germany. Smaller than the Charolais and Limousin, the Simmental was developed for milk, meat, and draft. It is yellowish-brown or red combined with characteristic white markings.

Beef cattle feed. Beef cattle can utilize roughages of both low and high quality, including pasture forage, hay, silage, corn fodder, straw, and grain by-products. Cattle also utilize nonprotein nitrogen in the form of urea and biuret feed supplements, which can supply from one-third to one-half of all the protein needs of beef animals. Nonprotein nitrogen is relatively cheap and abundant and is usually fed in a grain ration or in liquid supplements with molasses and phosphoric acid, or is mixed with silage at ensiling time; it may also be used in supplement blocks for range cattle or as part of range pellets. Other additions to diet include corn (maize), sorghum, milo, wheat, barley, or oats. Fattening cattle are usually fed from 2.2 to 3.0 percent of their live weight per day, depending on the amount of concentrates in the ration and the rate at which they are being fattened. Such cattle gain from 2.2 to 3.0 pounds (1,000–1,350 grams) per day and require from 1.3 to 3.0 pounds (600–1,350 grams) of crude protein, according to their weight and stage of fattening. Up until the early 1970s, when the practice was prohibited, fattening cattle were given the synthetic hormone diethylstilbestrol as a supplement in their feed or in ear implants. The use of this synthetic hormone results in a 10 to 20 percent increase in daily gain with less feed required per pound of gain. Synthetic vitamin A sources have become so cheap as to permit the use of 10,000 to 30,000 International Units per day for cattle being fattened for market (finished) in enclosures bare of vegeta-

Hormones for cattle

tion (drylots) used for this purpose. The economics of modern cattle finishing encourages the use of all-concentrate rations or a minimum of roughage, roughage substitutes including oyster shells, sand, and rough plastic pellets. Corn silage produces heavy yields per acre at a low cost and makes excellent roughage for beef-cattle finishing.

Beef cows kept for the production of feeder calves are usually maintained on pasture and roughages with required amounts of protein supplement and some grain being fed only to first-calf heifers or very heavy milking cows. Most beef cows tend to be overnourished and may become excessively fat and slow to conceive unless they happen to be exceptionally heavy milkers. Most pregnant cows go into the winter in satisfactory condition and need to gain only enough to offset the weight of the fetus and related membranes. They can therefore utilize coarser roughages, having a total daily crude protein requirement of from 1.3 to 1.7 pounds (585–765 grams). Daily vitamin A supplement at the rate of 18,000 to 22,000 International Units per cow is advisable unless the roughages are of a green, leafy kind and the fall pasture has been of excellent quality. Feed requirements for bulls vary with age, condition, and activity, from 2.0 to 2.4 pounds (900–1,080 grams) of crude protein per day; from 25,000 to 40,000 International Units of vitamin A; and during breeding periods nearly the same energy intake as calves or short yearlings being finished for market, the main feeding requirement being to prevent their becoming excessively fat.

All cattle require salt (sodium chloride) and a palatable source of both calcium and phosphorus, such as limestone and steamed bone meal. Most commercial salts carry trace minerals as relatively cheap insurance against deficiencies that occasionally exist in scattered locations.

Beef cattle management. Beef production has become highly scientific and efficient because of the high cost of labour, land, feed, and money. Most brood-cow herds, which require a minimum of housing and equipment, are managed so as to reduce costs through pasture improvement, and are typically found in relatively large areas and herds. Other aspects of management include performance testing for regular production of offspring that will gain rapidly and produce acceptable carcasses, and the use of preventive medicine, feed additives, pregnancy checks, fertility testing of sires, artificial insemination of some purebred and commercial herds, protection against insects and parasites, both internal and external, adequate but not excessive feed intakes, and a minimum of handling.

Calving

Calving of beef cows is arranged to occur in the spring months to take advantage of the large supplies of cheap and high-quality pasture forages. Fall calving is less common and occurs generally in regions where winters ale moderate and supplies of pasture forage are available throughout the year. Calves ale normally weaned at eight to ten months of age because beef cows produce very little milk past that stage and also because they need to be rested before dropping their next calf. Feeder calves sell by the pound, so that weight for age is even more important than conformation or shape. Consequently, crossbred cattle are used; their hybrid vigour results in greater breeding efficiency and milk production on the part of the dam, as well as greater birth weight, vigour, and gaining ability on the part of the offspring.

Beef cows are normally first bred at 15 to 18 months. The gestation period is 283 days, and the interval between estrus, or periods in which the dam is in heat, is 21 days. Cows should produce a living calf every 12 months. Pasture breeding, in which nature is allowed to take its course, calls for one mature bull for every 25 cows, whereas hand breeding, in which control is exercised by the breeder, requires half as many bulls. Artificial insemination permits one outstanding sire to produce thousands of calves annually.

Diseases of beef and dairy cattle. Dairy cattle are susceptible to the same diseases as beef cattle. Many diseases and pests plague the cattle industries of the world, the more serious ones being prevalent in the humid and less developed countries. One of the more common diseases to be found in the developed countries is brucellosis, which has been controlled quite successfully through vaccination and testing. This disease produces undulant fever in man through milk from infected cows. Leptospirosis, prevalent in warm-blooded animals and man, is caused by a spirochete and results in fever, loss of weight, and abortion. Bovine tuberculosis has been largely eliminated; where it has not, it can infect other warm-blooded animals, including man. Test and slaughter programs have proven effective. Rabies, caused by a specific virus that also can infect most warm-blooded animals, is usually transmitted through the bite of infected animals, either wild or domestic. Foot-and-mouth disease has been eliminated from most of North America, some Central American countries, Australia, and New Zealand. The rest of the world is still plagued by the disease, which attacks all cloven-footed animals. Man is mildly susceptible to this organism. Successful vaccinations have been developed for blackleg, malignant edema, infectious bovine rhinotracheitis (or red nose), and several other diseases. Anaplasmosis, common to most tropical and semitropical regions, is spread by the bite of mosquitoes and flies. Anthrax, caused by a generally fatal bacterial infection, has been largely eliminated in the U.S. and western Europe. Rinderpest, still common to Asia, the U.S.S.R., and Europe, is caused by a specific virus that produces high fever and diarrhea. An infectious fever sometimes called nagana, caused by the tsetse fly, attacks both cattle and horses and is prevalent in central and southern Africa, as well as in the Philippines. Grass tetany and milk fever both result from metabolic disturbances. Bloat, caused by rapid gas formation in the rumen, or first compartment of the stomach, is sometimes fatal unless relieved. Pinkeye is an infectious inflammation of the eyes spread by flies or dust and is most serious in cattle having white pigmentation around one or both eyes. Mastitis, an inflammation of the udder, is caused by rough handling or by infection. Vibriosis, a venereal disease that causes abortion; pneumonia, an inflammation of the lungs; and shipping fever all cause serious losses and are difficult to control except through good management. Broad-spectrum antibiotics (effective against various micro-organisms), as well as powerful and specific pharmaceuticals, are effective and profitable means of keeping cattle herds healthy. Vermifuges, which destroy or expel parasitic worms, and insecticides, which kill harmful insects, are also highly effective and much used.

Pigs. Pigs are believed to have been domesticated in eastern Asia as early as 2900 BC. The early Egyptians, Greeks, and Romans all raised pigs for their meat. Popular in medieval Europe, pigs were brought to the New World by Columbus on his second voyage in 1493; from Haiti they were taken to Cuba and Florida. Throughout ancient and medieval times pigs were scavengers and foragers. Scientific breeding did not begin until the 18th century. China has always favoured pork production. Hogs are excellent scavengers, are relatively easy to raise in confinement and can be slaughtered with a minimum of equipment because of their size and the many ways in which their carcasses can be processed into food and fat. Hogs are also quite efficient in converting feed to food.

Breeds. Pigs have a gestation period of 114 days with a 21-day interval between periods of estrus. A boar can mate with 15 to 45 sows per year. The average litter consists of seven pigs, each with a birth weight of 2.5 pounds (1,135 grams). Sows should each produce 1.5 litters or more per year. Most market pigs are produced from crossbred sows by a boar of a third breed. There are over 300 breeds or local varieties of pigs throughout the world. A brief description follows of the better-known commercial breeds that have figured prominently in improving and upgrading domestic breeds and crosses throughout the world.

The Hampshire pig, which originated from the Norfolk thin-rind breed of England, is black with a white belt completely encircling its body and both front legs and feet. There should be no white on the head or the ham.

The Yorkshire pig, which originated early in the 19th century in England, where it was considered a bacon type, is long, lean, and trim with white hair and skin. Found in most countries, this breed is probably the most widely-distributed in the world.

The Duroc-Jersey breed originated in the eastern U.S. from red pigs brought by Columbus and de Soto. The modern Duroc, originated from crosses of the Jersey Red of New Jersey and the Duroc of New York in the late 19th century, ranges from golden-red to mahogany-red in colour, with no black allowed. This breed proved particularly suitable for feeding in the U.S. Corn Belt (parts of Ohio, Indiana, Illinois, Wisconsin, Minnesota, South Dakota, Nebraska, Missouri, and Oklahoma; all of Iowa) and has been extensively used in Argentina, Canada, Chile, and Uruguay. The Poland China originated about 1860 in southern Ohio from a number of different breeds common to that area. The Spotted Poland China originated in Indiana about 1915 from crosses of the Poland China and the native spotted pigs. The Chester White, which originated in Chester County, Pennsylvania, after 1818, is restricted to the U.S. and Canada. The Berkshire, which originated in Berkshire, England, about 1770, is used for fresh pork production in England and Japan; a larger bacon type has been evolved in Australia and New Zealand. The Landrace is a white, lop-eared pig found in most countries in central and eastern Europe, with local varieties in Denmark, Germany, The Netherlands, and Sweden. World attention was first drawn to the Landrace by Denmark where, since 1890, by progeny testing (the selection of boars for breeding on the basis of the scientific assessment of their progeny), a superior pig, designed for Denmark's export trade in Wiltshire bacon to England, has been produced. Denmark no longer permits the export of Landrace for breeding. Sweden also has progeny tested from Landrace stock but for a shorter period. Pigs from Sweden were first exported to England in 1953 when prices of up to $1,000 were paid. This resulted in a worldwide Landrace explosion, and most major pig-producing countries have since taken stock.

The Landrace hog

The importance of the Asian pig breeds was recognized in the use of Chinese and "Siamese" pigs from southeastern Asia in the improvement of early European and North American breeds and is reflected in the name of the world-famous Poland China. China leads the world in pig numbers, and pork is traditional in the Chinese diet. Exports from China are substantial.

Feeding. Corn is a favourite feed for pigs, but wheat, sorghum, milo, barley, and oats are also used if the price is favourable. Wherever abundant and reasonable in price, soybean-oil meal is a favourite source of protein, and other oil meals and high-protein by-products are used in some countries. Antibiotics to control disease have become a standard ingredient in most pig rations. Improvements in breeding, disease control, management, and feed formulation have all contributed to faster gains and lower feed requirements per pound of weight gain. The use of antibiotics after World War II, especially in regions of less favourable sanitation, increased gain by as much as 20 percent.

Management. Pork production lends itself to mechanization and a minimum use of high-priced labour. The use of self-feeders and concentrated rations and construction of slotted floors and lagoons for waste disposal have become almost universal among large commerical producers in developed countries. Most commercial producers try to farrow pigs every two to three months of the year. Sows should have an 80 percent conception rate on first mating, eight or more pigs per litter, and average 1.6 litters per year. Most pigs are raised in confinement with various means of environmental control. Air-conditioned farrowing barns for excessively hot summers and heated floors and space heating or heat lamps for cold winters have become widespread.

Diseuses. Pigs are subject to many infectious and parasitic diseases. Among the more common of these is transmissible gastroenteritis, an infectious disease often fatal to young pigs. Cholera, formerly controlled by vaccination, is now controlled by slaughter of infected animals. Leptospirosis, common to pigs as well as man and most warm-blooded animals, can be controlled by vaccination. Necrotic enteritis and other infections of the intestinal tract are largely controlled by antibiotics. Atrophic rhinitis produces sneezing, crooked snouts, and little or no weight gain even with proper feeding (unthriftiness). Erysipelas, a bacterial infection of pigs, turkeys, and man, causes in pigs inflammation of the skin, a swelling and stiffness of joints, and unthriftiness; it can be controlled through a vaccination program. Parasitic diseases are controlled mainly through effective sanitation programs and are less of a problem under confinement raising than when pigs are raised in the open. Effective vermifuges, which kill or expel parasitic worms, have been developed for further control. Production in the United States, the Scandinavian countries, and Western Europe has been developed to a high state of efficiency, largely through effective control of infectious and parasitic diseases.

Sheep. Sheep were probably first domesticated in western Asia, perhaps as early as 10,000 BC. There is ample evidence of domesticated sheep in prehistoric Europe, where their skins were used for clothing and shelter, their meat and milk for food. Sheepherding is frequently mentioned in the Old Testament.

Sheep in the Middle Ages

In the Middle Ages England became known as the home of most medium and long-wooled mutton breeds, while Spain achieved worldwide recognition as the origin of the fine-wool Merino. These two countries became leaders in the production of both mutton and wool. In the 16th and 17th centuries, such New World countries as Argentina, with suitable climate and grazing land, and later such countries as Australia and New Zealand introduced sheep with the object of exporting wool, which is light in relation to its value and is relatively imperishable. The same considerations, plus the ability of sheep to subsist on sparse forage and limited water, made the western U.S. a natural sheep-raising region, especially in the early years of the 20th century when public lands provided vast ranges for flocks as large as 4,000 head. Later, sheep in the U.S. declined in favour of more profitable cattle, from over 56,000,000 in 1942 to fewer than 20,500,000 in 1970.

The gestation period for sheep is 147 days with 16.7 days between periods of estrus, which last 29 hours. The average number of lambs raised per hundred ewes is 91 and the average fleece weight per shearing is 8.34 pounds (3,783 grams).

Breeds. Of mole than 200 breeds of sheep in existence in the world, the majority are of limited interest except in the localities where they are raised. Sheep breeds are generally classified as medium wool, long wool, and fine wool. Of the medium wool breeds the Hampshire, Shropshire, Southdown, Suffolk, Oxford, and Dorset all originated in England. The Cheviot and Black Faced Highland originated in Scotland. The Panama, Columbia, and Targhee were developed in the U.S., and the Corriedale in New Zealand. After World War II, such larger breeds as the Suffolk and Hampshire increased in popularity at the expense of the smaller breeds.

The long wool breeds, including the Cotswold, Lincoln, Leicester, and Romney, were all developed in England and, in addition to mutton, produce wool of unusually long fiber length that is suitable for rugs and coarse fabrics.

The original fine-wool breed was the Merino, developed in Spain from parent stock native to that country before the Christian era. Though medieval Spain sought to preserve a monopoly on the Merino, the sheep gradually spread to France, Italy, and the rest of Europe. Today the Merino is prominent in Australia, the U.S., the U.S.S.R., South Africa, Argentina, France, and Germany; the breed is designated by various names such as Australian Merino in Australia, Merino Transhumante in Spain, and others. The Merino was the principal ancestor of the French Rambouillet, somewhat larger in size and less wrinkled than the Merino. This breed prospers in the Western ranges of the U.S., where two-thirds of that country's sheep are raised. The Corriedale breed, adapted to both farms and ranges, is especially valued in New

Zealand and Australia. Most commercial sheep today represent two-breed or three-breed crosses, with white-faced crossbred ewes preferred in the range areas and a black-faced sire, such as Suffolk or Hampshire, preferred for market lambs, which are either finished for slaughter or sold as breeding ewes.

Feeding. Sheep are excellent foragers and, being ruminants, can effectively utilize both pasture forage and harvested roughage. Selective in their grazing habits, they prefer short grass when available. Pregnant ewes can run on late pasture as long as it is available and abundant, but in winter subsist satisfactorily on well-cured legume hay or mixed hay carrying a high percentage of legume. Corn silage is relatively inexpensive and relished by sheep; lactating ewes and lambs being finished for market usually require some concentrate, with corn (maize) favoured because of its high energy content and reasonable cost.

Range sheep grazing selectively on native plants frequently develop mild deficiencies of protein, energy, phosphorus, and vitamin A, especially when plants are mature or dormant or are eaten by ewes in the later stages of pregnancy or lactation. Broad spectrum antibiotics at the rate of five to ten milligrams per pound of feed are normally used in all lamb finishing rations to prevent digestive disturbances and infections.

Management. Range sheep are normally white-faced crosses carrying both long-wool and Rambouillet breeding and are consequently very hardy and thrifty. They are wintered in bands, or flocks, of from 1,000 to 4,000 head at lower altitudes, and are moved in bands ranging from 1,000 to 1,500 head to summer range at much higher altitude, sometimes 300 miles (480 kilometres) from their winter quarters. Each flock is tended by a sheepherder and his dogs who move systematically from one grazing area to another. The herder often lives in a covered wagon or truck and may spend weeks at a time in complete solitude. The most famous sheepherders are the Basques, who emigrated widely from their home in Spain. The breeding ewes are mated to Suffolk or Hampshire rams and produce lambs during the late winter or early spring so that the lambs will be old enough to move to summer grazing without difficulty.

Shepherds

In many parts of the world small flocks are kept partly as scavengers to clean up fence rows, weeds, brush, and other undesirable forage, but this is a diminishing role. Large flocks are maintained partly for wool and partly for market lambs. Lambs are usually dropped in the spring and are sold at ages of from three to eight months and weights of around 40 pounds (18 kilograms) for Easter lambs, and 100 pounds (45 kilograms) for the usual market lambs. Sheep are sheared in the spring after the worst winter weather has passed. Some breeds are noted for producing a high percentage of twins, and others, such as the Dorset, for both high frequency of twins and heavy milk production. Shepherds frequently switch a newly born twin lamb to a ewe that has just lost a single lamb, thereby utilizing the extra milk. This practice requires skill and experience since mother ewes recognize their own lamb by both its smell and the sound of its call.

Diseases. Such internal parasites as the tapeworm and several species of roundworms that infest the gastrointestinal tract are perhaps the greatest scourge of sheep, but modern vermifuges are quite effective against these. Dips are used to combat such external parasites as ticks, lice, and mites. Foot rot, caused by an infection of the soft tissue between the toes, results in extreme lameness and even loss of the hoof. The more persistent type is caused by a specific organism that is difficult to treat. The pain and the restricted movement of infected sheep result in rapid loss of weight. Enterotoxemia, or pulpy kidney, affects lambs at two to six weeks of age, especially those starting on unusually lush or rich feeds. A vaccination is quite effective in preventing this otherwise costly ailment.

Goats. Probably first domesticated in the East, perhaps during prehistoric times, the goat has long been used as a source of milk, cheese, mohair, and meat. Its skin has been valued as a source for leather. In China, Great Britain, Europe, and North America, the goat is primarily a milk producer. By good management its limited (six months per year) breeding season and the consequent difficulty of maintaining a level supply of milk throughout the year, can be overcome. The goat is especially adapted to small-scale production of milk for the family table; one or two goats supply sufficient milk for a family throughout the year and can be maintained economically in quarters where it would not be practical to keep a cow.

Goat's milk

Pure-white goat's milk compares favourably with cow's milk in favour and keeping qualities under sanitary conditions. It has certain characteristics differing from cow's milk that make it more easily digested by infants, invalids, and persons with allergies. Goat flesh is edible, that from young kids being quite tender and more delicate in flavour than lamb, which it resembles. Goat flesh is much prized in the Mediterranean countries, particularly in Spain, Italy, the south of France, and Greece. The Angora and Cashmere breeds are famous for their fine wool or mohair.

The many breeds may be roughly grouped: the prick-eared—*e.g.*, Swiss goats; the eastern, or Nubian, with long, drooping ears; and the wool goat—*e.g.*, Angora. While it is usually easy to distinguish goats from sheep, certain hair breeds of the latter are, to the layman, only distinguishable from goats by the direction of the tail, upward in goats, downward in sheep.

Of the Swiss goats, from which many of the best modern breeds are derived, the Toggenburg and Saanen are most important. The French breeds have much Swiss blood. In Germany the many varieties trace to Swiss breeds, which are also popular throughout Scandinavia and the Netherlands.

The Maltese goat, an important source of milk on the island of Malta, probably contains eastern blood. Many goats are found in Spain, northern Africa, and Italy, among them the Murcian, Granada, and La Mancha.

Nubians are African goats, chiefly Egyptian. They are usually large, short-haired goats with large lop ears and Roman noses. They may be of solid colour, parti-coloured, or spotted. The goats in Israel and Syria have long hair and large lop ears and most commonly are solid black or with white spots. Most Indian varieties, the best of which come from the Yamuna River area, have lop ears.

In Britain, the native goat was small, with short legs, long hair—usually gray but of no fixed colour—and with no definite markings. The widespread use of pedigree males, mostly of Swiss extraction, to improve the milk yield, has resulted in the almost total disappearance of the native types.

Horses. Horses were among the last species of livestock to be domesticated. Domestication took place at least as early as 3000 BC, probably in the Near East. The wild ass, which when domesticated is usually called a donkey, was first domesticated in Egypt about 3400 BC.

Breeds. The Arabian, the oldest recognized breed of horse in the world, is thought to have originated in Arabia before AD 600. Though its history is lost in the past, the breed probably descended from the Libyan horse, which, in turn, was probably preceded by horses of similar characteristics in Assyria, Greece, and Egypt as early as 1000 BC. The Arabian may be bay, gray, chestnut, brown, black, or white in hair colour, but always has a black skin. It ranges from 14.1 to 15.1 hands (56–60 inches, or 1.4 to 1.5 metres) in height. The Arabian horse has one lumbar vertebra less than other breeds of horse and is characterized by the high carriage of its head, long neck, and spirited action.

The Thoroughbred racing horse is descended from three desert stallions brought to England between 1689 and 1724; all of the Thoroughbreds of the world today trace their ancestry to one of these stallions.

The American Saddle Horse, originating in the U.S., was formed by crossing Thoroughbreds, Morgans, and Standardbreds on native mares possessing an easy gait. The American Saddle Horse is 15 to 16 hands (60–64 inches, or 1.5–1.6 metres) in height. Its colours are bay, brown, black, gray, and chestnut. There are two distinct types of the American Saddle Horse: three-gaited and five-gaited. The three natural gaits are walk, trot, and

Horse gaits

canter. Three-gaited saddle horses are shown with a short tail and cropped mane. They often have slightly less style and finish than the five-gaited horse. The five-gaited saddle horse has the three natural gaits plus two man-trained gaits—the rock and slow gait, or running walk. The American Saddle Horse is also used as a fine harness horse mainly for show.

The American Quarter Horse traces to the Thoroughbred, and includes the blood of other breeds, such as the Morgan, the American Saddle Horse, and several strains of native horses. This fast, muscular horse has been raced, ridden in rodeos, and used for herding cattle.

The typical Quarter Horse is 15 to 16 hands tall and is of powerful build, suitable for both racing and the rough life of a cow pony. This horse is noted for its intelligence, easy disposition, and cow sense.

The Tennessee Walking Horse, or plantation horse, traces mainly to the Standardbred but also includes Thoroughbred and American Saddle Horse blood. The Tennessee Walking Horse is noted for its running walk, a slow-gliding gait in which the hind foot oversteps the print of the front foot by as much as 24 inches (600 millimetres). This breed is 15.2 to 16 hands high and is bay, black, chestnut, roan, or gray in colour.

The Morgan traces directly to "the Justin Morgan horse", foaled in 1793, of unknown breeding but no doubt tracing to Arabian stock. A dark bay in colour, Morgan stood 14 hands high and weighed 950 pounds (430 kilograms). He was a heavily muscled, short-legged horse of great style, quality, and endurance. He is the world's best example of prepotency, since he alone founded the Morgan breed. The Morgan is used for both riding and driving. It ranges from 14 to 16 hands in height and resembles the Arabian in size, conformation, quality, and endurance.

The American Standardbred originated around New York City during the first half of the 19th century from Thoroughbred, Morgan, Norfolk Trotter, Arabian, Barb, and pacers of mixed breeding. The modern Standardbred is smaller than the Thoroughbred, ranging from 15 to 16 hands in height and averaging about 15.2 hands. In racing condition it weighs from 900 to 1,000 pounds (410–450 kilograms). Stallions in stud condition average from 1.100 to 1,200 pounds (500–545 kilograms). Compared with the Thoroughbred, the Standardbred is longer-bodied, shorter-legged, heavier-boned, and stockier in build. Prevailing colours are bay, brown, and chestnut.

Draft horses

Draft horses have largely been supplanted by trucks and tractors in the developed countries of the world. Major draft breeds include the Percheron, developed in France; the Clydesdale of Scotland; the Shire of England; the Suffolk of England; and the Belgian of Belgium. These breeds range from 15% to 17 hands in height at the withers; at maturity the mares weigh from 1,600 to 2,000 pounds (720–900 kilograms) and the stallions from 1,900 to 2,200 pounds (860–1000 kilograms).

The more popular pony breeds are the Shetland, which originated in the Shetland Islands, and the Hackney, of English origin. Ponies must be under 14.2 hands in height at the withers and are used both for show and for children's pleasure.

Feeding. The specific and exact nutrient requirements of horses are poorly understood. Usually, these may be supplied economically from pasture forage, harvested roughages, and concentrates. Good quality grass-legume pastures, in addition to iodized or trace-mineralized salt, will supply adequate nutrients to maintain an adult horse at light work (such as pulling a small cart) or mares during pregnancy. Lush, early spring pasture is very high in water and protein contents and may need to be supplemented with a high-energy source, such as grain, to meet the needs of horses performing medium to heavy work (such as plowing). Conversely, late fall- and winter-pasture forage is low in water and protein and may require protein and vitamin A supplementation. High-quality legume hays, such as early bloom alfalfa, are preferred for horses, especially those that are growing or lactating. Moldy or dusty feeds should be avoided because horses are extremely susceptible to forage poisoning and respiratory complications. Grass hays, such as timothy, prairie grass, orchard grass, and bluegrass, were preferred by early horsemen, especially for race horses, because they were usually free from mold and dust and tended to slow down the rate of passage through the intestinal tract. These hays are low in digestible energy and protein, however, and must be adequately supplemented. Silages of all sorts should be avoided since horses and mules are extremely susceptible to botulism and digestive upsets.

Oats are the preferred grain for horses because of their bulk. Corn, barley, wheat, and milo can be used, however, whenever they are less expensive. Weanling foals require three pounds of feed per hundred pounds of live weight per day; as they approach maturity, this requirement drops to one pound of feed per hundred pounds of live weight daily. Horses normally reach mature weight at less than four years of age and 80 percent of their mature weight at less than two years of age.

A large and ever-growing number of horses stabled in cities and suburbs where sufficient roughages cannot be grown provide a large market for complete horse rations, including roughage, which are tailored to the total needs of specific animals according to their particular function at a given time, such as growth, pregnancy, lactation, or maintenance.

Horses will vary from the normal requirement in terms of weight, temperament, and previous nutrition. Foals will eat some pasture grass, forage, or hay when they are three days old and grain when they are three weeks old.

Management. Highly bred light horses are notoriously poor reproducers. Many horse farms consider a 60 percent foaling rate for a large band of mares to be average. Most large horse farms employ resident veterinarians to check for abnormalities or disease before breeding and to check mares for pregnancy 40 to 45 days after breeding. Because many mares conceive only every other year, expert assistance at foaling time is an absolute necessity, especially if the foal is sired by an expensive stallion out of a valuable mare.

Foaling mares

The gestation period of horses is 340 days. The period between estrus ranges from 18 to 28 days with an average of 22 days. Mature stallions can safely mate with from 50 to 100 mares per season, although the practice with expensive Thoroughbreds is to book no more than 35 to 40 mares.

The feet and legs of horses demand unusual attention. The old adage "no foot no horse" remains apt. Hooves should be trimmed regularly, beginning when the horse is a foal or only a few months old. Otherwise they may grow long and uneven, causing improper action, undue stress on joints, and broken or cracked hooves. Horses that are worked regularly, especially on hard and stony ground, as well as show horses and race horses in service, must be shod. Shod horses should have their hooves trimmed and their shoes refitted every four to six weeks. Tendency toward unsoundness is probably inherited but may be aggravated by poor hoof care and excessive stresses.

Diseases. Horses are especially susceptible to tetanus or lockjaw but can be given two-year protection through the use of a commonly accepted toxoid. There are two common types of abortion in horses: virus abortion, specifically viral rhinopneumonitis, and the *Salmonella* type. The former, which produces an influenza with pinkeye, catarrh, general illness, and abortion, affects both mares and foals, but all surviving horses develop natural resistance soon after infection. Pregnant mares thought to be subjected to infection may be given some protection by available vaccines the success of which was not fully proved in the early 1970s. The *Salmonella* type of abortion can be prevented completely by vaccination. Encephalomyelitis, or sleeping sickness, is prevented by vaccination. A specific vaccine is available for anthrax, which is prevalent in Asia. Hemolytic anemia of foals has become a problem. Foals so afflicted are born normal but soon become sluggish and progressively weaker; the membranes of their eyes, mouth, and lips become very pale and the heartbeat becomes rapid. This condition is caused by antibodies in the mare's milk that destroy the foal's red blood cells. These antibodies are caused by the difference in blood type between the foal and the mother. Newborn foals can be muzzled to avoid nursing while their blood is checked for reaction against the serum and milk of its mother. Where reactions are noted, the mare is hand-milked at hourly intervals for 12 to 24 hours, and the foal is fed milk from another suitable mare or a milk substitute. Horses are quite susceptible to various infections, but rotation of pastures, strict sanitation, and the use of suitable vermifuges are quite effective.

Asses and mules. The words donkey and ass are generally used interchangeably to denote the same animal, though ass is more often employed when the animal is wild and donkey is used for a domesticated beast. Wild asses inhabit arid, semidesert plains where the vegetation is sparse and coarse; the domestic donkey does well on coarse food and is hardy under rough conditions, hence its usefulness to man as a beast of burden in places where horses cannot flourish, such as the mountains of Ethiopia and other parts of northeast Africa. the high plains of Tibet, and the arid regions of Mongolia.

The donkey's occasional obstinacy in refusing work too heavy for it has become proverbial, but its equally proverbial stupidity has probably become legendary through its reaction to brutal treatment and neglect. It is naturally patient and persevering, responding to gentle treatment with affection and attachment to its master.

Mining mules

Mules are still used in some of the subtropical and tropical countries because of their ability to withstand most types of stress including heat, irregular feeding, and abuse. The mule is produced by crossing a jackass (Equus *asinus*) with a mare. The so-called Mammoth Jack was developed in America from European imports dating back to the late 18th century. It stands 15 to 16 hands (1.5 to 1.6 metres) in height and weighs from 900 to 1,150 pounds (410–520 kilograms) at maturity. The reverse cross of a stallion on a jenny, or female ass, is called a hinny but theoretically has the same characteristics as a mule. At one time many different types of mules were recognized, such as draft mules, farm mules, sugar mules, cotton mules, and mining mules in declining order of size. The mining mule, a small, rugged individual weighing as little as 600 pounds (270 kilograms), was used in pit mines. These small mules are usually produced by pony dams.

Mules are surer-footed than horses and also more intelligent. For this reason they are still used as saddle and pack mounts in precarious terrain. Unlike horses, they also refuse to damage themselves by overeating or by thrashing around when tangled up or in cramped quarters.

Buffalo and camels. Buffalo. The name buffalo is applied to several different cud-chewing (ruminant) mammals of the ox family (Bovidae). The true, or Indian, buffalo (*Bubalus bubalis*), also known as water buffalo, or arna, exists both as a wild and domestic animal; it has been domesticated in Asia from very early times and was introduced into Italy about the year 600. A large oxlike animal of massive and rather clumsy build with large horns that are triangular in cross section, the Indian buffalo, standing five feet (1.5 metres) at the shoulder, has a dull black body, often very sparsely covered with hair. The horns, which may be over six feet (1.8 metres) long, spread outward and upward, approaching each other toward the tips; they meet more or less in one plane above the rounded forehead and elongated face. Used for draft purposes, and also for milk and butter. the domesticated Indian buffalo is found throughout the warmer parts of the Old World from China to Egypt, and in Hungary, France, and Italy. Its cousin, the Cape, or African. buffalo (*Syncerus caffer*), a black animal of similarly massive build, has never been domesticated.

Camels. The term camel is applied to two species of the genus *Camelus*. The Arabian camel, *Camelus dromedarius*, has one hump, the Bactrian camel, *Camelus bactrianus*, has two. The limbs are long and the feet have no traces of the second or fifth toes; the wide-spreading soft feet are well adapted for walking upon sand or snow. Horny pads on the chest and knees support the camel's weight when kneeling.

Camels are known only as domestic animals and are not found in a truly wild state. The Bactrian camel occurs throughout the highlands of Central Asia from Turkistan to Mongolia and is an important beast of burden throughout that region. The Arabian camel, characteristic of India, the Near East, and North Africa, is likewise primarily important as a beast of burden, though it also provides wool, milk, hides. and meat. It is longer-legged, shorter coated, and more lightly built than the Bactrian camel, standing about seven feet (2.1 metres) tall at the shoulder.

Camel diets

Camels can flourish on the coarsest of sparse vegetation, feeding on thorny plants, the leaves and twigs of shrubs, and dried grasses that other animals would refuse, though camels ale not averse to more attractive food if it is available. When the feeding is good they accumulate stores of fat in their humps, upon which they are able to draw when conditions are adverse not only for sustenance but also for the manufacture of water by the oxidation of the fat; but they do not store water in the miscalled water cells. They are thus able to fast and go without drinking for several days; they have been known to go without water for 17 days and survive. Other adaptations that enable them to survive in deserts and other unfavourable environments include double rows of heavy protective eyelashes, haired ear openings, the ability to close their nostrils, and keen senses of sight and smell. The female produces one young at a birth after a gestation of 11 months and suckles it for a year; maturity is reached at the age of 10 to 12 years, and the life span is 30 to 40 years.

POULTRY

Chickens. Man first domesticated chickens of Indian origin for the purpose of cockfighting in Asia, Africa, and Europe. Very little attention was given to egg or meat production. Cockfighting was outlawed in England in 1849 and in most other countries thereafter. Exotic breeds and new standard breeds proliferated in the years to follow, and poultry shows became very popular. From 1890 to 1920 chicken raisers stressed egg and meat production, and commercial hatcheries became important after 1920.

Breeds. The breeds of chickens are generally classified as American, Mediterranean, English, and Asiatic. The American breeds of importance today are the Plymouth Rock, the Wyandotte, the Rhode Island Red, and the New Hampshire. The Barred Plymouth Rock, developed in 1865 by crossing the Dominique with the Black Cochin, has grayish-white plumage crossed with dark bars. It has good size and meat quality and is a good layer. The White Plymouth Rock, a variety of the Barred Plymouth Rock, has white plumage and is raised for its meat. Both varieties lay brown eggs. The Wyandotte, developed in 1870 from five or more strains and breeds, has eight varieties and is characterized by a plump body, excellent meat, and good egg production. Only the white strain is of any significance today because it is used in broiler crosses where its white plumage, quality of flesh, and rapid growth are highly desirable. Popular in England, this breed also lays a brown egg.

An American breed, the Rhode Island Red, developed in 1857 from Red Malay game fowl crossed with reddish-coloured Shanghais—with some brown Leghorn, Cornish, Wyandotte, and Brahma blood—is good for meat production and is one of the top meat breeds for the production of eggs. It has brilliant red feathers and lays brown eggs.

The New Hampshire, developed in the U.S. in 1930 from Rhode Island Red stock, is a meaty, early maturing breed with light-red feathers and lays large brown eggs. The only Mediterranean breed of importance today is the Leghorn. This breed, originated in Italy, has 12 varieties, the single-comb White Leghorn being more popular than all of the other types combined. This breed, the leading egg producer of the world, lays white eggs and is kept in large numbers in England, Canada, Australia, and the U.S. The White Minorca, a second Mediterranean breed, is often used in crossbreeding for egg production.

The only English breed of modern significance is the Cornish, a compact and heavily meated bird used in crossbreeding programs for broiler production. It is a poor producer of eggs, however.

The only Asiatic breed of significance today, the Brahma, which originated in India, has three varieties, the light Brahma being preferred because of its size.

Chicken breeding is an outstanding example of the application of basic genetic principles of inbreeding, linebreeding, and crossbreeding, as well as of intensive mass selection to effect faster and cheaper gains in broilers and maximum egg production for the egg-laying strains. Maximum use of heterosis, or hybrid vigour, through incrosses and crossbreeding has been made. Crossbreeding for egg production has used the single-comb White Leghorn, the Rhode Island Red, the New Hampshire, the Barred Plymouth Rock, the White Plymouth Rock, the Black Australorp, and the White Minorca. Crossbreeding for broiler production has used the White Plymouth Rock or New Hampshire crossed with White or Silver Cornish or incrosses utilizing widely diverse inbred strains within a single breed. Rapid and efficient weight gains, and high quality, plump, meaty carcasses have been achieved thereby.

Development of the **egg**

The male sperm lives in the hen's oviduct for two to three weeks. Eggs are fertilized within 24 hours after mating. Yolks originate in the ovary and grow to about 40 millimetres (1.6 inches) in diameter, after which they are released into the oviduct, where the thick white and two shell membranes are added. The egg then moves into the uterus where the thin white and the shell are added. This process requires a total of 24 hours per egg. The hatching of fertilized eggs requires 21 days, with the heavy breeds requiring a few more hours and the lighter breeds slightly fewer. Ideal hatching temperature approximates 100" F (38" C) with control of air flow, humidity, oxygen, and carbon dioxide being essential. Standardized egg-laying tests and official random sample tests have been used for many years to measure actual productivity.

Feeding. Chicken feeding is a highly perfected science that ensures a maximum intake of energy for growth and fat production. High quality and well-balanced protein sources produce a maximum amount of muscle, organ, skin, and feather growth. The essential minerals produce bones and eggs; 3 to 4 percent of the live bird being composed of minerals and 10 percent of the egg. Calcium, phosphorus, sodium, chlorine, potassium, sulfur, manganese, iron, copper, cobalt, magnesium, and zinc are all required. Vitamins A, C, D, E and K and all 12 of the B vitamins are also required. Water is essential, and antibiotics are almost universally used to stimulate appetite, control harmful bacteria, and prevent disease. Modern rations produce a pound of broiler on about two pounds (900 grams) of feed and a dozen eggs from 4½ pounds (2,025 grams) of feed.

Management. Among the world's agricultural industries, meat chicken breeding in the U.S. is one of the most advanced. It is presently considered the model for other animal industries, the broiler industry leading the way in advanced agricultural technology and efficiency. Intensive nutritional research and application, highly improved breeding stock, intelligent management, and scientific disease control have gone into the effort to give a modern broiler of uniformly high quality produced at ever-lower cost. Today, one man can care for 25,000 to 50,000 broilers that reach market weight in three months' time, giving him an annual output of from 100,000 to 200,000 broilers. A modern broiler chick gains over 43 times its initial weight in an eight-week period. Aggressive marketing methods increased the per capita consumption of broilers fourfold in the two decades beginning in 1950, with further substantial increases predicted for the future. Less than half as much feed is now required to produce a pound of broiler meat as was needed in 1940. While per capita consumption of eggs has declined, the feed requirement per dozen eggs is only slightly more than half as high as it was in the early 1900s. Annual egg production per hen has increased from 104 to 218 since 1910.

A carefully controlled environment that avoids crowding, chilling, overheating, or frightening is almost universal in chicken raising. Cannibalism, which expresses itself as toe picking, feather picking, and tail picking, is controlled by debeaking at one day of age and by other management practices. The feeding, watering, egg gathering, and cleaning operations are highly mechanized. Four-fifths of the 3,800,000,000 chicks hatched per year in the early 1970s were used for broiler production and one-fifth for egg production. In egg production feed and labour represent 75 percent of the cost. Pullet (immature hen) flocks predominate. Hens are usually housed in wire cages with two or three hens per cage and three or four tiers of cages superposed to save space. Cages for laying hens have been found to increase production, lower mortality, reduce cannibalism, lower feeding requirements, reduce diseases and parasites, improve culling, and reduce both space and labour requirements.

Other poultry. These include turkeys, ducks, geese, guinea fowl, and squabs.

***Turkey** production.* After World War II turkey production became highly specialized, with larger flocks predominating. Turkeys are raised in great numbers in Canada where their ancestors still live wild, as also in some parts of the U.S. Broad Breasted Bronze, Broad Breasted White, and White Holland are the most popular of the larger breeds, representing nearly three-fourths of the total production. The Beltsville Small White is the most popular of the smaller bleeds and composes the bulk of the remaining 25 percent. At 24 weeks of age the toms are 50 percent heavier than the hens. In breeding flocks, one tom is required per 8 or 10 hens. Tremendous improvements both in breeding and nutrition have been made in this century. Since 1910, the amount of feed required to produce a pound of turkey meat has fallen 40 percent, while the time required has been reduced 25 percent. Ninety to 100 pounds (40–45 kilograms) of feed will produce a turkey for market weight with from three to five pounds required per pound of gain on full-size turkeys, and 2½ to 2% pounds (1,125–1,250 grams) of feed per pound (450 grams) of gain for turkey broilers, which are marketed at from 12 to 15 weeks of age. Turkey poults are hard to start on feed. One method is to dip their beaks in water and then in feed. Another is to light the feed troughs very brightly and to use oatmeal or ground yellow corn sprinkled on top of the feed. Turkeys are given range, or open land, and automatic waterers, self-feeders, range shelters, heavy fencing, and rotated pastures are used. Successful marketing techniques have increased turkey consumption; *e.g.*, in the U.S., per capita consumption from 1950 to 1970 rose 250 percent.

Duck and goose production. Duck raising is practiced on a limited scale in nearly all countries, for the most part as a small-farm enterprise. The flocks once kept in England are much reduced, the demand for eggs being greatly lessened, though a limited market still exists. Khaki Campbell and Indian Runner ducks are prolific layers, each averaging 300 eggs per year. In Indonesia, where the labour supply is large, duck herders take a flock of ducks to the high country during the warmer seasons and work their way down the mountainsides to the lowlands. Ducks are easily transported, can be raised in close confinement, and convert some waste products and scattered grain (*e.g.*, by gleaning rice fields) to nutritious and very desirable eggs and meat. In developed countries, commercial plants have been built exclusively for duck meat production; an example is the large duckling industry of Long Island, N.Y. There are also local industries in The Netherlands and England, the favourite breed in England being the Aylesbury. This breed has white flesh and can reach eight pounds (3.6 kilograms) in eight weeks. The U.S. favourite is the Pekin duck, which is slightly smaller than the Aylesbury and yellow-fleshed.

Geese

Goose raising is a minor farm enterprise in practically all countries, but in Germany, Austria, some eastern European countries (notably Poland), parts of France, and locally elsewhere, there is important commercial goose production. The two outstanding meat breeds are the Toulouse, predominantly gray in colour, and the Embden (or Emden), which is white. Geese do not appear to have attracted the attention of geneticists on the same scale as the meat chicken and the turkey, and no change in the goose industry comparable to that in the others has occurred or seems to be in prospect. In some commercial plants, geese are fattened by a special process resulting in a considerable enlargement of their livers, which are sold as a delicacy, pâté de foie gras.

Guinea fowl and squabs. Guinea fowl are raised as a sideline on a few farms in many countries, and eaten as gourmet items. In Italy there is a fairly extensive industry. There the birds are raised in yards with open-fronted shelters. In England, guinea fowl are marketed at 16–18 weeks of age and in the U.S. at about 10–12 weeks. The market weight is usually about 2½–3½ pounds, but food conversion is poor.

Pigeons are raised not only as messengers and for sport but also for the meat of their squabs (nestlings), also a gourmet item. Squab production, carried on locally, is rare in most countries with established poultry industries.

Poultry diseases. Poultry are quite susceptible to a number of diseases; some of the more common are fowl typhoid, pullorum, fowl cholera, chronic respiratory disease, infectious sinusitis, infectious coryza, avian infectious hepatitis, infectious synovitis, bluecomb, Newcastle disease, fowl pox, avian leukosis complex, coccidiosis, blackhead, infectious laryngotracheitis, infectious bronchitis, and erysipelas. Strict sanitary precautions, the intelligent use of antibiotics and vaccines, and the widespread use of cages for layers and confinement rearing for broilers have made it possible to effect satisfactory disease control.

Parasitic diseases of poultry, including hexamitiasis of turkeys, are caused by roundworms, tapeworms, lice, and mites. Again, modern methods of sanitation, prevention, and treatment provide excellent control.

BIBLIOGRAPHY. The UNITED STATES DEPARTMENT OF AGRICULTURE, *Agricultural Statistics* (annual), provides reliable agricultural information for the United States and includes basic worldwide statistics. The UNITED NATIONS FOOD AND AGRICULTURE ORGANIZATION (FAO), *Production Yearbook* (annual), is the most authoritative source for worldwide agricultural statistics. See also the proceedings of the first world conference on animal production, *Efficiency in Animal Production* (1963); and *Dairying Throughout the World: Monographs on Dairying in a Number of Countries Throughout the World* (1966), published on the occasion of the 17th International Dairy Congress. Other works on this subject include: CHARLES S. PLUMB, *Types and Breeds of Farm Animals*, rev. ed. (1920), a classic history of breeds; FAO, *European Breeds of Cattle*, 2 vol. (1966), an excellent history and description of European dairy, dual-purpose, and beef breeds; *Stud Managers Handbook* (1958), authoritative, selected articles from early stud manager's courses at the University of Kentucky; *Swine Feed Additives* (1970), proceedings of a symposium by the College of Agriculture at the University of Kentucky that is an excellent summation of the physiological and economic effects of antibiotics on swine production and pork prices; and standard textbooks, such as A.L. NEUMANN and R.R. SNAPP, *Beef Cattle*, 6th ed. (1969); W.E. CARROLL and J.L. KRIDER, *Swine Production*, 4th ed. (1971); W.E. PETERSEN and A.M. FIELD, *Dairy Farming* (1953); A.R. WINTER and E.M. FUNK, *Poultry: Science and Practice*. 5th ed. (1960); and D.J. KAYS, *The Horse* (1953).

(W.P.G.)